Machine Translated by Google

Encyclopédie

Auteur et éditeur

Elzein Salah Elzein

Membre IEEE

- Articles en algorithmes

Éditeur IEEE, Royaume-Uni



Ce contenu est confidentiel

Machine Translated by Google

algorithmes

MDPI

Article

Outil de diagnostic numérique basé sur un réseau neuronal convolutif pour l'identification des maladies psychosomatiques

Marta Narigina , Andrejs Romanovs * et Yuri Merkuryev

Département de modélisation et de simulation, Université technique de Riga, 6A Kipsalas Street, LV-1048 Riga, Lettonie ; marta.narigina@rtu.lv (MN); jurijs.merkurjevs@rtu.lv (YM)
* Correspondance : andrejs.romanovs@rtu.lv

Résumé : Cet article évalue les capacités des modèles de réseaux neuronaux convolutifs (CNN) dans la détection des émotions à partir des expressions faciales, cherchant à faciliter le diagnostic des maladies psychosomatiques, généralement effectuées dans des contextes cliniques. À l'aide de l'ensemble de données FER-2013, deux modèles CNN ont été conçus pour détecter six émotions avec une précision de 64 %, bien qu'elles ne soient pas uniformément réparties ; ils ont démontré une plus grande efficacité pour identifier « heureux » et « surprise ». L'évaluation a été réalisée à l'aide de plusieurs mesures de performance (exactitude, précision, rappel et scores F1), en plus d'une validation plus approfondie avec un environnement clinique simulé supplémentaire pour des contrôles pratiques. Bien qu'elle montre des niveaux prometteurs pour une utilisation future, cette enquête met en évidence la nécessité d'études de validation approfondies en m Cette recherche souligne la valeur potentielle de l'IA en tant que complément aux approches de diagnostic traditionnelles tout en se concentrant sur des domaines de portée plus large (ensembles de données plus larges) et d'intérêt (intégration multimodale) à prendre en compte parmi les recommandations des études à venir. Cette étude souligne l'importance des modèles CNN dans le développement de diagnostics psychosomatiques et dans la promotion d'un développement futur basé sur l'éthique et les soins aux patients.

Mots-clés : réseaux de neurones convolutifs ; intelligence artificielle; reconnaissance des émotions; maladies psychosomatiques ; analyse des expressions faciales; détection des émotions en temps réel

check for updates

Citation : Narigina, M. ; Romanovs, A. ; Merkuryev, Y. Diagnostic numérique basé sur un réseau neuronal convolutif Outil d'identification des Maladies psychosomatiques. Algorithmes 2024, 17, 329. https://doi.org/10.3390/a17080329

Rédactrice académique : Gabriella Trucco

Reçu : 29 juin 2024
Révisé : 23 juillet 2024
Accepté : 23 juillet 2024
Publié : 30 juillet 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Le diagnostic précis des maladies psychosomatiques représente un défi important dans le cadre des soins de santé modernes. Les maladies psychosomatiques, dans lesquelles des facteurs psychologiques influencent considérablement les fonctions physiologiques, manifestent souvent des symptômes similaires à ceux de maladies purement physiques. Ce chevauchement peut conduire à des diagnostics erronés et, par conséquent, à des stratégies de traitement inappropriées [1–3]. Avec l'intégration croissante de l'intelligence artificielle dans les processus de diagnostic, il existe une opportunité croissante d'améliorer la précision du diagnostic à l'aide d'outils technologiques [4,5].

Les progrès récents dans l'apprentissage automatique, en particulier les réseaux de neurones convolutifs (CNN), ont donné des résultats prometteurs dans diverses applications médicales, notamment les diagnostics basés sur l'image et les prédictions comportementales [6]. La capacité de cette technologie à analyser et interpréter des modèles subtils dans les expressions faciales offre une nouvelle approche pour identifier les signaux émotionnels révélateurs de maladies psychosomatiques [7].

Cette étude vise à combler le fossé entre les évaluations psychologiques et les méthodes de diagnostic traditionnelles en intégrant un outil basé sur CNN pour analyser les expressions faciales, fournissant ainsi un moyen auxiliaire de diagnostic des maladies psychosomatiques. Cette recherche s'articule autour de plusieurs questions cruciales, notamment les suivantes : 1. Quelles

sont les limites actuelles du diagnostic des maladies psychosomatiques à l'aide des méthodes traditionnelles ?
méthodes traditionnelles [8] ?

2. Un modèle CNN formé pour reconnaître les expressions émotionnelles peut-il améliorer la précision du diagnostic des maladies psychosomatiques [9] ?

Machine Translated by Google

3. Dans quelle mesure un tel modèle peut-il être intégré efficacement aux flux de travail cliniques pour soutenir les professionnels de la santé sans remplacer leur jugement d'expert ?

En abordant ces questions, cet article explore la faisabilité et l'efficacité de l'utilisation d'une aide au diagnostic basée sur CNN en milieu clinique, améliorant ainsi la précision et l'efficacité globales du diagnostic des maladies psychosomatiques (10).

Les sections sont méticuleusement structurées pour fournir une analyse détaillée des différents aspects impliqués dans la construction d'outils de diagnostic basés sur CNN. Chaque partie est conçue pour approfondir chaque étape du processus de développement.

La section 2, Travaux connexes, examine différents travaux qui ont utilisé des méthodes d'apprentissage profond pour améliorer la reconnaissance des émotions du visage. Ces techniques ont été évaluées à l'aide de l'ensemble de données FER-2013 et ont montré des améliorations notables dans la précision de la reconnaissance des émotions.

La section 3, Méthodologie, détaille l'ensemble de données, l'architecture CNN, le processus de formation, les étapes de prétraitement et l'intégration multimodale de l'analyse faciale et des données de pouls pour prédire les troubles psychologiques.

La section 4, Évaluation des modèles CNN, évalue la fiabilité et l'efficacité des modèles CNN à l'aide de diverses mesures de performance et techniques de validation croisée dans des environnements cliniques simulés.

La section 5, Résultats et analyse, présente les mesures de performance des modèles CNN, les points forts, les domaines nécessitant des améliorations et la validation statistique au moyen de matrices de confusion et de graphiques de précision.

La section 6, Discussion, interprète l'importance des résultats, les biais potentiels, les limites, les considérations éthiques et les implications pratiques de l'intégration des outils d'IA dans les soins de santé.

La section 7, Conclusions, résume les principales conclusions et recommande d'élargir les ensembles de données, d'effectuer des validations dans le monde réel et de combiner des données multimodales pour améliorer les capacités de diagnostic, en mettant l'accent sur les considérations éthiques et l'impact potentiel de l'IA sur l'amélioration des soins aux patients.

Chaque section est créée pour partir de la précédente, le tout vers la construction d'une discussion approfondie qui souligne le grand potentiel des CNN pour changer la manière dont les diagnostics sont effectués dans les établissements de soins de santé.

2. Travaux connexes

Les progrès récents en matière d'apprentissage profond et de vision par ordinateur ont eu un impact significatif sur le domaine de la reconnaissance des émotions faciales (FER). Cette section passe en revue diverses études qui ont utilisé des techniques d'apprentissage en profondeur pour améliorer les performances des systèmes FER, en particulier celles évaluées à l'aide de l'ensemble de données FER-2013.

Proposé dans la référence [11], Georgescu et al. a présenté une approche dans laquelle les fonctionnalités apprises automatiquement par un certain nombre d'architectures CNN sont fusionnées avec des fonctionnalités fabriquées à la main via le modèle bag-of-visual-words (BOVW). La classification des émotions est obtenue via cette méthodologie hybride, adoptant un cadre d'apprentissage local et décrivant de meilleures performances en termes de précision grâce à l'utilisation combinée d'ensembles de fonctionnalités automatiques et artisanales.

Pecoraro et coll. [12] ont développé un module d'auto-attention à canal multi-tête local (LHC), considéré comme un composant complémentaire du réseau neuronal convolutif standard (CNN). Ce mécanisme évite les actions génériques d'attention et se concentre plutôt sur les parties du visage qui sont des acteurs clés de l'expression émotionnelle. Il a été démontré que si vous introduisez des modules LHC, vous pouvez augmenter la précision de la reconnaissance des expressions faciales en localisant les zones critiques du visage.

Fard et coll. [13] ont introduit la perte de corrélation adaptative (Ad-Corre), qui se compose de trois éléments, dont le discriminateur de caractéristiques, le discriminateur moyen et le discriminateur d'intégration. La perte Ad-Corre est utilisée en conjonction avec la perte d'entropie croisée habituelle pour augmenter les capacités discriminantes des vecteurs de caractéristiques, ce qui entraîne par conséquent

Machine Translated by Google

meilleure précision de classification. Cette nouvelle fonction de perte a été expérimentée sur Xception comme ainsi que les modèles Resnet50, et les résultats ont montré des améliorations majeures des performances.

Un nouveau modèle FER a été proposé par Vignesh et al. [14], qui est une fusion de blocs de segmentation basés sur U-Net dans l'architecture VGG-19. Cette approche unique, appelée La segmentation VGG-19 met l'accent sur les caractéristiques significatives d'une carte de caractéristiques qui, dans produit à son tour un meilleur résultat : une extraction de caractéristiques plus avancée et une plus grande précision dans reconnaître les émotions.

Mukhopadhyay et coll. [15] ont présenté une méthode FER basée sur l'apprentissage profond exploitant caractéristiques texturales telles que les modèles binaires locaux (LBP), les modèles ternaires locaux (LTP) et modèles binaires locaux complétés (CLBP). En formant un CNN sur ces caractéristiques texturales, le Le modèle a obtenu une précision améliorée dans la détection des émotions faciales.

Shahzad et coll. [16] ont proposé la méthode ZFER, basée sur le zonage et améliorant la reconnaissance des émotions sur le visage en fournissant des repères supplémentaires dont le les yeux, les sourcils, le nez, le front et la bouche. Cette approche utilise l'hybride VGG-16 modèle qui apprend en fonction de points de repère basés sur des zones pour augmenter la précision de la classification émotions faciales.

El Boudouri et Bohi [17] proposent EmoNeXt, qui est une conception ConvNeXt modifiée. spécifiquement pour FER. EmoNeXt comprend un STN (Spatial Transformer Network) pour concentrer sur les zones riches en fonctionnalités du visage, ainsi que des blocs de compression et d'excitation qui permettent capturer les dépendances au niveau des canaux. De plus, ils ont ajouté une régularisation de l'attention personnelle terme qui vise à générer efficacement des vecteurs de caractéristiques en les compactant. Résultats basés sur les expériences réalisées à l'aide de l'ensemble de données FER-2013 ont montré qu'EmoNeXt est meilleur que tous modèles de pointe existants en matière de précision de la classification des émotions.

L'ensemble de données FER-2013, fortement appliqué dans ces travaux, comprend 35 887 images en niveaux de gris qui sont classés en sept classes émotionnelles comme suit : colère, dégoût, peur, bonheur, neutralité, tristesse et surprise. Son grand volume et sa diversité de données font de cet ensemble de données un point de référence standard pour l'évaluation du système FER [18].

La combinaison de ces méthodes avancées d'apprentissage en profondeur (y compris l'auto-attention mécanismes et modèles hybrides dotés de fonctionnalités à la fois automatiques et artisanales) plus de nouvelles fonctions de perte ont considérablement amélioré les performances des systèmes conçus pour le soin du visage reconnaissance d'expressions. Ces changements soulignent la promesse de l'apprentissage profond dans la création Modèles FER solides et fiables : EmoNeXt constitue une avancée majeure dans ce domaine.

3. Méthodologie

Le principal ensemble de données utilisé dans cette étude était l'ensemble de données FER-2013 [19] qui, fourni dans le tableau 1, comprend 35 887 images de visages humains, chacune étiquetée avec l'un des sept expressions émotionnelles comme suit : colère, dégoût, peur, bonheur, neutralité, tristesse et surprendre. Cet ensemble de données est largement reconnu dans le domaine de la vision par ordinateur pour sa diversité et complexité, contenant des expressions faciales capturées dans diverses conditions du monde réel.

Tableau 1. Test de l'ensemble de données FER-2013 et images de formation.

Émotion	Images de formation	Images de test
En colère	3995	958
Dégoût	436	111
Peur	4097	1024
Heureux	7215	1774
Neutre	4965	1233
Triste	4830	1247
Surprendre	3171	831

Avant le traitement, les images ont été redimensionnées à une dimension uniforme de 48 ×48 pixels. Les valeurs des pixels ont été normalisées dans la plage [0, 1] pour garantir une entrée uniforme dans le système neuronal. réseau, facilitant la stabilisation du processus d'apprentissage et l'amélioration du modèle convergence [20].

Machine Translated by Google

Pour élargir la simulation de diverses expressions faciales et conditions pouvant être rencontrées par le modèle dans la vie réelle, différentes méthodes d'augmentation des données ont été utilisées en plus de la rotation, du zoom, du retournement et du déplacement horizontal. Cette étape vise à obtenir davantage de capacités de généralisation du modèle.

Même si l'ensemble de données FER-2013 était détaillé et complet, l'émotion « dégoût » a été exclue de cette étude car elle comportait un nombre d'échantillons beaucoup plus petit que les autres émotions. Dans l'ensemble d'entraînement, il y avait 436 images exprimant le « dégoût » et dans l'ensemble de test, il y en avait 111. Cette asymétrie au sein de l'ensemble de données pourrait entraîner un manque d' entraînement suffisant pour cette émotion spécifique, conduisant ainsi à un modèle peu fiable avec un biais. vers ces catégories riches en échantillons émotionnels. Notre décision d'exclure le « dégoût » visait à maintenir une approche plus équilibrée lors de la formation du modèle afin que les prédictions faites par notre modèle pour toutes les autres catégories d'émotions soient fiables.

3.1. Architecture et développement de CNN

Le processus de formation comprenait plusieurs époques avec un arrêt anticipé de la formation basé sur une perte de validation, évitant ainsi le surentraînement. Un planificateur personnalisé a été utilisé pour réguler le taux d'apprentissage de manière dynamique. À mesure que le nombre d'époques augmentait, le taux d'apprentissage diminuait, ce qui contribuait à affiner les poids du modèle et à améliorer la précision.

Les expériences ont été menées sur un ordinateur portable MSI Summit B15 A11 MT équipé d'un processeur Intel Core i7, de 64 Go de RAM et fonctionnant sous un système d'exploitation Linux. Chaque session de formation, comprenant 100 époques, a duré environ 10 heures.

L'architecture du réseau neuronal convolutif a été conçue pour extraire des caractéristiques progressivement plus détaillées des images faciales. •

Les couches initiales du réseau neuronal convolutif sont conçues pour extraire progressivement les caractéristiques détaillées des images faciales. Le réseau comprend deux couches convolutives avec 64 filtres chacune, suivies d'une couche de pooling maximum. Cette configuration est conçue pour capturer les caractéristiques fondamentales du visage, telles que les bords et les textures simples.

• Les couches intermédiaires comprennent des ensembles de couches convolutives avec un nombre croissant de filtres inclus, avec 128 et 256 filtres par ensemble. Chaque ensemble est ensuite suivi d' une couche de pooling maximale. Les couches susmentionnées sont conçues pour détecter des caractéristiques plus complexes, telles que les éléments constitutifs des expressions faciales [3].
• Les couches finales du réseau sont appelées « couches avancées ». Un ensemble final de couches convolutives profondes avec 512 filtres est utilisé pour extraire les caractéristiques de haut niveau qui sont cruciales pour la reconnaissance des nuances subtiles des expressions. • Le réseau est constitué de couches entièrement connectées qui agissent comme des classificateurs, c'est-à-dire des couches de classification. Afin d'éviter le surajustement, le modèle utilise une régularisation par abandon , tandis que la normalisation par lots est utilisée pour maintenir la sortie moyenne proche de 0 et l' écart type de sortie proche de 1 tout au long de la formation [3].

L'optimiseur Adam [21], connu pour son efficacité dans la gestion des gradients clairsemés et ses capacités de taux d'apprentissage adaptables, a été utilisé dans le processus de formation. Le processus de formation comprenait plusieurs époques, avec un arrêt précoce de la formation basé sur une perte de validation, évitant ainsi le surentraînement. Un planificateur personnalisé a été utilisé pour réguler le taux d'apprentissage de manière dynamique. À mesure que le nombre d'époques augmentait, le taux d'apprentissage diminuait, ce qui contribuait à affiner les poids du modèle et à améliorer la précision [22].

Le deuxième modèle suit une architecture similaire mais améliorée, incorporant des couches et des techniques supplémentaires pour améliorer les performances comme suit :

- Couches initiales : Semblables au premier modèle, les couches initiales comprennent deux couches convolutives avec 64 filtres chacune, suivies d'une normalisation par lots et d'une activation LeakyReLU [23], se terminant par une couche de pooling maximale et un abandon pour la régularisation. •
Couches intermédiaires : les couches intermédiaires se composent de deux ensembles de couches convolutives avec 128 et 256 filtres par ensemble, chacun suivi d'une normalisation par lots, d'une activation LeakyReLU, d'un regroupement maximal et de couches d'abandon.

Machine Translated by Google

• Couches avancées : ce modèle comprend un ensemble supplémentaire de couches convolutives profondes avec 256 filtres, suivis d'une normalisation par lots et d'une activation LeakyReLU, avant d'appliquer un pooling et un abandon maximum. •
Couches finales : après avoir aplati les cartes de caractéristiques, le modèle comprend une couche dense avec 256 unités, une normalisation par lots, une activation LeakyReLU et un abandon, se terminant par une couche softmax dense pour la classification.

Le modèle utilisait l'optimiseur Adam avec un taux d'apprentissage initial de 0,001, et un planificateur de taux d'apprentissage a été utilisé pour diminuer le taux d'apprentissage au fil du temps afin d'améliorer la stabilité et les performances de l'entraînement. Le modèle a été formé à l'aide d'un ImageDataGenerator [24] pour l'augmentation des données, qui inclut des techniques telles que la rotation, les changements de largeur et de hauteur, le cisaillement, le zoom et les retournements horizontaux pour améliorer les capacités de généralisation du modèle.

Le deuxième modèle intègre également des techniques de régularisation avancées, telles que la régularisation L2 [25], pour éviter davantage le surajustement et améliorer la robustesse du modèle.

3.2. Méthodologie de prédiction des troubles psychologiques à l'aide de l'analyse faciale et des données de pouls

La prédiction de la maladie psychologique est l'objectif principal de cette étude. Le système proposé utilise une technique multimodale intégrant l'analyse des expressions faciales à l'aide d'un modèle formé par CNN pour les données émotionnelles et extrayant la fréquence cardiaque des images vidéo pour les informations physiologiques. Cette méthode combine à la fois les principes de vision par ordinateur et d'ingénierie biomédicale.

Pour améliorer la capacité du système à diagnostiquer les conditions psychosomatiques, les données de pouls sont intégrées à la reconnaissance des émotions faciales [26]. Le matériel utilisé à cet effet comprend une webcam standard, en particulier la Logitech C920, qui capture des vidéos en temps réel à 30 images par seconde. Les images vidéo sont traitées pour extraire à la fois les expressions faciales et la fréquence du pouls du sujet.

3.2.1. Méthodologie de détection du pouls Le

mécanisme de détection du pouls est crucial pour intégrer de manière transparente les données physiologiques aux algorithmes de reconnaissance des expressions faciales en temps réel. Cette méthode implique une série de processus soigneusement planifiés pour garantir une collecte et un traitement de données précis et fiables .

Dans un premier temps, la région d'intérêt (ROI) sur le visage est choisie puis examinée pour détecter les variations d'intensité de couleur causées par la circulation sanguine. Le système utilise le classificateur Haar Cascade pré-entraîné d'OpenCV [27] pour identifier les visages dans chaque image vidéo. Ce classificateur est efficace car il peut détecter rapidement et précisément les caractéristiques du visage dans une vidéo en direct en analysant l'image et en sélectionnant les zones rectangulaires susceptibles de contenir des personnes. La position et la taille de chaque visage détecté sont fournies par des coordonnées.

Après avoir détecté le visage, le système choisit la zone du visage identifiée comme région d' intérêt (ROI). Cette région est essentielle pour des investigations plus approfondies car elle comprend les zones de la peau où les changements de couleur sont les plus visibles en raison du flux sanguin. Afin de garantir une sélection constante de la ROI (région d'intérêt) sur toutes les montures, le système utilise des repères faciaux, qui sont des endroits spécifiques du visage tels que les coins des yeux, le bout du nez et les coins de la bouche. . Ces repères servent à préserver les dimensions et l'emplacement de la ROI même en présence de mouvements mineurs du suje
Lorsque le système détecte de nombreux visages, il donne la priorité au visage le plus central et de plus grande taille. Cela garantit que la région d'intérêt (ROI) correspond au sujet principal.

Après avoir identifié de manière fiable le retour sur investissement, l'étape suivante consiste à évaluer la force du canal vert à l'intérieur de cette région spécifique. Le canal vert est sélectionné en raison de sa grande sensibilité aux variations du volume sanguin sous la peau, qui sert d'indicateur du pouls. Le modèle de couleur RVB classe les images en canaux rouge, vert et bleu Parmi ces canaux, le canal vert est particulièrement apte à détecter les changements de volume sanguin. En effet, l'hémoglobine présente dans le sang absorbe davantage la lumière verte.
Le système calcule l'intensité moyenne du canal vert pour chaque pixel à l'intérieur

Machine Translated by Google

la région d'intérêt (ROI) en prenant la moyenne des valeurs de couleur verte. Il en résulte une valeur d'intensité unique pour chaque image. Les mesures d'intensité sont recueillies sur une période de temps pour créer un ensemble de données de séries chronologiques, où chaque valeur correspond à une image particulière, représentant la variation continue de l'intensité du canal vert. Le système ajoute systématiquement ces valeurs d'intensité moyennes à une liste tout en enregistrant également des horodatages pour maintenir un séquençage temporel précis, ce qui est crucial pour une analyse précise des impulsions.

Afin de préparer l'analyse, le système reste inactif jusqu'à ce qu'il ait collecté un minimum de 100 lectures d'intensité. Cela garantit qu'il y a suffisamment de données disponibles pour l'analyse de fréquence. Dans l'étape suivante, les composantes fréquentielles des variations d'intensité sont examinées afin de déterminer la fréquence du pouls. Les données d'intensité sont ajustées en supprimant toutes tendances linéaires ou variations lentes qui ne sont pas liées au battement cardiaque, garantissant ainsi que le signal résiduel représente de manière appropriée les oscillations périodiques provoquées par le flux sanguin. Les valeurs d'intensité sans tendance sont transformées du domaine temporel au domaine fréquentiel à l'aide de la transformée de Fourier rapide (FFT) [28], décomposant le signal en ses fréquences individuelles. L'analyse FFT cible spécifiquement la plage de fréquences de 0,7 Hz à 2 Hz, ce qui correspond à la plage de fréquence cardiaque moyenne de 42 à 120 battements par minute (BPM). Le système détecte la fréquence la plus élevée dans cette plage, qui correspond à l'élément récurrent le plus important du signal associé au battement cardiaque. La fréquence maximale est traduite en battements par minute (BPM) en la multipliant par 60.

Le résultat final du processus de détection du pouls est la fréquence du pouls calculée. Si la quantité de points de données est insuffisante, le système persistera à collecter et à analyser les données jusqu'à ce qu'une fréquence cardiaque fiable puisse être déterminée. Le système peut extraire efficacement les informations sur la fréquence de pouls des images vidéo en analysant soigneusement l'intensité du canal vert au sein de la ROI et en employant des algorithmes sophistiqués de traitement du signal. Associées à la détection des émotions du visage, ces données constituent une base solide pour évaluer les maladies psychosomatiques en temps réel.

3.2.2. Intégration avec la reconnaissance des émotions

faciales La méthodologie combine les données de pouls avec les sorties de reconnaissance des expressions faciales pour fournir une évaluation approfondie de l'état psychologique de l'individu. Le processus comprend plusieurs composants interconnectés pour atteindre cet objectif.

Au cours de la phase d'identification des émotions, un modèle de réseau neuronal convolutif (CNN) analyse la zone spécifique du visage pertinente (appelée région d'intérêt du visage ou ROI) afin d'identifier et de classer les émotions. Ce modèle a suivi une formation approfondie pour identifier avec précision six émotions fondamentales, notamment la colère, la peur, le bonheur, la neutralité, le chagrin et la surprise.

En même temps, l'outil effectue une analyse du pouls pour calculer la fréquence cardiaque, fournissant ainsi une mesure physiologique supplémentaire. Cette approche simultanée garantit l' acquisition et l'analyse de données à la fois émotionnelles et physiologiques.

Un algorithme basé sur des règles est utilisé pour analyser les données combinées. Cette méthode met en corrélation des émotions et des plages de pouls particulières avec des troubles psychologiques suspectés. Par exemple, l'algorithme peut interpréter un pouls élevé ainsi que des sentiments comme « heureux » ou « surprise » comme le signe d'un état maniaque. En revanche, une diminution de la fréquence cardiaque observée en combinaison avec des sentiments de tristesse ou de peur pourrait indiquer un état dépressif.

En intégrant ces différentes sources de données, le système offre une perspective complète tive sur l'état mental de l'individu, permettant des évaluations plus précises et détaillées.

3.2.3. Architecture globale du système

L'architecture globale du système est conçue pour intégrer ces processus de manière transparente, garantissant une analyse et un retour d'informations en temps réel. Le système capture l'entrée vidéo, traite les données faciales et de pouls et interprète les résultats pour fournir une analyse détaillée de l'état émotionnel et physiologique du sujet. Cette approche intégrée améliore

Machine Translated by Google

l'exactitude et la fiabilité des évaluations de l'état psychosomatique, offrant un outil précieux pour le diagnostic clinique.

3.2.4. Conclusions

L'intégration des données de pouls avec la reconnaissance des émotions faciales à l'aide de techniques avancées de traitement du signal et d'apprentissage profond fournit un cadre solide pour l'évaluation des maladies psychosomatiques. Cette méthodologie améliore non seulement la précision du diagnostic, mais offre également une compréhension plus complète de l'état émotionnel et physiologique du patient . Les travaux futurs se concentreront sur l'amélioration de la robustesse du système et la validation de ses performances dans divers contextes cliniques.

3.2.5. Interprétation des conditions

Avec à la fois des données émotionnelles et pouls à portée de main, le système utilise un algorithme basé sur des règles pour interpréter l'état psychologique du sujet. Cet algorithme prend en compte à la fois l' émotion identifiée et la fréquence du pouls.

Si le pouls est trop élevé ou trop faible, cela peut nous indiquer un état maniaque ou dépressif lorsqu'il est associé à des émotions correspondantes. Par exemple, un pouls élevé indiquant « heureux » ou « surprise » pourrait suggérer un état maniaque.

Le diagnostic final (par exemple, « normal », « dépression », « maladie bipolaire ») est établi par combinant l'évaluation émotionnelle avec des données physiologiques [29].

Cet examen conjoint aide à démêler les conditions psychologiques qui pourraient échapper aux simples expressions faciales, offrant ainsi une base solide pour la prédiction des maladies psychosomatiques.

Le système est conçu pour s'intégrer de manière transparente dans les environnements cliniques existants, fonctionnant en arrière-plan pendant les consultations des patients. Cela permet aux prestataires de soins d'obtenir des informations diagnostiques augmentées sans perturber le processus de consultation.

4. Évaluation des modèles CNN

L'efficacité et la fiabilité du modèle de réseau neuronal convolutif (CNN) développé pour la reconnaissance des émotions ont été rigoureusement évaluées au moyen d'un processus structuré de test et de validation. Ce processus a été conçu pour évaluer l'exactitude et l'applicabilité du modèle en milieu clinique, notamment pour faciliter le diagnostic des maladies psychosomatiques.

4.1. Tests de performance

Les performances du modèle ont été évaluées principalement à l'aide de l'ensemble de données de test, composé d'images inédites de l'ensemble de données FER-2013 afin de garantir une évaluation impartiale. L'ensemble de données n'a pas été utilisé pendant la phase de formation et a donc fourni une base solide pour les tests. Les principales mesures utilisées pour évaluer les performances du modèle étaient les suivantes.

La précision du modèle a été évaluée. Cette métrique a été calculée en comparant les états émotionnels prédits par le modèle avec les étiquettes réelles fournies dans l'ensemble de données de test. Un taux de précision élevé indique que le modèle est efficace pour identifier correctement différentes émotions à partir des expressions faciales.

Une matrice de confusion a été générée pour fournir une représentation visuelle des performances du modèle à travers différentes émotions. Cette matrice a facilité l'identification d' émotions spécifiques pour lesquelles le modèle pourrait présenter des performances sous-optimales, telles que la confusion de tristesse et de dégoût, ce qui pourrait avoir une pertinence clinique significative.

Les mesures de précision, de rappel et de score F1 ont été calculées. Ces métriques ont été calculées pour chaque catégorie d'émotion afin de fournir une analyse plus détaillée de la performance du modèle. La précision est une mesure de l'exactitude des prédictions positives, le rappel est une mesure de la capacité à trouver toutes les instances positives et le score F1 fournit un équilibre entre précision et rappel.

4.2. Techniques de validation

De plus, les performances du modèle ont été validées selon les critères suivants :

Machine Translated by Google

• La validation croisée a été utilisée comme technique pour garantir la robustesse du modèle et réduire la probabilité de surajustement. Afin de garantir la robustesse du modèle et de réduire le risque de surajustement, une validation croisée k fois a été utilisée pendant la phase de formation. Cette technique impliquait de diviser l'ensemble de données de formation en « k » sous-ensembles et de former de manière itérative le modèle « k » fois (k = 5), chaque itération impliquant l'utilisation d'un sous-ensemble différent soumis à validation. • Tests en temps réel : pour simuler une utilisation réelle,
le modèle a également été testé dans un environnement clinique simulé, où il a traité des flux vidéo en direct pour détecter et interpréter les expressions émotionnelles en temps réel. Ces tests ont été d'une grande importance pour évaluer les performances du modèle dans des situations dynamiques similaires aux interactions réelles avec les patients. Les expériences, comprenant à la fois des phases de formation et de test, ont été menées sur un ordinateur portable MSI Summit B15 A11 MT équipé d'un processeur Intel Core i7 et de 64 Go de RAM, fonctionnant sous un système d'exploitation Linux. La durée de formation pour chaque session de 100 époques était d'environ 100 heures.

Afin d'entraîner le modèle CNN à reconnaître les signaux émotionnels liés aux phénomènes psychosomatiques Pour lutter efficacement contre les maladies, plusieurs stratégies ont été mises en œuvre,
notamment les suivantes : • Augmentation améliorée des données : pour refléter la variabilité observée dans les contextes cliniques réels , des techniques approfondies d'augmentation des données ont été utilisées, notamment la rotation, la mise à l'échelle et les changements d'éclairage. Cette approche a facilité la formation du modèle pour reconnaître avec précision les signaux
émotionnels dans diverses conditions. • Optimisation des couches : diverses configurations architecturales ont été testées, y compris des modifications du nombre et de la taille des couches convolutives et des filtres, dans le but d'optimiser le modèle pour la détection précise de signaux émotionnels subtils qui sont souvent révélateurs de maladies psychosomatiques.

L'intégration du modèle de reconnaissance des émotions dans les flux de travail cliniques s'est appuyée sur plusieurs bonnes pratiques, notamment les
suivantes : • Conception centrée sur l'utilisateur : l'interface de l'outil de reconnaissance des émotions a été conçue avec la contribution des professionnels de la santé pour garantir qu'elle est intuitive et améliore plutôt que perturbe le flux de travail
clinique. • Boucles de rétroaction : des commentaires continus ont été recherchés auprès des utilisateurs finaux pour faciliter les améliorations itératives du système. Un tel retour d'information était de la plus haute importance pour le raffinement de la conception de l'interface, l'amélioration de la précision du modèle et l' assurance de la pertinence de l'outil en milieu clinique
[30]. • Considérations éthiques et de confidentialité : Compte tenu de la sensibilité des évaluations psychologiques, l'intégration du modèle a été réalisée en mettant fortement l'accent sur les considérations éthiques et la vie privée des patients. Cela comprenait une gestion sécurisée des données et des processus de consentement clairs des patients [31].

5. Résultats et analyse

Les données utilisées pour entraîner les deux modèles CNN ont été conçues pour classer six émotions, notamment la colère, la peur, le bonheur, la neutralité, la tristesse et la surprise, en utilisant les images étiquetées. Dans ce processus de formation, les modèles ont été formés pendant 100 époques où les performances ont été mesurées à chaque époque avec à la fois la précision et la perte de formation et de validation enregistrées.

Les scores de précision, de rappel et F1 ont été utilisés pour évaluer les performances de chaque modèle. De plus, des matrices de confusion ont été utilisées pour décrire dans quelle mesure le modèle était capable de classer les instances et de visualiser les performances de classification du modèle. De plus, un test d'indépendance du chi carré a été utilisé pour déterminer s'il existait une relation significative entre les prédictions des modèles.

Une illustration montrant comment les étiquettes vraies et prédites sont réparties entre les six catégories d'émotions dans le premier modèle est présentée à la figure 1.

Machine Translated by Google



Figure 1. Matrice de confusion pour le premier modèle.

La performance de la formation et de la validation sur 100 époques dans le premier modèle est illustré à la figure 2. Cela a été réalisé en dessinant à la fois des courbes de précision et de perte.

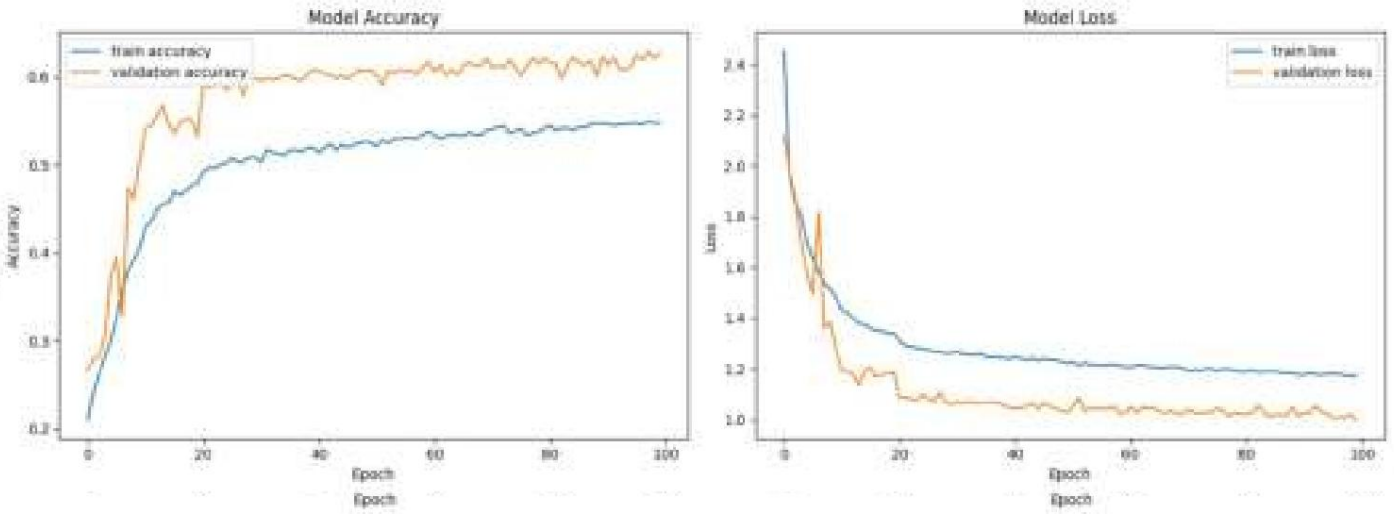


Figure 2. Précision et perte du modèle – premier modèle.

Le premier modèle a atteint une précision globale de 61,78 %. Le tableau 2 résume les précision, rappel et score F1 pour chaque émotion.

Tableau 2. Mesures de performances pour le premier modèle.

Émotions	Précision	Rappel	Score F1	Soutien
En colère	0,55	0,51	0,53	958
Peur	0,45	0,24	0,32	1024
Heureux	0,75	0,88	0,81	1774
Neutre	0,52	0,61	0,57	1233
Triste	0,47	0,49	0,48	1247
Surprendre	0,70	0,72	0,71	831
Précision			0,62	7067
Macro moyenne	0,60	0,59	0,59	7067
Moyenne pondérée	0,61	0,62	0,61	7067

La matrice de confusion du deuxième modèle (Figure 3) montre la distribution des étiquettes vraies et prédites dans les six catégories d'émotions.

Machine Translated by Google



Figure 3. Matrice de confusion pour le deuxième modèle.

Les courbes de précision et de perte du deuxième modèle (Figure 4) montrent l'entraînement et performances de validation pendant 100 époques.

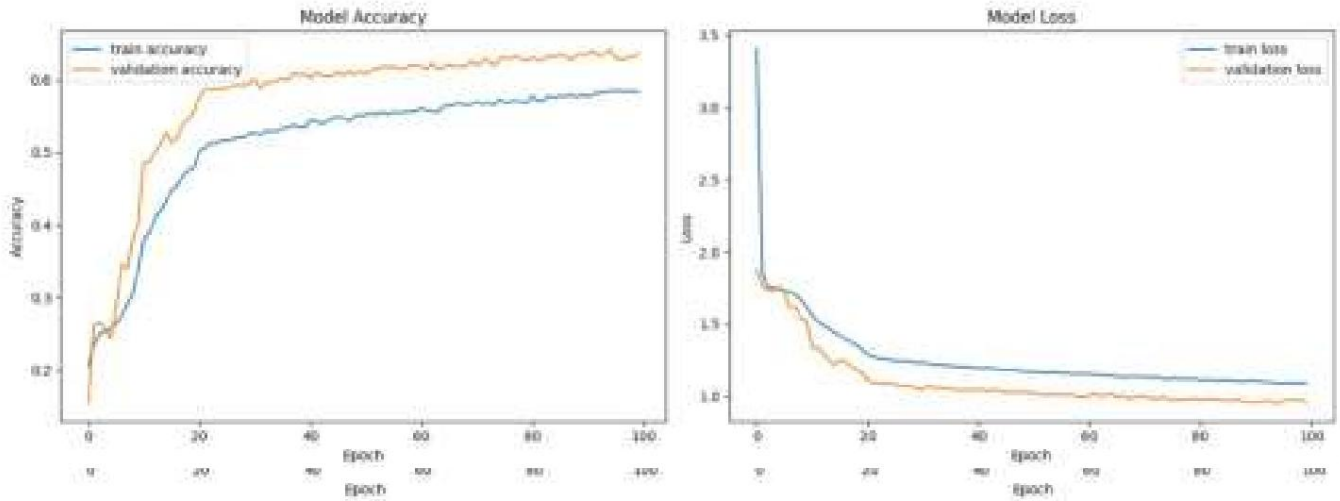


Figure 4. Précision et perte du modèle – deuxième modèle.

Le deuxième modèle a atteint une précision globale de 63,95 %. Le tableau 3 résume les précision, rappel et score F1 pour chaque émotion.

Tableau 3. Mesures de performances pour le deuxième modèle.

Émotions	Précision	Rappel	Score F1	Soutien
En colère	0,58	0,55	0,56	958
Peur	0,50	0,28	0,36	1024
Heureux	0,85	0,88	0,87	1774
Neutre	0,55	0,68	0,61	1233
Triste	0,49	0,53	0,51	1247
Surprendre	0,72	0,77	0,75	831
Précision			0,64	7067
Macro moyenne	0,62	0,62	0,61	7067
Moyenne pondérée	0,63	0,64	0,63	7067

Le deuxième modèle a surpassé le premier modèle, atteignant une précision globale plus élevée et des mesures de performances améliorées pour la plupart des catégories d'émotions. La précision et la perte les courbes indiquent que la formation des deux modèles a été efficace, le deuxième modèle montrant une meilleure généralisation, comme en témoigne sa plus grande précision de validation.

Machine Translated by Google

Les deux modèles ont montré une grande précision et un grand rappel de l'émotion « heureuse », reflétant la robustesse des modèles dans l'identification de cette catégorie. Cependant, les deux modèles présentaient des problèmes avec l'émotion de « peur », ce qui indique la nécessité d'améliorations supplémentaires et de données d'entraînement supplémentaires.

Nous avons effectué un test du chi carré (1) pour comparer les mesures de performance (précision, rappel et F1-Score) entre les deux modèles dans différentes catégories d'émotions.

Le chi carré (x^2) la statistique est calculée à l'aide de la formule :

$$x^2 = \sum \frac{(Oi - Ei)^2}{Ei} \quad (1)$$

où:

- Oi est la fréquence observée.
- Ei est la fréquence attendue.

Le résultat du test du chi carré pour le premier modèle est de 0,0025 avec une valeur p = 1,0000. Le Le résultat du test du chi carré pour le deuxième modèle est de 0,0036 avec une valeur p = 1,0000.

Les résultats du test du chi carré indiquent qu'il n'y a pas de différences statistiquement significatives dans les mesures de précision, de rappel et de score F1 entre les deux modèles dans les six catégories d'émotions.

5.1. Application au diagnostic des maladies psychosomatiques

L'utilisation de la classification des émotions peut grandement aider à identifier les conditions psychosomatiques suivantes.

La dépression se caractérise par des sentiments persistants de tristesse, une perte d'intérêt et un manque de énergie. Les émotions pertinentes incluent la tristesse, la neutralité et la colère.

Les troubles anxieux se caractérisent par une peur et une inquiétude intenses ainsi que par une sensation de nervosité. Les émotions pertinentes incluent la peur, la neutralité et la surprise.

Le trouble bipolaire comprend des sautes d'humeur extrêmes d'un côté de la manie à l'autre bout de la dépression. Les émotions pertinentes incluent la joie, la tristesse et la colère.

Le trouble de stress post-traumatique (SSPT) se développe après avoir été exposé à des événements qui évoquent des niveaux élevés de peur, entraînant des flashbacks associés à de la colère ou de la surprise au souvenir. Les émotions pertinentes incluent la peur, la colère et la surprise.

Le trouble obsessionnel-compulsif (TOC) se caractérise par des pensées intrusives et des envies d'accomplir des rituels. Cela implique souvent une concentration excessive sur les symptômes physiques, comme la douleur, qui peuvent être pénibles. La nature répétitive des symptômes du TOC peut amener les individus à se sentir piégés dans un cycle sans fin, contribuant ainsi à des sentiments d'impuissance et de désespoir. Faire face à ces types de symptômes rend souvent les individus encore plus frustrés car ils ont du mal à faire face à leurs activités quotidiennes.

5.2. Conclusions

Les deux modèles CNN sont des outils de diagnostic prometteurs pour détecter les troubles psychosomatiques basés sur les états émotionnels. Le deuxième modèle présentait une plus grande précision et de meilleures mesures de performance globales et semblait être plus robuste en termes de classification des émotions . Une reconnaissance précise des émotions pourrait améliorer considérablement le diagnostic et le traitement de divers troubles psychosomatiques, améliorant ainsi les résultats pour les patients. Les travaux futurs devraient se concentrer sur le perfectionnement de ces modèles, en intégrant des données de formation plus diversifiées et en explorant des améliorations algorithmiques pour accroître encore leur utilité clinique.

6. Discussion

Les résultats de cette étude suggèrent que les modèles de réseaux neuronaux convolutifs (CNN) sont très prometteurs pour la reconnaissance des émotions, un aspect essentiel du diagnostic des maladies psychosomatiques. Avec des précisions de test de 61,78 % et 63,95 %, les modèles démontrent une capacité supérieure à la moyenne à identifier et catégoriser correctement les états émotionnels sur la base des expressions faciales. Il convient de noter en particulier la capacité des modèles à identifier les émotions des

Machine Translated by Google

« bonheur » et « surprise », suggérant une base solide pour un perfectionnement ultérieur du modèle. Les performances constantes sur ces émotions suggèrent que les modèles ont une fiabilité niveau de précision et de sensibilité dans ces domaines, ce qui est crucial pour leur utilité potentielle dans milieux cliniques.

Nous n'avons pas effectué de comparaison directe avec les méthodes de diagnostic actuelles en clinique. pratique, mais les modèles CNN fournissent un niveau impartial et uniforme qui pourrait correspondre bien avec les approches traditionnelles. À l'heure actuelle, les évaluations de l'état émotionnel en clinique les paramètres sont largement subjectifs car ils dépendent souvent des déclarations du patient et de l'avis du clinicien. interprétation. Mais disposer d'un outil de reconnaissance des émotions introduirait de l'objectivité dans le processus. mesure, qui pourrait ensuite être utilisée pour porter des jugements cliniques basés sur les émotions (dans les cas où les patients ont des difficultés à exprimer leurs émotions ou ne connaissent pas leur états émotionnels).

6.1. Comparaison avec les modèles de base et les techniques de pointe

La reconnaissance des émotions faciales est un élément important de l'interaction homme-machine et l'informatique affective. Pour avancer dans le domaine, il faut comparer les nouveaux modèles avec des méthodes de base traditionnelles et des techniques de pointe. Cet article présente une analyse comparative de divers modèles, y compris la régression logistique, le vecteur de support machine (SVM), k-voisin le plus proche (k-NN) et CNN, en utilisant l'ensemble de données FER-2013.

Fournir une évaluation complète et précise de la reconnaissance des émotions faciales systèmes, nous avons comparé les performances du modèle CNN avec un modèle de base simple et des modèles plus complexes issus de la littérature actuelle. Le but de cette étude est d'évaluer l'efficacité de différentes architectures de modèles et d'identifier les domaines potentiels d'amélioration. améliorations futures.

Trois modèles de base ont été mis en œuvre pour servir de référence en matière de performances comparaison comme suit :

- La régression logistique est un modèle linéaire simple souvent utilisé pour les tâches de classification.
- La machine à vecteurs de support (SVM) est un modèle non linéaire puissant avec une base radiale. noyau de fonction (RBF).
- Les k-voisins les plus proches (k-NN) sont une méthode non paramétrique qui classe les données en fonction de proximité dans l'espace des fonctionnalités.

L'efficacité de ces modèles simples a été évaluée sur l'ensemble de données FER-2013. Le les résultats détaillés dans le tableau 4 ont servi de point de référence pour l'évaluation de nos modèles CNN.

Tableau 4. Comparaison de l'exactitude, de la précision, du rappel et du score F1 des méthodes proposées.

Modèle	Précision	Précision	Rappel	Score F1
Régression logistique	0,40	0,39	0,38	0,38
SVM	0,45	0,44	0,43	0,43
k-NN	0,42	0,40	0,39	0,39
CNN modèle 1	0,62	0,60	0,59	0,59
CNN modèle 2	0,64	0,62	0,62	0,61

Nous avons amélioré les résultats de notre modèle, atteignant désormais une précision de 64 %. Le problème réside dans l'ensemble de données, nous avons donc créé une méthodologie et prévoyons d'utiliser l'ensemble de données AffectNet, qui est plus efficace lorsque nous disposons de ressources supplémentaires. Nous avons expérimenté l'apprentissage par transfert techniques telles que VGG16, ResNet et EfficientNet, mais les résultats étaient d'environ 40 %, et le temps d'apprentissage sur la machine locale était trop long.

Pour contextualiser les performances de nos modèles CNN, nous les avons comparés aux modèles de pointe rapportés dans des études récentes. EmoNeXt, un modèle ConvNeXt adapté, et le réseau de masquage résiduel (ResMaskNet) ont été sélectionnés pour comparaison en raison de leur grande précision sur l'ensemble de données FER-2013. La précision comparative est résumée dans le tableau 5.

Machine Translated by Google

Tableau 5. Une comparaison des précisions des méthodes proposées.

Modèle	Précision
Réseau de masquage résiduel	0,74
ÉmoNeXt	0,71
CNN modèle 1	0,62
CNN modèle 2	0,64

L'analyse comparative met en évidence le fait que, bien que les modèles CNN proposés soient meilleurs que les simples modèles de base pour toutes les mesures de performance, ils sont toujours en deçà de ResMaskNet et EmoNeXt, qui sont des modèles de pointe plus avancés. Les deux derniers modèles ont des architectures et des méthodes plus compliquées ; c'est pourquoi leurs performances sont meilleures que les autres modèles. En particulier, ResMaskNet et EmoNeXt utilisent des techniques d'extraction de caractéristiques de haut niveau ainsi que des techniques de masquage, qui contribuent grandement à accroître la précision de la détection des émotions à partir des expressions faciales.

Une comparaison avec le modèle de référence et le modèle le plus récent a été effectuée pour établir l'efficacité relative du modèle proposé grâce à une évaluation complète. Bien que les modèles CNN proposés surpassent les modèles de base plus simples, ils ne répondent pas aux techniques de pointe en raison d'un manque d'optimisation. Cependant, cette comparaison détaillée répond à la demande de l'examinateur et offre un aperçu des domaines potentiels qui pourraient être encore améliorés dans le développement de systèmes de reconnaissance des émotions faciales.

6.2. Limites et biais potentiels

Cette étude souffre des limitations suivantes qui peuvent affecter la façon dont les résultats sont interprétés : •

Limitations de l'ensemble de données : L'ensemble de données FER-2013 est diversifié. Cependant, il ne rend pas compte de toute la gamme des émotions humaines et ne tient pas compte du contexte qui peut être présent dans un contexte clinique réel.

• Facteurs environnementaux : Les modèles n'ont pas été testés dans des conditions d'éclairage ou d'arrière-plan différentes. Cela pourrait affecter leurs performances dans des applications pratiques. •

Généralisation : La capacité des modèles à généraliser à différents groupes démographiques n'a pas été évaluée, ce qui est d'une importance cruciale pour un outil conçu pour un usage clinique.

6.3. Implications cliniques et considérations éthiques

Les résultats de cette recherche démontrent que l'intégration d'un modèle CNN de reconnaissance émotionnelle dans le domaine clinique permettra d'améliorer le diagnostic des maladies psychosomatiques. L'outil fournira des informations quantitatives supplémentaires sur l'état émotionnel des patients, ce qui pourra fournir aux cliniciens des informations qui pourraient être manquées ou mal interprétées.

Les améliorations suivantes peuvent être apportées afin d'améliorer l'outil proposé sur la base des résultats de cette étude : • Augmenter

la taille et la variabilité de l'ensemble de données de formation en collectant un plus large éventail d'expressions et de situations émotionnelles, ainsi qu'en incluant davantage d'instances pour affiner- ajuster et augmenter la précision du modèle. • Validation dans des

contextes réels : il est important de mener des expériences dans des contextes cliniques réels impliquant différentes données démographiques afin que les modèles appris puissent mieux s'adapter aux nuances du monde réel. • Pour améliorer la qualité

des informations fournies par les modèles, il est important de combiner leurs résultats avec les données supplémentaires pouvant être obtenues auprès des patients, telles que les antécédents médicaux et les informations verbales.

En conclusion, cette discussion reconnaît les limites potentielles et actuelles des modèles CNN. Il présente également des pistes concrètes pour les recherches futures visant à améliorer les modèles.

Machine Translated by Google

efficacité et utilité en milieu clinique, favorisant l'intégration des outils d'IA dans les soins de santé vers des soins aux patients plus empathiques et plus informés.

7. Conclusions

Le but de cette étude était d'explorer le potentiel de l'utilisation de modèles de réseaux neuronaux convolutifs (CNN) pour identifier les maladies psychosomatiques sur la base d'expressions émotionnelles avec des données basées sur des signaux faciaux. Cette étude a révélé que les modèles CNN étaient capables de détecter sept émotions différentes avec un niveau de précision légèrement supérieur à 59 % avec de bons taux de reconnaissance du bonheur et de la surprise.

7.1. Résumé des principales conclusions

Les principales conclusions de cette étude sont les suivantes : • Certains états émotionnels peuvent être identifiés de manière cohérente par les modèles CNN, ce qui pourrait servir de proxy pour les maladies psychosomatiques sous-jacentes. • Les émotions « heureux » et « surprise » ont été reconnues avec une grande précision : ces deux émotions jouent un rôle important dans la compréhension du bien-être du patient. • Sur la base de matrices de confusion et de mesures de performance, les domaines dans lesquels le modèle pourrait être amélioré ont été établis. L'un de ces domaines consiste à reconnaître les émotions sous- représentées dans l'ensemble de données de formation.

Les résultats ont des implications dans la pratique clinique et suggèrent que l'intelligence artificielle pourrait aider les professionnels de la santé en tant que source supplémentaire d' informations diagnostiques.

7.2. Recommandations pour les recherches futures

Les domaines suivants méritent d'être étudiés ultérieurement pour améliorer l'outil CNN : • Élargir et diversifier l'ensemble de données de formation pour avoir une plus grande couverture sur les expressions émotionnelles et la variabilité démographique. • Réaliser des études de validation clinique qui testent les modèles en situation réelle pour affiner leurs capacités prédictives. • Combiner des données multimodales – des signaux physiologiques et audio – dans le but de créer un système plus complet de reconnaissance des émotions.

Pour améliorer la précision de la reconnaissance des émotions et la résilience aux diverses distorsions des données, les recherches à venir doivent prendre en considération les éléments suivants : • Apprentissage par renforcement [32] : l'apprentissage par renforcement favorise la généralisation en fournissant un retour d'information en temps réel et en permettant un apprentissage par essais et erreurs, ce qui est essentiel pour obtenir des performances optimales. • Méthodes d'ensemble [33] : les méthodes d'ensemble, y compris le bagging, le boosting et l'empilement , aident à combiner des modèles pour améliorer les performances. • GAN [34] : les GAN génèrent des données synthétiques pour augmenter les ensembles de données de formation, ce qui est un élément important dans l'amélioration de la qualité globale d'un modèle. • RNN [35] : les RNN sont utilisés pour analyser les modèles temporels, contribuant ainsi à une meilleure compréhension.

états émotionnels permanents. • Modèles hybrides : la fusion de CNN avec des SVM ou des arbres de décision permet de gérer différents types de données, notamment en termes de complexité. • Recherche d'architecture neuronale (NAS) [36] : La recherche d'architecture neuronale implique l'automatisation du processus de découverte pour atteindre une structure de réseau neuronal optimale qui serait la plus appropriée en fonction de la nature de l'ensemble de données disponible.

L'utilisation de ces méthodes peut considérablement améliorer les modèles de reconnaissance des émotions et conduire à la création d'un bien meilleur outil de diagnostic clinique.

Les implications éthiques et l'acceptabilité des patients pourraient également faire partie de la portée des futures études sur les outils d'IA appliqués au diagnostic clinique. Cela garantira que ces technologies seront utilisées pour augmenter le jugement humain et non pour le remplacer.

Machine Translated by Google

L'utilisation des modèles CNN dans le diagnostic des maladies psychosomatiques contribue au pas important à franchir au carrefour entre technologie et santé. D'autres avancées technologiques, telles que l'outil de reconnaissance des émotions conçu dans le cadre de cette recherche, ont une valeur potentielle encore plus grande, car elles peuvent favoriser une précision accrue des diagnostics et une efficacité accrue du traitement, conduisant ainsi à des résultats favorables pour les patients et à une compréhension approfondie de la façon dont les états émotionnels sont liés à la santé physique (qui fournira des informations sur leur co-dépendance). Cependant, nous devons veiller à ce que ces outils soient développés dans le respect de l'éthique, de la confidentialité des patients et, ne l'oublions pas, des mécanismes susceptibles de renforcer le lien patient-médecin.

Contributions des auteurs : Conceptualisation, MN et AR ; Méthodologie, YM ; Validation, MN et YM ; Enquête, Minnesota ; Rédaction – ébauche originale, MN ; Rédaction : révision et édition, AR et YM ; Supervision, AR Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche n'a reçu aucun financement externe.

Déclaration de disponibilité des données : les ensembles de données présentés dans cet article ne sont pas facilement disponibles car ils font partie d'une étude de thèse de doctorat en cours. Les demandes d'accès aux ensembles de données doivent être adressées à Marta Narigina.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Références

1. Hong, AJ; DiStefano, D. ; Dua, S. Les CNN peuvent-ils classer avec précision les émotions humaines ? Une étude comparative. arXiv 2023, arXiv :2310.09473. Disponible en ligne : https://arxiv.org/pdf/2310.09473 (consulté le 1er juin 2024).
2. Romanovs, A. ; Sultanovs, E. ; Buss, E. ; Merkouriev, Y. ; Majore, G. Défis et solutions pour des services de télémédecine résilients. Dans les actes du 8e atelier de l'IEEE 2020 sur les progrès en matière d'ingénierie de l'information, électronique et électrique (AIEEE), Vilnius, Lituanie, 22-24 avril 2021 ; p. 1 à 7. [Référence croisée]
3. Navakauskiene, R. ; Treigyte, G. ; Borutinskaite, V.-V. ; Matuzevicius, D. ; Navakauskas, D. ; Magnusson, K.-E. Alpha-dystrobrevine et ses protéines associées dans les cellules de leucémie promyélocytaire humaine induites à l'apoptose. J. Protéom. 2012, 75, 3291-3303. [Référence croisée] [Pub Med]
4. Narigina, M. ; Kempelis, A. ; Romanovs, A. Prévision basée sur l'apprentissage automatique des données de capteurs pour une détection environnementale améliorée. WSEAS Trans. Système. 2023, 22, 543-555. [Référence croisée]
5. Valiulière, G. ; Treigyte, G. ; Savickiene, J. ; Matuzevicius, D. ; Alksne, M. ; Jarasiene-Burinskaja, R. ; Bukelskiene, V. ; Navakauskas, D. ; Navakauskiene, R. Modèles de modifications des histones dans les tissus et les tumeurs du modèle de xénogreffe de leucémie promyélocytaire aiguë en réponse à une thérapie épigénétique combinée. Bioméde. Pharmacother. 2016, 79, 62-70. [Référence croisée] [Pub Med]
6. Narigina, M. ; Osadcijs, E. ; Romanovs, A. Analyse des technologies de traitement des données médicales. Dans Actes de la 63e Conférence scientifique internationale 2022 sur les technologies de l'information et les sciences de gestion de l'Université technique de Riga (ITMS), Riga, Lettonie, 6 et 7 octobre 2022 ; p. 1 à 6. [Référence croisée]
7. Sultanovs, E. ; Strebko, J. ; Romanovs, A. ; Lektauers, A. Les technologies de l'information dans le mécanisme de contrôle des processus médicaux. Dans Actes de la 61e Conférence scientifique internationale 2020 sur les technologies de l'information et les sciences de gestion de l'Université technique de Riga (ITMS), Riga, Lettonie, 15-16 octobre 2020 ; p. 1 à 5. [Référence croisée]
8. Wei, Z. Une nouvelle méthode de reconnaissance des expressions faciales pour les applications en temps réel. Dans la série de conférences IOP : Science et ingénierie des matériaux ; Éditions IOP : Bristol, Royaume-Uni, 2019 ; Tome 612, p. 062048. Disponible en ligne : https://iopscience.iop.org/article/10.1 088/1757-899X/612/5/052048/pdf (consulté le 1er juin 2024).
9. Qu, D. ; Zheng, Y. ; Li, H. ; Wang, J. ; Chen, X. Reconnaissance des émotions faciales à l'aide de CNN dans PyTorch. arXiv 2023, arXiv :2312.10818. Disponible en ligne : https://arxiv.org/pdf/2312.10818 (consulté le 22 juillet 2024).
10. Verkruysse, W. ; Svaasand, LO; Nelson, JS Imagerie photopléthysmographique à distance utilisant la lumière ambiante. Opter. Exprimer 2008, 16, 21434-21445. [Référence croisée] [Pub Med]
11. Georgescu, M.-I. ; Ionescu, RT ; Popescu, M. Apprentissage local avec des fonctionnalités approfondies et artisanales pour la reconnaissance des expressions faciales. Accès IEEE 2019, 7, 64827-64836. [Référence croisée]
12. Pecoraro, R. ; Basile, V. ; Bono, V. Auto-attention locale à canaux multi-têtes pour la reconnaissance des expressions faciales. Informations 2022, 13, 419. [Référence croisée]
13. Fard, AP; Mahoor, MH Ad-Corre : Perte adaptative basée sur la corrélation pour la reconnaissance des expressions faciales dans la nature. Accès IEEE 2022, 10, 26756-26768. [Référence croisée]
14. Vignesh, S. ; Savithadevi, M. ; Sridevi, M. ; Sridhar, R. Un nouveau modèle de reconnaissance des émotions faciales utilisant la segmentation VGG-19 Architecture. Int. J. Inf. Technologie. 2023, 15, 1777-1787. [Référence croisée]
15. Mukhopadhyay, M. ; Dey, A. ; Kahali, S. Une méthode de reconnaissance des expressions faciales basée sur l'apprentissage profond utilisant des caractéristiques texturales. Calcul neuronal. Appl. 2023, 35, 6499-6514. [Référence croisée]

Machine Translated by Google

16. Shahzad, T. ; Iqbal, K. ; Khan, MA; Iqbal, N. Rôle du zonage dans l'expression faciale grâce à l'apprentissage profond. Accès IEEE 2023, 11, 16493-16508. [Référence croisée]

17. El Boudouri, Y. ; Bohi, A. EmoNeXt : Un ConvNeXt adapté pour la reconnaissance des émotions faciales. Dans les actes du 25e IEEE 2023 Atelier international sur le traitement du signal multimédia (MMSP), Poitiers, France, 27-29 septembre 2023. [CrossRef]

18. Goodfellow, IJ; Erhan, D. ; Transporteur, PL ; Courville, A. ; Mirza, M. ; Hamner, B. ; Cukierski, W. ; Tang, Y. ; Thaler, D. ; Lee, D.-H. Défis de l'apprentissage des représentations : un rapport sur trois concours d'apprentissage automatique. Dans Actes du traitement de l'information neuronale : 20e conférence internationale, ICONIP 2013, Daegu, République de Corée, 3-7 novembre 2013 ; Actes, partie III. Springer : Berlin/Heidelberg, Allemagne, 2013 ; pp. 117-124. [Référence croisée]

19. Sambare, M. FER 2013 : Ensemble de données de reconnaissance des expressions faciales [Ensemble de données]. Kaggle. Disponible en ligne : https://www.kaggle.com/ ensembles de données/msambare/fer2013/data (consulté le 1er juin 2024).

20. Chand, HV ; Chrisanthus, A. ; Thampi, AK Une revue de diverses approches basées sur CNN pour la reconnaissance des émotions faciales. Dans Actes du Congrès international sur les technologies de l'information et de la communication, Lalitpur, Népal, 26-28 avril 2023 ; Disponible en ligne : https://ieeexplore.ieee.org/document/10133947 (consulté le 1er juin 2024). 21. ¸Sen, SY; Özkurt, N. Réglage des hyperparamètres du réseau neuronal convolutif avec Adam Optimizer pour la classification ECG. Dans les actes de la conférence 2020 sur les innovations dans les systèmes et applications intelligents (ASYU), Istanbul, Turquie, 15-17 octobre 2020 ; p. 1 à 6. [Référence croisée]

22. Özdemir, MA ; Élagoz, B. ; Alkan, A. Reconnaissance des émotions en temps réel à partir des expressions faciales à l'aide de CNN. Dans Actes du Congrès 2019 sur les technologies médicales (TIPTEKNO), Izmir, Turquie, 3-5 octobre 2019 ; IEEE : Piscataway, NJ, États-Unis, 2019. Disponible en ligne : https://ieeexplore.ieee.org/document/8895215 (consulté le 1er juin 2024).

23. Lui, K. ; Zhang, X. ; Ren, S. ; Sun, J. Évaluation empirique des activations rectifiées dans un réseau de convolution. arXiv 2015. [CrossRef]

24. Chowdhury, S. ; Chowdhury, S. ; Cinquante, JT ; Khan, R. Détection et classification des véhicules à l'aide de réseaux de neurones profonds. Dans Actes de la Conférence internationale 2022 sur l'électricité et les technologies de l'information (IEIT), Malang, Indonésie, 15-16 septembre 2022 ; pp. 95-100. [Référence croisée]

25. Wang, Y. ; Bian, Z.-P. ; Hou, J. ; Chau, L.-P. Réseaux de neurones convolutifs avec régularisation dynamique. IEEETrans. Réseau neuronal. Apprendre. Système. 2021, 32, 2299-2304. [Référence croisée]

26. Li, X. ; Chen, J. ; Zhao, G. ; Pietikäinen, M. Mesure de la fréquence cardiaque à distance à partir de vidéos faciales dans des situations réalistes. Dans Actes de la conférence IEEE 2014 sur la vision par ordinateur et la reconnaissance de formes, Columbus, OH, États-Unis, 23-28 juin 2014 ; pages 4264 à 4271. [Référence croisée]

27. Pradeep, V. ; Madhushree, B. ; Sumukha, BS; Richards, GR ; Prashant, SP Détection des émotions faciales à l'aide de CNN et OpenCV. Dans les actes de la Conférence internationale 2024 sur les technologies émergentes en informatique pour les applications interdisciplinaires (ICETCS), Bangalore, Inde, 25-26 avril 2024 ; p. 1 à 6. [Référence croisée]

28. Yang, Y. ; Liu, F. ; Li, M. ; Jin, J. ; Weber, E. ; Liu, Q. ; Crozier, S. IRM de détection compressée basée sur la transformée de Fourier pseudo-polaire. IEEETrans. Bioméde. Ing. 2017, 64, 816-825. [Référence croisée]

29. Tarassenko, L. ; Villarroel, M. ; Guazzi, A. ; Jorge, J. ; Clifton, DA; Pugh, C. Surveillance des signes vitaux par vidéo sans contact utilisant lumière ambiante et modèles auto-régressifs. Physiol. Mesures. 2014, 35, 807-831. [Référence croisée]

30. De Haan, G. ; Jeanne, V. Fréquence d'impulsion robuste du rPPG basé sur la chrominance. IEEETrans. Bioméde. Ing. 2013, 60, 2878-2886. [Référence croisée]

31. Poh, M.-Z. ; McDuff, DJ ; Picard, RW Mesures automatisées du pouls cardiaque sans contact utilisant l'imagerie vidéo et la séparation aveugle des sources. Opter. Express 2010, 18, 10762-10774. [Référence croisée]

32. Chanson, C. ; Chen, C. ; Li, Y. ; Wu, X. L'apprentissage par renforcement profond s'applique à la classification des données d'électromyographie. Dans les actes de la conférence internationale IEEE 2018 sur les systèmes cyborg et bioniques (CBS), Shenzhen, Chine, 25-27 octobre 2018 ; pp. 505-510. [Référence croisée]

33. Gang, Z. ; Jia, C. ; Guo, C. ; Li, P. ; Gao, J. ; Zhao, L. Prédiction de la maladie pulmonaire obstructive chronique basée sur un cadre d'apprentissage d'ensemble composite en plusieurs étapes. Dans Actes de la Conférence internationale de l'IEEE 2023 sur la bioinformatique et la biomédecine (BIBM), Istanbul, Turquie, 5-8 décembre 2023 ; pages 1921-1924. [Référence croisée]

34. Chang, J. ; Zhang, Z. ; Wang, Z. ; Li, J. ; Meng, L. ; Lin, P. EEG d'auditeur génératif pour la reconnaissance des émotions vocales à l'aide de réseaux contradictoires génératifs avec détection compressée. IEEE J. Biomed. Informer sur la santé. 2024, 28, 2025-2036. [Référence croisée] [Pub Med]

35. Le, MD; Singh Rathour, V. ; Truong, QS; Mai, Q. ; Brijesh, P. ; Le, N. Réseau neuronal convolutif récurrent multi-module avec encodeur de transformateur pour la classification des arythmies ECG. Dans les actes de la conférence internationale IEEE EMBS 2021 sur l'informatique biomédicale et de santé (BHI), Athènes, Grèce, 27-30 juillet 2021 ; p. 1 à 5. [Référence croisée]

36. Zhao, F. ; Nie, J. ; Maman, M. ; Chen, X. ; Lui, X. ; Wang, B. ; Hou, Y. Évaluation du rôle de différentes régions hétérogènes dans l'IRM-DCE pour prédire les sous-types moléculaires du cancer du sein sur la base de la recherche d'architecture de réseau et du transformateur de vision. Dans les actes de la 45e conférence internationale annuelle 2023 de l'IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australie, 24-27 juillet 2023 ; p. 1 à 4. [Référence croisée]

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Lester : Animation de rotoscope via un objet vidéo Segmentation et suivi

Ruben Tous

Département d'architecture informatique, Universitat Politecnica de Catalunya, 08034 Barcelone, Espagne ; ruben.tous@upc.edu

Résumé : Cet article présente Lester, une nouvelle méthode pour synthétiser automatiquement des animations 2D de style rétro à partir de vidéos. La méthode aborde le défi principalement comme un problème de segmentation et de suivi d'objets. Les images vidéo sont traitées avec le Segment Anything Model (SAM) et les masques résultants sont suivis dans les images suivantes avec DeAOT, une méthode de propagation hiérarchique pour la segmentation d'objets vidéo semi-supervisée. La géométrie des contours des masques est simplifiée grâce à l'algorithme de Douglas-Peucker. Enfin, des traits du visage, une pixellisation et un effet de lumière de base peuvent être ajoutés en option. Les résultats montrent que la méthode présente une excellente cohérence temporelle et peut traiter correctement des vidéos avec différentes poses et apparences, des plans dynamiques, des plans partiels et des arrière-plans divers. La méthode proposée fournit une approche plus simple et déterministe que les pipelines de traduction vidéo à vidéo basés sur des modèles de diffusion, qui souffrent de problèmes de cohérence temporelle et ne gèrent pas bien les sorties pixellisées et schématiques. La méthode est également plus réalisable que les techniques basées sur l'estimation de pose humaine en 3D, qui nécessitent des modèles 3D personnalisés fabriqués à la main et sont très limitées en ce qui concerne le type de scènes qu'elles peuvent traiter.

Mots-clés : animation ; rotoscopie; segmentation; infographie; vidéo; intelligence artificielle

check for updates

Citation : Tous, R. Lester : Animation rotoscope via la segmentation et le suivi d'objets vidéo. Algorithmes 2024, 17, 330. https://doi.org/10.3390/a17080330

Rédacteur académique : Arslan Munir

Reçu : 24 juin 2024
Révisé : 23 juillet 2024
Accepté : 25 juillet 2024
Publié : 30 juillet 2024

Droit d'auteur : © 2024 par l'auteur. Licencié MDPI, Bâle, Suisse.

Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Dans des secteurs tels que le cinéma ou les jeux vidéo, la rotoscopie est utilisée pour exploiter les informations des séquences d'action en direct pour la création d'animations 2D. Cette technique a contribué à l'émergence d'une esthétique visuelle unique dans certains jeux vidéo emblématiques des années 80 et 90, comme « Prince of Persia » (1989), « Another World » (1991) ou « Flashback » (1992), et elle continue d'être utilisé dans des titres modernes comme « Lunark » (2023). Cependant, comme dans le cas de Lunark, cette technique nécessite encore un travail manuel au niveau du châssis, ce qui la rend très coûteuse.

Les progrès récents de l'intelligence artificielle (IA) ont ouvert la possibilité d'automatiser ce processus dans une certaine mesure. Les derniers progrès en matière de modèles génératifs conditionnels pour la traduction vidéo vers vidéo offrent une approche prometteuse pour résoudre ce problème. Cependant, ces méthodes ne sont pas encore capables de fournir systématiquement un résultat soigné et professionnel avec un style uniforme et une cohérence spatiale et temporelle. Leurs performances varient considérablement en fonction des différents aspects de la vidéo d'entrée, tels que la scène et la résolution. De plus , ils sont lents et nécessitent d'énormes ressources de calcul. Par conséquent, la nécessité d' une solution automatisée efficace et déterministe reste un problème ouvert.

Cet article décrit une nouvelle méthode, appelée Lester, pour automatiser ce processus. Un utilisateur saisit une vidéo et une palette de couleurs et le système génère automatiquement une animation 2D de style rétro dans un format directement utilisable dans un jeu vidéo ou dans un film d'animation (voir Figure 1 pour un exemple).

La méthode aborde le défi principalement comme un problème de segmentation d'objets. Étant donné un cadre représentant une personne, un masque pour chaque trait visuel différent (cheveux, peau, vêtements, etc.) est généré avec le Segment Anything Model (SAM) [1]. Pour identifier de manière univoque les masques

Machine Translated by Google

dans toutes les images, un algorithme de suivi d'objet vidéo est utilisé en utilisant l'approche de [2]. Les masques sont purgés pour résoudre les erreurs de segmentation et leur géométrie est simplifiée avec l'algorithme de Douglas-Peucker [3] pour obtenir le style visuel souhaité. Enfin, des traits du visage, une pixellisation et un effet de lumière de base peuvent être ajoutés en option.





Figure 1. Exemple d'image d'animation synthétisée (à droite) et son image d'origine associée (à gauche, le visage a été anonymisé) à partir d'un exemple de vidéo d'entrée.

2. Travaux connexes

L'invention du rotoscope par Max Fleischer en 1915 a transformé la capacité de créer des dessins animés en utilisant des images fixes ou animées. Des variantes de cette technique ont été utilisées jusqu'à ce jour dans des films d'animation (par exemple « A Scanner Darkly » (2006)), des séries télévisées d'animation (par exemple « Undone » (2019)) et des jeux vidéo (par exemple « Prince of Persia »). » (1989), « Un autre monde » (1991), « Flashback » (1992), « Lunark » (2023)). Les productions contemporaines intègrent des algorithmes de traitement d'images pour faciliter la tâche, mais elles reposent principalement sur un travail manuel coûteux. Parvenir à une automatisation complète avec un résultat professionnel soigné (style uniforme, cohérence spatiale et temporelle) reste actuellement irréalisable. Cependant, les progrès récents en matière d'intelligence artificielle (IA) ouvrent la possibilité de tenter de progresser vers une automatisation c

À notre connaissance, il existe deux stratégies principales pour résoudre ce problème. D'une part, le défi peut être relevé grâce aux modèles génératifs conditionnels, répandus aujourd'hui dans les tâches de traduction d'images. Les premières méthodes (par exemple, CartoonGAN [4], AnimeGAN [5] ou les travaux de [6]) utilisaient des réseaux contradictoires génératifs conditionnels (GAN) pour transformer des photos de scènes du monde réel en images de style dessin animé. Ces méthodes ont récemment été surpassées par celles basées sur des modèles de diffusion tels que Stable Diffusio Bien que la plupart de ces méthodes aient été conçues pour la génération de texte en image, elles ont été adaptées avec succès pour la traduction d'image en image en utilisant des techniques telles que SDEdit [9] ou ControlNet [10].

Des méthodes pures de traduction image à image avec des modèles de diffusion existent également (par exemple, SR3 [11] et Palette [12]), mais avec une configuration supervisée qui ne peut être appliquée qu'à des situations dans lesquelles une vérité fondamentale de paires d'images (originale, cible) comme la super-résolution ou la colorisation . Quoi qu'il en soit, même si nous disposons d'un bon modèle génératif conditionnel, il reste le défi de l'appliquer aux images d'une vidéo tout en préservant la cohérence temporelle. Les dernières tentatives (par exemple, [13]) combinent des astuces (par exemple, entrées en mosaïque) pour uniformiser la sortie du modèle génératif avec des techniques d'interpolation de style vidéo-guidée (par exemple, EbSynth [14]). Les résultats de ces méthodes se sont considérablement améliorés et, pour certains styles, les résultats sont si bons qu'ils sont déjà utilisés dans des productions indépendantes. Cependant, pour la plupart des styles d'art numérique, ces méthodes imposent un compromis indésirable entre la fidélité du style et la cohérence temporelle. Parmi les derniers travaux de ce type, Sora [15] d'OpenAI est particulièrement remarquable. L'analyse des échantillons fournis révèle que Sora excelle dans la compréhension du monde physique et possède les éléments fondamentaux d'un modèle mondial, ayant le potentiel d'évoluer vers un moteur physique dans le futur [16].

D'un autre côté, une approche complètement différente consiste à appliquer un algorithme d'estimation de pose humaine 3D pour suivre la pose 3D à travers les différentes images. La pose 3D est ensuite appliquée à un modèle de personnage 3D avec l'aspect souhaité (cheveux, vêtements, etc.). Généralement, ce modèle de personnage 3D est créé à la main avec un logiciel spécialisé. Enfin, un shader et une configuration de rendu appropriés sont configurés pour obtenir le style visuel souhaité. C'est l' approche adoptée dans [17–21]. Cette approche présente de nombreux inconvénients. Premièrement, cela dépend de l' état de l'art de l'estimation de la pose humaine en 3D, qui a beaucoup progressé mais échoue encore dans de nombre

Machine Translated by Google

situations (faible résolution, troncatures, poses inhabituelles) principalement dues à la difficulté d' obtenir suffisamment de données de terrain. Deuxièmement, cela nécessite la conception d'un modèle 3D approprié pour chaque personnage différent. Troisièmement, la pose 3D est basée sur un modèle donné, généralement SMPL. Pour l'appliquer à un modèle différent, cela nécessite une étape de reciblage de pose. Les rotations des articulations peuvent être facilement reciblées, mais les formes de mélange SMPL ne le peuvent pas, ce qui entraîne des déformations corporelles irréalistes. C'est pourquoi les démonstrations de ce type de travaux sont généralement réalisées avec des robots, car ceux-ci ne comportent pas de pièces déformables. Un avantage de cette approche par rapport aux modèles génératifs est qu'elle fournit une bonne cohérence temporelle, car les incohérences de l'étape d'estimation de la pose humaine 3D peuvent être correctement résolues par interpolation, comme cela est fait dans [20].

Dans ce travail, nous proposons une troisième approche, principalement basée sur la segmentation et le tracking. Contrairement à la première stratégie mentionnée (modèles génératifs), l'approche proposée est plus prévisible et offre une bien meilleure cohérence temporelle. Contrairement à la deuxième stratégie évoquée (estimation de la pose humaine en 3D), l'approche proposée est plus robuste et beaucoup plus pratique puisqu'elle ne nécessite pas d'avoir un modèle 3D personnalisé pour chaque nouveau personnage.

Ce n'est pas la première fois que ce genre de problème est abordé de cette manière. Dans [22], la segmentation a été utilisée pour styliser un portrait vidéo à partir d'un exemple d'image artistique, obtenant de meilleurs résultats que le transfert de style neuronal [23], qui prévalait à cette époque.

3. Méthodologie 3.1. Aperçu

L'entrée principale de la méthode (le code est disponible sur https://github.com/rtous/lester-code (consulté le 20 juillet 2024) ; l'ensemble de données et les résultats des expériences sont accessibles au public sur https://github.com/rtous/lester (consulté le 20 juillet 2024)) est une séquence vidéo cible T d'une performance humaine. Il n'y a pas de contraintes générales sur les caractéristiques de la personne représentée (sexe, âge, vêtements, etc.), la pose, l'orientation, les mouvements de caméra, l'arrière-plan, l'éclairage, les occlusions, les troncatures (prises de vue partielles du corps) ou la résolution.

Cependant, certains scénarios sont connus pour avoir un impact négatif sur les résultats (ce point sera abordé dans la section 4). Bien que la méthode soit principalement conçue pour les vidéos mettant en vedette une seule personne, son application théorique avec plusieurs personnes sera abordée dans la section 4. Lorsqu'elle fonctionne sans entrées supplémentaires, la méthode utilise par défaut des paramètres spécifiques pour la segmentation, les couleurs et d'autres attributs. Cependant, il est prévu que l' utilisateur préfère spécifier des configurations personnalisées, nécessitant des entrées supplémentaires telles que des invites de segmentation, des palettes de couleurs, etc., comme détaillé dans les sous-sections correspondantes ci-dessous. Le résultat de la méthode est un ensemble d'images PNG montrant le personnage animé résultant et sans arrière-plan (où le canal alpha est défini sur zéro).

La méthode comprend trois étapes de traitement indépendantes. Le principal est la segmentation et le suivi des différents traits visuels (cheveux, peau, vêtements, etc.) de la personne présentée dans les images vidéo. La deuxième étape se concentre sur la correction des erreurs de segmentation en purgeant les masques et en simplifiant leurs contours polygonaux. La troisième et dernière étape consiste à appliquer des détails de finition personnalisables, notamment les couleurs, l'éclairage des fausses jantes, les traits du visage et la pixellisation. La figure 2 montre le flux de travail global de la méthode. Chaque composant est développé dans les sous-sections suivantes.

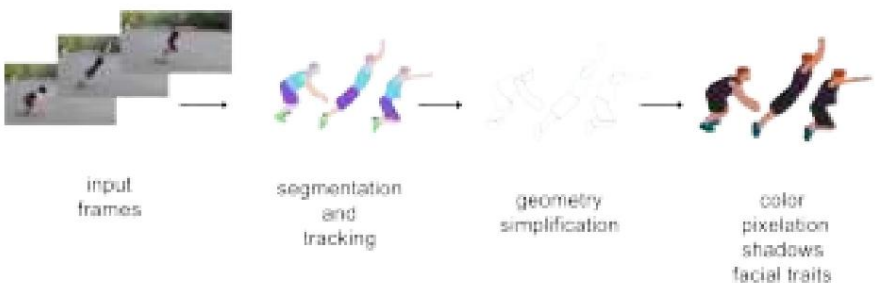


Figure 2. Aperçu du flux de travail du système (entrées à gauche, sortie à droite).

Machine Translated by Google

3.2. Segmentation et suivi

Afin de segmenter les différents traits visuels du personnage, le Segment Anything Model (SAM) [1] est utilisé. SAM est un modèle de segmentation rapide qui devient un modèle de base pour résoudre une variété de tâches. Étant donné une image d'entrée avec trois canaux (hauteur × largeur ×3) et une invite, qui peut inclure des informations spatiales ou textuelles, il génère un masque (hauteur ×poids ×3), avec chaque pixel attribué une couleur RVB. Chaque couleur représente une étiquette d'objet de l'espace d'étiquette d'objet L = l1, l2, . . . , déduit de l'invite, correspondant au type d'objet prédit à l'emplacement de pixel correspondant dans l'image d'entrée. L'invite peut être n'importe quelle information indiquant ce qu'il faut segmenter dans une image (par exemple, un ensemble de points de premier plan/arrière-plan, une boîte ou un masque approximatif ou un texte de forme libre). SAM se compose d'un encodeur d'image (basé sur MAE [24], un Vision Transformer pré-entraîné [25]) qui calcule une intégration d'image, d'un encodeur d'invite (étant le texte de forme libre basé sur CLIP [26]) qui intègre des invites, puis les deux sources d'informations sont combinées dans un décodeur de masque (basé sur les modèles de segmentation Transformer [27]) qui prédit les masques de segmentation.

L'application de SAM au niveau de l'image entraînerait des étiquettes différentes pour le même objet dans différentes images. SAM ne part pas d'une liste fermée d'étiquettes d'objets mais crée et attribue plutôt des étiquettes pendant l'exécution. Par conséquent, il n'y a aucune garantie que les mêmes étiquettes soient attribuées au même objet dans chaque image. Pour le faire de manière cohérente, nous utilisons SAM-Track [2]. Les résultats SAM pour la première image sont propagés à la deuxième image, avec les mêmes étiquettes, avec DeAOT [28], une méthode de propagation hiérarchique pour la segmentation d'objets vidéo semi-supervisée. La propagation hiérarchique peut propager progressivement les informations des trames passées vers la trame actuelle et transférer la caractéristique de la trame actuelle d'agnostique à l'objet à spécifique à l'objet. SAM est appliqué aux images suivantes pour détecter de nouveaux objets, mais la segmentation des objets déjà vus est effectuée par DeAOT.

Nous avons choisi des invites basées sur des mots clés pour spécifier la segmentation souhaitée. Nous fournissons des invites par défaut pour des scénarios courants (par exemple, « cheveux », « peau », « chaussures », « t-shirt », « pantalon »), mais il est prévu que dans de nombreux scénarios, l'utilisateur devra spécifier une modification. - indigène. Ces invites ne sont pas traitées avec les capacités de texte libre de SAM en raison de leurs limitations, mais sont traitées avec Grounding-DINO [29], un détecteur d'objets ouvert , qui les traduit en boîtes d'objets qui sont envoyées à SAM sous forme de boîtes. invite. Un problème que nous avons identifié est que certaines combinaisons d'objets conduisent à des masques incorrects lorsqu'ils sont spécifiés dans la même invite. Pour résoudre ce problème, nous autorisons la spécification de plusieurs invites traitées indépendamment. La figure 3 montre le flux de travail de l' étape de segmentation et de suivi.

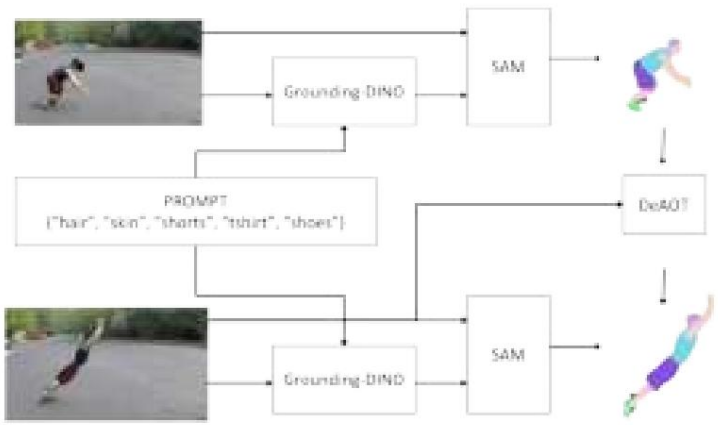


Figure 3. Workflow de l'étape de segmentation et de suivi pour les deux premières images (entrées à gauche, sortie à droite).

3.3. Simplification des contours

L'un des choix souhaités pour le style visuel cible est la possibilité d'afficher des géométries très simples, c'est-à-dire des contours de forme avec de courtes chaînes polygonales, afin d'imiter l'aspect visuel, par exemple, d'un « Autre Monde ». Pour y parvenir, une série d'étapes sont réalisées.

Machine Translated by Google

Pour chaque image, le résultat de l'étape de segmentation et de suivi est un masque avec chaque pixel attribué une couleur RVB représentant une étiquette d'objet. Le masque est d'abord divisé en sous-masques, un pour chaque couleur/étiquette. Chaque sous-masque est divisé, à son tour, en contours avec l'algorithme de [30]. Chaque contour individuel est représenté par un vecteur des coordonnées du sommet de les segments de ligne de démarcation. Le nombre de segments de ligne de chaque contour est réduit avec l'algorithme de Douglas – Peucker [3]. L'algorithme commence par joindre les extrémités du contour avec une ligne droite. Les sommets plus proches de la ligne qu'une valeur de tolérance donnée t sont supprimé. Ensuite, la ligne est divisée au point le plus éloigné de la ligne de référence, créant ainsi deux nouveaux lignes. L'algorithme fonctionne de manière récursive jusqu'à ce que tous les sommets situés dans la tolérance soient supprimés. La tolérance t est l'un des paramètres optionnels de la méthode proposée et permet à l'utilisateur contrôler le niveau de simplification. Comme chaque contour est simplifié indépendamment, il n'est pas souhaitable des espaces peuvent apparaître entre les géométries des contours limitrophes qui étaient auparavant parfaitement aligné. Pour éviter cela, et avant l'étape de simplification, tous les contours sont dilatés convolution du sous-masque avec un noyau 4 ×4. Il convient de mentionner que la simplification L'algorithme peut échouer pour certaines géométries de contour et valeurs de t. Dans ces situations (généralement très petits contours), le contour d'origine est conservé. Voir l'algorithme 1 et la figure 4.



(un) (b) (c) (d) (e)

Figure 4. Aperçu graphique de l'étape de simplification du contour d'une trame : (a) résultat de l' étape de segmentation et de suivi ; (b) détection des contours ; (c) dilatation des contours ; (d) simplification des contours ; (e) le cadre coloré résultant.

Algorithme 1 Algorithme de simplification des contours

Exiger : M une liste contenant des sous-masques, α une zone de contour minimale, t une valeur de tolérance pour simplification

Assurer : R une liste de contours où Ri,j est le contour simplifié j du sous-masque i

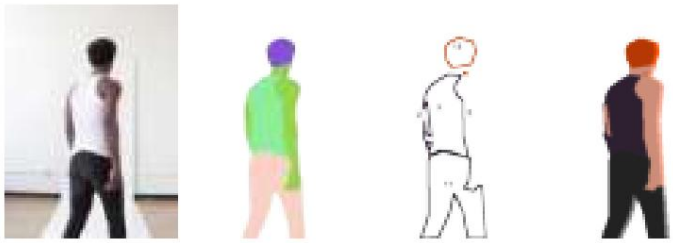
```
fonction SIMPLIFIER(M, α, t)
    pour je ← 0 à longueur(M) − 1 faire
        Ci ← getContours(Mi)
        pour j ← 0 à longueur(Ci) − 1 faire
            si aireO f(Ci,j) > α alors
                Di,j ← dilateContour(Ci,j)
                si contourCanBeSimpli fied(Di,j , t) alors
                    Ri,j ← simpli f yContour(Di,j , t)
                autre
                    Ri,j ← Di,j
                finir si
            finir si
        fin pour
fin pour
    retourner R
fin de fonction

contourCanBeSimpli fi ied renvoie false si l'algorithme de Douglas-Peucker échoue, comme cela peut être le cas.
se produire pour certaines géométries de contour et valeurs de t.
```

Machine Translated by Google

3.4. Détails de finition

La troisième et dernière étape de la méthode consiste à appliquer des détails de finition personnalisables, offrant aux utilisateurs une flexibilité dans la conception du style visuel du résultat. L'aspect le plus important est la couleur. La méthode génère un guide visuel avec des identifiants uniques pour chaque segment de couleur, comme le montre la figure 5. L'une des entrées de la méthode est une palette de couleurs, qui indique les valeurs RVB qui seront appliquées à chaque identifiant de segment.



(un) (b) (c) (d)

Figure 5. Processus de coloration : (a) trame d'entrée ; (b) la segmentation ; (c) un guide visuel avec des identifiants uniques pour chaque segment de couleur ; (d) image finale.

Un autre détail de finition, optionnel, est la possibilité d'ajouter quelques traits du visage d'aspect schématique. Les caractéristiques sont générées à partir des 68 repères faciaux pouvant être obtenus à l'aide du détecteur de visage frontal décrit dans [31], qui utilise un histogramme de gradients orientés (HOG) [32] combiné à un classificateur linéaire, une pyramide d'images et une fenêtre coulissante. schéma de détection. Une autre option est la possibilité d'ajouter un effet de lumière périphérique très simple en superposant la même image avec une luminosité différente et décalée horizontalement. Enfin, un effet de pixellisation est obtenu en redimensionnant l'image à l'aide d'une interpolation bilinéaire et en la redimensionnant avec l'interpolation du voisin le plus proche.
La figure 6 montre comment l'éclairage du contour, les traits du visage et la pixellisation sont progressivement appliqués à un exemple de monture.



(un) (b) (c) (d)

Figure 6. Détails de finition facultatifs : (a) Résultat de la méthode plate (b) Lumière de la jante (c) Caractéristiques du visage (d) Pixelation.

4. Expériences et résultats

Afin d'évaluer les performances de la méthode, nous avons collecté un ensemble de 25 courtes vidéos d'une durée de 5 à 20 s. Les vidéos présentent un sujet humain solitaire affichant différents styles vestimentaires et se livrant à diverses actions. Les vidéos ont été obtenues à partir de l'ensemble de données UCF101 Human Actions [33], de l'ensemble de données Fashion Video [34], de l'ensemble de données Kinetics [35] et de YouTube. L'ensemble de données assemblé comprend à la fois des vidéos basse et haute résolution, ainsi que des vidéos en milieu sauvage et contrôlées. L'ensemble de données et les résultats des expériences peuvent être trouvés dans [36]. Les figures 7 et 8 montrent quelques exemples de résultats.

Pour évaluer la qualité perceptuelle des animations générées, une étude utilisateur a été menée. Une enquête a été conçue comprenant des questions sur trois aspects visuels : la qualité de la forme, la cohérence des couleurs et la cohérence temporelle, ainsi que le score global. Les options de notation étaient basées sur l'évaluation de catégorie (ACR) décrite dans la recommandation ITU-T P.910 [37], sur une échelle allant de 1 = mauvais à 5 = excellent). L'évaluation subjective de la qualité des résultats est résumée dans le tableau 1. Nous rapportons le MOS (score d'opinion moyen décrit dans ITU-T P.800.1 [38]). Les vidéos sont regroupées en trois catégories, en basse résolution dans la nature

Machine Translated by Google

(ITW-LR), haute résolution en milieu sauvage (ITW-HR) et haute résolution à réglage contrôlé (C-HR).

Tableau 1. Valeurs MOS d'évaluation subjective de la qualité de l'ACR (1 = Mauvais à 5 = Excellent) de la forme qualité, cohérence des couleurs, cohérence temporelle et qualité globale. [un] ITW-LR →basse sauvage résolution; [b] ITW-HR →haute résolution sauvage ; [c] C-HR →réglage contrôlé haute résolution.

Sous-ensemble	#	Forme	Couleur	Temporel	Dans l'ensemble
ITW-LR [un]	5	3.20	3.24	3,84	3.36
ITW-RH [b]	14	3,81	4.03	4.33	3,99
C-HR [c]	6	4,70	4,60	4,67	4,67
tous		3,90	4.01	4.31	4.02

Points forts : Les résultats montrent que la méthode est capable de synthétiser des animations de qualité à partir de diverses vidéos d'entrée (un MOS moyen de 4,02 sur 5). La performance de manière significative améliore (4,67) sur les vidéos haute résolution prises dans des conditions contrôlées (vidéos 2, 11, 12, 13, 14 et 15). La méthode présente une très bonne cohérence temporelle et peut correctement traiter des vidéos avec différentes poses et apparences, des plans dynamiques (par exemple, vidéo 5), des vidéos avec des plans partiels (par exemple, le plan moyen en gros plan de la vidéo 7) et des arrière-plans variés.

Il convient de noter que, même si elle ne constitue pas l'objet principal de ce travail, la méthode peut traiter des vidéos contenant plusieurs personnes. Les objets associés à différents individus sont suivis indépendamment, chacun avec un identifiant unique, même si les objets sont du même type. Cela permet de mapper ces objets sur différentes couleurs, permettant ainsi à l'ensemble du pipeline d'être appliqué aux prises de vue à plusieurs personnes. La figure 9 montre un exemple de résultat d'une base multi-personnes.

Limites : les performances de segmentation et de suivi diminuent dans les vidéos avec basse résolution (toutes les vidéos de l'ensemble de données UCF101 Human Actions), arrière-plans complexes (vidéos 17 et 20), couleurs peu contrastées (vidéos 8, 10, 22 et 23), éclairage inapproprié (vidéos 14 et 25) et occlusions (vidéo 23). La figure 10 montre quatre exemples de cadres avec erreurs de quatre vidéos différentes (vidéos 22, 23, 25 et 8 de l'ensemble de données de test).

Coût de calcul : La méthode a été implémentée en utilisant le langage de programmation Python. Les expériences ont été réalisées sur un serveur haut de gamme doté d'un quadcore Intel i7-3820 à 3,6 GHz avec 64 Go de RAM DDR3 et 4 cartes GPU NVIDIA Tesla K40 avec 12 Go de GDDR5 chacun. L'utilisation maximale de la RAM par la méthode a été enregistrée à 3,1 Go. lors de l'exécution de la tâche de segmentation, ce qui explique également un pic de RAM GPU utilisation de 2,1 Go. Pour une seule image, l'étape 1 (segmentation et suivi) prend 0,56 s, l'étape 2 (simplification des contours) prend 0,43 s et l'étape 3 (finition des détails, principalement les traits du visage) prend 1,32 s. Au total, la méthode fonctionne à 0,43 FPS (1 FPS sans traits du visage).

Comparaison avec les modèles génératifs image à image : bien que nous n'ayons pas connaissance de D'autres méthodes qui ont tenté de résoudre le problème pour ce style visuel spécifique, une alternative pourrait impliquer l'utilisation de modèles génératifs conditionnels. Un pipeline potentiel, décrit dans [13], adapte Stable Diffusion [7] pour la traduction image à image à l'aide de ControlNet [10], et utilise des entrées en mosaïque ainsi qu'une interpolation de style vidéo-guidée (EbSynth [14]) pour maintenir la cohérence visuelle entre les images. Nous avons testé ce pipeline de traitement avec certains des vidéos de l'ensemble de données de test pour évaluer la faisabilité de cette approche. Pour atteindre le style souhaité, nous avons affiné la diffusion stable avec Dreambooth [39]. La figure 11 montre quatre exemples de trames obtenues avec cette approche, mais pour évaluer la cohérence temporelle, le les résultats peuvent être trouvés dans notre référentiel de résultats dans [36]. Les résultats révèlent plusieurs inconvénients par rapport à la méthode proposée. Premièrement, ils présentent une faible cohérence temporelle à travers cadres. Deuxièmement, la méthode nécessite de préciser le niveau de fidélité au style souhaité, où une fidélité plus élevée compromet la structure et les détails du cadre d'origine, conduisant à un compromis indésirable. Troisièmement, les performances varient considérablement en fonction des différents aspects de la vidéo d'entrée, tels que la scène et la résolution. Enfin, la méthode fonctionne nettement plus lent que la méthode proposée, avec une vitesse de traitement de seulement 0,3 FPS.

Machine Translated by Google

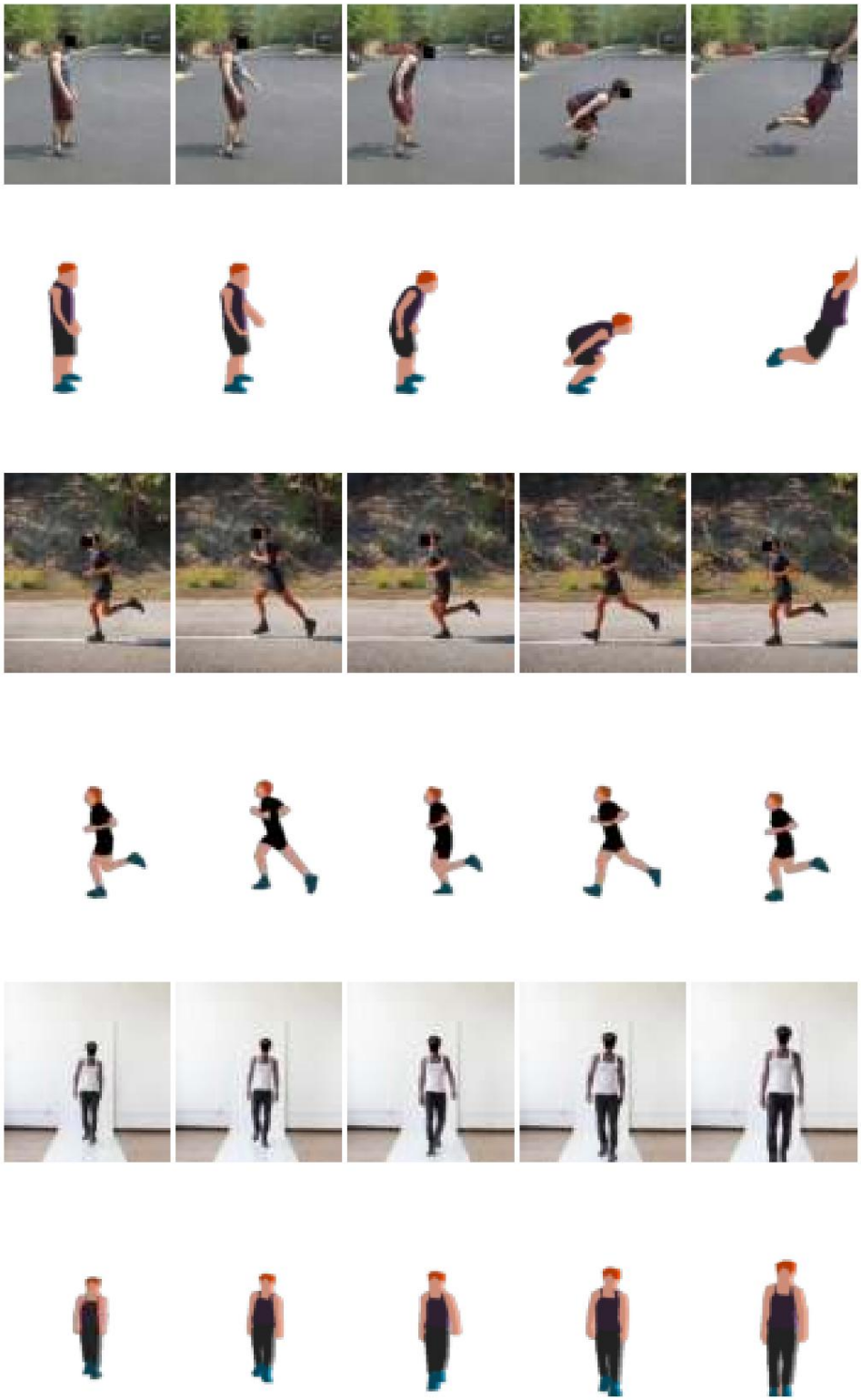


Figure 7. Exemples de résultats. Lignes impaires : entrées de vidéos (les visages ont été anonymisés). Lignes paires : résultats de la méthode.

Machine Translated by Google

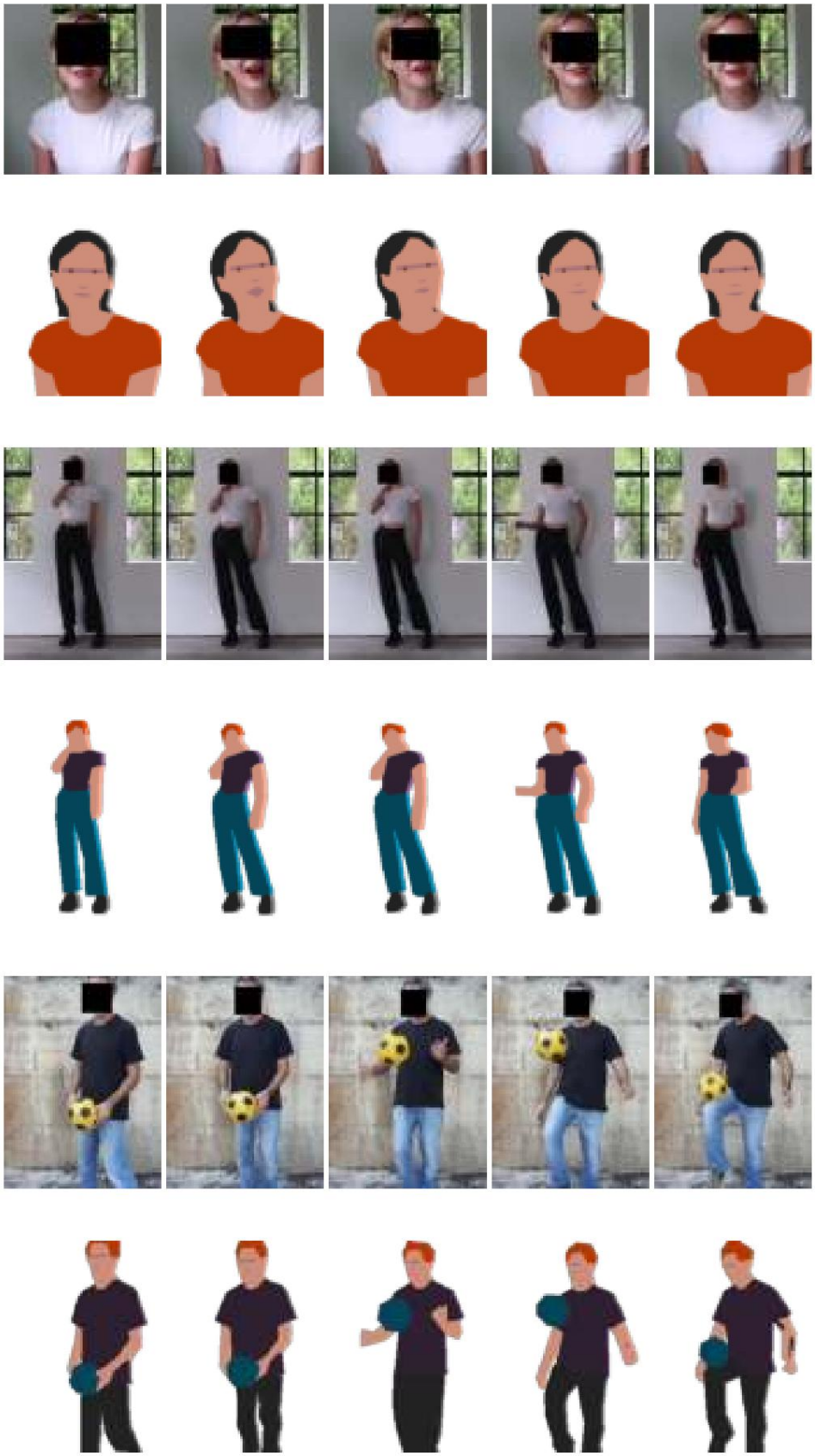


Figure 8. Exemples de résultats. Lignes impaires : entrées de vidéos (les visages ont été anonymisés). Lignes paires : résultats de la méthode.

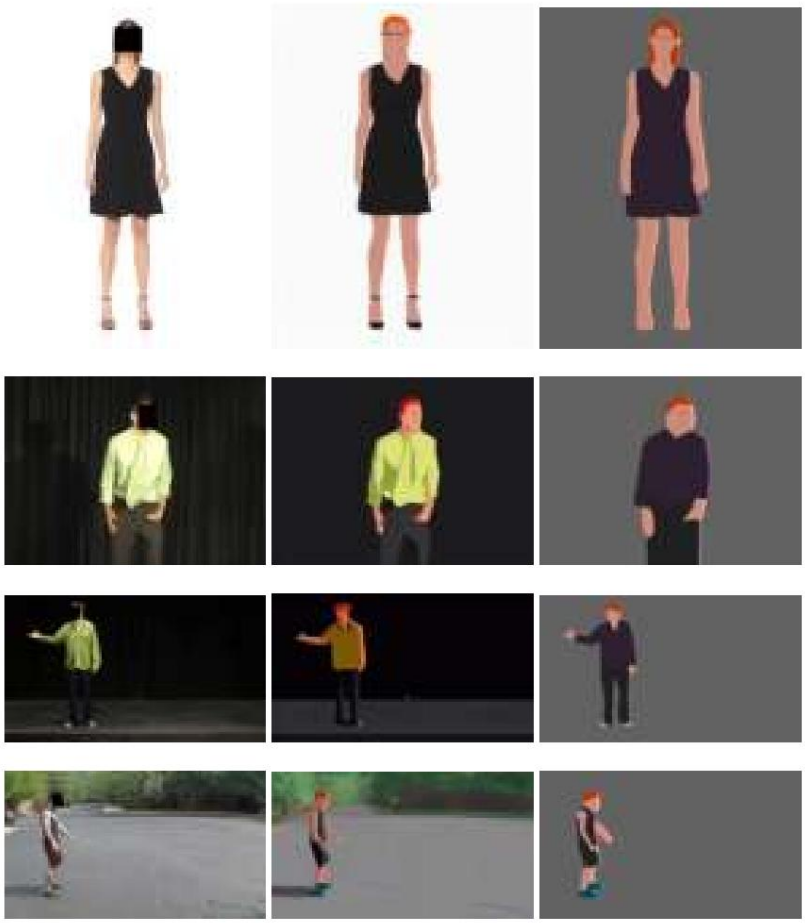
Machine Translated by Google



Figure 9. Exemple d'image d'animation synthétisée (à droite) et son image d'origine associée à partir d'un exemple saisir une vidéo à plusieurs personnes.



Figure 10. Exemples de cas de défaillance. Rangée supérieure : vidéos d'entrée (vidéos 22, 23, 25 et 8 de l'ensemble de données de test, les visages ont été anonymisés); rangée du bas : résultats.



(un) (b) (c)

Figure 11. Comparaison avec un pipeline génératif : (a) trame d'entrée ; (b) pipeline génératif décrit dans [13] ; (c) Lester (méthode proposée).

Machine Translated by Google

Discussion : Les résultats obtenus sont très satisfaisants et confirment que les dernières avancées en matière de segmentation et de suivi rendent possibles de nouvelles applications, également dans le domaine de la synthèse d'images. La méthode proposée s'est avérée être une méthode efficace pour générer des animations rotoscopiques à partir de vidéos. Les limitations détectées ne représentent pas un handicap puisqu'il est prévu que la méthode soit utilisée dans des vidéos sélectionnées ou créées expressément. Son coût de calcul non plus , puisque la méthode devrait être utilisée dans des tâches d'édition qui ne nécessitent pas d'exécution en temps réel. La méthode est principalement conçue pour les vidéos mettant en vedette une seule personne, mais pourrait être adaptée aux vidéos mettant en vedette plusieurs personnes en utilisant plusieurs paramétrages pour les invites de segmentation et les palettes de couleurs.

5. Conclusions

Dans ce travail, nous présentons Lester, une méthode de synthèse automatique de personnages animés 2D de style rétro à partir de vidéos. Plutôt que de relever le défi avec un modèle génératif image à image ou avec des outils d'estimation de la pose humaine, il est principalement abordé comme un problème de segmentation et de suivi d'objets. L'évaluation subjective de la qualité montre que l' approche est réalisable et que la méthode peut générer des animations de qualité à partir de vidéos avec différentes poses et apparences, des plans dynamiques, des plans partiels et des arrière-plans divers.
Bien que ce ne soit pas l'objectif, il convient de mentionner que les performances de segmentation et de suivi de la méthode sont moins cohérentes dans la nature, en particulier contre les vidéos basse résolution . Mais la méthode devrait être utilisée sur des vidéos sélectionnées ou réalisées sur mesure dans des environnements semi-contrôlés.

De par ses capacités, le procédé constitue un outil efficace qui réduit considérablement le coût de génération d'animations rotoscopiques pour les jeux vidéo ou les films d'animation. En raison de ses caractéristiques, il est plus simple et moins déterministe que les pipelines basés sur des modèles génératifs et plus pratique et flexible que les techniques basées sur l'estimation de la pose humaine.

Financement : Cette recherche est partiellement financée par le ministère espagnol de la Science et de l'Innovation sous le contrat PID2019-107255GB, et par le programme SGR 2021-SGR-00478 du gouvernement catalan.

Déclaration de disponibilité des données : l'ensemble de données et les résultats des expériences sont accessibles au public sur https://github.com/rtous/lester (consulté le 20 juillet 2024). Le code est disponible sur https://github. com/rtous/lester-code (consulté le 20 juillet 2024).

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Abréviations

Les abréviations suivantes sont utilisées dans ce manuscrit :

Cote de catégorie absolue ACR
Réseaux contradictoires génératifs du GAN
Histogramme HOG des dégradés orientés
Score d'opinion moyen MOS
Modèle SAM Segment Anything

Références

1. Kirillov, A. ; Mintun, E. ; Ravi, N. ; Mao, H. ; Rolland, C. ; Gustavson, L. ; Xiao, T. ; Whitehead, S. ; Berg, AC; Voila, WY ; et coll. Segmentez n'importe quoi. arXiv 2023, arXiv :2304.02643.
2. Cheng, Y. ; Li, L. ; Xu, Y. ; Li, X. ; Yang, Z. ; Wang, W. ; Yang, Y. Segmentez et suivez n'importe quoi. arXiv 2023, arXiv :2305.06558.
3. Douglas, DH ; Peucker, TK Algorithmes pour la réduction du nombre de points requis pour représenter une ligne numérisée ou sa caricature. Dans Classiques de la cartographie : réflexions sur les articles influents de Cartographica ; John Wiley & Sons : Hoboken, New Jersey, États-Unis, 2011 ; p. 15-28. [Référence croisée]
4. Chen, Y. ; Lai, Yukon ; Liu, YJ CartoonGAN : Réseaux contradictoires génératifs pour la caricature de photos. Dans les actes du CVPR. IEEE Computer Society, Salt Lake City, UT, États-Unis, 18-23 juin 2018 ; pages 9465 à 9474.
5. Chen, J. ; Liu, G. ; Chen, X. AnimeGAN : un nouveau GAN léger pour l'animation photo. Dans Algorithmes et applications d'intelligence artificielle, Actes du 11e Symposium international, ISICA 2019, Guangzhou, Chine, 16-17 novembre 2019 ; Springer : Berlin/Heidelberg, Allemagne, 2020 ; pp. 242-256. [Référence croisée]

Machine Translated by Google

6. Liu, Z. ; Li, L. ; Jiang, H. ; Porte-poisse.; Tu, D. ; Wang, S. ; Zha, Z. Caricature vidéo cohérente non supervisée avec cohérence perceptuelle du mouvement. Dans Actes de la trente-sixième conférence AAAI sur l'intelligence artificielle (AAAI-22), virtuel, 22 février-1er mars 2022. [CrossRef]

7. Rombach, R. ; Blattmann, A. ; Lorenz, D. ; Esser, P. ; Ommer, B. Synthèse d'images haute résolution avec modèles de diffusion latente. arXiv 2021, arXiv :2112.10752.

8. Esser, P. ; Kulal, S. ; Blattmann, A. ; Entézari, R. ; Müller, J. ; Saini, H. ; Lévi, Y. ; Lorenz, D. ; Sauer, A. ; Boesel, F. ; et coll. Mise à l'échelle des transformateurs de flux rectifiés pour la synthèse d'images haute résolution. arXiv 2024, arXiv :2403.03206.

9. Meng, C. ; Hé.; Chanson, Y. ; Chanson, J. ; Wu, J. ; Zhu, JY; Ermon, S. SDEdit : synthèse et édition d'images guidées avec des équations différentielles stochastiques. Dans Actes de la Conférence internationale sur les représentations d'apprentissage, virtuelle, 25-29 avril 2022.

10. Zhang, L. ; Rao, A. ; Agrawala, M. Ajout d'un contrôle conditionnel aux modèles de diffusion texte-image. Dans les actes du 2023 Conférence internationale IEEE/CVF sur la vision par ordinateur (ICCV), Paris, France, 2-6 octobre 2023.

11. Saharia, C. ; Ho, J. ; Chan, W. ; Salimans, T. ; Flotte, DJ ; Norouzi, M. Super-résolution d'image via raffinement itératif. arXiv 2021, arXiv :2104.07636.

12. Saharia, C. ; Chan, W. ; Chang, H. ; Lee, C. ; Ho, J. ; Salimans, T. ; Flotte, D. ; Norouzi, M. Palette : Modèles de diffusion image à image. Dans Actes de la conférence ACM SIGGRAPH 2022, New York, NY, États-Unis, 7-11 août 2022 ; SIGGRAPHE '22. [Référence croisée]

13. Yang, S. ; Zhou, Y. ; Liu, Z. ; Loy, CC Restituer une vidéo : traduction vidéo à vidéo guidée par texte Zero-Shot. Dans les actes du Actes de la conférence ACM SIGGRAPH Asia, Sydney, NSW, Australie, 12-15 décembre 2023.

14. Jamriska, O. ebsynth : synthèse d'image et transfert de style rapides basés sur des exemples. 2018. Disponible en ligne : https://github.com/jamriska/ebsynth (consulté le 20 juillet 2024).

15. Brooks, T. ; Peebles, B. ; Holmes, C. ; DePue, W. ; Guo, Y. ; Jing, L. ; Schnurr, D. ; Taylor, J. ; Luhman, T. ; Luhman, E. ; et coll. Modèles de génération vidéo comme simulateurs du monde. 2024. Disponible en ligne : https://openai.com/index/video-generation-models-as-world-simulators/ (consulté le 20 juillet 2024).

16. Chen, Z. ; Li, S. ; Haque, M. Un aperçu de Sora d'OpenAI et de son potentiel pour les jeux gratuits de moteur physique et la réalité virtuelle. Trans. approuvé par EAI. Robot. 2024, 3. [Réf. croisée]

17. Liang, D. ; Liu, Y. ; Huang, Q. ; Zhu, G. ; Jiang, S. ; Zhang, Z. ; Gao, W. Video2Cartoon : Générer un dessin animé 3D à partir d'une diffusion de football vidéo. Dans Actes de MULTIMEDIA '05, Singapour, 11 novembre 2005.

18. Ngo, V. ; Cai, J. Conversion d'une vidéo de football 2D en dessin animé 3D. Dans les actes de la 10e Conférence internationale sur le contrôle de 2008, Automatisation, robotique et vision, Hanoï, Vietnam, 17-20 décembre 2008 ; pp. 103-107. [Référence croisée]

19. Weng, CY; Sans boucles, B. ; Kemelmacher-Shlizerman, I. Réveil photo : animation de personnages 3D à partir d'une seule photo. Dans Actes de la conférence IEEE/CVF sur la vision par ordinateur et la reconnaissance de formes (CVPR), Long Beach, Californie, États-Unis, 15-20 juin 2019.

20. Tous, R. Pictonaut : Caricature de films utilisant l'estimation de la pose humaine en 3D et les GAN. Multimed. Outils Appl. 2023, 82, 1-15. [Référence croisée]

21. Rubén Tous, JN ; Igual, L. Achèvement de la pose humaine dans des prises de vue partielles de la caméra du corps. J. Exp. Théorique. Artif. Intell. 2023, 1-11. [Référence croisée]

22. Fiscer, J. ; Jamriska, O. ; Simons, D. ; Shechtman, E. ; Lu, J. ; Asente, P. ; Luká˘c, M. ; Sýkora, D. Synthèse basée sur des exemples de stylisés Animations faciales. ACM Trans. Graphique. 2017, 36, 1–11. [Référence croisée]

23. Gatys, Los Angeles ; Ecker, AS; Bethge, M. Un algorithme neuronal du style artistique. arXiv 2015, arXiv : 1508.06576.

24. Lui, K. ; Chen, X. ; Xie, S. ; Li, Y. ; Dollar, P. ; Girshick, R. Les auto-encodeurs masqués sont des apprenants en vision évolutive. Dans Actes de la conférence IEEE/CVF 2022 sur la vision par ordinateur et la reconnaissance de formes (CVPR), Nouvelle-Orléans, LA, États-Unis, 18-24 juin 2022 ; pages 15979 à 15988. [Référence croisée]

25. Dosovitskiy, A. ; Beyer, L. ; Kolesnikov, A. ; Weissenborn, D. ; Zhai, X. ; Unterthiner, T. ; Dehghani, M. ; Minderer, M. ; Heigold, G. ; Gelly, S. ; et coll. Une image vaut 16 x 16 mots : transformateurs pour la reconnaissance d'images à grande échelle. Dans Actes de la 9e Conférence internationale sur les représentations d'apprentissage, ICLR 2021, virtuel, 3-7 mai 2021. Disponible en ligne : https://openreview.net/ (consulté le 20 juillet 2024).

26. Radford, A. ; Kim, JW; Hallacy, C. ; Ramesh, A. ; Goh, G. ; Agarwal, S. ; Sastry, G. ; Askell, A. ; Michkine, P. ; Clark, J. ; et coll. Apprentissage de modèles visuels transférables à partir de la supervision du langage naturel. Dans Actes de la 38e Conférence internationale sur l'apprentissage automatique, virtuelle, 18-24 juillet 2021 ; pages 8748 à 8763.

27. Carion, N. ; Massa, F. ; Synnaève, G. ; Usunier, N. ; Kirillov, A. ; Zagoruyko, S. Détection d'objets de bout en bout avec des transformateurs. Dans Actes de Computer Vision–ECCV 2020 : 16e Conférence européenne, Glasgow, Royaume-Uni, 23-28 août 2020 ; pp. 213-229. [Référence croisée]

28. Yang, Z. ; Yang, Y. Fonctionnalités de découplage dans la propagation hiérarchique pour la segmentation d'objets vidéo. Dans Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Nouvelle-Orléans, LA, États-Unis, 28 novembre-9 décembre 2022.

29. Liu, S. ; Zeng, Z. ; Louer.; Li, F. ; Zhang, H. ; Yang, J. ; Li, C. ; Yang, J. ; Su, H. ; Zhu, J. ; et coll. Dinosaure à la terre : épouser un dinosaure avec une pré-formation à la terre pour la détection d'objets en espace ouvert. arXiv 2023, arXiv :2303.05499.

30. Suzuki, S. ; be, K. Analyse structurelle topologique d'images binaires numérisées par suivi de bordure. Calculer. Graphique de vision. Image Processus. 1985, 30, 32-46. [Référence croisée]

31. Kazemi, V. ; Sullivan, J. Alignement de visage en une milliseconde avec un ensemble d'arbres de régression. Dans Actes de la conférence IEEE 2014 sur la vision par ordinateur et la reconnaissance de formes, Columbus, OH, États-Unis, 23-28 juin 2014 ; pages 1867-1874. [Référence croisée]

Machine Translated by Google

32. Dalal, N. ; Triggs, B. Histogrammes de gradients orientés pour la détection humaine. Dans les actes de la conférence 2005 de l'IEEE Computer Society sur la vision par ordinateur et la reconnaissance de formes (CVPR'05), San Diego, Californie, États-Unis, 20-26 juin 2005 ; Volume 1, pages 886 à 893. [Référence croisée]
33. Soomro, K. ; Zamir, AR; Shah, M. UCF101 : un ensemble de données de 101 classes d'actions humaines à partir de vidéos dans la nature. arXiv 2012, arXiv : 1212.0402.
34. Zablotskaia, P. ; Siarohin, A. ; Zhao, B. ; Sigal, L. DwNet : réseau dense basé sur la distorsion pour la génération de vidéos humaines guidées par la pose. Dans Actes de la 30e British Machine Vision Conference 2019, BMVC 2019, Cardiff, Royaume-Uni, 9-12 septembre 2019 ; p. 51.
35. Kay, W. ; Carreira, J. ; Simonyan, K. ; Zhang, B. ; Hillier, C. ; Vijayanarasimhan, S. ; Viole, F. ; Vert, T. ; Retour, T. ; Natsev, P. ; et coll. Le Ensemble de données vidéo sur l'action humaine cinétique. arXiv 2017, arXiv :1705.06950.
36. Tous, R. Page d'accueil du projet Lester. Disponible en ligne : https://github.com/rtous/lester (consulté le 20 juillet 2024).
37. UIT. Méthodes subjectives d'évaluation de la qualité vidéo pour les applications multimédias. Recommandation UIT-T P.910. 2008. Disponible en ligne : https://www.itu.int/rec/T-REC-P.910 (consulté le 20 juillet 2024).
38. UIT. Terminologie du score d'opinion moyen (MOS). Recommandation UIT-T P.800.1 Méthodes d'évaluation objective et subjective de la qualité vocale et vidéo. 2016. Disponible en ligne : https://www.itu.int/rec/T-REC-P.800.1 (consulté le 20 juillet 2024).
39. Ruiz, N. ; Li, Y. ; Jampani, V. ; Pritch, Y. ; Rubinstein, M. ; Aberman, K. DreamBooth : Ajustement précis des modèles de diffusion texte-image pour la génération axée sur le sujet. arXiv 2023, arXiv :2208.12242.

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Une solution Swarm Intelligence pour les utilisateurs multi-véhicules rentables Problème de ramassage et de livraison

Abeer I. Alhujaylan 1,* et Manar I. Hosny 2

1 Département des technologies de l'information, Collège d'informatique, Université Qassim, Buraydah 51452, Arabie
2 Saoudite Département d'informatique, Collège des sciences informatiques et de l'information, Université King Saud, Riyad 11362, Arabie Saoudite ; mifawzi@ksu.edu.sa
* Correspondance : a.alhujaylan@qu.edu.sa

Résumé : Les applications de livraison connaissent une croissance significative, nécessitant des algorithmes efficaces pour coordonner le transport et générer des profits. Un problème qui prend en compte les objectifs des applications de livraison est le problème de ramassage et de livraison rentable multi-véhicules (MVPPDP). Dans cet article, nous proposons huit nouvelles métaheuristiques pour améliorer les solutions initiales du MVPPDP basées sur l' algorithme bien connu d'intelligence en essaim, Artificial Bee Colony (ABC) : K-means-GRASP-ABC(C)S1, K-means- GRASP-ABC(C)S2, K-means modifié-GRASP-ABC(C)S1, K-means modifié-GRASP-ABC(C)S2, ACO-GRASP-ABC(C)S1, ACO-GRASP-ABC(C)S2, ABC(S1) et ABC(S2). Toutes les méthodes ont obtenu des performances supérieures dans la plupart des cas en termes de temps de traitement. Par exemple, pour 250 clients, les temps moyens des algorithmes étaient de 75,9, 72,86, 79,17, 73,85, 76,60, 66,29, 177,07 et 196,09, ce qui était plus rapide que celui des méthodes de pointe qui prenaient 300 s. . De plus, tous les algorithmes proposés ont donné de bons résultats sur des instances de petite taille en termes de profit en obtenant treize nouvelles meilleures solutions et cinq solutions égales aux solutions les plus connues. Cependant, les algorithmes sont légèrement à la traîne dans les instances de taille moyenne et grande en raison de la stratégie aléatoire gourmande et de GRASP qui ont été utilisés dans la phase scout bee. De plus, nos algorithmes donnent la priorité à des solutions et à des itérations minimales pour un temps de traitement rapide dans les applications de commerce mobile quotidiennes, tandis que la réduction du nombre d'itérations et de la taille de la population réduit la probabilité d'obtenir une bonne qualité de solution.

Mots-clés : problème de ramassage et de livraison rentable multi-véhicules ; Colonie d'abeilles artificielles ; intelligence en essaim ; algorithmes métaheuristiques

check for updates

Citation : Alhujaylan, AI ; Hosny, Michigan Une solution Swarm Intelligence pour le problème de ramassage et de livraison rentable sur plusieurs véhicules . Algorithmes 2024, 17, 331. https://doi.org/10.3390/a17080331

Rédacteur académique : Vittorio Maniezzo

Reçu : 2 juillet 2024
Révisé : 17 juillet 2024
Accepté : 25 juillet 2024
Publié : 31 juillet 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Au cours des cinq dernières années, les applications de livraison ont connu une croissance substantielle, en particulier les applications de livraison de nourriture, où le COVID-19 a propulsé l'industrie de quelques années alors que des millions de personnes confinées commandaient de la nourriture et d'autres services essentiels en ligne. La plateforme de livraison de courses Instacart a indiqué que ses objectifs pour 2022 avaient été atteints au cours de la troisième semaine de confinement. En 2022, la Chine est devenue le plus grand marché de livraison de nourriture, avec une valeur marchande stupéfiante de 42,5 milliards de dollars. D'ici 2029, la taille du marché de l'industrie des applications de livraison de nourriture dans son ensemble devrait atteindre 165 milliards de dollars [1]. De plus, 85 % des Américains passent des commandes de nourriture en ligne via des applications de commande de nourriture. De plus , 80 % des restaurants aux États-Unis offrent à leurs clients la commodité de la commande mobile [2]. De même, les applications de livraison de nourriture en Arabie Saoudite fournissent des services instantanés et une expérience utilisateur simplifiée permettant aux individus de profiter de la nourriture. En conséquence, de nouvelles applications qui rivalisent pour fournir des services, de la commodité et du confort à plus de 35 millions de personnes émergent périodiquement. Des exemples de ces applications devenues très populaires en Arabie Saoudite sont Mrsool, Uber Eats, Talabat, Hunger station, Ninja, Jahez, To You, The Chifz, Shgardi et NANA [3]. La demande croissante d' applications de livraison de nourriture a parallèlement accru la pression sur les entreprises de livraison, qui ont besoin de

Machine Translated by Google

des algorithmes expéditifs pour coordonner le transport des marchandises afin de générer des profits et garantir la satisfaction des clients. L'un des problèmes présentés dans la littérature qui imite l'objectif des applications de livraison de nourriture, à savoir augmenter les profits, atteindre la satisfaction des utilisateurs et réduire les coûts, est le problème de ramassage et de livraison rentable multi-véhicules (MVPPDP).

Gansterer et coll. [4] ont introduit le MVPPDP, qui est un sous-type du problème de ramassage et de livraison sélectifs (SPDP) et une nouvelle variante du problème bien connu de ramassage et de livraison (PDP). Dans les contraintes d'un horizon de planification d'une journée, le MVPPDP cherche à identifier les itinéraires les plus rentables en minimisant les frais de déplacement et en maximisant la collecte des revenus. Le MVPPDP est un problème statique caractérisé par une flotte de véhicules standardisée, un dépôt central, un ensemble de clients et une quantité prédéterminée de demandes et de produits. Un binôme client, composé du client retrait et du client livraison, définit chaque demande. Ce faisant, les produits sont acheminés d'un client de retrait à un client de livraison afin de générer une source de revenus pour l'opération.

Les applications réelles de MVPPDP incluent des applications mobiles pour les services de livraison de nourriture et les entreprises de livraison.

Le MVPPDP est l'une des applications du problème d'optimisation combinatoire (CO) qui reçoit une attention particulière de la part des chercheurs en informatique et en recherche opérationnelle.

La résolution d'un problème de CO nécessite principalement de rechercher la solution optimale ou la meilleure possible parmi un ensemble limité de solutions candidates réalisables. De nombreux algorithmes peuvent être utilisés pour résoudre des problèmes de CO, et ces algorithmes peuvent être classés en deux catégories : les algorithmes exacts, qui garantissent l'obtention de solutions optimales dans un laps de temps adéquat, et les algorithmes approximatifs, généralement appelés heuristiques ou métaheuristiques, qui ne sont pas garanti de fournir une solution optimale mais de fournir une bonne solution dans un délai raisonnable. Les mots heuristique et méta appartiennent à la langue grecque, où heuristique vient de trouver' ou ng 'to heuriskein', signifiant ng 'to ng' découvrir' de nouvelles méthodes et stratégies au- delà, à un niveau supérieur. Métaheuristique ng à résoudre des problèmes, tandis que le préfixe méta signifie que x méthodes peuvent être classées en fonction du nombre de solutions utilisées en même temps en deux catégories : les métaheuristiques basées sur une solution unique (métaheuristiques S) et les métaheuristiques basées sur une population (métaheuristiques P). En S-métaheuristique, l'accent est mis sur la modification et l'amélioration d'une solution candidate unique. D'un autre côté, les P-métaheuristiques préservent et améliorent plusieurs solutions candidates au lieu d'une seule solution. Ils sont basés sur l'amélioration itérative des solutions au sein d'une population et sur l'évolution de l'ensemble de la population jusqu'à ce qu'un ensemble de bonnes solutions soit atteint. Les caractéristiques orientées vers l'exploration des métaheuristiques P les ont rendues meilleures que les métaheuristiques S pour diversifier l'espace de recherche. Les exemples les plus célèbres de métaheuristiques P sont des algorithmes inspirés du comportement naturel, tels que les algorithmes génétiques et l'intelligence en essaim (SI) [5].

Généralement, deux phases doivent être appliquées pour obtenir une bonne solution au problème : la construction de la solution et l'amélioration de la solution. La construction de solutions implique la création d'une ou plusieurs solutions initiales possibles qui peuvent être considérées comme un point de départ pour explorer l'espace de recherche. L'amélioration de la solution consiste à modifier les solutions de départ pour les améliorer, selon des stratégies prédéfinies, jusqu'à ce qu'une solution de haute qualité soit atteinte ou que les critères d'arrêt soient remplis (par exemple, un nombre maximum d'itérations ou une durée spécifique) [5].

Étant donné que le MVPPDP est un problème NP-difficile (puisqu'il s'agit d'un sous-type du SPDP, qui est un cas particulier du PDP) [4], des algorithmes métaheuristiques peuvent être utilisés pour localiser efficacement les solutions optimales dans une quantité raisonnable. temps, en particulier pour les problèmes de tailles moyennes à importantes. Dans cet article, nous introduisons de nouveaux algorithmes pour résoudre ce problème, dans le but d'améliorer la solution initiale proposée dans [6,7]. Notre approche est basée sur l'utilisation de l'algorithme Artificial Bee Colony (ABC), qui est l'un des algorithmes d'intelligence artificielle largement reconnus, connu sous le nom de SI. Diverses stratégies ont été mises en œuvre avec ABC, chacune démontrant son efficacité pour accélérer la procédure de recherche et réduire le temps de traitement.

Les principaux objectifs de ce document sont (1) d'aider les applications de livraison de nourriture et les entreprises de livraison à optimiser leurs services grâce à une sélection accélérée des clients et

Machine Translated by Google

planification d'itinéraires afin de maximiser les profits ainsi que la satisfaction des clients ; (2) une planification efficace des transports qui peut contribuer à conserver les ressources énergétiques en réduisant les effets néfastes sur l'environnement des émissions de gaz ; et que (3) pour certains cas testés, les approches proposées obtiennent des résultats qui dépassent les méthodes de pointe documentées dans la littérature pour le MVPPDP.

Le reste de cet article est organisé comme suit : une revue de certains travaux connexes est présentée dans la section 2. La section 3 introduit la définition formelle du problème. La méthode proposée est décrite en détail dans la section 4. La section 5 illustre les résultats expérimentaux.
Enfin, les conclusions et les travaux futurs sont présentés dans la section 6.

2. Études connexes

Le MVPPDP est classé comme un type de SPDP et son objectif est de maximiser les profits tout en minimisant les frais de déplacement en visitant uniquement les paires de clients rentables. Atteindre un itinéraire favorable qui maximise la différence entre les revenus totaux gagnés et le coût total du défilement est l'objectif à la fois du MVPPDP et du problème des circuits rentables (PTP). La principale distinction est que dans le PTP, l'itinéraire du véhicule n'est soumis à aucune contrainte, y compris, mais sans s'y limiter, le temps de trajet maximum, les limites de capacité du véhicule ou les contraintes de préséance [8]. Les recherches portant sur le PTP sont rares dans la littérature scientifique [9]. Par conséquent, nous commençons par présenter les recherches relatives au MVPPDP.

Gansterer et coll. [4] ont introduit les quatre algorithmes initiaux. La métaheuristique de recherche de voisinage variable (VNS) a été implémentée dans [10]. L'heuristique de construction gloutonne (C1) et l'heuristique d'insertion la moins chère en deux étapes (C2) ont été utilisées pendant la phase de construction du VNS. Au cours de la phase d'amélioration, les opérateurs de voisinage suivants ont été mis en œuvre : l'échange arrière par paire, l'échange aller par paire, l'insertion avant, les paires de relocalisation, l'insertion arrière, l'échange inter-tours, l'insertion inter-tours, l'insertion inter-factices, l'échange inter-factices. et l'échange du centre de gravité. Les auteurs ont mis en œuvre deux approches pour appliquer les opérateurs de quartier : une méthode séquentielle (Seq), qui utilise une séquence prédéterminée de quartiers, et une méthode auto-adaptative (Sea), qui s'ajuste en fonction de l'état de recherche pour déterminer la séquence de quartiers. . Afin de comparer la méthode proposée avec un algorithme alternatif, un algorithme basé sur une métaheuristique de recherche locale guidée (GLS) a finalement été Une évaluation de l'algorithme proposé a été menée sur 36 instances randomisées de différentes tailles. En termes de qualité de solution, les résultats expérimentaux ont démontré que les deux variantes du VNS ont surpassé le GLS.

Haddad [11] a introduit un nouvel algorithme pour résoudre le MVPPDP. En attendant une solution au problème rentable du ramassage et de la livraison, ils ont proposé un algorithme connu sous le nom d'IPPD, dérivé d'une combinaison d'une recherche locale itérée et d'une descente de quartier variable aléatoire. Lors de la recherche, l'algorithme IPPD ne s'est pas limité à accepter uniquement des solutions viables. Pour la construction de la solution initiale, l'heuristique C1 utilisée par Gansterer et al. [4] a été mis en œuvre. Pour améliorer la solution, plusieurs mouvements de quartier ont été mis en œuvre : échange de paires ou déplacement de paires, échange de blocs ou déplacement de blocs, déplacement de ramassage ou de livraison, échange inter-blocs ou déplacement inter-blocs, échange de centre de gravité, opérateur d'insertion et opérateur de retrait, et inter -échange de paires ou décalage inter-paires. Sur les instances de référence suggérées par Gansterer et al. [4], l'algorithme proposé a été évalué. Il a démontré son efficacité dans le traitement des dossiers de petite et moyenne taille en identifiant de nouvelles solutions optimales pour six dossiers. Cependant, pour les grandes instances, les performances de l' algorithme proposé étaient insuffisantes.

Dans Alhujaylan et Hosny [6], la phase de construction de la procédure de recherche adaptative randomisée gloutonne (GRASP) largement reconnue a été utilisée pour développer des solutions préliminaires pour le MVPPDP. Ils ont utilisé le même ensemble de données créé par [4]. Ils ont effectué une comparaison des performances entre la méthode et deux heuristiques de construction précédemment documentées dans la littérature pour construire des solutions initiales au MVPPDP. Les résultats expérimentaux ont démontré que la méthode proposée a surpassé

Machine Translated by Google

heuristiques de construction alternatives, en particulier dans les instances de petite taille, où huit nouvelles meilleures solutions initiales au problème ont été obtenues.

Dans [7], Alhujaylan et Hosny ont présenté de nouvelles heuristiques pour construire des solutions initiales pour le MVPPDP. Les heuristiques sont basées sur le premier regroupement de l'espace de recherche du MVPPDP à l'aide de trois méthodes de regroupement : les K-moyennes, les K-moyennes adaptatives et l'optimisation des colonies de fourmis (ACO). Ensuite, une version modifiée du GRASP a été utilisée pour construire la solution MVPPDP initiale basée sur les résultats de chaque algorithme de clustering. La solution proposée a été appliquée dans l'ensemble de données présenté par les chercheurs de [4]. Ils ont comparé leurs résultats avec ceux des autres heuristiques de construction précédemment utilisées pour le MVPPDP. Les résultats expérimentaux ont prouvé l'efficacité des algorithmes proposés tant en termes de qualité de solution que de temps de traitement.

Dans [12], les auteurs ont incorporé des fenêtres temporelles avec des temps de trajet incertains dans le MVPPDP. Les chercheurs ont utilisé un sous-ensemble de la base de données créée par [4], qui comprenait vingt et cinquante requêtes, correspondant respectivement à deux et trois véhicules. L'objectif du client est d'identifier une solution qui optimise le bénéfice net, calculé comme la différence entre les revenus perçus, le coût de l'itinéraire et les dépenses engagées en raison du non-respect des plages horaires. Ils ont pris en compte de larges fenêtres horaires et des fenêtres horaires standard. En supposant que les temps de trajet sont des variables aléatoires indépendantes distribuées Gamma et dans une perspective sans risque, ils ont démontré que les coûts anticipés de ponctualité et d'arrivée anticipée peuvent être calculés sous une forme fermée. Les auteurs ont suggéré d'utiliser une approche de recherche graphique comme méthode de résolution. Des expériences ont démontré que des solutions précises peuvent être obtenues dans un délai raisonnable.

3. Définition du problème

Le MVPPDP est un problème statique prédéfini comportant plusieurs constituants. Il est postulé qu'un dépôt centralisé est chargé de recevoir les demandes des clients. Ces demandes concernent les clients retrait et livraison par paires. Pour répondre à ces demandes, nous exploitons une flotte de véhicules identiques qui transportent une variété de produits identiques d'un sous-ensemble de clients de ramassage à leurs clients de livraison respectifs, générant ainsi un profit spécifique. Les limitations suivantes du MVPPDP doivent être prises en compte : • La contrainte d'appariement stipule qu'un couple client prédéfini est attribué à chaque cl demande (enlèvement et livraison).

• Contrainte de préséance : La priorité doit être donnée à la visite du client retrait par rapport au client livraison. • La limite de temps de trajet : Chaque véhicule est soumis à une limitation de durée de trajet journalière. cela ne doit pas être dépassé lors du service des clients.

Une limitation est la capacité de chaque véhicule, qui ne doit pas être dépassée pendant collecte des produits auprès des clients de ramassage.

• Chaque véhicule doit partir et arriver au dépôt avec un chargement vide. • Une seule visite est autorisée par client.

Le MVPPDP peut être formellement défini comme suit [10] : Soit G = (V, A) un graphe dans lequel V = {0, . . . , 2n + 1} définit l'ensemble de sommets, où le sommet (0, 2n + 1) correspond à un dépôt central, tandis que les sommets restants représentent les clients de collecte P = {1, . . . , n}, et les clients de livraison D = {n + 1, . . . , 2n}. A = {(i, j) : i, j V, i = j} définit l'ensemble d'arcs, où chaque arc est associé à un coût de routage non négatif cij. Chaque sommet de livraison i a un revenu ri à gagner en le visitant. De plus, chaque sommet i a une offre qi (récupération, qi > 0) ou une demande (livraison, qn+i = −qi). Au départ, il n'y a ni offre ni demande dans le dépôt (q0 = 0). De plus, chaque sommet i a une durée de service di . Il existe un ensemble de véhicules K = {1, . . . , m}, et chaque véhicule a une capacité de charge C et une durée de parcours maximale T.

Machine Translated by Google

La fonction objectif (OF) du MVPPDP vise à maximiser le profit total en soustrayant le total des frais de voyage du total des revenus. Il peut être formulé ainsi :

$$\text{Maximiser} \sum_{i \ V} \sum_{j \ V} \sum_{k \ K} ri - cij \ xijk, \quad (1)$$

Le modèle mathématique du MVPPDP peut être formulé comme dans [4,6] comme suit :

$$\sum_{je \ V} \sum_{k \ K} xijk \le 1 \ j \quad V \quad (2)$$

$$\sum_{j \ V} \sum_{k \ K} xijk \le 1 \ je \quad V \quad (3)$$

$$xi0k = 0 \ je \quad V, k \quad K \quad (4)$$

$$x2n+1,jk = 0 \ j \quad V, k \quad K \quad (5)$$

$$\sum_{je \ V} xijk - xjik = 0 \ j \quad V \{0, 2n + 1\}, k \quad K \quad (6)$$

$$\sum_{j \ V} xijk - xn+i,jk = 0 \ je \quad P, k \quad K \quad (7)$$

$$\sum_{j = V} x0jk = \sum_{je = V} xi,2n+1,k = 1 \ k \quad K \quad (8)$$

$$xijk = 1 \quad Qjk = Qik + qj \ je \quad V, j \quad V \{0\}, k \quad K \quad (9)$$

$$Qik \le C \ je \quad V, k \quad K \quad (10)$$

$$Q0k = 0k \quad K \quad (11)$$

$$Bik \le Bn+i,k \quad je \quad P \quad (12)$$

$$Bjk \ge Xijk \ Bik + di + tij \quad je \quad V, j \quad V, k \quad K \quad (13)$$

$$B2n+1,k \le T \ k \quad K \quad (14)$$

$$Bok = 0 \ k \quad K \quad (15)$$

$$xijk \quad 0, 1 \ je \quad V, j \quad V, k \quad K \quad (16)$$

$$Qik, Bik \ge 0 \ je \quad V, k \quad K \quad (17)$$

Les contraintes (2) et (3) exigent que chaque sommet soit visité au plus une fois. Les contraintes (4) et (5) empêchent tout véhicule d'entrer dans le dépôt d'origine (point de départ) ou de quitter le dépôt de destination (point d'arrivée). La contrainte (6) indique la conservation du flux. La contrainte (7) stipule que chaque couple de clients (enlèvement et livraison) doit être desservi par le même véhicule.
La contrainte (8) impose que tous les véhicules démarrent et reviennent au dépôt. La contrainte (9) précise que le dépassement de la capacité du véhicule n'est pas autorisé. La contrainte (10) précise que la charge d'un véhicule est limitée par C. La contrainte (11) précise que tous les véhicules démarrent avec une cargaison vide. La contrainte (12) indique que le client livreur ne peut pas être visité avant le client retrait. La contrainte (13) stipule que l'heure la plus précoce pour commencer le service à un sommet j est déterminée en ajoutant le début de l'heure de service au sommet i à l'heure de service à et au temps de trajet entre i et j. La contrainte (14) indique que le temps maximum pour chaque véhicule est limité par T. La contrainte (15) précise que chaque véhicule commence au dépôt à l'instant zéro. Enfin, les variables de décision sont définies par (16) et (17).

4. Méthode proposée

Dans la présente étude, nous avons utilisé l'algorithme de colonie d'abeilles artificielles (ABC) proposé pour la première fois par Karaboga [13]. Il imite le comportement de recherche de nourriture collaborative et collective d'une colonie d'abeilles. Dans l'algorithme ABC, il existe trois types différents d'abeilles :

Machine Translated by Google

les abeilles éclaireuses, les abeilles employées et les abeilles observateurs. Chaque type joue un rôle différent dans l'exploration et l'exploitation des sources de nourriture. Une source de nourriture représente la solution à un problème, et les meilleures solutions appartiennent aux sources de nourriture contenant plus de nectar. L'algorithme ABC fonctionne comme suit : Dans un premier temps, les abeilles éclaireuses sont assignées au hasard à l'espace de recherche pour trouver les premières sources de nourriture. Ensuite, les abeilles employées sont envoyées exploiter les lieux découverts par les abeilles éclaireuses, chaque abeille employée étant affectée à une source de nourriture. Après cela, les abeilles employées commencent une recherche de quartier pour trouver une meilleure source de nourriture à proximité. Après avoir trouvé une meilleure source de nourriture, les abeilles employées abandonnent toutes les sources de nourriture précédentes et n'exploitent que les meilleures. Ces abeilles employées retournent ensuite à la ruche et partagent leurs informations sur les sources de nourriture découvertes avec les abeilles spectatrices à travers ce qu'on appelle la danse frétillante. La décision de savoir si les abeilles observateurs suivront les abeilles employées dépend de la richesse des sources de nourriture. Si les abeilles spectatrices décident de suivre, elles deviennent alors des abeilles employées et répètent la procédure des abeilles employées. Après plusieurs itérations, la source de nourriture peut être épuisée ; à ce stade, les abeilles employées deviennent des abeilles éclaireuses et commencent à chercher au hasard une nouvelle source de nourriture pour remplacer la précédente. L' organigramme de l'algorithme ABC est illustré dans la figure 1 [14].

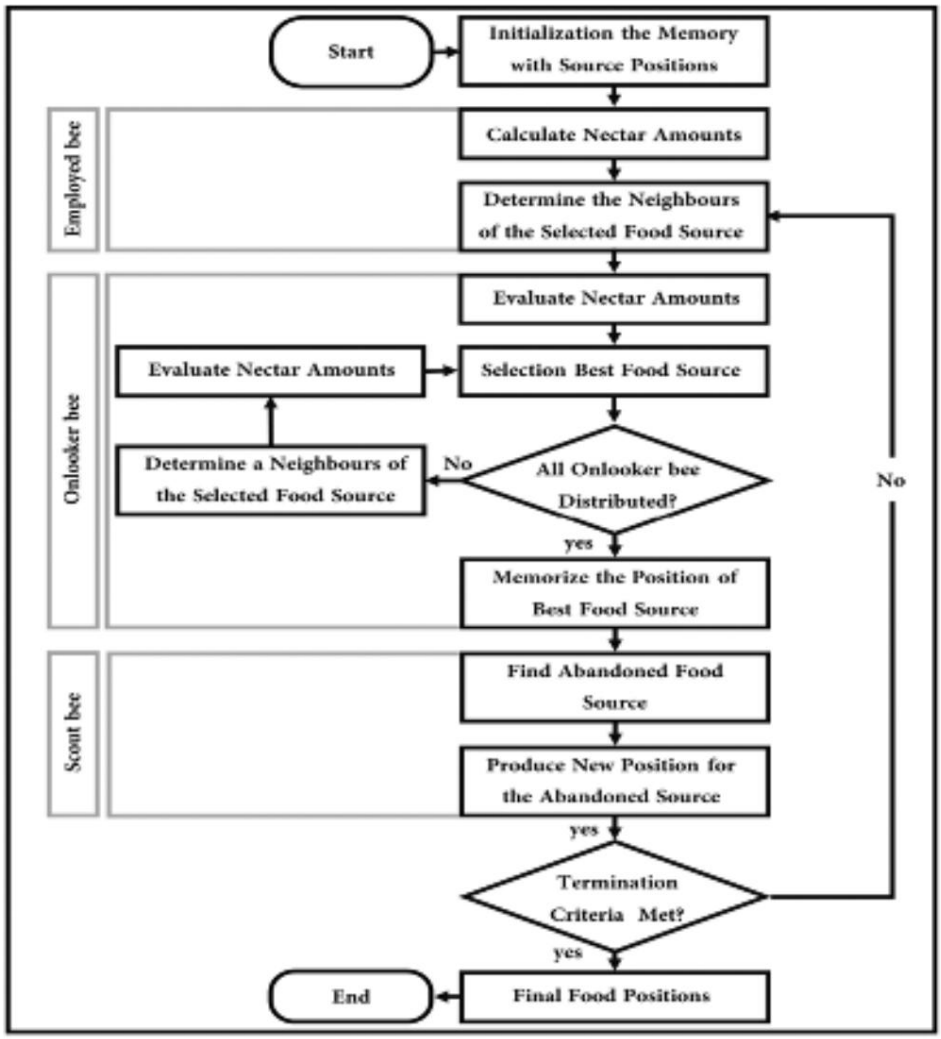


Figure 1. Organigramme de l'algorithme de colonie d'abeilles artificielles [14].

L'heuristique ABC a été utilisée pour résoudre plusieurs variantes de VRP, comme celle utilisée par Szetoa et al. [15] pour résoudre le problème des tournées de véhicules capacitaires, par Wu et al. [16] pour résoudre le VRP avec des fenêtres temporelles, et par Yao et al. [17] pour résoudre le VRP périodique.

tt

Machine Translated by Google

Dans cet article, nous visons à améliorer les solutions initiales proposées dans [7]. De nombreuses métaheuristiques ont été utilisées pour améliorer la qualité des solutions, telles que la recherche locale (LS), le recuit simulé (SA) et VNS. Dans notre étude, nous utilisons également une métaheuristique combinant recherches gourmandes et randomisées et basée sur le comportement des abeilles, à savoir l'ABC. La principale raison de choisir l'ABC plutôt que d'autres algorithmes pour les abeilles domestiques est sa popularité, le nombre d'études ayant utilisé l'ABC par rapport à d'autres variantes atteignant 54 %, comme le rapporte l'enquête de Karaboga [18]. Pour aborder le MVPPDP, deux versions de l'ABC sont proposées : ABC sans clustering (ABC) et ABC avec clustering (ABC(C)), comme cela sera expliqué ci-après.

4.1. ABC sans clustering (ABC)

Dans cette partie de la recherche, nous avons cherché à améliorer les solutions initiales obtenues en utilisant le GRASP amélioré dans [7]. La métaheuristique ABC se compose de trois phases : la phase de l'abeille employée, la phase de l'abeille spectatrice et la phase de l'abeille éclaireuse. L'objectif principal des deux premières phases est d'augmenter l'intensification de l'espace de recherche pour obtenir de bonnes solutions, tandis que la dernière phase vise à accroître la diversification et à éviter de rester coincé dans des optima locaux. Les détails de notre algorithme proposé sont illustrés ci-dessous.

1. Phase des abeilles employées :

après avoir attribué les solutions initiales aux abeilles employées, elles lancent une recherche de quartier pour trouver de meilleures solutions à proximité en sélectionnant au hasard l'un des opérateurs de quartier. Les opérateurs du quartier visent à apporter une petite modification à la solution pour l'améliorer. Ainsi, pour chaque solution, un itinéraire est sélectionné aléatoirement pour appliquer un opérateur également sélectionné aléatoirement. L'ensemble des opérateurs de quartier est décrit ci-dessous : • Déplacer un client pick-up uniquement : Choisissez un client pick-up au hasard et déplacez sa position, de telle sorte que la nouvelle position pick-up ne doive pas venir après son client livreur.

• Déplacer un client livreur uniquement : Choisissez un client livreur au hasard et déplacez sa position, de telle sorte que la nouvelle position de livraison ne doive pas précéder son client de retrait. • Déplacer
une paire de clients : Choisissez une paire de clients au hasard et déplacez leurs positions , en préservant la contrainte de préséance (ie, retrait avant livraison). • Échangez
deux paires de clients : sélectionnez deux paires de clients au hasard et échangez leurs positions. •
Opérateur Two-Opt : Sélectionnez deux paires consécutives de clients (une paire consécutive est celle dans laquelle un client pick-up est suivi directement par son client livraison) et réordonnez les bords entre les paires en tenant compte de la contrainte de préséance. •
Supprimez une paire de clients de manière aléatoire et insérez-en une non visitée : supprimez une paire de clients de manière aléatoire, sélectionnez une paire de clients non visités qui a un taux d'insertion (IR) élevé à l'aide de l'équation (18) et insérez-la dans la meilleure position de l'itinéraire.

$$\text{IRcustomerpair} = \frac{\text{rpaire client}}{\text{c(dépôt,PC)}}, \tag{18}$$

où r est le revenu de la paire de clients et c(dépôt, PC) est la distance entre le dépôt et le client de retrait de cette paire. • Supprimez une paire de

clients avec un IR faible et insérez une paire non visitée avec un IR élevé : calculez l'IR pour toutes les paires visitées à l'aide de l'équation (18) et supprimez une paire de clients avec la valeur IR la plus basse. Ensuite, recalculez l'IR pour toutes les paires de clients non visitées, sélectionnez-en une avec la valeur IR la plus élevée et insérez-la dans la meilleure position de l'itinéraire. • Insérez
une paire de clients avec un IR élevé : calculez l'IR pour toutes les paires de clients non visitées, sélectionnez-en une avec la valeur IR la plus élevée et insérez-la dans la meilleure position de l'itinéraire.

Après avoir appliqué l'opérateur de voisinage sélectionné, si la nouvelle solution S' est réalisable, l'adéquation de S' est calculée à l'aide de l'équation (1) et comparée à l'adéquation de l'ancienne solution S. Pour échapper aux optima locaux, nous avons décidé d'utiliser l'opérateur de voisinage sélectionné. algorithme démon (DA) plutôt

Machine Translated by Google

qu'un simple algorithme de recherche locale, puisque le DA a la capacité d'accepter une nouvelle solution réalisable et non améliorée. Le DA est une variante du SA présenté par Creutz [19]. La fonction d'acceptation du DA est basée sur la valeur du démon (crédit) qui est initialisée au départ avec une valeur D donnée . Dans le cas où la nouvelle solution S' n'est pas une amélioration, la différence (ΔE) entre la fitness de S et S' est comparée à la valeur du démon, et S' est acceptée si cette différence est inférieure ou égale à D. Ensuite, si S' est acceptée, elle est mémorisée, et la valeur de D est diminuée ; sinon, l'ancienne solution S est mémorisée, la valeur D est augmentée, et un compteur de non-amélioration pour cette solution est augmenté de un. Le DA est plus simple et plus rapide que le SA car ce dernier nécessite plus de temps de calcul pour générer un nombre aléatoire et une fonction exponentielle. De plus, il ne contient qu'un seul paramètre qui doit être réglé. Le DA s'est avéré être un algorithme efficace lorsqu'il a été utilisé pour résoudre plusieurs problèmes réels, obtenant de bons résultats en termes de qualité de solution et de temps, par rapport au SA [19].

2. Phase d'abeille spectatrice :

Seules les solutions qui ont de bonnes valeurs OF ont une forte probabilité d'être attribuées à des agents spectateurs. Plusieurs stratégies ont été proposées pour sélectionner des solutions, telles que la sélection à la roulette, la sélection en tournoi, l'échantillonnage universel stochastique et la sélection basée sur le classement [20]. Nous avons adopté la stratégie la plus populaire, qui est la stratégie de sélection de la roulette , dans laquelle la probabilité de sélection Pi est calculée comme suit :

$$Pi = \frac{OFi}{\sum_{j=1}^{n} OFj}, \quad (19)$$

où OFi est la fonction objectif pour une solution i, et $\sum_{j=1}^{n} OFj$ est la somme des valeurs de la fonction objectif pour toutes les solutions d'une population. Une fois qu'une solution est sélectionnée, l'un des itinéraires de solution est sélectionné de manière aléatoire et tous les opérateurs de quartier utilisés dans la phase d'abeille employée lui sont appliqués séquentiellement, puis l'adéquation de la nouvelle solution est mesurée à l'aide de l'équation (1). Pour augmenter le processus d'intensification et se concentrer uniquement sur les zones prometteuses de l'espace de recherche, le critère d'acceptation est basé sur un simple algorithme LS (local search). Autrement dit, si S' est meilleur que S, mémorisez S' et il n'est pas nécessaire de compléter les opérateurs restants ; sinon, essayez l'opérateur suivant. Si l'ancienne solution n'est améliorée par aucun opérateur, l'ancienne solution est mémorisée, et un compteur de non-amélioration pour cette solution est augmenté de un.

3. Phase d'abeille scoute : La

phase d'abeille scoute vise à accroître la diversification et à élargir l'espace de solution, évitant ainsi de rester bloqué dans les optima locaux. La phase Scout Bee démarre lorsqu'un compteur de non-amélioration d'une solution Si atteint un nombre prédéfini d'essais, appelé limite. La limite fait référence au nombre de tentatives qui ont été faites pour améliorer la solution Si , respectivement dans les phases d'abeille employée et d'abeille observateur. Trois actions importantes doivent être effectuées lorsqu'un compteur de non-amélioration pour la solution S atteint la limite : (1) effacer le compteur de non-amélioration, (2) effacer la solution Si et (3) reconstruire la solution Si .

Une stratégie de reconstruction appropriée joue un rôle important dans la phase de reconnaissance, où elle contribue à améliorer la qualité de la solution ou laisse la solution inchangée. Ainsi, nous avons proposé deux stratégies pour reconstruire la solution Si . Les détails de ces heuristiques de reconstruction sont présentés comme suit : Première stratégie (ABC(S1)) : L'une des trois

- heuristiques suivantes est sélectionnée au hasard :

1. Heuristique d'insertion aléatoire :

Tant que la contrainte de temps n'est pas violée, les couples de clients sont sélectionnés aléatoirement et insérés directement dans l'itinéraire. Chaque client retraité est suivi de son client livreur ; il n'est donc pas nécessaire de vérifier la capacité du véhicule et les contraintes de préséance.

2. Heuristiques gourmandes et aléatoires :

Machine Translated by Google

Tant que la contrainte de temps n'est pas violée, les paires de clients sont sélectionnées au hasard et insérées aux meilleures positions de l'itinéraire. Outre la contrainte de temps, les contraintes de capacité et de préséance des véhicules ne doivent pas être violées.

3. Heuristique d'insertion gourmande :

Le taux d'insertion est calculé pour toutes les paires de clients à l'aide de l'équation (18). La paire de clients ayant la valeur IR la plus élevée est ensuite insérée dans l'itinéraire dans la meilleure position. Les contraintes de temps, de capacité des véhicules et de préséance ne doivent pas être violées. •

Deuxième stratégie (ABC(S2)) : Cette stratégie fonctionne exactement comme le GRASP de [7].

Une fois la solution Si reconstruite, l'OF est calculé à l'aide de l'équation (1), et Si est mémorisé en tant que nouvelle solution et utilisé pour recommencer le processus d'amélioration en utilisant les phases d'abeille employée et spectatrice.

Les grandes lignes de la phase de routage de notre ABC sont présentées dans l'algorithme 1, où chaque notation est définie comme suit : Pop Size représente la taille de la population ; Max Iter : le nombre maximum d'itérations ; Nombre de véhicules : la quantité de véhicules ; US : couples de clients non servis ; ΔE : la différence entre les valeurs OF pour une ancienne et une nouvelle solution ; D : valeur démoniaque ; et Non − amélioration : un compteur pour compter le nombre de tentatives infructueuses d'amélioration de la solution.

Algorithme 1 : Le pseudocode de l'ABC proposé.

Pop = {S1 , S2, S3, .., Sn} ; /* Solutions initiales générées à l'aide de GRASP(V2)*/ Limit = valeur limite initiale, D = valeur démoniaque initiale, Non-amélioration = 0, Global_Best = Meilleure solution dans Pop ; Pour M = 1 à Max_Iter /*** Phase abeille employée ***/ Pour i = 1 à Pop_Size Sélectionner au hasard une voie de solution
Si ; Appliquer un opérateur de
voisinage sélectionné au hasard à l'itinéraire actuel
de la solution Si pour obtenir une
nouvelle solution S Si S est réalisable Puis calculer l' OF de S i End
If

S'_i ; S'_i

Si $DE(S'_i) > DE(S_i)$)
Mémorisez la nouvelle solution S'_i ;
Sinon si ($\Delta E \le Di$)
Mémorisez la nouvelle solution S'_i
; mettre à jour la valeur du démon : Di = Di − ΔE ;
Autre
Mémorisez l'ancienne solution Si ;
mettre à jour la valeur du démon : Di = Di + ΔE ;
Non-amélioration(i) = Non-amélioration(i) + 1 ;
Fin si
Fin pour
/*** Phase d'abeille spectatrice ***/
Sélectionnez N solutions parmi les abeilles employées en utilisant la stratégie de sélection de la roulette ;
Pour i = 1 à N
Sélectionner au hasard une voie de solution Si ;
Pour j = 1 au nombre d'opérateurs de quartier
Appliquer un opérateur de quartier sélectionné (j) à l'itinéraire actuel de la solution
Si pour obtenir une nouvelle solution S'_i ;
Si S'_i est réalisable Alors calculez le OF de S i S'_i ;
Fin si
Si $DE(S'_i) > DE(S_i)$)
Mémorisez la nouvelle solution S'_i ;
Casser;
Terminer si
Fin pour

Machine Translated by Google


```
Si la solution Si ne s'est pas améliorée après application des opérateurs de voisinage
Non-amélioration(i) = Non-amélioration(i) + 1 ;
Fin si
Fin pour
/*** Phase de l'abeille éclaireuse ***/
Pour i = 1 à N
Si Non amélioration(i) > Limite
Non-amélioration(i) = 0 ;
Si = 0 ;
Reconstruire Si en utilisant la stratégie S1 ou S2 de la phase scout bee ; Calculer le OF pour Si ;
Terminer si
Fin pour
Attribuez la solution qui a le maximum de OF à la Best_Solution ;
Si (Global_Best < Meilleure_Solution)
Global_Meilleure = Meilleure_Solution ;
Fin si
Fin pour
Résultat : la solution Global_Best
```

4.2. ABC avec clustering (ABC(C))

Dans cette partie de l'étude, la méthode ABC décrite dans la section 4.1 a été utilisée pour améliorer les solutions initiales qui ont été construites à l'aide des différentes méthodes GRASP(C) rapportées dans [7], à savoir K-means-GRASP(C), modifié- K-signifie GRASP(C) et ACO-GRASP(C). Encore une fois, certaines modifications doivent être prises en compte lors de l'utilisation de l'ABC avec ces algorithmes. Lorsque vous utilisez les opérateurs suivants, le processus d'insertion doit être limité aux paires de clients à l'intérieur du même cluster uniquement : supprimez une paire de clients de manière aléatoire et insérez-en une non visitée, supprimez une paire de clients avec un IR faible et insérez une paire non visitée avec un IR élevé. , et insérez une paire de clients avec un IR élevé. En outre, la restriction d'insertion doit être appliquée lors de la phase scout bee dans les deux stratégies (S1 et S2). Nous avons utilisé les mêmes valeurs de paramètres que celles utilisées dans l'ABC (sans regroupement) pour améliorer les solutions initiales dans cette partie de l'étude. De plus, les mêmes stratégies de la phase scout bee (S1 et S2) ont été utilisées dans les algorithmes ABC(C)). Par conséquent, au total, nous avons six algorithmes qui utilisent ABC(C) : (1) K-means-GRASP(C)-ABC(C)S1, (2) K-means-GRASP(C)-ABC(C) S2, (3) K-means-GRASP(C)-ABC(C)S1 modifié, (4) K-means-GRASP(C)-ABC(C)S2 modifié, (5) ACO-GRASP(C)- ABC(C)S1, et 6) ACO-GRASP(C)- ABC(C)S2. Par exemple, le premier algorithme fait référence à la génération de la solution initiale à l'aide du GRASP après avoir regroupé l'espace de recherche à l'aide de l'algorithme K-means, puis à l'améliorer à l'aide de l'ABC avec la stratégie de reconstruction S1.

5. Expériences informatiques

Tous les algorithmes ont été implémentés dans MATLAB (R2017b) sur un ordinateur portable alimenté par un processeur Intel Core i7-4510U fonctionnant à 2,00 GHz (2 601 MHz, deux cœurs, quatre processeurs logiques). Les sous-sections suivantes fournissent des descriptions des ensembles de données utilisés et des spécificités du réglage des paramètres avant de présenter les résultats expérimentaux.

5.1. Instances de test

Pour tester les performances de nos méthodes proposées, nous avons utilisé les mêmes instances que celles utilisées par [4]. L'ensemble de données comprend 36 instances, qui ont été classées en trois groupes distincts : petite taille (composée de 20 et 50 clients, respectivement, desservis par deux et trois véhicules), taille moyenne (composée de 100 et 250 clients, desservis par 4 et 5 véhicules). , respectivement) et de grande taille (respectivement 500 et 1000 clients servis par 6 et 8 véh
De plus, douze instances composent chaque groupe. Ces événements présentent une variabilité en termes de temps et de revenus. Pour générer des itinéraires courts et longs, la limite de temps totale est soit petite, soit grande, avec une plage de 2 500 à 15 000 s.

Machine Translated by Google

De plus, les montants des recettes sont établis de trois manières : de manière aléatoire, proportionnel à la demande, ou égal pour tous les clients. Les revenus sont acquis à partir de la livraison client postérieurement à l'expédition de marchandises provenant du client pick-up.

La quantité de marchandises est spécifiée sous la forme d'un entier positif compris entre 1 et 50.

5.2. Réglage des paramètres

5.2.1. Réglage des paramètres d'ABC sans clustering

L'algorithme ABC possède trois paramètres : le nombre d'itérations (Max_Iter), la taille de la population (Pop_Size), et la valeur initiale de la limite. De plus, puisque nous avons utilisé le DA dans la phase d'abeille employée, la valeur initiale du démon (D) doit être ajustée.

Les détails du test de chaque paramètre sont présentés ci-dessous.

Nombre d'itérations

Chaque instance de données de 100 clients a été testée avec différents nombres d'itérations : 10, 50, 100, 500 et 750. Comme le montre le tableau 1, les résultats des tests ont montré qu'il y avait aucune amélioration des valeurs de la fonction objectif après 500 itérations. Ainsi, le maximum le nombre d'itérations a été fixé à 500.

Tableau 1. Résultats du réglage du nombre d'itérations de l'ABC.

Ensemble de données / Instances	Nombre d'itérations				
	10	50	100	500	750
13-100-FS	32 274,04	35 432,98	39 437,98	39 438,58	39 438,58
14-100-FL	63 311,86	63 311,86	63 311,86	63 455,13	63 455,13
15-100-PS	68 625,48	68 625,48	71 185,28	72 032,22	72 032,22
16-100-PL	115 792,8	115 792,8	124 608,2	131 404,9	131 404,9
17-100-RS	67 974,56	74 246,27	74 246,27	74 330,95	74 330,95
18-100-RL	112 986,9	112 986,9	116 775,2	119 991,9	119 991,9

Taille de la population

Les mêmes instances de jeux de données ont été testées à nouveau pour sélectionner le nombre approprié de solutions dans la population. Au début, la taille de la population était fixée à 30 personnes, car elle ont fourni les meilleurs résultats donnés dans [7]. Ensuite, nous avons augmenté la population à différentes tailles : 40, 50 et 60 solutions. Le tableau 2 présente les valeurs de l'OF après avoir fixé le nombre de itérations à 500. Les résultats ont montré que l'augmentation de la taille de la population à plus de 50 n'a pas conduit à une amélioration de la valeur de la fonction objectif. Ainsi, la taille de la population a été estimé à 50 ans.

Tableau 2. Résultats du réglage de la taille de la population de l'ABC.

Ensemble de données / Instances	Taille de la population			
	30	40	50	60
13-100-FS	39 438,58	39 438,58	41 472,86	41 472,86
14-100-FL	63 455,13	66 453,66	66 453,66	66 453,66
15-100-PS	72 032,22	72 032,22	72 170,54	72 170,54
16-100-PL	131 404,9	131 404,9	131 404,9	131 404,9
17-100-RS	74 330,95	77 484,06	77 484,06	77 484,06
18-100-RL	119 991,9	120 249,7	126 504,3	126 504,3

Valeur limite

La valeur limite du compteur de non-amélioration est un autre paramètre important de l'algorithme ABC car il affecte la conversion du processus de recherche. Ainsi, la valeur La limite sélectionnée doit être adaptée au nombre d'itérations. Nous avons donc testé la limite en utilisant deux valeurs : 100 et 200. Les résultats du tableau 3 montrent que l'augmentation de la

Machine Translated by Google

une valeur limite supérieure à 100 n'a pas contribué à une amélioration de la fonction objectif valeur. La valeur limite a donc été fixée à 100.

Tableau 3. Résultats du réglage de la valeur limite de l'ABC.

Ensemble de données Instances	Limite	
	100	200
13-100-FS	41 472,86	41 472,86
14-100-FL	66 453,66	66 453,66
15-100-PS	72 170,54	72 170,54
16-100-PL	131 404,9	131 404,9
17-100-RS	77 484,06	77 484,06
18-100-RL	126 504,3	126 504,3

Valeur démoniaque

Plusieurs valeurs ont été testées pour sélectionner la meilleure valeur initiale du démon (D) : 1000, 2 500, 5 000 et 7 500. Les résultats du tableau 4 n'indiquent aucune amélioration de la fonction objective. si la valeur du démon était supérieure à 5000. Ainsi, la valeur du démon D a été considérée comme étant de 5000.

Tableau 4. Résultats du réglage de la valeur démoniaque de l'ABC.

Ensemble de données Instances	Valeur démoniaque			
	1000	2500	5000	7500
13-100-FS	41 472,86	41 596,33	41 954,86	41 954,86
14-100-FL	66 453,66	66 683,84	66 683,84	66 683,84
15-100-PS	72 170,54	72 449,56	72 449,56	72 449,56
16-100-PL	131 404,9	131 404,9	131 404,9	131 404,9
17-100-RS	77 484,06	77 484,06	77 484,06	77 484,06
18-100-RL	126 504,3	126 504,3	126 504,3	126 504,3

5.3. Réglage des paramètres de ABC(C)

Puisque nous avons utilisé le même algorithme ABC, avec peu de modifications, dans ABC avec le approche de clustering, nous avons adopté les mêmes valeurs de paramètres que celles utilisées dans le ABC pour notre ABC(C).

5.4. Résultats expérimentaux

Comme mentionné précédemment, nous avons proposé deux versions de l'ABC pour améliorer la version initiale solutions : l'ABC utilisé avec GRASP, et l'ABC(C) utilisé avec les méthodes de clustering, c'est-à-dire K-means-GRASP(C), K-means-GRASP(C) modifié et ACO-GRASP(C). Nous avons également deux stratégies distinctes pour la phase des abeilles scoutes. S1 construit la solution au hasard en sélectionnant l'une des trois stratégies suivantes : insertion aléatoire, insertion gourmande et aléatoire ou insertion gourmande. Après avoir testé cette stratégie, nous avons constaté que l'heuristique d'insertion aléatoire faisait n'améliore pas les résultats par rapport aux autres, nous avons donc éliminé ce choix et conservé l'autre deux. La deuxième stratégie, S2, fonctionne comme le GRASP de [7]. Nous en avons donc huit au total méthodes proposées nommées comme suit : (1) ABC(S1), (2) ABC(S2), (3) K-means-GRASP(C)-ABC(C)S1, (4) K-means-GRASP(C)-ABC(C)S2, (5) K-means-GRASP(C)-ABC(C)S1 modifié, (6) K-means-GRASP(C)-ABC(C)S2 modifié, (7) ACO-GRASP(C)-ABC(C)S1 et (8) ACO- GRASP(C)-ABC(C) S2. Tous les algorithmes proposés ont été exécutés cinq fois, comme avec les algorithmes précédents dans [6] et [7]. Nous avons comparé nos algorithmes proposés et les algorithmes de pointe présentés dans [10,11], qui sont C1-Seq VNS, C2-Seq VNS, C1-Self Adaptive VNS, C2-Self Adaptive VNS et IPPD, pour résoudre le MVPPDP en termes de le temps de traitement et les bénéfices gagnés.

Premièrement, en ce qui concerne le temps de traitement, tous les algorithmes précédents ont été exécutés pendant 5 minutes pour chaque instance. Nous notons cependant que ce délai n'est pas considéré comme acceptable pour le applications de commerce mobile (par exemple, Hungerstation, MRSOOL, etc.) utilisées quotidiennement et s'appuyant sur le

Machine Translated by Google

vitesse de réponse, donc dans notre cas, nous avons utilisé la condition d'arrêt de 500 itérations pour mesurer le temps de traitement. Le tableau 5 présente les résultats de comparaison entre notre proposition algorithmes et algorithmes de pointe en termes de temps (en secondes). Aussi, la moyenne de chaque catégorie (lignes en surbrillance dans le tableau 5) a été utilisée pour comparer les performances des algorithmes comme le montre la figure 2. D'après le tableau 5 et la figure 2, il est évident que notre Les algorithmes basés sur le clustering proposés surpassent tous les algorithmes précédents en ce qui concerne au délai de traitement dans tous les cas, à l'exception des cas impliquant 1 000 clients. Aussi, les autres algorithmes proposés qui ne sont pas basés sur le clustering (ABC(S1) et ABC(S2)) étaient plus rapides que les algorithmes précédents pour les instances de petite et moyenne taille. Le La raison pour laquelle nos algorithmes n'excellent pas dans les instances de grande taille est probablement due à la stratégies d'abeilles scoutes qui ont été utilisées, à savoir la méthode gourmande dans la première stratégie et le GRASP dans le second. Dans la méthode gloutonne, les paires de clients sont sélectionnées en fonction sur une mesure spécifique puis inséré dans la meilleure position qui conduit à une diminution les frais de déplacement. Cette approche donne de bons résultats mais prend du temps, surtout dans les instances de grande taille. Dans la deuxième stratégie, les paires de clients non visités sont sélectionnées en fonction de calculer l'IR, ce qui entraîne également une perte de temps dans les instances de grande taille. Dans l'ensemble, cependant, le regroupement de l'espace de recherche a clairement aidé à traiter des instances de grande taille et a rendu le temps de traitement plus rapide par rapport à d'autres algorithmes qui traitent tous les clients sans regroupement.

Tableau 5. Comparaison entre les algorithmes de l'état de l'art et les algorithmes proposés basés sur temps (secondes).

Ensemble de données	Précédent Algo-rithmétiques(5 min Durée d'exécution)	K-Means-SAISIR-ABC(C)S1	K-Means-SAISIR-ABC(C)S2	Modifié K-Means-SAISIR-ABC(C)S1	Modifié K-Means-SAISIR-ABC(C)S2	ACO-SAISIR-ABC(C)S1	ACO-SAISIR-ABC(C)S2	ABC(S1)	ABC(S2)
1-20-FS	300	14,48	14,47	15,20	14,58	16,78	12,12	31,04	30h65
2-20-FL	300	12,13	12,42	13,82	13,17	13,51	11,45	30,20	30.56
3-20-PS	300	16,62	17,48	17,83	17,17	18,72	15,85	33,77	39.28
4-20-PL	300	12	12,01	12,61	12,24	15,63	12,98	24,43	28h45
5-20-RS	300	14,87	15,21	14,97	14,57	16,67	14,57	30,79	31.17
6-20-RL	300	11,32	12,38	13,47	13,71	13,63	12,25	22,62	21h56
Moyenne-20	300	13,57	14	14,65	14,24	15,82	13,20	28,81	30.28
7-50-FS	300	23,24	24,29	25,67	25,49	25,45	22,03	50,10	54,80
8-50-FL	300	19,79	21,24	21,53	21,45	22,18	20,65	43,81	46.63
9-50-PS	300	25,67	24,64	26,58	25,15	25,87	20,49	50,55	54,72
10-50-PL	300	21,63	22,64	24,41	22,96	23,34	20,90	44,77	48,77
11-50-RS	300	26,31	26,58	26,86	26,05	27,49	20,23	48,90	51,94
12-50-RL	300	22,73	20,93	23,37	21,87	22,77	20,91	45,94	48.34
Moyenne-50	300	23,23	23,39	24,74	23,83	24,52	20,87	47,34	50,86
13-100-FS	300	37,87	33,57	36,58	35,10	36,89	33,36	80,65	83,81
14-100-FL	300	31,42	33,67	34,59	34,38	35,55	33,14	73,90	81,74
15-100-PS	300	31,03	32,03	36,98	35,55	36,75	33,23	76,39	82,69
16-100-PL	300	33,54	34,33	36,50	34,57	36,20	32,41	74,19	82,93
17-100-RS	300	38	35,92	37	35,81	38,19	34,03	82,59	86,96
18-100-RL	300	36,19	35,37	34,38	33,34	37,49	33,63	77,02	81.01
Moyenne-100	300	34,68	34,15	36,01	34	36	33 .30	.46	83.19
19-250-FS	300	73 .50	72,85	78,10	0,79	0,85	67,79	176,81	204.37
20-250-FL	300	79,65	73,28	80,50	73,30	74,19	65,10	181,10	195
21-250-PS	300	74,20	72,09	75,23	76,28	77,66	63,51	169,26	195.17
22-250-PL	300	73,82	72,718	79,53	72,88	74,05	65,44	179	193,60
23-250-RS	300	74,67	72,99	77,63	74,40	77,86	66,44	177,25	193,93
24-250-RL	300	79,55	73,21	84,04	72,38	76,21	69,46	178,99	194.47
Moyenne-250	300	75,90	7 2,86	79,17	73,84	79,62	66,29	177,07	196.09
25-500-FS	300	163,89	155,32	172,74	73,85	76,60	142,52	388,62	474.16
26-500-FL	300	167,03	155,73	177,18	158,36	162,15	148,98	415,25	466.38
27-500-PS	0	160,91	146,27	172,04	160,25	170,25	140,62	386,55	465,87
28-500-PL	300	168,60	144,23	180,26	158,45 158,74	162,76 169,89	141,34	435,34	464.62

Machine Translated by Google

Tableau 5. Suite

Ensemble de données	Précédent Algo-rithmétiques(5 min Durée d'exécution)	K-Means-SAISIR-ABC(C)S1	K-Means-SAISIR-ABC(C)S2	Modifié K-Means-SAISIR-ABC(C)S1	Modifié K-Means-SAISIR-ABC(C)S2	ACO-SAISIR-ABC(C)S1	ACO-SAISIR-ABC(C)S2	ABC(S1)	ABC(S2)
29-500-RS	300	170,15	140,16	170,50	158,58	162,42	139,80	383,89	468.15
30-500-RL	300	169,78	141,24	174,22	161,07	169,30	143,36	428,65	464,89
Moyenne-500	300	166,73	147,16	174,49	159,24	166,13	142,77	406,38	467,35
31-1000-FS	300	432,51	354,73	448,29	408,67	426,32	397,22	1040,22	1406.22
32-1000-FL	300	439,85	360,11	447,72	423,01	461,18	430,54	1197,38	1438.43
33-1000-PS	300	434,34	364,52	432,93	418,51	391,01	416,76	1128,14	1398.49
34-1000-PL	300	448,06	378,57	440,70	435,69	389,93	430,39	1204,97	1416.64
35-1000-RS	300	451,79	404,20	425,83	421,47	381,24	411,31	1110,80	1296.40
36-1000-RL	300	420,15	410,26	446,23	418,03	393,29	420,10	1222,33	1286.07
Moyenne-1000	300	437,78	378,73	440,28	420,90	407,16	417,72	1150,64	1373.71

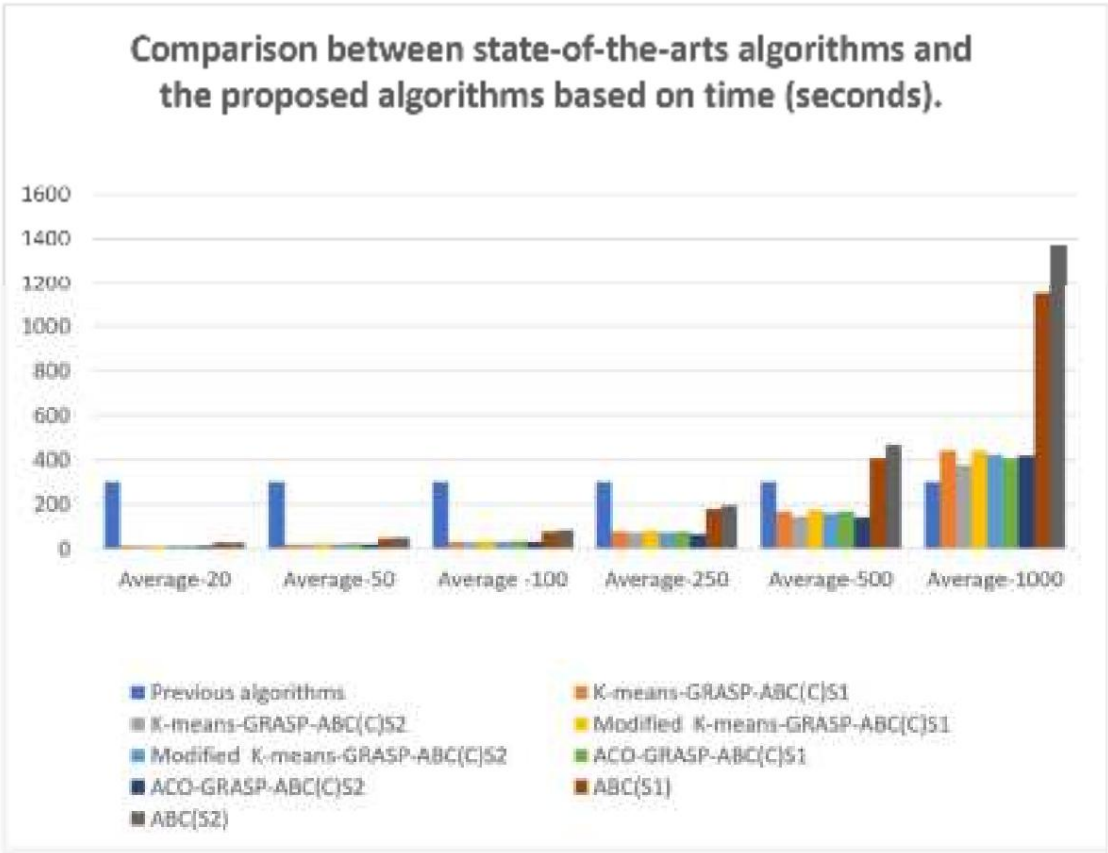


Figure 2. Comparaison entre les algorithmes de l'état de l'art et les algorithmes proposés basés sur temps (secondes).

D'autre part, nous avons comparé nos algorithmes proposés et les algorithmes précédents en termes de profit tel que calculé dans l'équation (1). Le tableau 6 présente le bénéfice moyen de cinq des essais, tandis que le tableau 7 présente le meilleur profit obtenu sur cinq essais et l'écart entre les meilleur profit et profit le plus connu en utilisant l'équation suivante :

$$\text{Écart} = \frac{\text{Meilleur profit sur cinq courses} - \text{Meilleur profit connu}}{\text{Bénéfice le plus connu}} * 100 \quad (20)$$

Machine Translated by Google

Tableau 6. Comparaison de nos algorithmes proposés avec les méthodes précédentes basées sur le profit moyen.

Ensemble de données	C1—Séquence VNS [9]	C2—Séquence VND [9]	C1 – Auto-même Adaptatif VND [9]	C2 – Auto-même Adaptatif VND [9]	IPPD [10]	K-Means-SAISIR-ABC(C)S1	K-Means-SAISIR-ABC(C)S2	Modifié K-Means-SAISIR-ABC(C)S1	Modifié K-Means-SAISIR-ABC(C)S2	ACO-SAISIR-ABC(C)S1	ACO-SAISIR-ABC(C)S2	ABC(S1)	ABC(S2)
1-20-FS	30 315,5	30 315,5	30 315,5	30 315,5	30 315,5	30 267,75	30 427,58	27 007,58	27 007,58	25 490,76	25 490,76	30 395,92	30 294,77
2-20-FL	38 765,6	38 765,6	37 866,3	38 765,6	39 179,66	37 642,38	37 468,82	39 773,34	39 240,95	36 342,53	36 322,33	39 459,15	36 153,93
3-20-PS	51 473,4	52 128,4	52 565,1	52 565,1	52 566	51 481,8	43 340,71	51 481,8	43 009,55	42 917,34	34 107,64	52 684,98	52 565,1
4-20-PL	58 365,6	58 365,6	58 365,6	58 365,6	58 777	57 350,35	57 086,36	57 350,35	57 301,88	56 523,32	47 367,54	58 888,64	56 808,21
5-20-RS	36 469,5	36 469,5	36 469,5	36 469,5	36 469,5	34 090,58	34 090,58	30 237,02	29 843,58	31 741,08	31 741,08	36 469,52	36 469,52
6-20-RL	43 368,6	43 368,6	43 368,6	43 368,6	43 781,81	42 215,53	42 086,44	44 401,45	43 736,53	40 769,67	40 450,55	44 244,27	42 143,26
Moyenne-20	43 126,37	43 235,53	43 158,43	43 308,32	43 514,91	42 174,73	40 750,08	41 708,59	40 023,35	38 964,12	35 913,32	43 690,41	42 405,8
Standard Déviation	10 289,94	10 399,10	10 559,14	10 475,04	10 560,13	10 436,42	9349.23	11 822,81	10 907,82	10 656,36	7500.01	10 573,40	10 312,25
7-50-FS	22 306	22 306	22 306	22 306	22 307,15	18 662,83	21 603,83	18 662,83	21 603,83	21 281,88	25 123,15	24 975,25	24 998,45
8-50-FL	58 050,8	58 050,8	57 303,4	57 996,7	58 050,8	50 772,64	50 873,68	50 881,55	50 773,54	46 699,31	47 534,79	53 555,25	53 268,82
9-50-PS	57 651,7	57 651,7	57 651,7	57 651,7	57 651,7	58 996,17	53 258,83	58 996,17	53 064,14	55 643,25	55 643,25	57 651,71	57 651,71
10-50-PL	132 648	132 648,0	128 353	128 334,8	129 640,3	109 726,8	117 770,8	116 372	116 149,7	97 740,02	106 491,9	124 819,3	128 477,1
11-50-RS	33 340,4	33 340,4	33 340,4	33 340,4	33 340,4	28 865,48	28 907,33	28 865,48	28 907,33	31 251,34	29 969,16	33 054,57	33 098,58
12-50-RL	79 350,3	79 350,3	77 655,5	79 350,3	77 832,56	76 422,62	76 503,28	76 309,46	76 408,91	64 359,55	64 519,79	76 582,15	75 732,86
Moyenne-50	63 891,2	50 139,84	62 768,33	63 163,32	63 137,16	57 241,09	58 152,95	58 347,92	57 817,91	52 829,22	54 880,35	61 773,03	62 204,58
Standard Déviation	39 248,45	22 523,36	37 635,69	37 749,54	38 077,61	33 048,63	35 097,31	35 184,37	34 547,72	27 041,68	29 384,58	35 937,41	37 167,44
13-100-FS	41 464,84	41 955,2	41 523,5	42 174	41 896,86	35 386,24	36 055,01	35 303,67	36 625,9	33 147,92	33 187,57	40 596,77	40 543,56
14-100-FL	81 271,2	81 971	80 080,4	80 500,2	79 479,88	58 704,7	60 780,74	57 607,07	60 035,43	51 106,58	47 074,98	64 937,47	67 237,63
15-100-PS	70 387,4	70 398,5	71 113,8	71 082,8	69 436,25	54 197	57 206,6	49 566,36	59 481,84	60 614,78	45 995,09	68 874,02	71 290,33
16-100-PL	124 288	123 931	122 691,8	122 183,2	125 057	98 140,3	104 031,6	100 122,4	110 099	92 681,38	96 365,75	112 690,5	124 487,1
17-100-RS	76 885,4	78 418,8	76 954,4	79 651,2	77 012,75	61 579,81	61 579,81	61 960,78	61 698,84	52 689,7	52 163,72	77 334,8	75 925,47
18-100-RL	131 052,6	130 670,4	130 403,4	130 366,6	131 058,6	98 522,81	100 291,3	104 771,9	109 450,2	92 423,54	93 867,6	115 108,7	122 320,4
Moyenne-100	87 558,24	87 890,82	87 127,88	87 659,67	87 323,56	67 755,14	69 990,84	68 222,03	72 898,54	63 777,32	61 442,45	79 923,7	83 634,09
Standard Déviation	34 099	33 674,36	33 546,8	33 090,45	34 316,01	25 383,15	26 636,59	28 060,86	30 021,67	24 032,04	26 834,7	29 022,89	33 175,38
19-250-FS	73 259,94	72 445,3	72 703	72 287,9	70 799,19	36 045,94	32 341,35	43 729,29	38 692,81	34 038,31	33 942,57	48 988,62	46 216,19
20-250-FL	132 996	132 438	132 394,4	133 404	129 485,3	54 144,64	63 245,12	70 325,74	81 370,79	51 239,48	57 273,9	71 397,91	81 288,1
21-250-PS	111 334,4	111 817,8	109 449,4	109 355,8	110 440,3	62 245,05	71 066,44	74 931,5	83 992,55	72 198,2	69 410,97	86 624,32	93 568,56
22-250-PL	197 899,8	197 214,2	194 131,2	194 549	194 338,6	112 454,5	131 917,5	133 567,4	155 750,3	95 906,32	127 789,5	146 104	160 301,3
23-250-RS	162 410,8	162 448,2	158 164,4	159 129,6	158 414,6	99 900	90 042,33	109 717,8	93 812,47	102 500,1	99 734,68	125 201,2	120 776
24-250-RL	269 072,8	268 713,8	266 559	267 291,4	266 204,2	148 156,7	176 579,9	180 807,7	177 336	146 270,4	164 012,2	204 037,4	203 859,8
Moyenne-250	157 829	157 512,9	155 566,9	156 003	154 947	85 491,14	94 198,78	102 179,9	105 159,2	83 692,14	92 027,3	113 725,6	117 668,3
Standard Déviation	69 184,1	69 165,71	68 327,76	68 666,69	68 777,74	42 022,43	52 050,37	49 795,51	51 632,15	40 174	48 193,98	56 662,62	57 036,7

Machine Translated by Google

Tableau 6. Suite

Ensemble de données	C1—Séquence VNS [9]	C2—Séquence VND [9]	C1 – Auto-même Adaptatif VND [9]	C2 – Auto-même Adaptatif VND [9]	IPPD [10]	K-Means-SAISIR-ABC(C)S1	K-Means-SAISIR-ABC(C)S2	Modifié K-Means-SAISIR-ABC(C)S1	Modifié K-Means-SAISIR-ABC(C)S2	ACO-SAISIR-ABC(C)S1	ACO-SAISIR-ABC(C)S2	ABC(S1)	ABC(S2)
25-500-FS	158 425,4	159 146,2	157 399,8	159 526,6	155 683,9	59 737,27	71 001,31	51 581,38	73 939,03	54 762,97	51 117,39	75 731,5	79 862,1
26-500-FL	243 210	244 591,6	245 600,8	245 183,6	242 173,1	101 753,4	120 779,3	106 505,3	129 200,4	62 041,19	88 256,7	112 172	121 824,9
27-500-PS	201 195,4	202 102,2	197 227,6	198 740	203 932,9	105 678,8	146 631,6	107 797,6	143 918,5	102 996,9	105 667,8	130 361,2	143 638,1
28-500-PL	319 239,2	321 181,6	315 854,6	317 610,8	329 642,8	229 525,6	241 331,4	225 812,4	243 361,8	152 573,5	191 169,5	201 973,8	222 042,1
29-500-RS	271 154,4	272 715,2	271 264,2	273 095	260 505,8	126 298,6	156 901,1	108 998,5	162 660,4	127 089,1	123 505,3	166 111,6	166 226,9
30-500-RL	402 557,2	406 689,6	399 990,8	404 517	408 819,4	259 683,6	272 509	241 454	282 087	167 819,5	205 885,1	239 675,7	255 261,1
Moyenne-500	265 963,6	267 737,7	264 556,3	266 445,5	266 793	147 112,9	168 192,3	140 358,2	172 527,9	111 213,8	127 600,3	154 337,6	164 809,2
Standard Déviation	69 184,1	69 165,71	68 327,76	68 666,69	68 777,74	42 022,43	52 050,37	49 795,51	51 632,15	40 174	48 193,98	56 662,62	57 036,7
31-1000-FS	150 920,4	151 943,2	151 990	153 820,6	152 358,7	33 184,03	48 758,35	39 530,22	48 457,57	23 266,5	20 723,05	33 125,19	41 340,89
32-1000-FL	218 211	218 213	223 999,8	222 072,8	222 089,1	92 970,35	93 386,23	92 142,15	95 500,88	24 110,85	44 359,35	57 137,24	68 705,69
33-1000-PS	439 681,8	441 120,2	433 353,8	429 853	457 759,4	272 126,9	288 095,1	236 797,1	280 912,6	178 654,3	201 178,4	240 944,9	263 540,2
34-1000-PL	678 396,2	651 754,6	677 511,6	669 212,8	681 370,5	498 835,3	519 115,5	492 753,1	507 105,4	346 244,9	366 004,4	397 930	426 372,5
35-1000-RS	455 054,6	461 394	459 167,4	462 999,2	461 919,6	207 134	255 835,4	192 030,8	242 210,5	124 538,1	176 577,1	212 979,1	228 716
36-1000-RL	668 655	662 605,4	668 615,4	667 184,8	676 159,7	464 549,3	450 834,3	407 703,7	448 210,5	297 212,8	322 568,7	364 643,1	372 491,4
Moyenne-1000	435 153,2	431 171,7	435 773	434 190,5	441 942,8	261 466,7	276 004,1	243 492,9	270 399,6	165 671,2	188 568,5	217 793,2	233 527,8
Standard Déviation	219 981,4	212 861,8	218 443,1	216 320	221 371,6	190 345	187 196	176 803,2	183 520,7	135 712,2	140 506,7	151 318,2	155 901,2

Tableau 7. Comparaison entre les algorithmes proposés et les méthodes précédentes basées sur la meilleure solution et les valeurs d'écart.

Ensemble de données	Meilleur Connu Profit	C1—Séquence VNS [4]		C2—Séquence VND [4]		C1—VND auto-adaptatif [4]		C2—VND auto-adaptatif [4]		IPPD [10]		K-Means-GRASP-ABC(C)S1	
		Meilleur Profit	Écart	Meilleur Profit	Écart	Meilleur Profit	Écart	Meilleur Profit	Écart	Meilleur Profit	Écart	Meilleur Profit	Écart
1-20-FS	30 315,5	30 315,5		30 315,5		30 315,5		30 315,5		30 315,5		30 267,75	−0,15
2-20-FL	39 340,3	38 765,6	0	38 765,6	0	38 765,6	0	38 765,6	0	39 179,66	0	37 750,35	−4
3-20-PS	52 855,2	51 473,4	−1,5 2,7	52 561,1	−1,5	52 565,1	−1,5	52 565,1	−1,5	52 566	−0,41	51 515,53	−2,5
4-20-PL	58 365,6	58 365,6	0	58 365,6	−0,6 0	58 365,6	−0,6 0	58 365,6	−0,6 0	58 777	−0,55 0,7	57 350,35	−1,7
5-20-RS	36 469,5	36 469,5	0	36 469,5	0	36 469,5	0	36 469,5	0	36 469,5	0	34 090,58	−6,5
6-20-RL	43 764,3	43 368,6	−0,9	43 368,6	−0,9	43 368,6	−0,9	43 368,6	−0,9	43 781,81	0,04	42 353,35	−3,2
Moyenne-20	43 518,4	43 126,37 −0,85 22		43 307,65 −0,50 22		43 308,32	−0,50	43 308,32	−0,50	43 514,91	−0,04	42 221,32	−2,9
7-50-FS	22 441	306 −0,6 58 050,8		306 −0,6 58 050,8		22 306	−0,6	22 306	−0,6	22 307,15	−0,6	18 662,83	−16,8
8-50-FL	58 050,8		0		0	58 050,8	0	58 050,8	0	58 050,8	0	51 040,14	−12

Machine Translated by Google

Tableau 7. Suite

Ensemble de données	Meilleur Connu Profit	C1—Séquence VNS [4]		C2—Séquence VND [4]		C1—VND auto-adaptatif [4]		C2—VND auto-adaptatif [4]		IPPD [10]		K-Means-GRASP-ABC(C)S1	
		Meilleur Profit	Écart	Meilleur Profit	Écart	Meilleur Profit	Écart	Meilleur Profit	Écart	Meilleur Profit	Écart	Meilleur Profit	Écart
9-50-PS	57,651.7 57,651.7 132,648	0		57 651,7	0	57 651,7		57 651,7		57 651,7		58 996,17	2.3
10-50-PL	132,648 33,340.4	0		132 648	0	132 648		132 648		132 648		114 962,9	−13,3
11-50-RS	33,340.4 79,350.3 79,350.3	0		33 340,4	0	33 340,4		33 340,4		33 340,4		28 907,33	−13,2
12-50-RL	63,913.7 63,891.20 −0.1	0		79 350,3	0	79 350,3		79 350,3		77 832,56	0 0 0	76 479,84	−3,6
Moyenne-50	44,402.3 42,174.4 −5.3 81,999.4			63 891,20 −0,10 42		63 891,20	0 0 0 0	63 891,20	0 0 0 0	63 638,44	−1,95	58 174,87	−8,9
13-100-FS	81,999.4 0 71,512.4 70,836 −1 125,032			174,4 −5,3 81 999,4 0		42 174,4	−0,10	42 215,5	−0,10	41 956,25	−0,43	36 322,94	−18,1
14-100-FL	124,770 80,878.8 80,878.8 131,845			71 326,4 −0,3 124		81 621	−5,3	80 848,1	−5,2	88 891,4		60 767,96	−25,8
15-100-PS	131,473 89,278.32 97,991.44 −1.13			770 78 818 131 845		71 399,2	−0,5 −0,2	71 399,2	−0,2 −0,2	71 234,58	−5,83 0 −0,39	54 428,34	−23,8
16-100-PL	75,309.1 73,610.9 −2.3	−0,2		88 488,87	−0,2	123 766	−1	124 910	−0,1	125 786,7	−0,6	99 860,9	−20,1
17-100-RS	136,594 134,889 113,886	0		−1. 40	−2,6	80 878,8	0	80 878,8	0	80 878,8	0	61 647,98	−23,7
18-100-RL	112,427 200,402 200,402	−0,3		73 575,7	0	131 339	−0,4	131 339	−0,4	131 845	0	102 403,8	−22,3
Moyenne-100	164,070 162,960 274,563 271,185			−2,4 134 186 112 480		88 529,73	−1,23	88 598,43	−1,02	90 098,79	−1,14	69 238,65	−22,4
19-250-FS	160,804 159,245.65 − 1,13 167 080			198 611 164 070 271		73 835,5	2	73 544,7	−2,4	72 038,6	−4,54	39, 672,94	−47,3
20-250-FL	161 063 −3,7 26-500-FL	−1,3		193 159	−1,8	132 870	−2,8	135,62	−0,7	1 31 012,9	−4,26	54 144,64	−60,3
21-250-PS	259 319 249 171 27-500-	−1,3		019,28	−1,3	112 428	−1,3	2 110	−3,5	112 747,3	−1,01	69 574,44	−38,9
22-250-PL	PS 206 788 203 832 −1,5	0		−1,27	−0,9	195 135	−2,7	062 196	−2	198 241,2	−1,09	112 454,5	−43,8
23-250-RS	28-500-PL 331 291 321	−0,7		161 040	0	162 115	−1,2	524 162	−1,2	161 804,7	−1,4	103 456	−36,9
24-250-RL	471 −3,1 29-500-RS	−1,2		−3,8 249	−1,2	272 957	−0,6	115 269	−1,8	269 734,7	−1,79	148 156,7	−46
Moyenne-250	281,7 11 279 105 −0,9 30-500-RL 417			661 205 480 −0,6 −1,8		158 223,42	−1,10	790 174	−1,93	157 596,57	−2,35	87 909,87	−45,3
25-500-FS	282 407 689 −2,4 Moyenne- 500 277			325 522 278 233 −		159 064	−5	822,60	−3,5	160 592,1	−4.04	59 737,27	−64,2
245,2 270 388,50 −2,61 31-1000-FS 173 826		−4.1		1,3 409	−3,9	247 986	−4,6	161 375	−4	249 105,7	−4.1	107 852,1	−58,4
154 196 −12,7 32-1000-FL 247 385 221 958 −11,5 33-1000-				755 −1,8 271 615,17		200 764	−3	249 312	−2,6	208 287,7	0,72	105 678,8	−48,8
PS 464 580 450 349 −3,2 34-1000-PL 712 918 693 407 −2,8				151 943,2	− 2,20	320 872	−3,2	201 590	−3,2	337 358,9	1,82	229 525,6	−30,7
35-1000-RS 492 083 463 625 −6,1 36-1000-RL 705 167 673				−12,2 221 322 −11,8		274 419	−2,7	321 155	−1,9	265 040	−6,29	126 298,6	−55,1
951 −4,6 Moyenne-1 000 465 993,2 405 958 −6,8				450 337 −3,2 672 910		407 849	−2,3	276 392	−2	411 967,6	−1,29	259 683,6	−37,7
				−5,9 472 626 −4,1 669		268 492,33	−3,47	409 262 269	−2,87	272,05 8,67	−2,20	165 807,7	−40,1
				926 −5,3 439 844,03		154 357	−12,6	847,67	−11,4	152 358,7	−11,8	33 184,03	−80,9
				−7,08		224 999	−9,9	156 106	−9	222 089,1	−10,1	92 970,35	−62,4
						442,31	−5	226 913,	−5,1	457 759,4	1,47	272 126,9	−41,4
						1 684	−4,2	218 683	−4,3	681 370,5	−3,26	498 835,3	−30
						121 465	−5,8	211 469	−4,9	461 919,6	−4,90	207 134	−57,9
						022 676	−4,4	162 684	−3,1	676 159,7	−2,90	464 549,3	−34,1
						720 441 255	−6,98	217 443 638	−6,3	441 942,83	−5,2	261 466,65	−43,8

Machine Translated by Google

Tableau 7. Suite

Ensemble de données	K-Means-GRASP-ABC(C)S2		K-Means modifiés-GRASP-ABC(C)S1		K-Means modifiés-GRASP-ABC(C)S2		ACO-GRASP-ABC(C)S1		ACO-GRASP-ABC(C)S2		ABC(S1)		ABC(S2)	
	Meilleur profit	Écart	Meilleur profit	Écart	Meilleur écart de profit		Meilleur écart de profit		Meilleur profit	Écart	Meilleur écart de profit		Meilleur écart de profit	
1-20-FS	30 427,58	0,3	27 007,58	−10,9	27 007,58	−10,9	25 490,76	−15,9	25 490,76	−15,9	30 554,62	0,78	30 315,52	0
2-20-FL	37 556,07	−4,5	40 104,23	1.9	39 240,95	−0,2	36 423,3	−7,4	36 322,33	−7,6	40 104,23	1.9	36 234,48	−7,8
3-20-PS	44 334,2	−16,1	51 515,53	−2,5	44 334,2	−16,1	42 944,78	−18,7	34 107,64 47	−35,4	53 164,53	0,58	52 565,1	−0,5
4-20-PL	57 350,35	−1,7	57 350,35	−1,7	57 350,35	−1,7	56 603,63	−18,1	−3	402,91 31	−18,7	59 140,98	1.32	57 286,89	−1,8
5-20-RS	34 090,58	−6,5	30 827,18	−15,4	29 843,58	31	741,08 0,04 41	−12,9	741,08 40	−12,9	36 469,52		36 469,52	0
6-20-RL	42 169,13	−3,6	44 764,07	2,2	43 780,72	071,62	−7,4 39 045,86	−6,1	600,21 35	−7,2	44 707,23	0	42 873,71	−2
Moyenne-20	40 987,99	−5,8	41 928,16	−3,6	40 259,56	−3,7	21 281,88	−12,3	−10,2	944,16 25	−17,4	44 023,52	2,1	42 624,2	−2
7-50-FS	21 603,83	−3,7	21 603,83	−3,7	21 603,83	49 128,45	−7,9 55	−5,1	123,15 50	11,9	24 975,25	1,16	24 998,45	11.3
8-50-FL	51 001,04	−12,1	51 001,04	−12,1	50 896,3	,643,25	−11,8 100 393,3	−15,3	061,24 55	−13,7	53 710,99	11,2	54 288,08	−6,4
9-50-PS	54 165,39	−6	54 165,39	−6	53 064,14	−13,2	31 251,34 −3,4	−3,4	643,25 106	−3,4	57 651,71		57 651,71	0
10-50-PL	119 099,6	−10,2	119 099,6	−10,2	116 868,6	65 792,21	−9,2 53	−24,3	544,8 29	−19,6	126 360	−7,4 0	−4,7 133 743,2	0,82
11-50-RS	28 907,33	−13,2	28 907,33	−13,2	28 907,33	915,07	−13,6 33 565,46	−6,2	969. 16 65	−10,1	33 058,21	−0,8	33 159,12	−0,5
12-50-RL	76, 697,65	−3,3	,697,65 58	−3,3	04,84 57	−25,1	52 328,66 60	−17	752,8 55	−17,1	67,34 62	−2,3	76,5 17,36	−3,5
Moyenne-50	58 579,14	−8,3	579,14 35	−8,3	990,84 38	614,78	−10,6 95 836,25	−15,6	515,73 33	−13,1	203,92 41	−2,6	63 392,99	−0,8
13-100-FS	36 188,32	−18,4	937,96 60	−19	323,6 61	−23 . 3	52 689,7 −16 95	−24,4	566,06 49	−24,4	812,08 66	−5,8	41 636,8	−6,2
14-100-FL	63 488,24	−22,5	851,27 54	−25,7	351,99 60	363,69		−36,1	789,52 47	−39,2	375,92 69	−19	72 595,41	−11,4
15-100-PS	60 815,37	−14,9	190,18 102	−24,2	554,49 111	−15,3		−15,2	498,28 101	−33,5	275,79 115	−3.1	71 400,63	−0,1
16-100-PL	104 786,4	−16,1	121,7 61	−18,3	728,2 61			−23,3	142,5 52	−19,1	579,2 77	−7,5	128 473,2	2.7
17-100-RS	61 647,98	−23,7	962,79 105	−23,3	962,79 110			−34,8	689,7 97	−34,8	484,06 119	−4.1	77 484,06	−4.1
18-100-RL	101 013,6	−23,3	175,3	−20,2	683,8			−27,6	568,94	−25,9	554,3	−9,3	126 871,5	−3,7
Moyenne-100	71 323,32	−20,1	70 039,87	−21,5	74 100,81	−17 65 066,42		−27,1	63 709,17	−28,6	81 680,23	−8,5	86 410,27	−3,2
19-250-FS	35 443,33	−52,9	48 079,56	−36,1	39 848,03	−47 36 309,86	−35,8 51	−51,7 36 612,19	−62,4	−51,3	36 309,86	−51,7 36 612,19	−62,4	−51,3
20-250-FL	67 014,68	−50,9	70 325,74	−48,5	87 575,52	239,48 73 654,78	−19,4	62 831,21 73 143,68		−54	51 239,48	62 831,21	−35,3 73	−54
21-250-PS	77 941,82	−31,5	81 510,78	−28,4	92 124,73	−19,1	95 906,32	−35,3	131 668,8	−35,7	73 654,78	143,68	−52,1 131 668,8	−35,7
22-250-PL	135 164,3	−32,5	133 567,4	−33,3	161 377,2	−41,8 107 628,9	146	−52,1	101 863,1	−34,2	95 906,32	−34,4 101 863,1	−46,7	−34,2
23-250-RS	97 231,36	−40,7	113 547,4	−30,7	95 471,16	270,4		−34,4	−46,7 166	−37,9	107 628,9	166 523,5		−37,9
24-250-RL	187 891,9	−31,5	180 807,7	−34,1	180 803	−34,1		523,5		−39,3	146 270,4			−39,3
Moyenne-250	100 114,6	−37,7	104 639,8	−34,9	109 533,3	−31,8	85 168,2	−47	95 440,4	−40,6	85 168,2	−47	95 440,4	−40,6
25-500-FS	74 692,46	−55,2	51 581,38	−69,1	76 481,48	−54,2 61 718,84	−49,2	−63	54 247,37	−67,5	78 432,78	−53 86 250,21	−56,7	−48,3
26-500-FL	124 478,4	−51,9	106 505,3	−58,9	131 572,4	72 388,37	−27,1 109 238	−72	93 308,66	−64	112 172	126 203,2	−34,1 148	−51,3
27-500-PS	152 566,6	−26,2	107 797,6	−47,8	150 645,3			−47,1	112 111,9	−45,7	136 168,2	384,5		−28,2

Machine Translated by Google

Tableau 7. Suite

Ensemble de données	K-Means-GRASP-ABC(C)S2		K-Means modifiés-GRASP-ABC(C)S1		K-Means modifiés-GRASP-ABC(C)S2		ACO-GRASP-ABC(C)S1		ACO-GRASP-ABC(C)S2		ABC(S1)		ABC(S2)	
	Meilleur profit	Écart	Meilleur profit	Écart	Meilleur écart de profit		Meilleur écart de profit		Meilleur profit	Écart	Meilleur écart de profit		Meilleur écart de profit	
28-500-PL	249 759,7	−24,6	225 812,4	−31,8	247 180,6	−25,3	152 573,5	−53,9	196 969,2	−40,5	201 973,8	−39	225 040,5	−32
29-500-RS	172 397,2	−38,8	108 998,5	−61,3	175 267,1	−37,7	141 743,7	−29,4	−49,6	128 570,6	−54,3	178 693,9	−36,5	173 003,3	−38,5
30-500-RL	275 488,2	−33,9	241 454	−42,1	294 424,7	167 819,5		−59,7	214 081,3	−48,6	239 675,7	−42,5	261 517,2	−37,3
Moyenne-500	174 897,1	−36,9	140 358,2	−49,3	179 261,9	−35,3	117 580,3	−57,5	133 214,8	−51,9	157 852,7	−43	170 066,5	−38,6
31-1000-FS	51 188,29	−70,5	39 530,22	−77,2	51 585,38	−70,3	23 266,5	−60,4	−86,6	25 563,9	−85,2	33 125,19	−80,9	42 197,08	−76,9	−75,7
32-1000-FL	97 093,51	−60,7	92 142,15	−62,7	97 772,35	24 110,85	−38	178 654,3	−90,2	47 620,03	−80,7	57 137,24	73	126,9	−48,1	267	−70,4
33-1000-PS	293 900,2	−36,7	236 797,1	−49	287 774,9	−24,9	346 244,9	−49,3	−61,5	209 372,1	−54,9	240 944,9	814,2	−44,1	430 894,4	−42,3
34-1000-PL	536 199,8	−24,7	492 753,1	−30,8	534 689,9	124 538,1	−34,2	297	−51,4	381 116,4	−46,5	397 930	−56,7	232 298,7	−48,2	−39,5
35-1000-RS	264 295,2	−46,2	192 030,8	−60,9	249 078,6	212,8			−74,6	181 237	−63,1	212 979,1	380 268,9			−52,7
36-1000-RL	458 126,7	−35	407 703,7	−42,1	463 915,7				−57,8	335 935,6	−52,3	364 643,1				−46
Moyenne-1000	276 004,15	−40,7	243 492,8	−47,7	280 802,8	−39,7	165 671,2		−64,4	196 807,5	−57,7	217 793,3	−53,2	237 766,7		−48,9

Machine Translated by Google

Le profit le plus connu a été calculé dans [9] en exécutant chaque instance pendant 20 h, puis le meilleur profit a été enregistré en tant que profit le plus connu. Comme le montrent les tableaux 6 et 7, les algorithmes proposés fonctionnent bien en termes de profit pour les instances de petite taille, l'ABC(S1) renvoyant six nouvelles meilleures solutions et deux autres solutions égales aux solutions les plus connues. L'ABC(S2) a également obtenu trois nouvelles meilleures solutions et trois solutions égales à la solution la plus connue. En outre, la K-means-GRASP(C)-ABC(C)S1 modifiée a permis d'obtenir deux nouvelles meilleures solutions. De plus, les K-means-GRASP(C)-ABC(C)S1, K-means-GRASP(C)-ABC(C)S2, K-means-GRASP(C)-ABC(C)S2 modifié, et ACO-GRASP(C)-ABC(C)S2 ont chacun produit une nouvelle meilleure solution. Dans l'ensemble, les performances générales de tous les algorithmes proposés étaient acceptables et proches des algorithmes précédents en termes de profit moyen sur ces instances.

Cependant, pour les instances de taille moyenne et grande, les algorithmes proposés n'ont pas donné de bons résultats par rapport aux méthodes précédentes, comme en témoigne le pourcentage d'écart moyen élevé par rapport aux résultats les plus connus présentés dans le tableau 6. Il existe de nombreuses raisons possibles pour cela. ce:

• Encore une fois, les méthodes utilisées dans la phase de reconnaissance de l'ABC semblent avoir un effet sur les performances des algorithmes proposés. La première est la stratégie aléatoire et gourmande, qui sélectionne une paire de clients aléatoire, ce qui signifie que la pire paire (soit avec de faibles revenus , soit séparée par une grande distance) peut être sélectionnée, ce qui entraîne une mauvaise qualité de solution. De plus, dans certains cas, les opérateurs de quartier ne peuvent pas améliorer suffisamment la qualité de la solution en raison de dépassements de temps ou de contraintes de charge. La deuxième méthode utilise le GRASP, qui construit des solutions en sélectionnant au hasard dans la liste restreinte de candidats (RCL), qui fait la moitié de la taille de l'ensemble de solutions candidates (CSS). Ainsi, la sélection des clients non visités n'est pas un problème pour les instances de petite taille car le nombre de paires de clients est faible, ce qui signifie que les chances de sélectionner la meilleure paire de clients (c'est-à-dire celle avec la valeur IR2 la plus élevée) sont élevées. En revanche, dans les instances de taille moyenne et grande, où la taille du RCL est grande, les chances de sélectionner la meilleure paire de clients diminuent, ce qui affecte la qualité des solutions. • Le nombre de solutions dans une population, le nombre d'itérations et la valeur limite jouent des rôles importants dans les performances des algorithmes. Dans nos algorithmes, nous nous appuyons sur un petit nombre de solutions et d'itérations, car un temps de traitement rapide est une exigence critique pour les applications de commerce mobile quotidiennes. Ces diminutions du nombre d'itérations et de la taille de la population réduisent cependant les chances d'obtenir une qualité de solution satisfaisante.

• Tous les algorithmes précédents étaient basés sur VNS, qui a la capacité de modifier la taille de la recherche de voisinage chaque fois que la qualité de la solution n'est pas améliorée. Ainsi, l'expansion systématique de la structure du quartier aide le VNS à explorer pleinement le quartier et à obtenir de bonnes solutions. En revanche, la fonctionnalité de diversification de notre approche est basée soit sur la fonctionnalité gourmande (insérer la paire de clients avec l'IR le plus élevé dans la meilleure position), soit sur une combinaison de fonctionnalités gourmandes et aléatoires (insérer une paire de clients aléatoire dans la meilleure position ou insertion d'une paire de clients aléatoires de RCL dans la meilleure position). Cependant, ces stratégies semblent moins efficaces en matière de diversification que les VNS, qui s'appuient sur un mécanisme organisé d'exploration des quartiers.

La figure 3 compare nos algorithmes proposés sur la base de la moyenne des meilleures solutions pour chaque catégorie d'instances (lignes en surbrillance dans le tableau 6). Il est clair que l'ABC dans les deux versions (S1 et S2) a toujours la préférence sur tous les autres algorithmes pour les instances comprenant 20, 50, 100, 250 et 500 clients. En revanche, les deux versions de K-means-GRASP(C)-ABC(C) et de K-means-GRASP(C)-ABC(C) modifié surpassent le GRASP(V2)-ABC dans les deux versions (S1 et S2) pour les instances de 1 000 clients. Les principales raisons des mauvaises performances des algorithmes de clustering dans les instances de petite et moyenne taille sont la répartition des paires de clients en K-means et K-means modifiés, où le processus de clustering peut aboutir à placer un petit nombre de clients dans un cluster. , ce qui signifie que le véhicule a encore du temps à servir, mais qu'il n'y a plus de clients

Machine Translated by Google

ff

ff

au sein de ce cluster. Ainsi, nous concluons que les algorithmes de clustering (sauf ACO) fonctionnent bien uniquement dans les instances de grande taille.

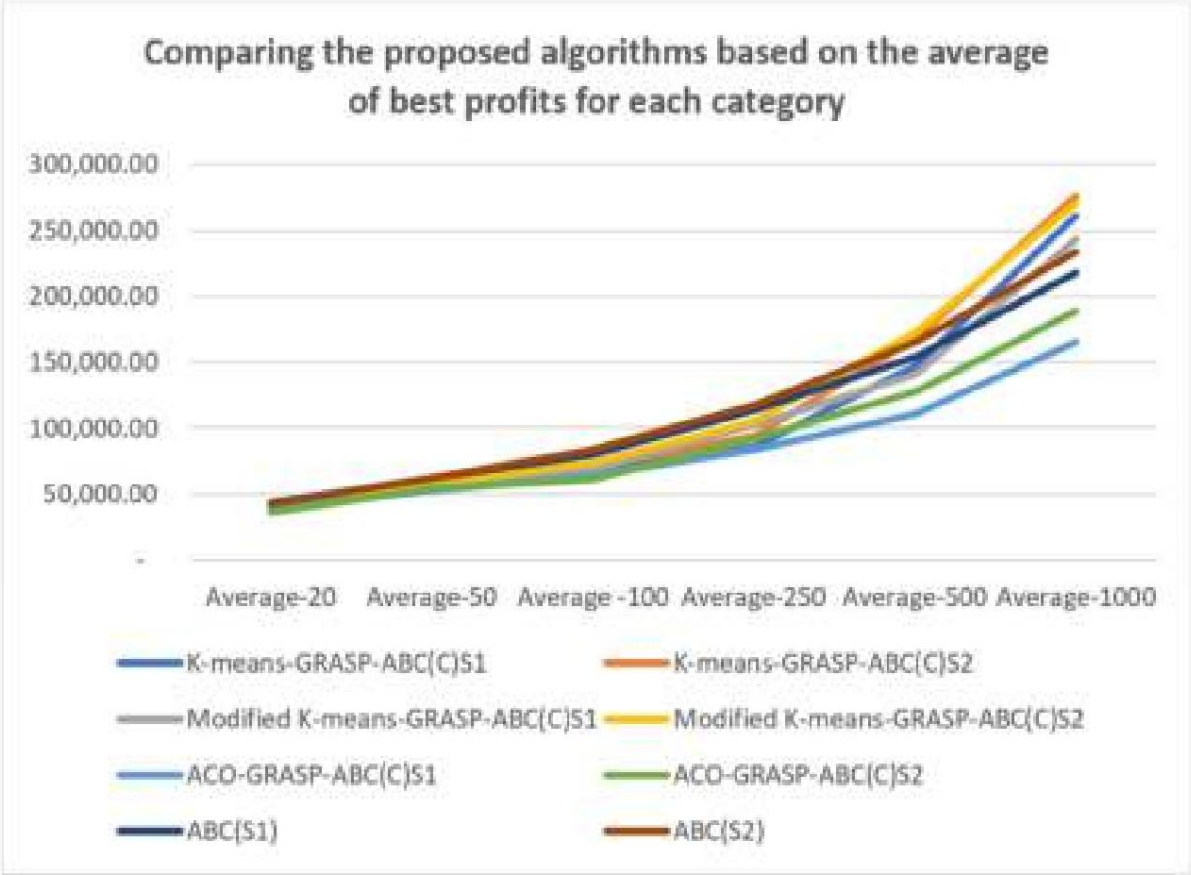


Figure 3. Comparaison des algorithmes proposés sur la base de la moyenne des meilleurs bénéfices pour chaque catégorie.

Concernant l'ACO-GRASP(C)-ABC(C)(S1 et S2), il est clair qu'il a les pires performances parmi tous les algorithmes. Encore une fois, la méthode de clustering de l'ACO a un effet significatif sur les résultats. Plus précisément, la fonction aléatoire d'ACO contribue à un regroupement inexact des paires de clients.

De plus, comme le montrent les lignes mises en évidence dans le tableau 7, nous avons remarqué que la deuxième stratégie pour les abeilles éclaireuses surpasse la première dans de nombreux cas, car la première stratégie inclut la sélection soit des meilleures paires de clients, soit des paires aléatoires, et l' option de sélection aléatoire peut affectent la qualité de la solution, tandis que la deuxième stratégie sélectionne les paires de clients de manière aléatoire dans la RCL, qui contient les meilleures paires de clients dans le CSS.

6. Conclusions

Dans cet article, nous présentons huit nouvelles métaheuristiques pour améliorer les solutions initiales au problème de collecte et de livraison rentables multi-véhicules (MVPPDP). Nous avons proposé deux versions de l'ABC (Artificial Bee Colony) pour améliorer les solutions initiales : l'ABC avec GRASP et l'ABC (C) avec GRASP et clustering, c'est-à-dire K-means-GRASP (C), K-means-GRASP modifié. (C) et ACO-GRASP (C). Nous avons également deux stratégies distinctes pour la phase des abeilles scoutes. La première stratégie, S1, construit la solution en sélectionnant aléatoirement l'une des deux stratégies suivantes : insertion gourmande et aléatoire ou insertion simplement gourmande. La deuxième stratégie, S2, est basée sur le GRASP. Ainsi, nous avons huit méthodes différentes proposées. Par rapport aux méthodes de pointe en termes de temps de traitement, les huit méthodes ont la préférence dans toutes les instances (à l'exception des instances de 1 000 clients). De plus, tous les algorithmes proposés ont donné de bons résultats pour les instances de petite taille en termes de profits en atteignant treize

Machine Translated by Google

nouvelles meilleures solutions et cinq solutions égales aux solutions les plus connues. Cependant, les algorithmes sont légèrement à la traîne dans les instances de taille moyenne et grande en raison de la stratégie aléatoire gourmande et du GRASP qui ont été utilisés dans la phase scout bee. De plus, nos algorithmes donnent la priorité à des solutions et à des itérations minimales pour des temps de traitement rapides dans les applications de commerce mobile quotidiennes, tandis que la réduction du nombre d'itérations et de la taille de la population réduit la probabilité d'obtenir une bonne qualité de solution.

Pour les recherches futures, davantage d'opérateurs de quartier seront soigneusement étudiés lors de la phase d'exploration afin d'être applicables dans le cadre des contraintes du MVPPDP, augmentant ainsi les chances de produire des solutions de haute qualité. De plus, nous essaierons d'autres métaheuristiques basées sur la population, telles que les GA, les algorithmes évolutionnaires ou d'autres métaheuristiques basées sur les essaims, qui pourraient contribuer à de nouvelles solutions pour le MVPPDP.

Contributions des auteurs : Conceptualisation, AIA ; méthodologie, AIA; analyse formelle, AIA; conservation des données , AIA ; Rédaction – ébauche originale, AIA ; acquisition de financement, AIA; Supervision, MIH ; rédaction – révision et édition, MIH Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche n'a reçu aucun financement externe.

Déclaration de disponibilité des données : http://prolog.univie.ac.at/research/MVPPDP/instances.zip (consulté le 17 juillet 2024).

Remerciements : Les chercheurs souhaitent remercier le doyen des études supérieures et de la recherche scientifique de l'Université Qassim pour son soutien financier (QU-APC-2024-9/1).

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Références

1. Curry, D. Statistiques sur les revenus et l'utilisation des applications de livraison de nourriture. 2024. Disponible en ligne : https://www.businessofapps.com/data/ marché des applications de livraison de nourriture/ (consulté le 24 janvier 2024).
2. Khalid, H. Avantages de l'application de livraison de nourriture pour les restaurants et les clients. Disponible en ligne : https://enatega.com/benefits-of-application-de-livraison-de-nourriture/ (consulté le 1er février 2024).
3. Équipe, DJ Top Food Delivery Apps in Saudi Arabia 2023. Disponible en ligne : https://www.digitalgravity.ae/blog/top-food- delivery-apps-in-saudi-arabia/ (consulté le 2 janvier 2024).
4. Gansterer, M. ; Kuçüktepe, M. ; Hartl, RF Le problème rentable du ramassage et de la livraison multi-véhicules. OU Spectre. 2017, 39, 303-319. [Référence croisée]
5. Talbi, EG Métaheuristiques : de la conception à la mise en œuvre ; John Wiley & Sons : Hoboken, New Jersey, États-Unis, 2009 ; ISBN978-0-470-27858-1.
6. Alhujaylan, AI ; Hosny, MI Une approche de construction de solution basée sur GRASP pour le ramassage et la livraison rentables de plusieurs véhicules problème. Int. J.Adv. Calculer. Sci. Appl. 2019, 10, 111-120. [Référence croisée]
7. Alhujaylan, AI ; Hosny, MI Algorithmes de clustering hybrides avec GRASP pour construire une solution initiale pour le MVPPDP. Calculer. Maître. Continuer. 2020, 62, 1025-1051. [Référence croisée]
8. Archetti, C. ; Speranza, MG; Vigo, D. Problèmes d'acheminement des véhicules avec bénéfices. Dans Routage des véhicules : problèmes, méthodes et applications, 2e éd. ; SIAM : Philadelphie, PA, États-Unis, 2014 ; Chapitre 10 ; pp. 273-297. ISBN978-1-61197-358-7.
9. Toth, P. ; Vigo, D. Routage des véhicules : problèmes, méthodes et applications, 2e éd. ; Société de mathématiques industrielles et appliquées : Philadelphie, Pennsylvanie, États-Unis, 2015 ; ISBN1611973589.
10. Küçüktepe, M. Un algorithme général de recherche de quartier variable pour le problème de ramassage et de livraison rentable multi-véhicules. Université de Vienne 2014. Disponible en ligne : https://scholar.google.com/scholar?hl=en&as_sdt=0,5&q=A+General+Variable+Algorithme+de+recherche+de+quartier+pour+le+multi-véhicule+rentable+ramassage+et+livraison+problème.+&btnG= (consulté le 22 juin 2024).
11. Haddad, MN Une heuristique efficace pour les problèmes de ramassage et de livraison individuels. doctorat Thèse, IC/UFF, Niterói, Brésil, 2017.
12. Bruni, MOI ; Toan, DQ Le problème d'acheminement rentable de ramassage et de livraison multi-véhicules avec des temps de trajet incertains. Transp. Rés. Procedia 2021, 52, 509-516. [Référence croisée]
13. Karaboga, D. Une idée basée sur Honey Bee Swarm pour l'optimisation numérique. Rapport technique-tr06, Université Erciyes, Faculté d'ingénierie, Département de génie informatique. 2005. Disponible en ligne : https://scholar.google.com/scholar?hl=en&as_ sdt=0,5&q=Une+idée+basée+sur+du+miel+abeille+essaim+pour+l'optimisation+numérique&btnG= (consulté le 22 juin 2024).
14. Alazzawi, AK ; Rais, HM; Basri, S. ; Alsariera, YA PhABC : Une stratégie de colonie d'abeilles artificielles hybrides pour la génération d'une suite de tests par paires avec prise en charge de contraintes. Dans Actes de la conférence étudiante de l'IEEE 2019 sur la recherche et le développement (SCOReD), Bandar Seri Iskandar, Malaisie, 15-17 octobre 2019 ; pp. 106-111, ISBN 9781728126135.
15. Szeto, Wyoming ; Wu, Y. ; Ho, SC Un algorithme de colonie d'abeilles artificielles pour le problème d'acheminement des véhicules capacitifs. EUR. J. Opér. Rés. 2011, 215, 126-135. [Référence croisée]

Machine Translated by Google

16. Wu, B. ; Shi, Z. Un algorithme de clustering basé sur l'intelligence en essaim. Dans Actes des conférences internationales 2001 sur Info- Tech et Info-Net. Actes (Cat. No.01EX479), Pékin, Chine, 29 octobre-1er novembre 2001 ; pp. 58-66. ISBN0-7803-7010-4.
17. Yao, B. ; Hu, P. ; Zhang, M. ; Wang, S. Algorithme de colonie d'abeilles artificielles avec stratégie de numérisation pour l'itinéraire périodique des véhicules problème. Simulation 2013, 89, 762-770. [Référence croisée]
18. Karaboga, D. ; Akay, B. Une enquête : algorithmes simulant l'intelligence des essaims d'abeilles. Artif. Intell. Rév.2009 , 31, 61-85. [Référence croisée]
19. Creutz, M. Simulation microcanonique de Monte Carlo. Phys. Le révérend Lett. 1983, 50, 1411. [Réf. croisée]
20. Bansal, JC ; Sharma, H. ; Jadon, SS Algorithme de colonie d'abeilles artificielles : une enquête. Int. J.Adv. Intell. Paradig. 2013, 5, 123-159. [Référence croisée]

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Preptimize : automatisation du prétraitement et des prévisions des données de séries chronologiques

Mehek Usmani [1], Zulfiqar Ali Memon [1], Adil Zulfiqar [2] et Rizwan Qureshi [3,*]

[1] Fast School of Computing, Université nationale d'informatique et des sciences émergentes, Karachi 65200, Pakistan ; k200984@nu.edu.pk (MU); zulfiqar.memon@nu.edu.pk (ZAM)
[2] Département de génie électrique, Université nationale d'informatique et des sciences émergentes, Campus de Faisalabad, Faisalabad 38000, Pakistan ; adil.zulfiqar@nu.edu.pk
[3] Centre de médecine régénérative et de santé, Institut des sciences et de l'innovation de Hong Kong, Académie chinoise des sciences, Parc scientifique, Hong Kong SAR 999077, Chine
* Correspondance : engr.rizwanqureshi786@gmail.com

Résumé : L'analyse des séries chronologiques est essentielle à la prise de décision commerciale et financière, en particulier avec l'intégration croissante de l'Internet des objets (IoT). Cependant, l'exploitation des données de séries chronologiques à des fins de prévision nécessite un prétraitement approfondi pour résoudre des problèmes tels que les valeurs manquantes, l'hétéroscédasticité, la saisonnalité, les valeurs aberrantes et le bruit. Différentes approches sont nécessaires pour les séries temporelles univariées et multivariées, les séries temporelles gaussiennes et non gaussiennes et les séries temporelles stationnaires ou non stationnaires. La seule gestion des données manquantes est complexe et nécessite des solutions uniques pour chaque type. L'extraction de caractéristiques statistiques, l'identification des problèmes de qualité des données et la sélection de techniques de nettoyage et de prévision appropriées nécessitent des efforts, du temps et une expertise considérables. Pour rationaliser ce processus, nous proposons une stratégie automatisée appelée Preptimize, qui intègre des techniques statistiques et d'apprentissage automatique et recommande des modèles de modèles de prédiction, suggérant les approches les plus appropriées pour un ensemble de données donné comme première étape vers une analyse plus approfondie. Preptimize lit un échantillon d'un grand ensemble de données et recommande le modèle de plan basé sur l'optimisation, le rendant facile à utiliser même pour les non-experts. Les résultats de diverses expériences ont indiqué que Preptimize surperformait ou avait des performances comparables aux modèles de référence dans plusieurs secteurs, notamment les cours des actions, les cryptomonnaies et la prévision de la consommation d'énergie. Cela démontre l'efficacité du cadre à recommander des modèles de prévision appropriés pour divers ensembles de données de séries chronologiques, soulignant sa large applicabilité dans différents domaines de la prévisio

Mots-clés : automatisation ; optimisation; séries chronologiques ; univarié et multivarié ; techniques statistiques; apprentissage automatique ; prétraitement ; prévision; interpolation; imputation simple et multiple ; données manquantes

check for updates

Citation : Usmani, M. ; Mémon, ZA; Zulfiqar, A. ; Qureshi, R. Preptimize : Automatisation des données de séries chronologiques Prétraitement et prévision. Algorithmes 2024, 17, 332. https://doi.org/10.3390/a17080332

Rédacteur académique : Vittorio Maniezzo

Reçu : 18 juin 2024
Révisé : 22 juillet 2024
Accepté : 26 juillet 2024
Publié : 1 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Les données de séries chronologiques, caractérisées par des points de données séquentiels enregistrés à des horodatages spécifiques, présentent des défis uniques en science des données, exigeant des techniques de prétraitement spécialisées pour garantir la qualité des données. L'analyse des séries chronologiques englobe la collecte, le nettoyage , le prétraitement et la prévision des données, adaptés à différents types de séries chronologiques, qu'elles soient numériques ou textuelles, univariées ou multivariées, saisonnières ou non saisonnières. Les recherches existantes se concentrent principalement sur les séries chronologiques univariées, tandis que les séries chronologiques multivariées posent des complexités supplémentaires telles que le bruit, les valeurs manquantes, les valeurs aberrantes et la saisonnalité. Pour relever ces défis, un cadre de prétraitement automatisé intégrant diverses techniques et algorithmes d'apprentissage automatique est proposé. L'analyse des séries chronologiques, cruciale dans les domaines de l'IoT et de la finance, nécessite un nettoyage et un prétraitement approfondis des données en raison de ses caractéristiques et de ses défis distinctifs. Comprendre les propriétés des données, telles que la stationnarité et la distribution, est essentiel pour sélectionner les techniques de prétraitement appropriées. De plus , les algorithmes d'apprentissage profond tels que les GAN et les encodeurs automatiques montrent des résultats prometteurs en matière de prétraitement et de prévision des données de séries chronologiques.

Machine Translated by Google

La recherche du modèle de prévision le plus précis avec des taux d'erreur minimes dans l'analyse des séries chronologiques numériques nécessite une expérimentation exhaustive de diverses techniques et de leurs combinaisons. Le type de technique applicable à un type particulier de données doit également être connu [1]. Cependant, ce processus prend du temps et demande beaucoup de travail.

Différentes méthodes d'imputation sont utilisées en fonction des propriétés des séries chronologiques, telles que la distribution et la saisonnalité, pour gérer efficacement les valeurs manquantes. Avec la croissance exponentielle des données de séries chronologiques générées par diverses sources telles que les bourses [2], la consommation d'énergie [3], les systèmes de surveillance de la santé et les systèmes d'enregistrement météorologique, les analystes sont confrontés à des charges de travail croissantes. Les séries chronologiques, caractérisées par des données séquentielles enregistrées sur des intervalles de temps, présentent des défis uniques tels que le bruit de gestion, la saisonnalité et les valeurs manquantes, qui nécessitent des techniques de prétraitement spécialisées pour garantir la qualité des données et améliorer la précision des prévisions.

Le prétraitement, une étape cruciale dans l'analyse des données et l'apprentissage automatique, implique une série d'étapes visant à améliorer la qualité des données et à simplifier la formation des modèles. Cependant, le prétraitement prend beaucoup de temps, ne suit pas une séquence prédéfinie et varie selon les domaines. Dans le contexte de l'analyse de séries chronologiques, le prétraitement devient plus complexe en raison de la nature en temps réel et des dépendances séquentielles des données. Les défis incluent la gestion des données manquantes, en double et incorrectes, y compris les valeurs aberrantes. La littérature existante se concentre principalement sur la gestion des données manquantes et la détection des valeurs aberrantes lors du prétraitement des séries chronologiques.

Les valeurs aberrantes, qui s'écartent considérablement de la majorité des points de données, peuvent être univariées ou multivariées, ces dernières posant une plus grande complexité. La détection multivariée des valeurs aberrantes nécessite souvent des approches d'apprentissage automatique en raison de l'espace de fonctionnalités multidimensionnel impliqué. Les techniques de détection des valeurs aberrantes englobent des méthodes statistiques et d'apprentissage automatique. Les techniques statistiques incluent le lissage exponentiel, le score z, les modèles de régression linéaire et l'intervalle interquartile, tandis que les techniques d'apprentissage automatique comprennent des méthodes supervisées telles que les réseaux neuronaux et les arbres de décision, et des méthodes non supervisées telles que l'analyse groupée.

Des techniques de mesure de distance, des méthodes basées sur la théorie de l'information et des algorithmes tels que KNN et COF sont également utilisés.

La gestion des valeurs manquantes dans les données de séries chronologiques implique soit de supprimer les enregistrements incomplets, soit d'imputer les valeurs manquantes. Les techniques d'imputation pour les données de séries chronologiques, y compris l'interpolation spline et les moyennes mobiles, diffèrent de celles utilisées dans les données non chronologiques en raison des propriétés temporelles. Les méthodes traditionnelles peuvent ne pas suffire, nécessitant des techniques avancées telles que l'interpolation, les encodeurs automatiques et les réseaux contradictoires génératifs (GAN) pour l'imputation et la détection des valeurs aberrantes.

L'analyse des données de séries chronologiques prend en compte des propriétés telles que la saisonnalité et la tendance, en classant les séries chronologiques en types en fonction de leurs caractéristiques. Les défis liés aux données de séries chronologiques s'étendent au-delà des valeurs aberrantes et des valeurs manquantes pour inclure la dérive des concepts, le déséquilibre des classes, la tendance de traitement, ainsi que les problèmes de saisonnalité et de cohérence des données.

Selon [4], l'analyse des séries chronologiques est principalement réalisée à l'aide de méthodes statistiques ; cependant, les techniques d'apprentissage automatique ont pris de plus en plus d'importance dans l'analyse de séries chronologiques , promettant des résultats plus précis avec un prétraitement adéquat. Cependant, le prétraitement reste vital pour améliorer la qualité des données et garantir des prédictions fiable

Des techniques telles que SCREEN [6] offrent des méthodes de nettoyage basées sur des contraintes pour les données de séries chronologiques, améliorant ainsi la précision et l'efficacité de l'analyse des données. Des études de recherche ont exploré diverses approches pour améliorer les techniques de prétraitement des données de séries chronologiques. Une étude [7] a utilisé le méta-apprentissage pour sélectionner des alternatives de prétraitement, en se concentrant sur les effets de transformation sur les performances des algorithmes sur des ensembles de données tels que l'automobile, le crédit, l'iris et le vote.

Bien que les arbres de régression aient montré une efficacité limitée en tant que méta-apprenant, les transformations ont considérablement amélioré les résultats. Une autre étude [8] a introduit une méthode itérative de réparation minimale et de prédiction des paramètres pour la réparation des anomalies, permettant une amélioration substantielle de la vitesse d'exécution sur les ensembles de données, notamment le GPS, les données de laboratoire Intel et les données météorologiques de Chine . Dans une enquête distincte [9], des séries chronologiques Landsat combinées à une analyse de trajectoire ont été utilisées pour détecter les instabilités forestières au Myanmar. Cette étude a mis en évidence la corrélation entre les zones d'instabilités et les arbres récoltés, indiquant

Machine Translated by Google

applications potentielles de l'exploitation forestière sélective. La détection des changements dans les séries temporelles a été analysée de manière approfondie [10], couvrant les techniques de prétraitement telles que la correction atmosphérique et la détection des nuages, ainsi que les algorithmes de prétraitement automatisés populaires tels que LEDAPS et Fmask. L'augmentation des réseaux entièrement convolutifs avec LSTM-RNN a été proposée [11] pour la classification des séries temporelles, démontrant des performances améliorées avec les mécanismes d'attention. De plus, des études [12] ont abordé les problèmes de profilage des données, proposant des solutions logicielles visuelles pour les contrôles de qualité des données et des techniques de détection des valeurs aberrantes en utilisant des méthodes telles que la technique de la fenêtre coulissante K et les réseaux de neurones [13].

Une revue [14] a souligné la nécessité de recherches plus approfondies sur les outils et systèmes de nettoyage des séries chronologiques, soulignant l'importance du prétraitement pour maximiser le potentiel des données. Les domaines potentiels de travaux futurs comprennent l'analyse des types de données d'erreur, les algorithmes de nettoyage de séries chronologiques multidimensionnelles, la détection d'anomalies et les techniques de nettoyage de données d'application spécifiques. Plusieurs systèmes innovants ont été proposés pour relever les défis liés au nettoyage des données de séries chronologiques dans divers domaines. Cleanits [15], par exemple, se concentre sur le nettoyage des données de séries chronologiques industrielles, en utilisant les bandes de Bollinger pour la détection des anomalies et le LSTM pour la réparation des erreurs. Bien qu'il soit principalement conçu pour les données industrielles, il existe un potentiel d'application dans d'autres domaines, nécessitant des améliorations en termes de flexibilité.

Dans le secteur de l'énergie, des systèmes d'extraction de connaissances ont été proposés [16] utilisant des techniques de clustering, les k-Medoids montrant des résultats prometteurs sur les séries temporelles énergétiques. La science médicale bénéficie également de l'analyse des données de séries chronologiques, ce qui incite au développement de systèmes comme Clairvoyance [17], facilitant le prototypage rapide et l'optimisation des systèmes de décision clinique. Des modèles et des systèmes hybrides comme AutoMTS ont été introduits pour nettoyer les données de séries chronologiques multivariées [18,19], démontrant leur importance dans l'amélioration de la qualité des données grâce à l'imputation et à la détection des valeurs aberrantes. AutoML et des frameworks comme DAEMON [20,21] se concentrent sur l'automatisation de l'ajustement des modèles d'apprentissage automatique et de la détection des valeurs aberrantes dans les données de séries chronologiques, atteignant une grande précision et démontrant le potentiel des services de détection d'anomalies en ligne [22].

Les données de séries chronologiques avec saisonnalité posent des défis uniques, tels que la multi-saisonnalité et l'hétéroscédasticité, qui compliquent l'analyse et la prévision. La recherche [23,24] a relevé ces défis en prétraitant les données, en créant des modèles et en les adaptant aux mesures originales, suivi d'une détection d'anomalies. D'autres études visent à améliorer la prévision des tendances locales et la correction des valeurs aberrantes dans l'analyse des séries chronologiques. La non-stationnarité a été abordée dans des études axées sur le nettoyage des données volatiles et non volatiles [25].
Des techniques telles que les k-voisins les plus proches et les fenêtres glissantes ont été explorées [26], les fenêtres glissantes étant plus performantes, en particulier pour les données volatiles comme les séries chronologiques boursières [27]. Des efforts ont été déployés pour mesurer la qualité des données à l'aide d'indices tels que l' indice de qualité de la valeur des attributs (AVQI) et l'indice de qualité de la valeur des données structurées (SDVQI), permettant la comparaison et l'amélioration des mesures de la qualité des données.

Des algorithmes d'optimisation tels que les algorithmes génétiques (AG) ont été utilisés pour automatiser la sélection des caractéristiques lors du prétraitement, améliorant ainsi l'efficacité des modèles de régression [28]. Alors que les solutions existantes telles que SCREEN, CLEANITS, DAEMON et Clairvoyance répondent à des défis de prétraitement spécifiques, il existe un besoin pour une approche générique couvrant plusieurs domaines.
Preptimize se démarque en proposant une solution intégrée de prétraitement et de recommandation de modèles , la rendant accessible aux non-experts et capable de gérer divers défis de séries chronologiques. SCREEN et DAEMON se concentrent sur la détection et le nettoyage des anomalies, tandis que CLEANITS se spécialise dans l'imputation. Clairvoyance est avant tout un outil de prévision utilisant des techniques avancées d'apprentissage automatique. La large applicabilité de Preptimize et son approche conviviale le différencient de ces outils spécialisés. Diverses bibliothèques open source et produits SaaS propriétaires automatisent la génération et l'optimisation de pipelines d'apprentissage automatique, rationalisant ainsi le processus de prétraitement [29,30].

Les valeurs des séries chronologiques à prévoir peuvent être nominales, catégorielles, textuelles et numériques. La prévision numérique est un problème de régression qui peut être résolu en utilisant diverses techniques telles que les moyennes mobiles, l'autorégression et le lissage exponentiel, ainsi que des méthodes avancées comme ARIMA, SARIMA, VAR et LSTM [31]. Approches hybrides

Machine Translated by Google

la combinaison de modèles statistiques et d'apprentissage automatique est également courante [32-34]. Cependant, aucune technique unique ne convient à toutes les séries chronologiques en raison de leurs propriétés variables, ce qui rend le processus de sélection long et complexe. Pour rationaliser le processus d'analyse d'une série chronologique, une procédure standard doit être adoptée pour automatiser le profilage, le prétraitement, la prévision optimisée du modèle, l'évaluation et la sélection du modèle le plus performant avec un minimum d'erreur.

La principale contribution de ce travail est l'introduction de Preptimize, une stratégie automatisée d'analyse de données de séries chronologiques qui intègre des techniques statistiques et d'apprentissage automatique pour recommander des modèles d'analyse optimaux. En lisant un échantillon d'un vaste ensemble de données, Preptimize suggère des modèles basés sur l'optimisation, incluant la meilleure combinaison de techniques de prétraitement et de prévision, améliorant ainsi la précision des prévisions sur différents types de données de séries chronologiques. Preptimize rationalise le processus d'analyse des séries temporelles univariées et multivariées avec des valeurs manquantes. Sur la base de la complexité impliquée dans l'analyse des séries chronologiques, cette recherche souligne l'importance de choisir la bonne technique de prévision et la bonne stratégie de prétraitement en fonction du type de données de séries chronologiques pour accélérer le processus d'analyse des données de séries chronologiques. De plus, ce travail a démontré le potentiel du cadre pour une large applicabilité dans la prévision de séries chronologiques, confirmant son efficacité dans la recommandation de modèles de prévision appropriés.

2. Matériels et méthodes

Sur la base de la discussion de la section précédente, il peut être déterminé qu'un cadre complet est essentiel pour définir les étapes standard lors du prétraitement et de la prévision des séries temporelles. La section suivante contient une explication détaillée du cadre proposé appelé Preptimize. L'objectif de Preptimize est d'automatiser le processus d' analyse des séries chronologiques en proposant le premier modèle pour une analyse plus approfondie. Le framework proposé a été implémenté en utilisant le langage de programmation Python.

Preptimize

Preptimize commence par lire les données de séries chronologiques et générer un profil complet comprenant l'analyse de la distribution, la saisonnalité et la détection des tendances. Il prétraite ensuite les données en résolvant les problèmes de qualité des données tels que les valeurs manquantes et les valeurs aberrantes à l'aide de diverses techniques et évalue les données prétraitées à l'aide de plusieurs modèles de prévision pour identifier le modèle le plus performant. Les étapes sont également mentionnées dans la figure 1. Étant donné que la conception du cadre est une tâche importante, cet article se concentre sur les données numériques et les problèmes de données manquantes. Ce travail vise à expliquer la plupart des solutions disponibles pour gérer les données manquantes dans les séries chronologiques et donne la meilleure solution possible. Le cadre proposé est testé sur différents ensembles de données, notamment des données de crypto-monnaie, des données boursières et des données générées par un réseau IoT.

De nombreuses techniques et algorithmes sont utilisés dans la mise en œuvre de Preptimize, comme indiqué dans le tableau 1. Les règles suivies par Preptimize pour recommander des modèles de modèles de prédiction potentiels sont dérivées de la littérature existante. Par conséquent, l'organisation et la rationalisation de ces règles et résultats constituent une contribution significative de ce travail. Le pseudocode de base du cadre proposé est donné dans la figure 2. La section suivante met en évidence uniquement les techniques les plus courantes utilisées à différentes étapes du processus.

Tableau 1. Détails des techniques utilisées dans la mise en œuvre de Preptimize.

Scène	Sous-scène	Détails des techniques utilisées
Profilage	Statistiques	Variance, écart type, minimum, maximum, médiane, mode, pourcentage manquant, quartiles, stationnarité, distribution, kurtosis, asymétrie, cardinalité, corrélation
	Visualisation	Graphique de distribution, nuage de points, tracé linéaire, diagramme en boîte, histogramme, graphique de décomposition des tendances et des saisons, carte thermique de corrélation, tracé de statistiques glissantes , ACF, PACF

Machine Translated by Google

Tableau 1. Suite

Scène	Sous-scène	Détails des techniques utilisées
Techniques de prétraitement	Transformation	Transformation pour la stationnarité utilisant la différence, la différence logarithmique, la décomposition exponentielle, additive, la décomposition multiplicative, la moyenne mobile et l'écart type, les statistiques ADF
		Transformation pour une distribution normale à l'aide de Box – Cox
	Manque	MCAR, MAR ou MNAR
	Effacement	Par liste, par paire
	Univarié	Moyennes mobiles incluant SMA, exponentielle, CMA
	Imputation	Remplissage en avant (LOCF), remplissage en aval (NOCB)
		GAIN (basé sur des réseaux contradictoires génératifs)
	Interpolation	Linéaire, akima, barycentrique, cubique, spline cubique, à partir de dérivés, d'indice, Krogh, le plus proche, pad, pchip, polynôme, polynôme par morceaux, quadratique, slinéaire, spline, temps, zéro
	Imputation multivariée	DataWig, Bayesian Ridge, forêt aléatoire, K-voisins, arbre de décision, Arbres supplémentaires, KNNImputer, HotDeck
Prévision	Univarié	Techniques statistiques incluant les moyennes mobiles (SMA, exponentielle, CMA) et d'autres techniques saisonnières et non saisonnières, notamment AR, ARIMA, SES, SARIMA, HWES
		MLP, SVR-RBF, SVR-POLY, arbre de décision, forêt aléatoire, RNN, GRU, GRU-bidirectionnel, Vanilla LSTM, Stacked LSTM, Bidirectionnel LSTM, CNN-LSTM
	Multivarié	Techniques statistiques dont VAR, VMA, VARMA
		MLP, arbre de décision, forêt aléatoire, RNN, GRU, GRU-bidirectionnel, LSTM empilé

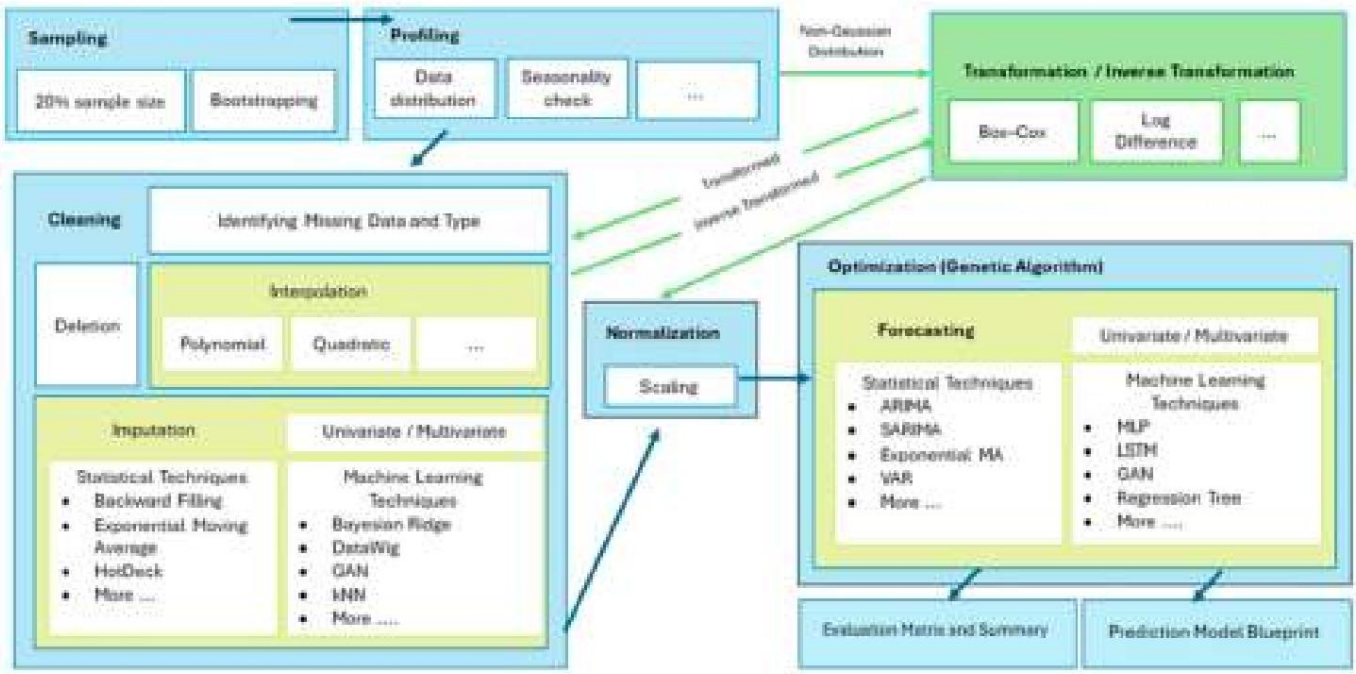


Figure 1. Preptimize : représentation visuelle de l'approche systématique intégrant le prétraitement, la sélection du modèle et les étapes d'évaluation approfondies, améliorant ainsi l'efficacité de l'analyse des séries chronologiques.

Machine Translated by Google


```
Initialiser l'ensemble de données

Sauvegardez l'ensemble de données pour gérer les valeurs nulles ultérieurement
# Profiler l'ensemble de données

Initialiser le profil en tant que dictionnaire

Si l'ensemble de données comporte plusieurs colonnes
    Définir le profil comme multivarié

    Imprimer la corrélation par paire de l'ensemble de données

    Afficher la carte thermique des corrélations
Autre
    Définir le profil comme univarié

Calculer le nombre total d'enregistrements et le type de série chronologique (univariée/multivariée)
# Préparer le profil de chaque colonne
Pour chaque colonne de l'ensemble de données
    Initialiser dataAttribute comme données de colonne

    Si Univariate, définissez UnivariateColumnName

    Vérifier les données manquantes dans la colonne
    Calculer les valeurs nonNullValues et nullValues

    S'il existe des données manquantes

        Calculer le pourcentage manquant

        Supprimez les valeurs nulles pour une analyse plus approfondie
    Calculer l'asymétrie et l'aplatissement de dataAttribute

    Déterminer le type de distribution (normal/pas normal)

    Afficher le profil de la colonne avec les statistiques calculées

    Tracer un graphique de décomposition saisonnière

    Tracer des graphiques de profil de base

    Enregistrer le profil pour la colonne dans le profil
# Restaurer l'ensemble de données d'origine à partir de la sauvegarde
Initialiser les données manquantes et manquantColCorrelationCompleteCases
Déterminer le type d'absence et la corrélation
Initialiser les jeux de données traités
# Gérer les séries chronologiques univariées
Si univarié
    Récupérer le profil de la colonne univariée

    S'il existe des données manquantes

        Si le type manquant est MCAR et le pourcentage manquant <= 10 %

            Appliquer la suppression par liste

        Autre

            Si les données sont numériques et non biaisées

                Appliquer des techniques d'imputation

            Sinon, si les données ne sont pas normalement distribuées

                Appliquer la transformation Box-Cox
```

Figure 2. Suite.

Machine Translated by Google

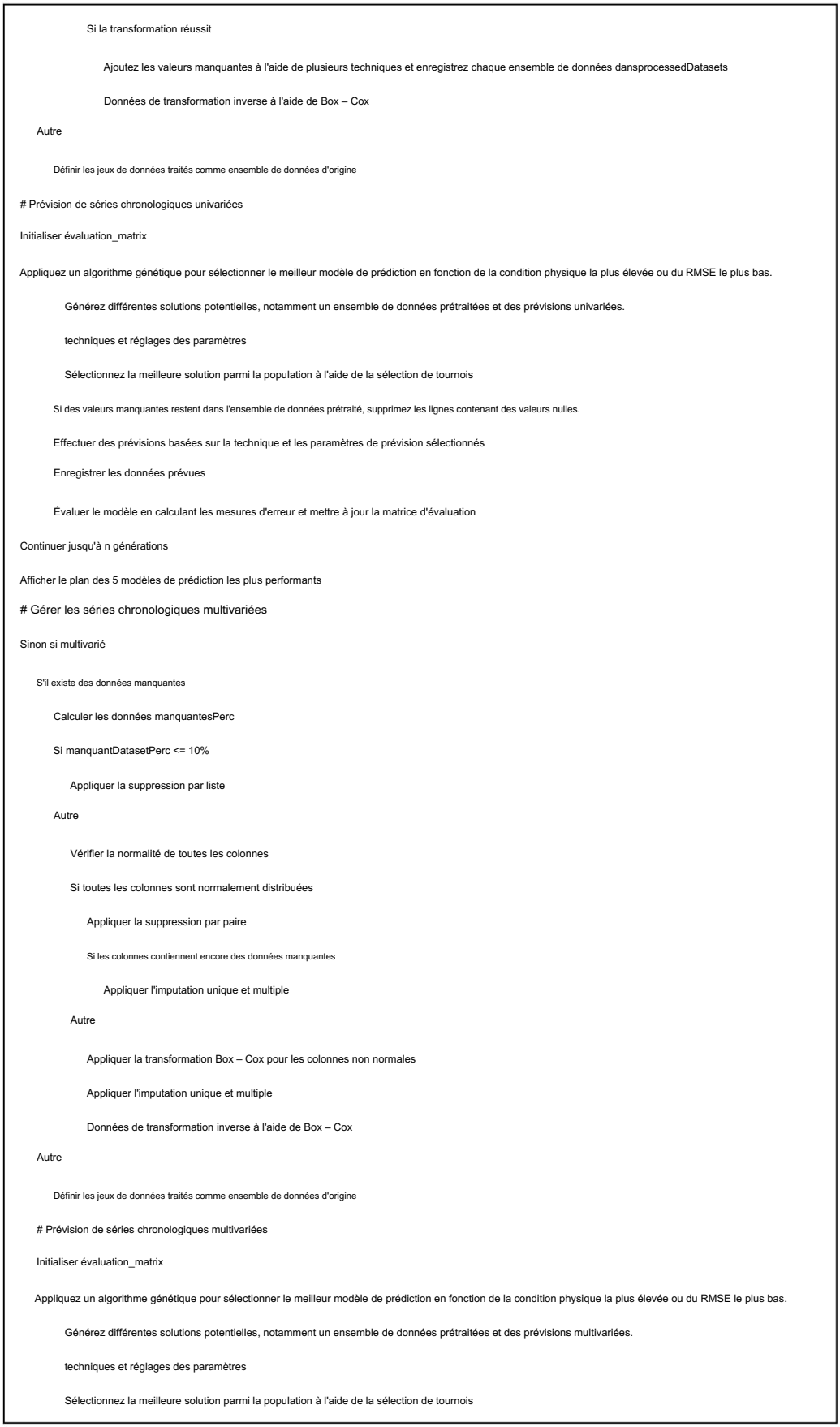


Figure 2. Suite.

Machine Translated by Google

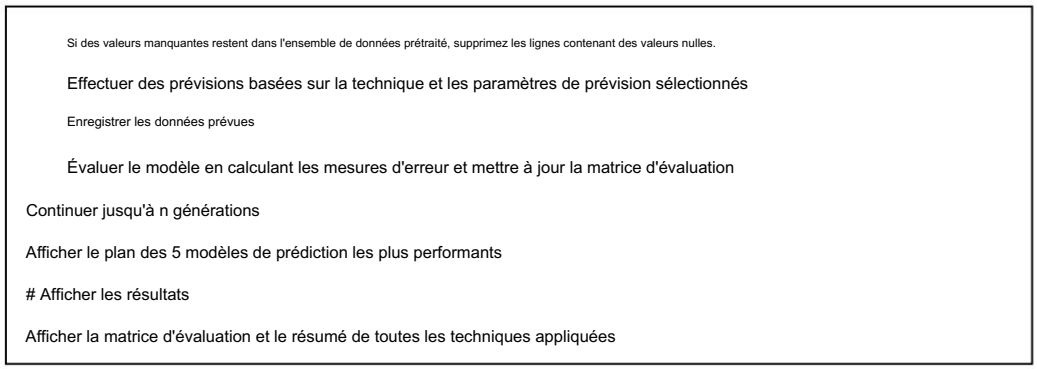


Figure 2. Preptimize : flux de travail séquentiel pour le prétraitement des données de séries chronologiques et la sélection de modèles. Ici, # représente le commentaire.

2.1. Échantillonnage des données

Dans la première étape, Preptimize génère un échantillon à partir des données pour représenter le comportement d'origine tout en réduisant la taille à 20 %. Le bootstrapping a été appliqué et l'index a été maintenu.

2.2. Profilage des données

À cette étape, différentes propriétés des données sont calculées et un profil est créé pour l' attribut. Ces propriétés incluent la moyenne, la médiane, la valeur minimale et maximale, la variance et l'écart type des données. Les données sont explorées plus en détail en trouvant le type de distribution, l'asymétrie et l'aplatissement des données. La stationnarité des données est également calculée, ce qui inclut la saisonnalité et la tendance. Une brève description de ces techniques est donnée ci-dessous.

2.2.1. Distribution des données

La distribution des données spécifie toutes les valeurs probables pour une variable et quantifie également le taux d'occurrence relatif indiquant la fréquence à laquelle une valeur apparaît. Il existe différentes distributions possibles, qui peuvent être normales, binomiales, de Poisson ou de Bernoulli . La plupart des situations suivent la distribution normale, qui s'exprime sous la forme d' une courbe en forme de cloche et symétrique autour de la ligne x = μ. Cela indique qu'un nombre égal de valeurs sont présentes des deux côtés du centre. La distribution gaussienne avec une moyenne de 0 et un écart type de 1 est la distribution gaussienne standard.

Les deux propriétés liées à la distribution sont l'asymétrie et l'aplatissement. L'asymétrie est une mesure de régularité. La série de données est symétrique si elle apparaît identique des deux côtés du point médian. L'aplatissement mesure le taux d'apparition de valeurs aberrantes au niveau de la queue. Les données avec des queues lourdes indiquent davantage de valeurs aberrantes et la valeur de l'aplatissement est élevée. L'aplatissement d'une distribution normale est de 3. Si inférieur à 3, il y a peu de valeurs aberrantes, et si supérieur à 3, il y a plus de valeurs aberrantes. L'asymétrie peut être négative, ce qui spécifie que la queue est sur le côté gauche ; sinon, la queue est à droite. Si l'asymétrie est nulle, alors la distribution est parfaitement symétrique. La distribution peut être considérée comme normale si les valeurs d'asymétrie se situent entre −3 et +3, et si l'aplatissement est compris entre −10 et +10 lorsque l'on utilise l' erreur type de mesure [35].

2.2.2. Stationnarité

Une série temporelle dont les caractéristiques ne sont pas soumises au temps est stationnaire. Ainsi, la présence d'une tendance ou d'une saisonnalité indique que la série n'est pas stationnaire. La propriété observée à différents moments est affectée par la tendance et la saisonnalité. Une série chronologique comporte trois composantes : la saison, la tendance et le bruit. La saison peut être définie comme le cycle répétitif à court terme de la série. La tendance est la pente des données, qui peut augmenter ou diminuer avec le temps, tandis que le bruit est le changement aléatoire dans la série. Il est souvent considéré comme difficile de traiter des données de séries chronologiques contenant des saisons ou des tendances. Le tracé de la série chronologique donne des informations instantanées sur la présence d'une saison et d'une tendance.

Machine Translated by Google

Il existe deux approches courantes pour traiter les séries chronologiques saisonnières. Une méthode consiste à diviser la série en supprimant la composante saisonnière et la tendance, puis à utiliser la série pour l'analyse ou à utiliser toute autre technique de transformation pour obtenir la représentation non saisonnière de la série. L'autre façon consiste à utiliser des techniques de prévision spécialement conçues pour les séries saisonnières comme SARIMA.

2.2.3. Identifier le type de données manquantes

Lors du traitement des données manquantes, trois catégories sont identifiées : manquantes complètement au hasard (MCAR), manquantes au hasard (MAR) et manquantes non aléatoires (MNAR). Dans l'analyse univariée, traiter le temps comme une variable ajoute à la complexité, nécessitant la prise en compte des traits des séries chronologiques et de la covariance pour l'imputation. Pour les données MCAR, l'imputation univariée est relativement simple. Cependant, en analyse multivariée, l'association entre les colonnes est prise en compte. Les données MCAR sont manquantes de manière aléatoire et peuvent souvent être remplacées par des valeurs moyennes ou médianes, car les données manquantes sont considérées comme récupérables. MCAR est donné par (1), où R correspond aux données manquantes et Y correspond aux données complètes.

$$P(Ri|Yi) = P(Ri) \quad (1)$$

Les données MCAR manquent d'association significative selon des mesures telles que le chi carré, la règle de Cramer ou le coefficient de corrélation, et contiennent généralement moins de valeurs nulles. Pour les grands ensembles de données, les instances comportant des données manquantes peuvent être omises ou imputées. Les méthodes de suppression pour MCAR incluent la suppression par liste, la suppression des enregistrements avec n'importe quelle valeur nulle, adaptée lorsque les données manquantes sont minimes, et la suppression par paire, la conservation des données pour les paires de quantités avec des points de données disponibles, préservant ainsi davantage de données. L'imputation devient nécessaire lorsque les méthodes de suppression ne conviennent pas. L'imputation consiste à remplacer les données manquantes par les valeurs attendues, en appliquant diverses techniques selon la nature des données. Dans le cas de MAR, les données manquantes peuvent suivre un modèle ou une séquence, apparaissant parfois en dehors de la séquence. Lorsqu'une relation entre des données manquantes et existantes est identifiée, elle est considérée comme MAR. Des techniques d'imputation telles que la moyenne, la médiane, le mode ou l'interpolation peuvent être utilisées pour gérer MAR, avec plusieurs techniques d'imputation disponibles pour les données hors séries chronologiques et l'interpolation pour les données de séries chronologiques. MAR est donné par (2)

$$P(Ri|Yi\ ) = P(Ri|Yi,o) \quad (2)$$

Dans MNAR, des instances spécifiques partagent des données manquantes, ce qui rend la prévision difficile car elle repose sur des informations non observées. Si les valeurs manquantes dans une colonne montrent peu ou pas de corrélation avec d'autres colonnes et qu'aucune association forte n'est trouvée lors de l'enquête, et que la fréquence des valeurs manquantes est élevée, cela est considéré comme MNAR. Le MNAR n'est pas ignorable, et il n'existe aucune méthode simple pour le gérer autre que l'obtention des données manquantes. MNAR a généralement plus de valeurs nulles que MCAR et ne peut être identifié que grâce à la connaissance du domaine. MNAR est donné par (3) :

$$P(Ri|Yi\ ) \quad = P(Ri|Yi\ ,o) \quad (3)$$

2.3. Visualisation des données

L'utilisation de divers graphiques et tracés constitue une approche précieuse dans l'analyse de séries chronologiques pour comprendre le comportement et les propriétés des données. Preptimize génère une gamme de visualisations, telles que des nuages de points, des tracés linéaires et des graphiques de décomposition, pour fournir un aperçu complet des données de séries chronologiques. Ces outils visuels aident à comprendre le comportement des données, à identifier les tendances et la saisonnalité et à sélectionner les techniques de prétraitement appropriées.

Le premier ensemble de graphiques offre une perspective globale des données, notamment des nuages de points, des graphiques linéaires, des diagrammes en boîte et des histogrammes. Chaque attribut d'une série chronologique multivariée est tracé individuellement pour capturer l'essence des données.

Un autre graphique crucial affiche la distribution des données, facilitant l'évaluation de normalité. Cette évaluation est essentielle pour déterminer les techniques statistiques appropriées.

Machine Translated by Google

Un graphique illustrant la décomposition des saisons et des tendances est particulièrement utile pour identifier les tendances saisonnières au sein de la série chronologique. Cette idée guide l'application de techniques appropriées adaptées aux données saisonnières, nécessitant potentiellement des transformations pour une meilleure analyse.

Pour évaluer la stationnarité des données, Preptimize génère un graphique traçant la moyenne mobile et l'écart type mobile par rapport aux données d'origine. La stationnarité est indiquée si ces valeurs restent constantes dans le temps [36].

La visualisation de la carte thermique représente les corrélations entre plusieurs variables, avec des niveaux de corrélation plus élevés représentés par des couleurs distinctes. Cela aide à comprendre la force des relations entre les variables.

Des graphiques supplémentaires se concentrent sur les données manquantes, notamment une matrice et un graphique à barres, qui fournissent des informations sur la nature des données manquantes et leurs relations potentielles avec d'autres attributs.

2.4. Nettoyage des

données La phase de prétraitement dans l'analyse des séries chronologiques implique le nettoyage des données résolvant divers problèmes de qualité des données tels que les valeurs manquantes, les valeurs aberrantes et le bruit. Contrairement à d'autres formes d'analyse de données, le prétraitement des séries chronologiques nécessite des techniques spécifiques adaptées à leurs caractéristiques uniques. Preptimize utilise une gamme de techniques d'imputation pour gérer les données manquantes, y compris, sans s'y limiter, l'interpolation linéaire, l'interpolation spline et les méthodes d'apprentissage automatique telles que les GAN. Le choix de la méthode d'imputation repose sur les caractéristiques spécifiques de l'ensemble de données, garantissant ainsi la meilleure qualité de données possible pour les prévisi Avec près de 25 techniques d'interpolation différentes disponibles, le cadre vise à explorer l'approche la plus adaptée à chaque ensemble de données. Les techniques de transformation sont appliquées lorsque les données s'écartent de la normalité ou présentent une non-stationnarité. La transformation Box – Cox est utilisée pour traiter la non-normalité, tandis que des méthodes telles que la décomposition saisonnière sont utilisées pour supprimer les tendances et la saisonnalité des données non stationnaires.

Les approches d'apprentissage automatique, notamment les réseaux contradictoires génératifs (GAN), sont intégrées au cadre pour améliorer la prédiction des données manquantes. Pour les séries chronologiques multivariées, des modèles d'imputation itérative et DataWig sont utilisés pour résoudre les problèmes de données manquantes.

Pour déterminer la technique d'imputation optimale, une évaluation complète est réalisée , en tenant compte des performances de prévision et des paramètres d'évaluation. Alternativement, une approche par force brute peut être utilisée pour tester toutes les techniques d'imputation possibles.

Dans l'ensemble, la phase de prétraitement joue un rôle crucial dans la préparation des données de séries chronologiques à des fins d'analyse et de prévision, nécessitant des techniques spécialisées adaptées à leurs caractéristiques et défis uniques.

2.4.1. Remplissage en avant et en arrière

La dernière observation reportée (LOCF) est une technique d'imputation unique largement utilisée qui impute la dernière valeur mesurée par remplissage ultérieur. Il doit y avoir au moins une mesure post-référence dans LOCF. L'observation suivante reportée en arrière (NOCB) est une méthode analogue à LOCF mais fonctionne dans le sens inverse par remplissage en arrière ; c'est-à-dire qu'il utilise la première valeur enregistrée après la valeur indisponible et la reporte en arrière.

2.4.2. Interpolation Les

techniques d'interpolation offrent un moyen de prédire les points de données manquants en estimant les valeurs entre les points de données connus. L'interpolation linéaire, la méthode la plus simple, relie les points adjacents par des lignes droites, en supposant que les points sur la ligne représentent les valeurs manquantes. L'interpolation polynomiale, plus polyvalente, utilise des fonctions polynomiales d' ordres variables pour représenter les relations entre les attributs. L'interpolation polynomiale de Lagrange et l'interpolation polynomiale de Newton sont des exemples de méthodes d'interpolation polynomiale. L'interpolation de Lagrange, bien que polyvalente, peut rencontrer des difficultés avec des ensembles de données plus

Machine Translated by Google

L'interpolation polynomiale de Newton utilise un processus de division récursive pour construire des polynômes complexes, ce qui peut être efficace pour la prédiction mais peut conduire à des complexités numériques et à une sensibilité aux valeurs aberrantes. L'interpolation spline, quant à elle, utilise des polynômes par morceaux pour ajuster des courbes plus douces à des groupes plus petits de points de données, l'interpolation spline cubique étant souvent préférée pour sa douceur et sa stabilité.

Les autres méthodes d'interpolation utilisées dans le cadre incluent le polynôme par morceaux, Akima, Barycentric, Krogh et Pchip, chacune offrant différentes approches pour estimer les valeurs manquantes dans les données de séries chronologiques.

2.4.3. Moyenne mobile Les

moyennes mobiles, telles que les moyennes mobiles simples (SMA), les moyennes mobiles cumulées (CMA) et les moyennes mobiles exponentielles (EMA), sont couramment utilisées pour l'imputation dans l'analyse des séries chronologiques. SMA calcule la moyenne d'un nombre spécifié de points de données, à l'aide d'une formule de sommation simple. CMA calcule la moyenne de toutes les entrées jusqu'à l' heure actuelle, chaque valeur étant également pondérée. L'EMA utilise des facteurs de pondération décroissants de façon exponentielle pour calculer la moyenne, la valeur étant déterminée à tout moment par un coefficient et la valeur EMA précédente. Ces moyennes mobiles peuvent faciliter l'imputation des données, à condition que la série chronologique soit stationnaire. La stationnarité, indiquant un comportement cohérent dans le temps, peut être évaluée à l'aide de tests statistiques comme le test Augmenté de Dickey-Fuller (ADF) [37] ou en observant des statistiques glissantes, telles que des tracés de moyenne mobile et d'écart type.

2.4.4. Algorithmes d'apprentissage profond

L' imputation de données à l'aide de techniques d'apprentissage profond, en particulier les réseaux neuronaux récurrents (RNN) et les réseaux contradictoires génératifs (GAN), est apparue comme une approche prometteuse dans la littérature récente pour l'analyse des séries chronologiques. Les RNN et leurs variantes excellent dans la modélisation des données de séquence, tandis que les GAN sont capables de générer et d'imputer des valeurs manquantes. Preptimize exploite un GAN pour les imputations simples et multiples. Le GAN fonctionne à travers deux modèles, le générateur (G) et le discriminateur (D), qui fonctionnent de manière antagoniste pour générer des échantillons de données réalistes et distinguer respectivement les échantillons réels et faux . La formation se poursuit jusqu'à ce que le discriminateur soit incapable de faire la différence entre les échantillons réels et générés, ce qui indique la capacité du générateur à produire des exemples crédibles.

Dans les GAN, D essaie d'attribuer l'étiquette correcte aux échantillons avec une probabilité maximale, tandis que G essaie de minimiser log (1 − D(G(z))). D'après [38], D et G jouent au jeu minimax à deux joueurs avec la fonction de valeur V (G, D), comme le montre (4).

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(Z)))] \quad (4)$$

Dans Preptimize, un modèle présenté dans [39] est utilisé pour imputer les données manquantes à l'aide d'un GAN. La méthode est appelée Generative Adversarial Imputation Nets (GAIN). Dans la méthode, G observe les données réelles, impute l'élément manquant et génère les données finalisées.
D prend alors une donnée finalisée et tente de découvrir quels éléments sont observés et lesquels sont imputés.

2.4.5. Imputation multiple Les

imputations multiples (IM) offrent une approche globale de la gestion des valeurs manquantes en utilisant l'ensemble complet de fonctionnalités pour approximer les valeurs manquantes. MI génère plusieurs ensembles de données en imputant des valeurs manquantes, ce qui permet de remplacer les valeurs manquantes par plusieurs valeurs possibles. Ces valeurs imputées sont estimées à partir de chaque ensemble de données, puis agrégées dans un ensemble de données final. MI est applicable dans les cas où les données manquent complètement au hasard (MCAR), lorsqu'elles manquent de manière aléatoire (MAR), et même lorsque les données manquent de manière non aléatoire (MNAR).

L'imputation itérative, quant à elle, implique la modélisation de chaque caractéristique en fonction des autres caractéristiques dans un processus séquentiel. Les valeurs imputées passées sont utilisées pour prédire les caractéristiques suivantes jusqu'à ce que toutes les valeurs manquantes soient remplies. Preptimize utilise la bibliothèque IterativeImputer dans

Machine Translated by Google

Python, qui propose divers estimateurs tels que le régresseur Bayesian Ridge, le régresseur de forêt aléatoire, le régresseur K-voisins, le régresseur d'arbre de décision et le régresseur Extra Trees.

De plus, la technique KNNImputer, fournie par la même bibliothèque Python, utilise l'approche des k-voisins les plus proches pour imputer les valeurs manquantes en considérant les valeurs des observations pour les points de données voisins.

DataWig, une autre méthode utilisée pour l'imputation, s'est révélée efficace dans la littérature pour imputer les valeurs manquantes. DataWig [40] est une bibliothèque d'apprentissage en profondeur qui exploite MXNet comme backend pour générer des prédictions pour les données manquantes.

De plus, la méthode hot deck [41] , connue pour son efficacité de calcul, est utilisée pour l'imputation. Cette méthode impute les valeurs manquantes en sélectionnant un enregistrement similaire de manière aléatoire dans une matrice de données et en utilisant les valeurs existantes de l'objet donneur. Le hot deck fonctionne sur un attribut à la fois et sert de technique d'imputation unique.

2.5. Prévisions II

existe diverses approches de prévision dans la littérature pour les séries chronologiques.
Les exemples incluent le lissage exponentiel, la moyenne mobile intégrée autorégressive (ARIMA) avec des variantes saisonnières et non saisonnières, plusieurs techniques d'apprentissage en profondeur et bien d'autres. Il existe des techniques distinctes pour prévoir une série temporelle univariée et multivariée. Preptimize utilise un algorithme génétique [42] pour explorer le vaste espace de modèles et de configurations possibles afin de trouver celui le plus approprié pour un ensemble de données donné. Chaque individu de la population représente une solution potentielle comprenant une stratégie de prétraitement, un modèle de prévision et des paramètres. Chaque individu est évalué sur la base d'une fonction de fitness, qui est dans ce cas le RMSE. Plus le RMSE est bas, meilleur est le modèle.

Les individus sont sélectionnés en fonction de leurs scores de condition physique pour devenir parents de la prochaine génération. La mutation est effectuée sur la base du schéma bit-flip ; 3 individus sont sélectionnés au hasard parmi la population par sélection en tournoi. Cette approche permet d'automatiser le processus de sélection du modèle, ce qui permet d'économiser potentiellement beaucoup de temps et d'efforts tout en obtenant une précision de prévision optimale ou quasi optimale.

2.5.1. Prévision univariée

La prévision univariée consiste à prédire les valeurs futures d'une variable unique sur la base de données historiques provenant d'intervalles de temps consécutifs. Le choix de la technique de prévision dépend du fait que la série chronologique présente ou non des composantes de saisonnalité et de tendance.

Il existe deux principaux types de méthodes de prévision univariées : les méthodes statistiques et les méthodes avancées basées sur l'apprentissage automatique. Les méthodes statistiques comprennent l'autorégression (AR), la moyenne mobile (MA), la moyenne mobile autorégressive (ARMA), la moyenne mobile intégrée autorégressive (ARIMA) et le lissage exponentiel simple. Les techniques de moyenne mobile telles que la moyenne mobile simple, la moyenne mobile exponentielle et la moyenne mobile cumulative sont courantes.

Pour les séries chronologiques avec saisonnalité et tendance, les options incluent la moyenne mobile intégrée autorégressive saisonnière (SARIMA), la moyenne mobile intégrée autorégressive saisonnière avec régresseurs exogènes et le lissage exponentiel Holt-Winters (HWES), ARIMA et SARIMA étant les plus efficaces.

ARIMA combine l'autorégression, la moyenne mobile et la différenciation pour atteindre la stationnarité . Une identification correcte des paramètres ARIMA (p, d, q) est cruciale, généralement réalisée en ajustant des modèles avec différents ordres à l'aide de critères tels que le critère d'information d'Akaike (AIC) pour sélectionner le meilleur ordre. L'AIC mesure les performances du modèle par rapport au nombre de paramètres. La fonction d'autocorrélation (ACF) et la fonction d'autocorrélation partielle (PACF) ne peuvent pas être utilisées pour identifier des valeurs fiables pour p et q [43]. L'analyse des résidus est essentielle dans la modélisation ARIMA pour garantir que les résidus ressemblent à du bruit blanc, indiquant l'absence d'autocorrélation. Le test de Ljung – Box est couramment utilisé pour évaluer l'autocorrélation résiduelle, où les valeurs de p supérieures à un seuil ne suggèrent aucune autocorrélation significative.

La prévision de séries chronologiques a suscité une attention considérable dans le domaine de l'apprentissage automatique et de l'apprentissage profond. Preptimize prend également en compte ces méthodes de prévision basées sur le ML. Le

Machine Translated by Google

La littérature présente diverses techniques à cet effet, notamment la mémoire à long terme (LSTM), la régression vectorielle de support (SVR), le perceptron multicouche (MLP), les arbres de décision, la forêt aléatoire, le réseau neuronal récurrent (RNN) et ses variante la Gated Recurrent Unit (GRU).

Parmi ces techniques se distingue le LSTM, connu pour son efficacité dans le traitement des données séquentielles. Cependant, les RNN traditionnels sont confrontés à des défis tels que la disparition et l'explosion des gradients, entraînant des temps de formation plus longs et une précision moindre. Ces problèmes sont résolus par LSTM et GRU, qui sont plus efficaces dans la gestion des problèmes de gradient.

GRU diffère des RNN par son utilisation de portes de mise à jour et de réinitialisation pour contrôler le flux d'informations et résoudre les problèmes de gradient. Ces portes déterminent la quantité d'informations précédentes qui sont conservées ou supprimées, améliorant ainsi l'efficacité du modèle. Des variantes de GRU, y compris des versions bidirectionnelles, sont utilisées dans le cadre proposé.

LSTM, une autre variante de RNN, comprend plusieurs unités LSTM, chacune équipée de portes de cellule, d'entrée, de sortie et d'oubli. Ces portes jouent un rôle crucial dans le modèle LSTM, facilitant le traitement séquentiel des données et la génération de prédictions.

Ce cadre utilise diverses configurations de réseaux neuronaux pour la prévision, notamment Vanilla LSTM, Stacked LSTM, Bidirectionnel LSTM et CNN-LSTM. Vanilla LSTM se compose d'une couche cachée et d'une couche de sortie, tandis que Stacked LSTM implique plusieurs couches cachées. Le LSTM bidirectionnel traite les données dans les deux sens, puis combine les interprétations. CNN-LSTM intègre un CNN pour comprendre les parties de la séquence, puis les transmet au LSTM pour interprétation, formant ainsi un modèle hybride.

En comparant GRU et LSTM, aucun gagnant clair ne se dégage pour tous les paramètres de performance . GRU s'exécute plus rapidement avec moins de paramètres et moins d'utilisation de la mémoire, tandis que LSTM fait preuve d'une plus grande précision, en particulier sur des ensembles de données plus volumineux.

MLP sert de réseau neuronal artificiel fondamental, comprenant des couches de rétroaction entièrement connectées, adaptées aux tâches de classification et de régression. Il comprend au moins une couche d'entrée, une couche cachée et une couche de sortie, utilisant des fonctions d'activation non linéaires pour gérer des données linéairement non séparables.

Pour les tâches de régression, la régression vectorielle de support (SVR) offre une flexibilité dans la définition de la tolérance d'erreur, en cherchant à minimiser la norme L2 du vecteur de coefficients tout en autorisant les écarts avec les variables de marge. Les arbres de décision, y compris les arbres de régression, sont utilisés pour prévoir des séries chronologiques, tandis que la forêt aléatoire combine plusieurs arbres de régression pour améliorer la précision des prévisions grâce au regroupement et au caractère aléatoire des caractéristiques.

2.5.2. Prévisions multivariées

Dans le cadre proposé, la prévision de séries chronologiques multivariées utilise une gamme de méthodes statistiques et d'apprentissage automatique. Bien que les approches statistiques pour les séries chronologiques multivariées soient limitées, le cadre intègre des techniques clés telles que l'autorégression vectorielle (VAR), la moyenne mobile vectorielle (VMA) et la moyenne mobile d'autorégression vectorielle (VARMA).

Contrairement aux techniques univariées telles que AR et ARIMA, les méthodes vectorielles comme VAR, VMA et VARMA sont bidirectionnelles, permettant aux variables de s'influencer mutuellement. Cependant, ces techniques nécessitent une stationnarité, souvent confirmée par le test ADF. VAR établit des relations entre les variables via un système d'équations, avec une sélection de décalage guidée par des critères tels que le critère d'information d'Akaike (AIC). VMA fonctionne avec une matrice de moyenne mobile et un vecteur à k dimensions, tandis que VARMA intègre à la fois des ordres autorégressifs et à moyenne mobile. Pour les prévisions basées sur l'apprentissage automatique, le cadre utilise des techniques telles que MLP, arbre de régression, forêt aléatoire, RNN, GRU et LSTM, chacune détaillée dans la section précédente.

3. Résultats expérimentaux et discussion

Cette étude évalue Preptimize, un cadre conçu pour l'analyse de données de séries chronologiques, dans divers domaines et scénarios, notamment les données boursières, les données de crypto-monnaie et les données de consommation d'énergie générées par l'IoT. Chaque expérience implique un prétraitement du

Machine Translated by Google

données, en appliquant diverses techniques de prévision et en évaluant les performances à l'aide de mesures telles que MAE, MSE, RMSE et MAPE. Les détails des ensembles de données utilisés sont mentionnés dans le tableau 2. Le cadre utilise des techniques de prétraitement et de prévision adaptées à chaque type d'ensemble de données, la sélection étant optimisée via un algorithme génétique (GA). Preptimize est conçu pour tirer parti de GA pour optimiser la sélection de modèles. Pour garantir une compréhension globale des fonctionnalités du framework, tous les modèles possibles ont été évalués. L' AG a été adaptée pour inclure tous les modèles potentiels, identifiant à la fois les meilleurs et les moins performants. Ceci a été réalisé en attribuant la valeur la plus élevée aux modèles les plus performants. Dans les applications pratiques, cependant, l'AG recommande généralement les cinq modèles les plus performants sur la base du RMSE le plus bas, équilibrant efficacité et efficacité. Les mesures d'évaluation incluent MAE, MSE, RMSE et MAPE, avec une analyse détaillée fournie dans un rapport complet généré par Preptimize. Les expériences sont menées séparément pour des ensembles de données univariées et multivariées, couvrant les données du marché boursier, des cryptomonnaies et de la consommation électrique à domicile. Chaque expérience est méticuleusement décrite, y compris les détails des ensembles de données, les procédures de formation/test et les résultats de l'évaluation. Les détails et les résultats de ces expériences sont donnés ci-dessous.

Tableau 2. Détails des ensembles de données utilisés dans les expériences.

Nom	Description
Expérience 1 – Série chronologique univariée	
Bourse	Symbole : MSFT (Microsoft) Colonnes : Date, Clôture Nombre d'enregistrements : 503 Dates : du 03/08/2020 au 01/08/2022 Intervalle : quotidien
Bourse	Identique à 1-A, mais avec des données manquantes
Crypto-monnaie	Crypto-monnaie : BTC-USD (Bitcoin) Colonnes : Date, Clôture Nombre d'enregistrements : 748 Dates : du 01/08/2020 au 17/08/2022 Intervalle : quotidien
Expérience 2 – Série chronologique multivariée	
IoT : consommation d'énergie	Domaine : Accueil Colonnes : Date, Global_active_power, Global_reactive_power, Tension, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3 Nombre d'enregistrements : 10 081 Dates : 2007-04-24 au 2007-04-30 Intervalle : minutes
Bourse	Symbole : MSFT (Microsoft) Colonnes : Date, Ouvert, Haut, Bas, Fermer Nombre d'enregistrements : 503 Dates : du 03/08/2020 au 01/08/2022 Intervalle : quotidien

3.1. Expérience 1 : série univariée

Preptimize a subi des tests initiaux sur des données de marché, notamment les actions et les crypto-monnaies , un type courant de données de séries chronologiques caractérisées par la fluctuation des prix des inscriptions au fil du temps. L'expérience 1 est divisée en trois parties. La première partie a utilisé des données boursières complètes pour évaluer les performances de Preptimize sur des données de haute qualité. La partie 2 a introduit des données intentionnellement manquantes pour évaluer la gestion par le cadre des ensembles de données incomplets. La troisième partie a utilisé des données de crypto-monnaie avec des valeurs manquantes pour tester davantage les capacités de Preptimize. Ces expériences visaient à évaluer l'efficacité du cadre dans divers scénarios et qualités de données dans le domaine des prévisions financières.

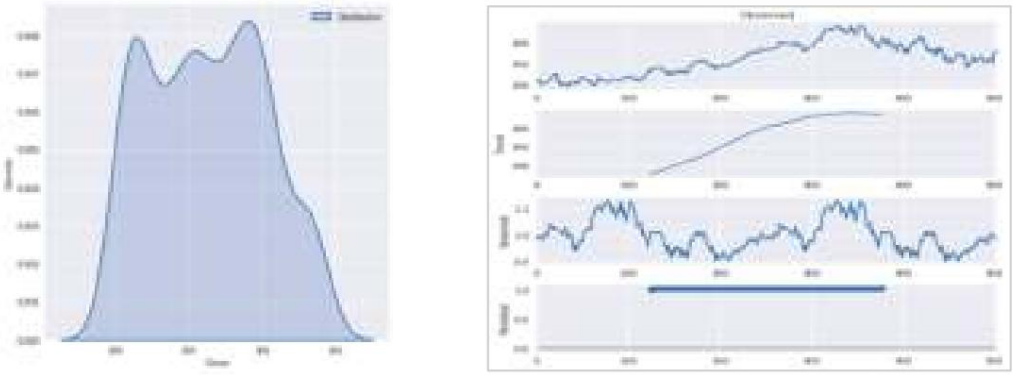
Dans la première partie de l'expérience, cette étude explore les données boursières de Yahoo Finance, en se concentrant sur la société Microsoft Corporation (MSFT), cotée à la bourse NASDAQ.

Machine Translated by Google

À l'aide de la bibliothèque yfinance Python, des données de marché d'août 2020 à août 2022 ont été collectées, en examinant spécifiquement les cours de clôture de chaque jour de bourse, hors week-end. Il s'agit d'un ensemble de données de séries chronologiques univariées, comprenant environ 500 enregistrements, divisés en ensembles de formation et de test. Notamment, l'ensemble de données était complet et exempt de problèmes de qualité, évitant ainsi la nécessité d'un prétraitement dans la phase initiale de l'expé

Au total, 139 modèles ont été utilisés pour prévoir la série chronologique, le rapport généré fournissant un aperçu détaillé du processus d'analyse des données. Malgré les limitations d'espace, le rapport comprend des fonctionnalités essentielles telles que les caractéristiques de base du profil, des graphiques de moyenne mobile et d'écart type, des nuages de points, des boîtes à moustaches, des graphiques de distribution et des graphiques de décomposition des tendances. De plus, le rapport comprend des journaux de formation, des calculs d'erreurs et des évaluations de performances pour tous les modèles. Les figures 3a et b illustrent la distribution des données, indiquant la normalité, et décrivent la décomposition de la tendance et de la saisonnalité au sein de la série chronologique. Des analyses complètes, y compris les valeurs p et les statistiques Augmentées de Dickey – Fuller (ADF), suggèrent la stationnarité de la série chronologique sans saisonnalité

Le fichier de synthèse des performances du modèle contenait les résultats d'environ 139 modèles, parmi lesquels quelques-uns sont mentionnés dans le tableau 3, montrant uniquement les 5 meilleurs et les 5 moins performants.



(un) (b)

Figure 3. (a) Le graphique montre la distribution normale des données des prix de clôture MSFT sur 2 ans. (b) Graphique de décomposition des cours de clôture pour diviser la composante de tendance et de saisonnalité. Le graphique représente une nette tendance à la hausse et aucune saisonnalité.

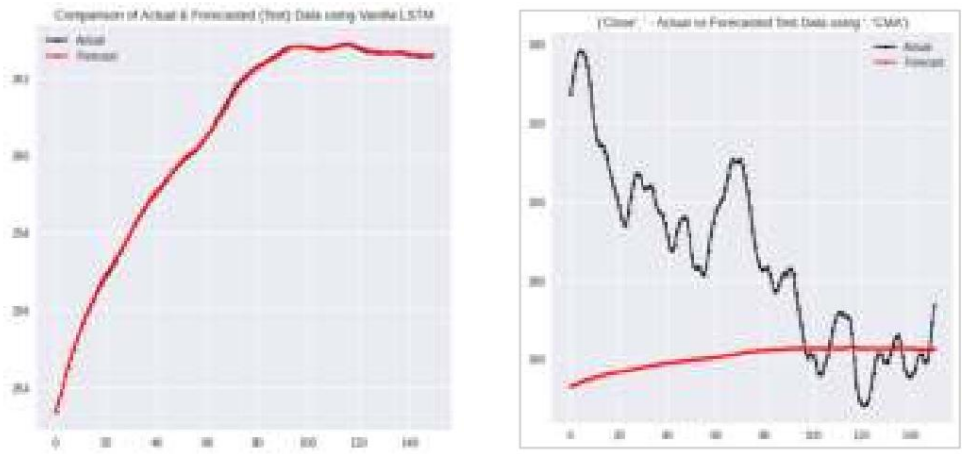
Diverses techniques de prévision ont été appliquées à un ensemble de données boursières univariées, englobant à la fois l'apprentissage automatique et des méthodes statistiques. Des techniques d'apprentissage automatique telles que LSTM, GRU et MLP, ainsi que des méthodes statistiques telles que la moyenne mobile simple, la moyenne mobile exponentielle et ARIMA, ont été testées avec différents hyperparamètres, notamment des couches cachées, des neurones et des époques. De même, diverses variantes de techniques statistiques ont été évaluées, telles que différents ordres pour ARIMA. Ces modèles ont été évalués sur l'ensemble de données de test à l'aide de diverses métriques.

Les résultats ont indiqué que le modèle d'apprentissage profond Vanilla LSTM fonctionnait le mieux sur l'ensemble de données, atteignant un RMSE de 0,0294 avec 50 époques et 40 unités. Les valeurs prédites générées par ce modèle ont été comparées aux valeurs attendues sur l'ensemble de test, comme le montre la figure 4a. Parmi les techniques statistiques, la moyenne mobile exponentielle (EMA) a obtenu les meilleurs résultats, avec un MAE de 2,0925 et un RMSE de 2,5814. EMA a été classée 127ème sur 139 modèles. À l'inverse, le modèle basé sur la moyenne mobile cumulée (CMA) a présenté les pires performances, avec un MAE de 27,2085 et un RMSE de 35,1176. La figure 4b présente une comparaison entre les valeurs réelles et les valeurs prévues produites par le modèle CMA.

Machine Translated by Google

Tableau 3. Résumé des performances du modèle de l'expérience basée sur des séries chronologiques univariées.

Rang	Prétraitement	Modèle	Paramètres	MAE	RMSE
Partie 1 : Séries chronologiques boursières sans problèmes de qualité des données					
1	N / A	Vanille LSTM	Époque = 50, unités = 40	0,02	0,03
2	N / A	GRU	Époque = 75, unités = 50	0,02	0,03
3	N / A	Vanille LSTM	Époque = 25, unités = 50	0,02	0,03
4	N / A	GRU	Époque = 50, unités = 40	0,03	0,04
127	N / A	Exponentiel	-	2.09	2,58
135	N / A	Régresseur d'arbre de décision	Split = MSE	7.57	8.03
136	N / A	Aléatoire Forêt	100 arbres, split = MSE	7,69	8h15
138	N / A	RA	-	21.11	26,88
139	N / A	RMR	-	27.21	35.12
Partie 2 : Séries chronologiques boursières avec données manquantes					
1	RMR	ARIMA	(7, 1, 2)	0,57	0,61
2	Spline 3	ARIMA	(7, 1, 7)	0,65	0,84
	Exponentiel	Exponentiel	-	1,92	2.43
6 59	Temps	RVS	Noyau = poly	3,64	4,73
64	Produits dérivés	RVS	Noyau = RBF	3,68	4,75
69	Polynôme par morceaux	LSTM empilé	Époque = 100, unités = 50	3,66	4,76
70	Produits dérivés	RNN	Époque = 25, unités = 50	3,67	4,76
1269	Polynôme 7	RMR	-	27.82	35,76
1270	Polynôme 9	RMR	-	28h46	36.52
Partie 3 : Série chronologique de crypto-monnaie avec données manquantes					
	GAGNER	LSTM bidirectionnel	Époque = 100, unités = 50	9,09	10.03
1	Puce	LSTM empilé	Époque = 100, unités = 50	8,86	10.64
	d'indexation	Vanille LSTM	Époque = 100, unités = 50	10,82	12.11
2	Exponentiel	GRU-bidirectionnel	Époque = 25, unités = 50	12,95	14.67
3 10	Cubique	RA	-	243,97	291,00
693	GAGNER une	Régresseur DT	Split = MSE	260,87	324.39
703	puce électronique	Forêt aléatoire	100 arbres, split = MSE	292,03	362.15
729 824	RMR	RVS	Noyau = polynôme	2575,77	2576.76
892	Spline 5	RMR	-	7431.54	9344.14



(un) (b)

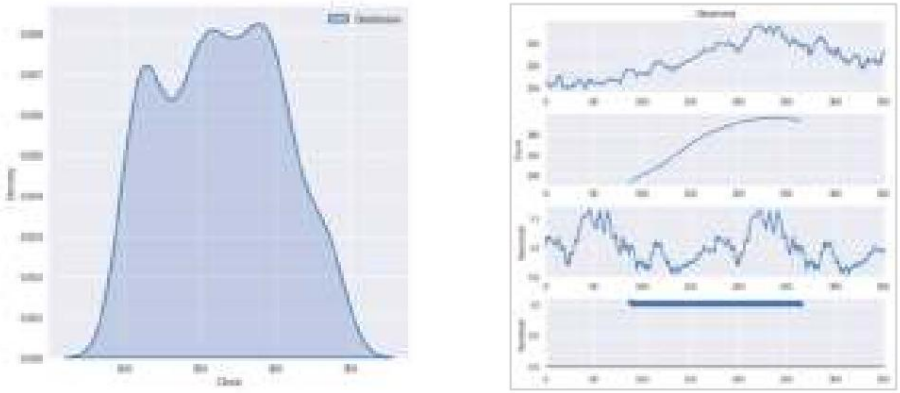
Figure 4. (a) Comparaison des données réelles et prévues générées par Vanilla LSTM en utilisant 50 époques et 40 unités. (b) Comparaison des données réelles et prévues générées par moyenne mobile cumulée (CMA).

ARIMA se distingue comme l'une des techniques les plus efficaces pour les séries chronologiques prévision. Dans Preptimize, ARIMA a été implémenté avec un ordre optimisé de (2, 1, 2), validé par le test de Ljung – Box. Cependant, malgré ces optimisations, le résultat Le RMSE de 3,6001 n'était pas particulièrement impressionnant.

Machine Translated by Google

L'ensemble de données boursières a été utilisé dans la deuxième partie de l'expérience concernant les problèmes de données manquantes. Comme mentionné précédemment, les données ont été intentionnellement supprimées pour tester les performances du framework en les comparant aux données d'origine.

Preptimize a été appliqué à cet ensemble de données contenant les tuples manquants ; le cadre a généré un rapport de plus de 3 500 pages avec une description détaillée des étapes appliquées à l'ensemble de données donné, y compris le prétraitement et l'analyse finale. Entre autres détails, le rapport comprenait des détails sur le type de données manquantes. Le cadre résumait les données comme étant manquantes complètement au hasard (MCAR), normalement distribuées et stationnaires, sans saison ni tendance. Les figures 5a et b montrent respectivement la tendance et la répartition saisonnière de la série chronologique et la distribution des données.



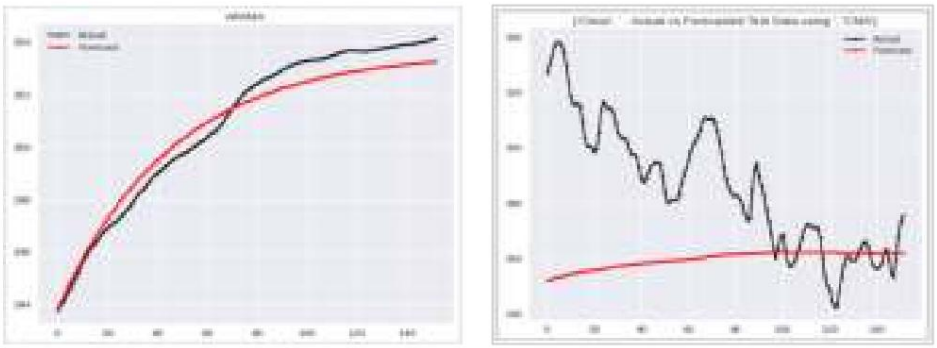
(un) (b)

Figure 5. (a) Graphique de décomposition du cours de clôture pour diviser la composante de tendance et de saisonnalité. Le graphique montre une légère tendance à la hausse et aucune saisonnalité. (b) Le graphique montre la distribution normale des données sur les prix de clôture.

Le résumé des performances de tous les modèles appliqués aux données a été calculé et stocké sur une feuille. La fiche récapitulative montrait qu'environ 1 270 modèles avaient été appliqués. Ces 1 270 modèles ont été générés à la suite de différentes combinaisons formées par des techniques de prétraitement et de prévision. Par exemple, d'abord, une stratégie de prétraitement telle que le remplissage aval a été appliquée pour imputer les données manquantes, puis toutes les autres techniques de prévision ont été appliquées pour la prévision univariée, y compris les techniques statistiques et d'apprentissage automatique, puis chaque modèle a été évalué à l'aide de quatre mesures différentes d'évaluation des performances. .

L'expérience impliquait d'essayer plusieurs techniques et modèles de prétraitement, notamment l'interpolation, le remplissage avant et arrière, les moyennes mobiles et les techniques d'apprentissage en profondeur telles que le GAN pour l'imputation. Chaque modèle a été évalué sur l'ensemble de données de test à l'aide de diverses métriques. Les résultats ont indiqué qu'ARIMA (p, d, q) avec l'ordre (7, 1, 2) fonctionnait mieux lorsque les données manquantes étaient imputées à l'aide de la CMA. Le test de Ljung – Box a confirmé l'absence d'autocorrélation dans le bruit blanc et le modèle. Ce modèle ARIMA a donné un MAE de 0,56997 et un RMSE de 0,6099. Les cinq principaux modèles appartenaient principalement à ARIMA, utilisant différentes techniques d'imputation telles que l'ordre spline 3, l'ordre polynomial 5, le remplissage vers l'arrière et le remplissage. Après ARIMA, la moyenne mobile exponentielle a donné de bons résultats avec plusieurs techniques d'imputation, notamment les techniques linéaires, slinéaires, polynomiales par morceaux et dérivées. Les 85 premiers modèles, qui ont donné de bons résultats, étaient tous des modèles statistiques. À l'inverse, le modèle basé sur la CMA traité par un polynôme de degré 9 a donné les pires résultats, avec un MAE de 28,4643 et un RMSE de 36,5159. La figure 6 illustre les performances des meilleurs et des pires modèles de l'expérience, la figure 6a affichant les valeurs prédites par le modèle ARIMA (7,1,2) sur l'ensemble de données imputées par CMA par rapport aux valeurs d'origine de l'ensemble de test, tandis que La figure 6b présente les performances du modèle CMA. La régression vectorielle de support avec un modèle de noyau polynomial imputé à l'aide d'une interpolation temporelle s'est révélée être la meilleure dans la catégorie de l'apprentissage automatique, donnant un MAE de 3,6430 et un RMSE de 4,7306. Les erreurs calculées pour les modèles mentionnés sont présentées dans le tableau 2.

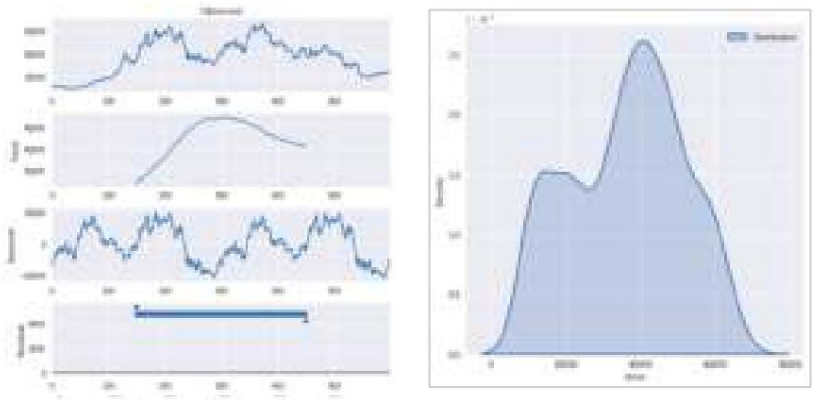
Machine Translated by Google



(un) (b)

Figure 6. (a) Comparaison des données réelles et prévues générées par ARIMA (7,1,2) imputées par RMR. (b) Comparaison des données réelles et prévues générées par la RMR imputée par polynôme interpolation de degré 9.

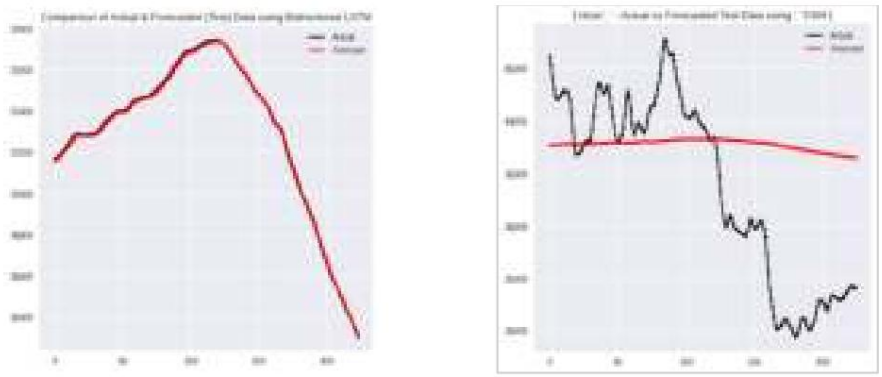
La troisième partie de l'expérience s'est concentrée sur l'analyse des données de crypto-monnaie, en particulier Bitcoin (BTC-USD) de la bourse Coinbase. La crypto-monnaie fonctionne indépendamment des systèmes bancaires traditionnels, en utilisant la technologie blockchain pour les transactions. l'authentification et la tenue des dossiers. L'ensemble de données, s'étendant sur environ deux ans à partir de D'août 2020 à août 2022, il s'agissait de 747 enregistrements de prix Bitcoin en USD. Pour évaluer le performances du framework dans la gestion des données manquantes, 30 % de l'ensemble de données a été aléatoire supprimé. Preptimize a traité les données, les identifiant comme stationnaires sans aucune trace perceptible. saisonnalité ou tendance. Malgré la nature numérique de l'ensemble de données et l'absence d'asymétrie, il présentait une variance considérable, indiquant une volatilité importante des prix au cours de la période. Figures 7a,b montrent respectivement la tendance et la répartition saisonnière de la série chronologique et la distribution des données. Diverses techniques d'imputation, notamment des méthodes d'apprentissage profond, d'interpolation et de moyennes mobiles, ont été utilisées pour corriger les valeurs manquantes, l'ensemble de données étant ensuite divisé en ensembles de formation et de test pour les prévisions. Au total, 892 modèles ont été formés et évalué, avec LSTM bidirectionnel combiné à l'imputation GAIN émergeant comme le meilleur performant, atteignant un RMSE de 10,028 et un MAE de 9,0879. Parmi les techniques statistiques, l'autorégression (AR) s'est montrée prometteuse lorsqu'elle est associée à une interpolation cubique, se classant à 693. À l'inverse, le modèle basé sur la CMA avec interpolation spline et polynôme du cinquième degré a présenté les performances les plus médiocres, avec un RMSE de 9344,1381 et un MAE de 7431,54. Notamment , les algorithmes d'apprentissage automatique basés sur des arbres ont fonctionné de manière sous-optimale, avec la décision le régresseur d'arbres et les modèles de forêts aléatoires se classent respectivement à 703 et 729, comme le montre Tableau 2. Les graphiques de performances des meilleurs et des pires modèles sont fournis dans la figure 8.



(un) (b)

Figure 7. (a) Graphique de décomposition des cours de clôture du Bitcoin (b) Distribution des données : le graphique montre distribution normale des données du cours de clôture du Bitcoin en USD.

Machine Translated by Google



(un) (b)

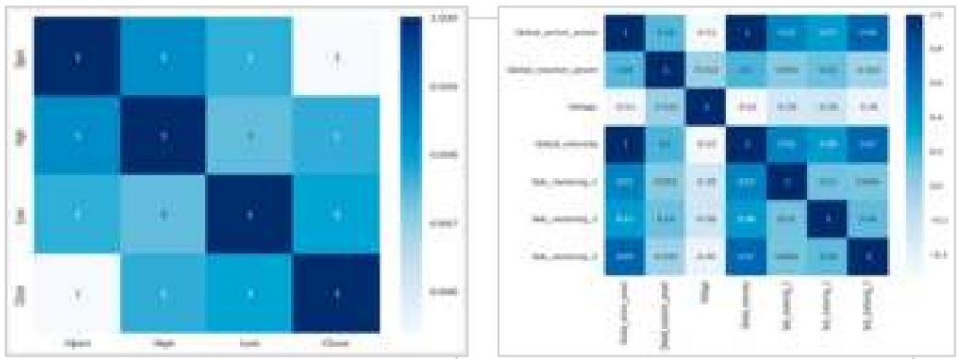
Figure 8. (a) Comparaison des données réelles et prévues générées par le LSTM bidirectionnel imputé par GAIN. (b) Comparaison des séries réelles et prévues générées par les RMR imputées par spline de Polynôme du 5ème degré.

Les résultats indiquent que Preptimize fonctionne bien sur différents types de séries chronologiques. données. Par exemple, le modèle Vanilla LSTM a affiché la meilleure performance en bourse données, avec un RMSE de 0,0294, tandis que les modèles ARIMA avec diverses techniques d'imputation les meilleurs résultats sur les ensembles de données avec des valeurs manquantes, soulignant la polyvalence du cadre et l'efficacité.

3.2. Expérience 2 : séries multivariées

La deuxième expérience s'est concentrée sur l'évaluation des performances de Preptimize sur des données de séries chronologiques multivariées, avec et sans valeurs manquantes. Les données complètes du stock marché ont été utilisés dans la première partie, tandis que la deuxième partie a introduit les instances manquantes en utilisant un ensemble de données sur la consommation électrique individuelle des ménages.

Dans la première partie, l'ensemble de données boursières a capturé les performances de Microsoft (MSFT) sur la base des valeurs d'ouverture, de haut, de bas et de clôture du 3 août 2020 au 1er août 2022, totalisant 503 enregistrements. Preptimize a commencé par le profilage des données, révélant une forte corrélation par paire parmi les colonnes, indiquant la dépendance. Cependant, les colonnes « ouvert » et « haut » étaient s'avèrent non stationnaires, contenant potentiellement des composantes de tendance et de saisonnalité, comme représenté sur la figure 9 ; la carte thermique illustre la corrélation par paire.



(un) (b)

Figure 9. (a) Carte thermique de corrélation montrant la corrélation entre les paires dans un ensemble de données multivariées du marché boursier. (b) Carte thermique de corrélation montrant la corrélation entre les paires de consommation d'énergie ensemble de données multivariées.

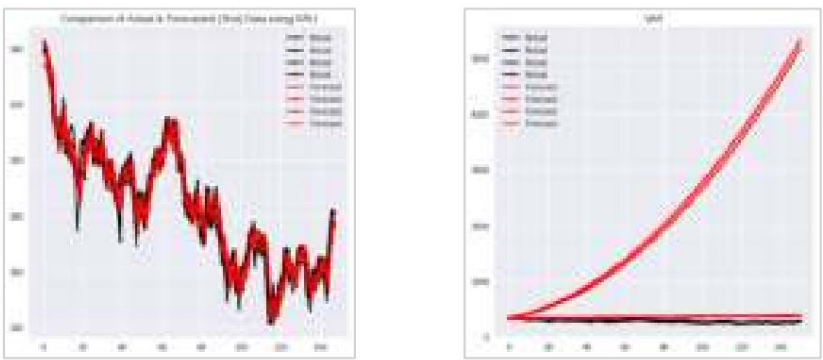
Passant à la prévision, aucun prétraitement n'étant nécessaire, Preptimize a appliqué diverses techniques, notamment VAR, VMA, VARMA, MLP, régresseur d'arbre de décision, randomisation. forêt, RNN, GRU, GRU-bidirectionnel et Stacked LSTM. La transformation a été appliquée à colonnes « ouvertes » et « hautes » pour les rendre stationnaires avant la prévision. Parmi les modèles

Machine Translated by Google

mentionné dans le tableau 4, le GRU avec 50 unités et 100 époques est apparu comme le plus performant, avec 3,9242 MAE et 5,1312 RMSE. La figure 10a présente les performances du GRU modèle sur l'ensemble de test multivarié. Cependant, les modèles statistiques comme VAR, VMA et VARMA ont montré de mauvaises performances, ne parvenant pas à comprendre le modèle de base de la série multivariée . La figure 10b présente les performances du modèle VAR sur les données multivariées. En outre, l'analyse des performances des colonnes individuelles a indiqué que les catégories « ouverte » et Les colonnes « hautes », initialement considérées comme non stationnaires, posaient des problèmes de prédiction. Malgré en appliquant la transformation par différence, ces colonnes présentaient des valeurs prédites divergentes significativement par rapport aux données originales, conduisant à des valeurs RMSE élevées.

Tableau 4. Résumé des performances du modèle de l'expérience basée sur des séries multivariées.

Rang	Prétraitement	Modèle	Paramètres	MAE	RMSE
Partie 1 : Séries chronologiques multivariées boursières sans problèmes de qualité des données					
	N / A	GRU	Époque = 100, unités = 50	3,92	5.13
1	N / A	RNN	Époque = 100, unités = 50	3,93	5.13
2	N / A	RNN	Époque = 25, unités = 50	3,95	5.19
3	N / A	MLP	Époque = 100, unités = 20 100	4,04	5.19
4	N / A	Forêt aléatoire	arbres, MSE divisé (2, 1,	5,47	7.08
17	N / A	VARMA	3) (0, 1,	407,98	657.08
23	N / A	VMA	1)	365,41	752,85
24 25	N / A	VAR	Commande = 5	963,64	1678.42
Partie 2 : Série chronologique multivariée sur la consommation électrique individuelle des ménages avec données manquantes					
	Exponentiel	GRU	Époque = 100, unités = 20	0,12	0,64
1	RMR	GRU	Époque = 100, unités = 20	0,13	0,64
	Crête bayésienne	MLP	Époque = 100, unités = 50	0,13	0,65
2	Forêt aléatoire	LSTM empilé	Époque = 100, unités = 50	0,13	0,65
3 4	KNNImputateur 5 (0,1,1)	VMA		0,43	1,88
330	KNNImputer 5 (2,1,3) polynôme_3	VARMA		0,43	1,88
331		RNN	Époque = 25, unités = 20	9,35	29h31
449 450	polynôme_3	Régresseur DT	MSE fractionné	14,53	58,72



(un) (b)

Figure 10. (a) Comparaison des performances du GRU pour les données de séries chronologiques multivariées pendant les tests. (b) VAR comparaison des performances du modèle pour les données de séries chronologiques multivariées pendant les tests.

La deuxième partie de l'expérience en série multivariée a testé Preptimize en utilisant le ensemble de données sur la consommation électrique des ménages individuels, visant à comprendre son comportement avec des données de séries chronologiques multivariées incomplètes [44]. Cet ensemble de données comprend sept colonnes, hors date et heure, enregistrant la consommation d'énergie sur quatre ans avec un taux d'échantillonnage d'une minute. Avec environ 2 075 259 instances, l'ensemble de données présentait des éléments manquants valeurs principalement en avril 2007. Les détails sont donnés dans le tableau 5.

En raison de la taille de l'ensemble de données, seule une semaine de données contenant des instances manquantes a été traitée . Avec 40 stratégies de prétraitement et 25 modèles de prévision, soit un total de 1 000 modèles ont été évalués. Cependant, en raison de limitations informatiques, l'expérience a été divisée en pièces, prenant plusieurs jours pour générer des résultats. Après avoir mené des expériences précédentes, 18 techniques de prétraitement réussies ont été combinées avec 25 modèles de prévision pour évaluation.

Machine Translated by Google

Tableau 5. Ensemble de données sur la consommation électrique des ménages individuels utilisé dans l'expérience 2.

Non.	Nom de la colonne	Unités	Description
1	global_active_power	Kilowatts	La puissance active totale consommée par le ménage
2	global_reactive_power	Kilowatts	La puissance réactive totale consommée par le ménage
3	tension	Volts	Tension moyenne
4	global_intensity	Ampères	Courant moyen
5	sub_metering_1	Wattheures	Énergie active pour la cuisine
6	sub_metering_2	Wattheures	Énergie active pour la lessive
7	sub_metering_3	Wattheures	Énergie active pour les systèmes de climatisation

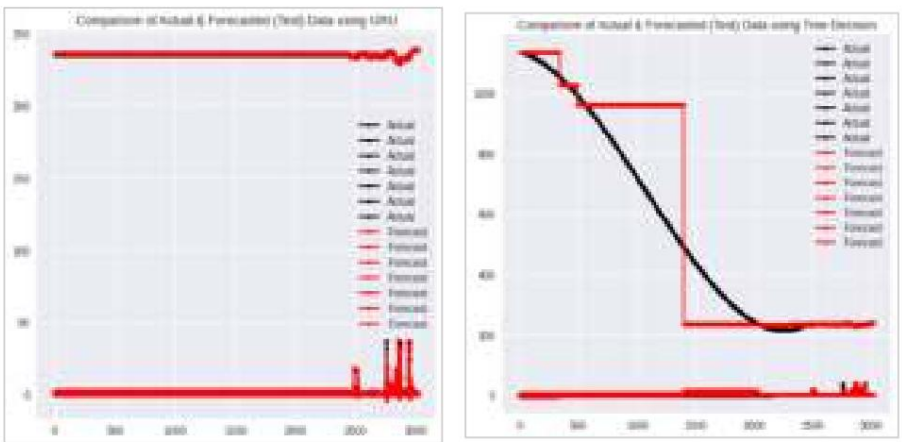
Le rapport complet généré par Preptimize a commencé par profiler les données, en analysant chaque attribut individuellement. Une carte thermique de corrélation, représentée précédemment sur la figure 8b, illustre les relations par paires entre les colonnes. Notamment, l'intensité mondiale et la mondialisation la puissance active présentait une forte corrélation positive, tandis que le sous-comptage 3, associé à systèmes de contrôle climatique, fortement corrélés à la puissance active mondiale et à l'intensité mondiale. En revanche, la puissance réactive globale n'a montré aucune corrélation avec d'autres attributs, en particulier tension et sous-comptage 3.

En outre, le rapport évalue le pourcentage de données manquantes, à environ 37 %, identifier les colonnes telles que le sous-compteur 1 et le sous-compteur 2 comme non normalement distribuées avec des valeurs d'aplatissement élevées. Preptimize a classé les quatre colonnes avec des données manquantes comme Type MNAR. Des tentatives de normalisation des distributions des sous-comptages 1 et des sous-compteurs 2 ont été gêné par la possibilité de valeurs nulles ou négatives, excluant la transformation Box – Cox.

Après la décomposition des tendances et des saisons, diverses techniques d'imputation, y compris des méthodes itératives, ont été appliquées. Chaque ensemble de données imputées a été évalué à l'aide de 18 critères différents. modèles de prévision, aboutissant à une évaluation complète des performances.

Preptimize a généré une fiche d'évaluation regroupant 450 modèles, détaillant les techniques de prétraitement utilisées lors de l'imputation. Comme indiqué précédemment, le tableau 3 présente les mesures de performance de certains modèles, en mettant en évidence le modèle le plus performant comme GRU imputé avec une moyenne mobile exponentielle, donnant un RMSE de 0,6393 et MAE de 0,1225. À l'inverse, le régresseur de l'arbre de décision, imputé par interpolation polynomiale de degré 3, a présenté les performances les plus médiocres, avec un RMSE de 14,5308 et un MAE. de 58,7152.

Remarquablement, les modèles statistiques ont démontré des performances constantes lorsque les données étaient imputé à l'aide de KNNImputer avec deux voisins, avec classement VMA, VARMA et VAR étroitement. Cependant, certains modèles d'apprentissage automatique s'en sortent moins bien. Graphiques de performances pour Les modèles de régresseur GRU et d'arbre de décision sont présentés respectivement dans les figures 11a et 11b.



(un) (b)

Figure 11. (a) GRU imputé avec une moyenne mobile exponentielle : comparaison des performances pour la puissance données de série de consommation pendant les tests. (b) Arbre de décision imputé avec interpolation polynomiale de degré 3 : comparaison des performances des séries de consommation d'énergie pendant les tests.

Machine Translated by Google

En résumé, l'expérience a révélé que la moyenne mobile exponentielle, la crête bayésienne, les techniques basées sur KNN et DataWig sont des méthodes d'imputation efficaces, tandis que l'interpolation polynomiale de degré 3, l'interpolation spline de degré 3 et HotDeck ont donné des résultats sous-optimaux.

4. Conclusions et discussion

L'analyse des séries chronologiques est devenue de plus en plus complexe en raison de la multitude de techniques disponibles. Le choix de la méthode de prévision et de la stratégie de prétraitement appropriées en fonction des caractéristiques de la série chronologique est crucial pour obtenir des résultats fiables. Des étapes appropriées doivent être suivies dans l'analyse des séries chronologiques pour identifier les meilleurs modèles applicables dans des contextes pratiques. Les modèles statistiques et d'apprentissage automatique, ainsi que diverses techniques d'imputation pour gérer les données manquantes, offrent diverses options d'analyse. Avec de nombreuses recherches détaillant les performances de chaque technique sur des ensembles de données spécifiques, il existe un besoin pour un système capable d'appliquer efficacement ces techniques pour identifier le meilleur modèle grâce à l

Grâce à des expériences sur divers ensembles de données de séries chronologiques, l'efficacité du cadre proposé, Preptimize, est évidente. Preptimize recommande plusieurs modèles de modèles de prédiction, servant d'étape préliminaire à l'analyse de la série chronologique donnée et permettant d'économiser beaucoup de temps et d'efforts. Les chercheurs peuvent encore améliorer ces modèles en créant des combinaisons hybrides, en ajustant les paramètres et en ajoutant davantage de fonctionnalités pour améliorer la précision des prévisions. Preptimize génère des rapports détaillés couvrant toutes les étapes de traitement des données, ainsi que des tableaux, des graphiques, des statistiques et une feuille d'évaluation complète présentant les performances de tous les modèles testés sur l'ensemble de données à l'aide de différentes mesures d'erreur.

Différentes expériences ont été menées sur divers ensembles de données pour évaluer les performances de ce framework. L'efficacité du cadre peut être jugée par la mesure dans laquelle le modèle de plan correspond à ceux utilisés dans la littérature et en comparant la précision basée sur le RMSE. Les performances de référence pour les modèles de prévision boursière sont basées sur des modèles LSTM, SVR et hybrides combinant des méthodes statistiques avec des techniques d'apprentissage automatique, rapportant souvent des valeurs RMSE aussi basses que 0,02 pour les cours des actions. Preptimize a également suggéré LSTM et ARIMA pour la prévision du cours des actions. Les performances de référence pour la prévision boursière de Microsoft (MSFT) ont rapporté un RMSE de 0,05 avec un modèle hybride combinant LSTM et RNN [45]. Le modèle de plan proposé avec Preptimize a atteint un RMSE de 0,029 en utilisant Vanilla LSTM. La forte volatilité du Bitcoin, avec des fluctuations de prix importantes sur de courtes périodes, rend la tâche difficile pour les modèles de prévision traditionnels. La prévision du prix du Bitcoin a rapporté des valeurs RMSE allant de 4 à 7 pour les modèl Preptimize a suggéré le LSTM (RMSE 10.02) pour la prédiction Bitcoin, conformément aux conclusions de la littérature. Une étude [46] a examiné divers modèles de prévision, notamment les réseaux neuronaux à action directe, les techniques LSTM, GRU, ARIMA et IA, sur un ensemble de données de consommation électrique d'un ménage individuel pour comparer leurs performances. Le RMSE moyen sur 12 modèles différents était de 0,83, le RMSE le plus bas de 0,75 étant atteint à la fois par LSTM et GRU. En revanche, Preptimize a recommandé d'utiliser GRU prétraité par lissage exponentiel , ce qui a abouti à un RMSE de 0,639. Sur la base de la discussion, nous pouvons conclure que Preptimize recommande efficacement des modèles de prédiction appropriés pour un ensemble de données donné, facilitant ainsi une prise de décision rapide avec un minimum d'effort.

Preptimize présente certaines limites qui doivent être reconnues. Premièrement, il est gourmand en ressources, nécessitant une puissance et un temps de calcul importants en raison de son approche globale consistant à appliquer de nombreuses techniques sur un échantillon de données pour identifier les méthodes de prétraitement et de prévision optimales. Deuxièmement, Preptimize prend principalement en charge les ensembles de données avec une distribution normale. Lorsqu'il s'agit de distributions non normales, il transforme la série pour qu'elle se rapproche de la normalité, ce qui peut ajouter de la complexité et des inexactitudes potentielles, surtout si la transformation n'atteint pas complètement une distribution normale. Nous visons à résoudre ces problèmes dans les travaux futurs.

Bien que le cadre proposé se concentre sur la résolution du problème des données manquantes dans l'analyse des séries chronologiques, de nombreux autres défis tels que la détection des valeurs aberrantes, la dérive des concepts et l'hétéroscédasticité persistent. Relever ces défis nécessite davantage d'automatisation

Machine Translated by Google

d'analyse de séries chronologiques, une tâche qui sera poursuivie dans les phases futures de la recherche. Les travaux futurs pourraient explorer l'extension de Preptimize pour gérer des types plus complexes de données de séries chronologiques, telles que celles comportant plusieurs saisons ou des intervalles irréguliers. De plus, l'intégration de modèles plus hybrides basés sur des techniques d'apprentissage automatique et l'extension de l'applicabilité du cadre à d'autres domaines tels que les soins de santé et la surveillance environnementale pourraient encore améliorer son utilité.

En outre, même si nous avons testé la robustesse et l'exhaustivité du cadre par rapport à plusieurs scénarios, nous reconnaissons que des scénarios supplémentaires, tels que des données saisonnières univariées, des pourcentages élevés de valeurs manquantes, des distributions non normales, des données multisaisonnières et des tendances à la baisse, pourraient générer d'autres résultats précieux. connaissances. Cependant, il n'était pas possible d'inclure tous les scénarios possibles dans les travaux initiaux. Nous encourageons d'autres chercheurs à explorer ce cadre avec différents ensembles de données pour combler cette lacune en matière de recherche.

Contributions de l'auteur : conceptualisation, méthodologie, validation, analyse formelle, visualisation, rédaction – préparation du projet original, MU ; supervision, ZAM; révision et édition, RQ et AZ Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche n'a reçu aucun financement externe.

Déclaration du comité d'examen institutionnel : sans objet.

Déclaration de consentement éclairé : sans objet.

Déclaration de disponibilité des données : Les ensembles de données utilisés dans cette étude sont disponibles en ligne. L' ensemble de données sur la consommation d'énergie est disponible sur https://archive.ics.uci.edu/ml/datasets/individual+household+ consommation électrique+énergie+ (consulté le 1er juin 2024) L'ensemble des données financières de la bourse est disponible sur https://aroussi.com/ (consulté le 1er juin 2024). L'ensemble de données de crypto-monnaie est disponible sur https://pypi.org/project/Historic-Crypto/ (consulté le 1er juin 2024).

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Références

1. Liu, Z. ; Zhu, Z. ; Gao, J. ; Xu, C. Méthodes de prévision pour les données de séries chronologiques : une enquête. Accès IEEE 2021, 9, 91896-91912. [Référence croisée]
2. Asadi, S. ; Esmaïl, H. ; Farhad, M. ; Masoud, NM Hybridation des réseaux neuronaux et des données évolutives de Levenberg – Marquardt prétraitement pour la prévision boursière. Connaître. Système basé. 2012, 35, 245-258. [Référence croisée]
3. Di Persio, L. ; Fraccarolo, N. Prévisions de consommation d'énergie par arbres de régression boostant le gradient. Mathématiques 2023, 11, 1068. [Référence croisée]
4. Crier, JD ; Kellet, N. Analyse des séries chronologiques : avec applications dans R, 2e éd. ; Springer : New York, NY, États-Unis, 2008 ; Tome 2, p. 31.
5. Zhou, Q. ; Ooka, R. Influence du prétraitement des données sur les performances du réseau neuronal pour la reproduction de simulations CFD de distribution du flux d'air intérieur non isotherme. Construction énergétique. 2021, 230, 110525. [Réf. croisée]
6. Chanson, S. ; Zhang, A. ; Wang, J. ; Yu, PS SCREEN : Nettoyage des données de flux sous contraintes de vitesse. Dans les actes de l'ACM 2015 Conférence internationale SIGMOD sur la gestion des données, Melbourne, Australie, 31 mai-4 juin 2015.
7. Bilalli, B. ; Abelló, A. ; Aluja-Banet, T. ; Wrembel, R. Prétraitement automatisé des données via méta-apprentissage. Dans Actes de la Conférence internationale sur l'ingénierie des modèles et des données, Almería, Espagne, 21-23 septembre 2016 ; Springer : Cham, Suisse, 2016.
8. Zhang, A. ; Chanson, S. ; Wang, J. ; Yu, PS Nettoyage des données des séries chronologiques : de la détection des anomalies à la réparation des anomalies. Proc. Fondation VLDB. 2017, 10, 1046-1057. [Référence croisée]
9. Shimizu, K. ; Ponce-Hernandez, R. ; Ahmed, OS ; Ota, T. ; Gagner, ZC ; Mizoué, N. ; Yoshida, S. Utilisation des séries chronologiques Landsat pour détecter les perturbations forestières dans les forêts tropicales exploitées sélectivement au Myanmar. Peut. J. Pour. Rés. 2017, 47, 289-296. [Référence croisée]
10. Zhu, Z. Détection de changement à l'aide de séries chronologiques Landsat : un examen des fréquences, du prétraitement, des algorithmes et des applications. ISPRS J. Photogramme. Télésens. 2017, 130, 370-384. [Référence croisée]
11. Karim, F. ; Majumdar, S. ; Darabi, H. ; Chen, S. Réseaux entièrement convolutifs LSTM pour la classification des séries chronologiques. Accès IEEE 2017, 6, 1662-1669. [Référence croisée]
12. Gschwandtner, T. ; Erhart, O. Connaissez votre ennemi : identification des problèmes de qualité des données de séries chronologiques. Dans les actes de l'IEEE Symposium sur la visualisation du Pacifique (PacificVis), Kobe, Japon, 10-13 avril 2018.
13. Jeenanunta, C. ; Abeyrathna, KD; Dilhani, MS ; Hnin, SW; Phyo, PP Détection des valeurs aberrantes des séries chronologiques pour la charge électrique à court terme prévision de la demande. Int. Sci. J. Ing. Technologie. (ISJET) 2018, 2, 37-50.
14. Wang, X. ; Wang, C. Nettoyage des données de séries chronologiques : une enquête. Accès IEEE 2019, 8, 1866-1881. [Référence croisée]
15. Ding, X. ; Wang, H. ; Su, J. ; Li, Z. ; Li, J. ; Gao, H. Cleanits : Un système de nettoyage de données pour les séries chronologiques industrielles. Proc. Fondation VLDB. 2019, 12, 1786-1789. [Référence croisée]

Machine Translated by Google

16. Ruiz, L. ; Pégalajar, M. ; Arcucci, R. ; Molina-Solana, M. Une méthodologie de regroupement de séries chronologiques pour l'extraction de connaissances dans le domaine de l'énergie données de consommation. Système expert. Appl. 2020, 160, 113731. [Réf. croisée]
17. Jarrett, D. ; Yoon, J. ; Bica, moi; Qian, Z. ; Ercole, A. ; Schaar, MVD Clairvoyance : une boîte à outils de pipeline pour les séries chronologiques médicales. Dans Actes de la Conférence internationale sur les représentations d'apprentissage, virtuelle, 3-7 mai 2021.
18. Desai, V. ; Dinesha, HA Une approche hybride des méthodes de prétraitement des données. Dans Actes de la Conférence internationale de l'IEEE pour l'innovation technologique (INOCON), Bangalore, Inde, 6-8 novembre 2020.
19. Sousa, R. ; Amado, C. ; Henriques, R. AutoMTS : Traitement entièrement autonome de données de séries chronologiques multivariées provenant de réseaux de capteurs hétérogènes. Dans Actes de la Conférence internationale sur les réseaux hétérogènes pour la qualité, la fiabilité, la sécurité et la robustesse, virtuelle, 29-30 novembre 2020 ; Springer : Cham, Suisse, 2020.
20. Chen, X. ; Deng, L. ; Huang, F. ; Zhang, C. ; Zhang, Z. ; Zhao, Y. ; Zheng, K. Daemon : Détection et interprétation d'anomalies non supervisées pour les séries chronologiques multivariées. Dans les actes de la 37e conférence internationale de l'IEEE 2021 sur l'ingénierie des données (ICDE), La Canée, Grèce, du 19 au 22 avril 2021 ; pp. 2225-2230.
21. Chauhan, K. ; Jani, S. ; Thakkar, D. ; Dave, R. ; Bhatia, J. ; Tanwar, S. ; Obaidat, MS Apprentissage automatique automatisé : la nouvelle vague d'apprentissage automatique. Dans Actes de la 2e Conférence internationale sur les mécanismes innovants pour les applications industrielles (ICIMIA), Bangalore, Inde, 5-7 mars 2020.
22. Sarafanov, M. AutoML pour Time Series : définitivement une bonne idée. Disponible en ligne : https://towardsdatascience.com/automl-for-time-series-definitelya-good-idea-c51d39b2b3f (consulté le 1er septembre 2023).
23. Soleil, B. ; Maman, L. ; Shen, T. ; Geng, R. ; Zhou, Y. ; Tian, Y. Une méthode robuste basée sur les données pour la multisaisonnalité et l'hétéroscédasticité dans Prétraitement des séries chronologiques. Fil. Commun. Foule. Calculer. 2021, 2021, 6692390. [CrossRef]
24. Zhang, médecin généraliste ; Min, Q. Prévisions de réseaux neuronaux pour les séries chronologiques saisonnières et tendancielles. EUR. J. Opér. Rés. 2005, 160, 501-514. [Référence croisée]
25. Ranjan, KG; Prusty, BR; Jena, D. Comparaison de deux méthodes de nettoyage de données appliquées aux séries chronologiques volatiles. Dans Actes de la Conférence internationale sur les applications et la technologie de l'électronique de puissance dans le scénario énergétique actuel (PETPES), Mangalore, Inde, 29-31 août 2019.
26. Ranjan, KG; Tripathie, DS ; Prusty, BR; Jena, D. Une détection et une correction améliorées des valeurs aberrantes basées sur la prédiction par fenêtre glissante pour les séries chronologiques volatiles. Int. J. Numéro. Modèle. Électron. Réseau. Champs des appareils 2021, 34, e2816. [Référence croisée]
27. Lv, P. ; Wu, Q. ; Xu, J. ; Shu, Y. Prédiction de l'indice boursier basée sur la décomposition des séries chronologiques et le modèle hybride. Entropie 2022, 24, 146. [Référence croisée]
28. Brunel, B. ; Alsamad, F. ; Piot, O. Vers l'apprentissage automatique automatisé en spectroscopie vibrationnelle : utilisation et paramètres d' algorithmes génétiques pour le prétraitement et l'optimisation de la régression. Chimio. Intell. Laboratoire. Système. 2021, 219, 104444. [Réf. croisée]
29. Kumar, S. 8 Bibliothèques AutoML pour automatiser le pipeline d'apprentissage automatique. 2020. Disponible en ligne : https://medium.com/swlh/8-bibliothèques-automl-pourautomatiser-machine-learning-pipeline-3da0af08f636 (consulté le 1er août 2023).
30. Jang, W.-J. ; Lee, ST ; Kim, J.-B. ; Gim, G.-Y. Une étude sur le profilage des données : se concentrer sur l'indice de qualité des valeurs d'attribut. Appl. Sci. 2019, 9, 5054. [Réf. croisée]
31. Ghaderpour, E. ; Pagiatakis, SD; Hassan, QK Une enquête sur la détection des changements et l'analyse des séries chronologiques avec des applications. Appl. Sci. 2021, 11, 6141. [CrossRef]
32. Zou, H. ; Yang, Y. Combinaison de modèles de séries chronologiques pour la prévision. Int. J. Prévisions. 2004, 20, 69-84. [Référence croisée]
33. Lim, B. ; Zohren, S. Prévisions de séries chronologiques avec apprentissage profond : une enquête. Philos. Trans. R. Soc. Un Mathématique. Phys. Ing. Sci. 2021, 379, 20200209. [Référence croisée]
34. Abbasimehr, H. ; Paki, R. ; Bahrini, A. Une nouvelle approche basée sur la combinaison de modèles d'apprentissage profond avec des méthodes statistiques pour Prévisions de séries chronologiques COVID-19. Calcul neuronal. Appl. 2022, 34, 3135-3149. [Référence croisée] [Pub Med]
35. Brown, TA Analyse factorielle confirmatoire pour la recherche appliquée ; The Guilford Press : New York, NY, États-Unis, 2006.
36. Haroon, D. Différence entre séries chronologiques. Dans Python Machine Learning Case Studies : cinq études de cas pour le Data Scientist ; Apress : New York, NY, États-Unis, 2017 ; p. 110.
37. Agiakloglou, C. ; Newbold, P. Preuves empiriques sur les tests de type Dickey-Fuller. J. Time Ser. Anal. 1992, 13, 471-483. [Référence croisée]
38. Goodfellow, I. ; Pouget-Abadie, J. ; Mirza, M. ; Xu, B. ; Warde-Farley, D. ; Ozaïr, S. ; Courville, A. ; Bengio, Y. Réseaux contradictoires génératifs. Commun. ACM 2020, 63, 139-144. [Référence croisée]
39. Yoon, J. ; Jordanie, J. ; Schaar, M. Gain : Imputation de données manquantes à l'aide de réseaux contradictoires génératifs. Dans les Actes de l'Internationale Conférence sur l'apprentissage automatique, Stockholm, Suède, 10-15 juillet 2018.
40. Biessmann, F. ; Rukat, T. ; Schmidt, P. ; Naidu, P. ; Schelter, S. ; Taptunov, A. ; Lange, D. ; Salinas, D. DataWig : Imputation de valeurs manquantes pour les tables. J. Mach. Apprendre. Rés. 2019, 20, 1–6.
41. Joenssen, DW; Bankhofer, U. Méthodes Hot Deck pour imputer les données manquantes. Dans Actes de l'atelier international sur l'apprentissage automatique et l'exploration de données dans la reconnaissance de formes, Berlin, Allemagne, 13-20 juillet 2012.
42. Holland, JH Adaptation dans les systèmes naturels et artificiels : une analyse introductive avec des applications à la biologie, au contrôle et à l'artificiel Intelligence; MIT Press : Cambridge, MA, États-Unis, 1992.
43. McDowall, D. ; McCleary, R. ; Bartos, BJ Analyse des séries chronologiques interrompues, 21e éd.; SAGE : New York, NY, États-Unis, 1980 ; p. 41.
44. Dua, D. ; Graff, C. Référentiel d'apprentissage automatique UCI ; Université de Californie, École d'information et d'informatique : Irvine, CA, États-Unis, 2019. Disponible en ligne : http://archive.ics.uci.edu/ml (consulté le 1er juin 2024).

Machine Translated by Google

45. Sourav, D. ; Apan, P. ; Sayan, S. ; Sayan, G. ; Oudatea, D. ; Chandra, D. ; Shilpi, B. Un nouveau modèle hybride utilisant Lstm et Rnn pour Prédiction boursière. Int. J. Ing. Rés. Technologie. 2024, 13. [Réf. croisée]
46. Gasparine, A. ; Slobodan, L. ; Cesare, A. Apprentissage profond pour la prévision de séries chronologiques : le cas de charge électrique. CAAI Trans. Intell. Technologie. 2022, 7, 1–25. [Référence croisée]

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Booléen basé sur le Deep Learning, séries chronologiques, détection d'erreurs et

Analyse prédictive dans les opérations de grues à conteneurs

Amruta Awasthi, Lenka Krpalkova et Joseph Walsh *

École des sciences, technologies, ingénierie et mathématiques (STEM), Université technologique de Munster (MTU), Kerry Campus, V92 HD4V Kerry, Irlande ; amruta.awasthi@research.ittralee.ie (AA) ; lenka.krpalkova@mtu.ie (LK)
* Correspondance : joseph.walsh@mtu.ie

Résumé : L'apprentissage profond est crucial dans la détection, le diagnostic et la prédiction des erreurs de la logistique maritime et des grues à conteneurs. Une nouvelle technique d'apprentissage en profondeur utilisant la mémoire à long terme et à court terme (LSTM) a détecté et anticipé les erreurs dans un système avec des données déséquilibrées. Le modèle LSTM a été formé sur des données d'erreur opérationnelle réelles provenant de grues à conteneurs. L'algorithme personnalisé utilise la technique de suréchantillonnage des minorités synthétiques (SMOTE) pour équilibrer les données déséquilibrées pour les erreurs de données opérationnelles (c'est-à-dire trop peu d'échantillons de classe minoritaire). Python était utilisé pour programmer. Des matrices de corrélation et des matrices de covariance de Pearson, Spearman et Kendall sont présentées. La perte de formation et de validation du modèle est affichée et les données restantes sont prédites. L'ensemble de test (30 % des données réelles) et les données prévues avaient des RMSE de 0,065. Une carte thermique d'une matrice de confusion a été créée à l'aide de Matplotlib et Seabo De plus, les sorties d'erreur pour la série chronologique pour les n secondes suivantes ont été projetées, les n secondes étant saisies par l'utilisateur. L'exactitude était de 0,996, la précision était de 1,00, le rappel était de 0,500 et le score f1 était de 0,667, selon les critères d'évaluation produits. Les expériences ont démontré que la technique est capable d'identifier les éléments critiques. Ainsi, les futures tentatives amélioreront la structure du modèle pour prévoir les erreurs du Big Data industriel. Cependant, l'avantage est qu'il peut gérer des données déséquilibrées, ce qui est généralement le cas de la plupart des secteurs. Avec des données supplémentaires, le modèle peut être encore amélioré.

check for updates

Citation : Awasthi, A. ; Krpalkova, L. ; Walsh, J. Basé sur l'apprentissage profond Booléen, séries chronologiques, détection d'erreurs, et analyse prédictive dans les opérations de grues à conteneurs. Algorithmes 2024, 17, 333. https://doi.org/10.3390/a17080333

Rédacteurs académiques : Mateus Mendes et Balduíno Mateus

Reçu : 16 juin 2024
Révisé : 20 juillet 2024
Accepté : 24 juillet 2024
Publié : 1 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

Mots-clés : analyse prédictive ; apprentissage profond ; analyse de données ; données booléennes de séries chronologiques ; détection d'erreurs ; réseaux neuronaux profonds ; Réseaux de neurones récurrents (RNN) ; Long Court Terme Réseaux de mémoire (LSTM); Technique de suréchantillonnage des minorités synthétiques (SMOTE); grues à conteneurs; industrie maritime

1. Introduction

1.1. Contexte Le

marché du transport maritime par conteneurs représente un marché de 116,04 milliards de dollars en 2024 et devrait atteindre 134,03 milliards de dollars d'ici 2029 [1]. À mesure que le volume du commerce mondial augmente, les impératifs d'une productivité et d'une efficacité opérationnelle accrues au sein des installations portuaires s'intensifient simultanément. Parallèlement à ces exigences, garantir la sécurité et la fiabilité des opérations des grues à conteneurs reste primordiale [2]. Les grues à conteneurs jouent un rôle central dans la logistique maritime moderne, facilitant le mouvement efficace des marchandises entre les navires et les côtes dans les terminaux à conteneurs du monde entier [3]. Les grues à portique à tour sont indispensables pour charger et décharger les conteneurs des navires et pour les déplacer dans la cour du termin

Les grues à conteneurs, qui opèrent le long des quais, utilisent des technologies complexes telles que des palonniers pour transférer les conteneurs entre les navires et les terminaux, contribuant ainsi à l'efficacité globale du port. Cette opération consiste à agencer précisément la grue pour qu'elle s'adapte aux différentes baies du navire, permettant ainsi l'accès aux conteneurs situés sur toute la longueur du navire. De plus, en plus des opérations sur les navires, des grues à conteneurs peuvent être utilisées pour empiler des conteneurs dans la cour du terminal, permettant ainsi un stockage provisoire ou ultérieur.

Machine Translated by Google

transport [4]. Les grues génèrent d'énormes quantités de données qui peuvent être analysées et des informations précieuses peuvent être obtenues pour une prise de décision efficace, meilleure et plus rapide par les industries.

Récemment, de plus en plus de recherches se sont concentrées sur les méthodes de diagnostic des pannes basées sur les données, en particulier celles basées sur les réseaux neuronaux. Cela est dû aux progrès de la technologie des capteurs et à l'augmentation de la capacité de stockage des données, qui ont amélioré les capacités de traitement des données des méthodes basées sur les réseaux neuronaux. Le fait de ne pas identifier les défauts du processus peut avoir des effets néfastes sur la productivité, la sécurité, la qualité de la production et l'économie du processus Un équipement de diagnostic automatisé efficace et avancé est essentiel dans les entreprises modernes pour détecter, diagnostiquer et réparer les comportements anormaux des processus. L'objectif principal de la détection et de l'isolation des défauts est d'alerter rapidement les opérateurs des problèmes potentiels, leur permettant de prendre les mesures nécessaires pour éviter une panne du système suite à l'apparition de problèmes [5].

La détection et la prévision rapides des erreurs sont cruciales pour optimiser les performances des grues et garantir la sécurité des opérations des grues à conteneurs.

L'importance d'une détection rapide des erreurs peut être comprise à travers plusieurs perspectives vitales : 1. Optimisation de la

productivité et de l'efficacité : La détection et la prévention rapides des erreurs optimisent les performances et l'efficacité opérationnelle de la grue. En identifiant et en traitant rapidement les erreurs, les opérateurs peuvent minimiser l'impact sur la productivité et le débit, garantissant ainsi que les grues fonctionnent à une capacité et une efficacité maximales [6]. Ceci peut être réalisé en prévoyant les erreurs à temps afin qu'elles ne se transforment pas en défauts et n'entraînent pas d'échecs. Cette optimisation des performances contribue à la productivité et à la rentabilité globales des opérations du terminal à conteneurs.

2. Prévention des perturbations opérationnelles : la détection rapide des erreurs permet d' identifier rapidement les problèmes ou dysfonctionnements potentiels dans les opérations de la grue [7]. En identifiant rapidement les erreurs, les opérateurs peuvent prendre des mesures correctives pour éviter les perturbations opérationnelles, telles que des pannes de grue ou des retards dans la manutention des marchandises. Cette approche proactive contribue à maintenir le bon déroulement des opérations au sein des terminaux à conteneurs, en minimisant les temps d'arrêt et en garantissant une utilisation efficace des ressources [8].
3. Amélioration des protocoles de sécurité : la détection précoce des erreurs permet aux opérateurs de mettre en œuvre des protocoles et des procédures de sécurité pour atténuer les risques et les dangers associés aux opérations de grue. En traitant rapidement les erreurs, les opérateurs peuvent prévenir les accidents, les blessures et les dommages aux marchandises ou aux infrastructures, en protégeant le bien-être du personnel et en préservant l'intégrité des installations du terminal [9].
4. Réduction des coûts de maintenance : La détection, le diagnostic et la prévision précoce des erreurs peuvent aider à réduire les coûts de maintenance associés aux opérations de grue. En identifiant les erreurs potentielles avant que lesdites erreurs ne se transforment en pannes importantes, les opérateurs peuvent mettre en œuvre des mesures de maintenance préventive et prescriptive pour résoudre les problèmes sous-jacents et prolonger la durée de vie des équipements [10]. Cette approche de maintenance proactive contribue à minimiser les temps d'arrêt, les coûts de réparation et les perturbations opérationnelles, conduisant finalement à des économies de coûts pour les opérateurs de terminaux [11].
5. Conformité aux normes réglementaires : la détection rapide des erreurs garantit la conformité aux normes réglementaires et aux réglementations de l'industrie régissant les opérations de grue. En identifiant et en corrigeant rapidement les erreurs, les opérateurs peuvent maintenir le respect des directives de sécurité et des protocoles opérationnels mandatés par les autorités réglementaires. Cette conformité garantit non seulement la sécurité du personnel et des actifs, mais permet également d'éviter d'éventuelles amendes et responsabilités juridiques associées à la non-conformité [12,13].

En résumé, l'identification et la prévision des erreurs en temps opportun sont essentielles pour optimiser les performances des grues, garantir la sécurité des opérations des grues à conteneurs et bénéficier aux parties prenantes de l'ensemble de la chaîne d'approvisionnement maritime.

1.2. Objectifs de recherche, motivation et contributions Un aspect

essentiel de la garantie de la sécurité, de l'efficacité et de la fiabilité des opérations des terminaux à conteneurs est la détection et l'atténuation des erreurs pouvant survenir lors des opérations de grue.

Machine Translated by Google

Les erreurs dans les opérations des grues à conteneurs à quai peuvent avoir des conséquences importantes, allant des dommages potentiels aux marchandises et aux infrastructures jusqu'aux perturbations de la logistique de la chaîne d'approvisionnement. De plus , les erreurs humaines, les dysfonctionnements des équipements et les aléas environnementaux contribuent à la complexité de la détection des erreurs et posent des défis pour garantir des opérations ininterrompues [14]. Les pannes du système peuvent être évitées en identifiant et en anticipant les erreurs, ce qui donne lieu à un système plus robuste et durable. La prédiction des erreurs dans les opérations des grues à conteneurs implique l'utilisation de méthodologies d'analyse de données, d'apprentissage automatique et d'apprentissage profond pour prédire les erreurs ou anomalies potentielles pouvant survenir lors du fonctionnement des grues à conteneurs [6].

Cette recherche vise à développer des stratégies proactives de détection et de prédiction des erreurs en exploitant la puissance de l'apprentissage profond, de l'analyse des données et de la modélisation prédictive, améliorant ainsi la sécurité, l'efficacité et la fiabilité des opérations des terminaux à conteneurs. Résoudre ce problème nécessite une compréhension globale de la dynamique opérationnelle des terminaux à conteneurs et l'application de technologies et de méthodologies avancées d'IA pour la détection, le diagnostic et la prédiction des erreurs [15].

En analysant diverses sources de données, notamment les données des capteurs de grue, les journaux d'opérations passées, les conditions météorologiques et d'autres aspects pertinents, des modèles prédictifs peuvent être créés pour prévoir et minimiser les problèmes à l'avance. La mise en œuvre de cette stratégie proactive améliore la sécurité, l'efficacité et l'efficience globale des opérations des terminaux à conteneurs.

Les résultats de la recherche dans ce domaine ont des implications significatives pour l'industrie maritime , les autorités portuaires, les opérateurs de terminaux et d'autres parties prenantes impliquées dans les opérations des terminaux à conteneurs. L'objectif est d'améliorer la sécurité opérationnelle, de minimiser les temps d'arrêt, d'optimiser l'utilisation des ressources et d'améliorer les performances globales du terminal en détectant et en atténuant efficacement les erreurs dans les données opérationnelles. La portée de cet article consiste à explorer comment l'apprentissage profond peut améliorer la détection des erreurs, la prévision et l'analyse des séries chronologiques dans les opérations de grues à conteneurs, ce qui constitue actuellement un besoin critique dans l'industrie maritime et contribuera à améliorer l'efficacité opérationnelle et la sécurité.

Le potentiel de cette recherche s'étend au-delà des grues à conteneurs et s'étend à d'autres secteurs confrontés à des problèmes de données déséquilibrées, où des événements ou des anomalies rares se produisent et peuvent en bénéficier de manière significative. D'autres secteurs industriels peuvent bénéficier de manière significative de cette recherche, comme le secteur manufacturier, en identifiant les pannes ou les défauts des équipements dès le début des lignes de production ; le secteur de l'énergie, pour prévoir les anomalies dans les systèmes de production ou de distribution d'électricité ; et le transport, pour détecter les irrégularités dans les performances des véhicules ou les systèmes de sécurité. De plus, les organisations peuvent cibler les interventions et les mesures préventives en identifiant les éléments critiques associés aux erreurs.

Par exemple:

Maintenance : donnez la priorité aux programmes de maintenance pour les composants les plus susceptibles de provoquer des erreurs ;

Optimisation des processus : ajuster les procédures opérationnelles en fonction des modèles d'erreur ; Allocation des ressources : pour minimiser les perturbations, allouez les ressources (personnel, matériel) de manière stratégique.

Cette recherche est motivée par l'urgence d'améliorer la reconnaissance et la prévision des erreurs dans les grues à conteneurs dans l'industrie du transport de conteneurs multimilliardaire, qui constitue l'épine dorsale de l'économie mondiale actuelle. Les techniques traditionnelles basées sur des règles échouent souvent car elles ne peuvent pas gérer des paramètres dynamiques complexes et une distribution inégale des données. L'apprentissage profond, notamment les réseaux LSTM, apporte une solution viable à ces lacunes. En capturant les connexions temporelles et en apprenant des données précédentes, les LSTM peuvent modéliser des modèles complexes et augmenter la précision.

1.3. Présentation du Deep Learning

L'apprentissage en profondeur est une branche de l'intelligence artificielle (IA) qui consiste à entraîner des systèmes informatiques à effectuer des tâches telles que reconnaître des modèles, prendre des décisions et résoudre des problèmes sans être explicitement programmés pour le faire. Il utilise des algorithmes appelés réseaux de neurones, inspirés de la structure et du fonctionnement du cerveau humain [16].

La surveillance de l'état des machines basée sur les données est de plus en plus répandue dans les environnements modernes.

Machine Translated by Google

systèmes de production car des capteurs à faible coût sont largement déployés et connectés à Internet. L'apprentissage profond offre des méthodes précieuses pour traiter et évaluer des données opérationnelles à grande échelle [17]. L'apprentissage profond permet aux ordinateurs de reconnaître des modèles, de classer des informations et de faire des prédictions très précises en apprenant à partir de grandes quantités de données.
L'apprentissage profond utilise des réseaux de neurones constitués de nombreuses couches, appelées architectures profondes, pour acquérir des connaissances à partir des entrées et générer des prédictions. Cette technologie a des applications dans divers domaines, notamment la reconnaissance d'images et de parole, le traitement du langage naturel, les véhicules autonomes et, dans ce cas-ci, l'industrie manufacturière.

Plusieurs facteurs clés rendent l'utilisation de l'apprentissage profond primordiale dans la détection, le diagnostic et la prévision des erreurs dans diverses industries, y compris les opérations de grues à conteneurs sur quai : 1. Les algorithmes d'apprentissage profond, tels que les réseaux neuronaux convolutifs (CNN) et les réseaux neuronaux récurrents (RNN).), ont démontré une capacité exceptionnelle à traiter des données complexes et de grande dimension [18]. Dans les opérations de grues à conteneurs, où les flux de données provenant de divers capteurs, systèmes de contrôle et journaux opérationnels sont abondants et diversifiés, les modèles d'apprentissage profond peuvent traiter efficacement ces données pour identifier des modèles subtils et des anomalies indiquant des erreurs.

2. Les modèles d'apprentissage profond excellent dans la capture des dépendances temporelles et des relations dynamiques au sein des données séquentielles [19]. Dans les opérations de grues à conteneurs, où les mouvements des grues, les processus de manutention des marchandises et les conditions environnementales évoluent, les algorithmes d'apprentissage profond peuvent analyser les données historiques pour prédire les états futurs, permettant ainsi une détection et une prévention proactives des erreurs [20].
3. L'évolutivité et l'adaptabilité des modèles d'apprentissage profond les rendent bien adaptés à la surveillance et à la prise de décision en temps réel et hors ligne dans des environnements opérationnels dynamiques. En apprenant continuellement de nouvelles données et en s'adaptant aux conditions changeantes, les systèmes de détection d'erreurs basés sur l'apprentissage profond peuvent améliorer la sécurité opérationnelle, optimiser l'utilisation des ressources et atténuer les risques potentiels liés aux opérations des grues à conteneurs [21,22].

Dans l'ensemble, l'importance de l'utilisation du deep learning pour la détection, le diagnostic et la prédiction des erreurs dans les opérations des grues à conteneurs réside dans sa capacité à exploiter la puissance de l'analyse avancée des données pour améliorer la sécurité opérationnelle, l'efficacité et la résilience de la logistique maritime. À mesure que la technologie progresse, l'exploitation des algorithmes d'apprentissage profond deviendra de plus en plus indispensable pour garantir le fonctionnement transparent et sécurisé des terminaux à conteneurs dans l' écosystème mondial de la chaîne d'approvisionnement.

En tirant parti de techniques avancées d'apprentissage en profondeur, la recherche vise à résoudre les complexités inhérentes à la détection et à la prédiction des erreurs dans les opérations des grues à conteneurs, où les données opérationnelles sont représentées sous forme de valeurs booléennes discrètes au fil du temps. L'algorithme proposé cherche à capturer efficacement les dépendances temporelles et les modèles subtils au sein des données de séries chronologiques booléennes, permettant une anticipation précise des erreurs potentielles avant leur apparition lors des opérations de grue. Grâce à une validation empirique et à une évaluation des performances, cet article s'efforce de démontrer l'efficacité et l'applicabilité de l' algorithme d'apprentissage en profondeur proposé pour améliorer la sécurité opérationnelle, l'efficience et la fiabilité au sein de la gestion des terminaux à conteneurs.

1.4. Flux de travail

Les sections suivantes de l'analyse du flux de travail examineront en profondeur l' état de l'art existant en explorant les innombrables facettes de la détection des erreurs dans les opérations des grues à conteneurs . En outre, les approches méthodologiques utilisées pour relever ce défi seront examinées parallèlement aux résultats empiriques tirés d'études pertinentes. Cette analyse complète élucidera une compréhension nuancée de la prédiction des erreurs dans le contexte des opérations de grues à conteneurs, mettant en lumière ses implications et les pistes potentielles d' avancement dans le discours universitaire et de recherche.

La section suivante fournira des informations sur la méthodologie et les données utilisées dans cette recherche . Il présente un nouvel algorithme personnalisé basé sur Python qui intègre SMOTE

Machine Translated by Google

et les techniques LSTM pour traiter les données déséquilibrées grâce à un prétraitement personnalisé, une analyse et l'incorporation de mesures d'évaluation. La structure du flux de travail est illustrée dans la figure 1 ci-dessous :

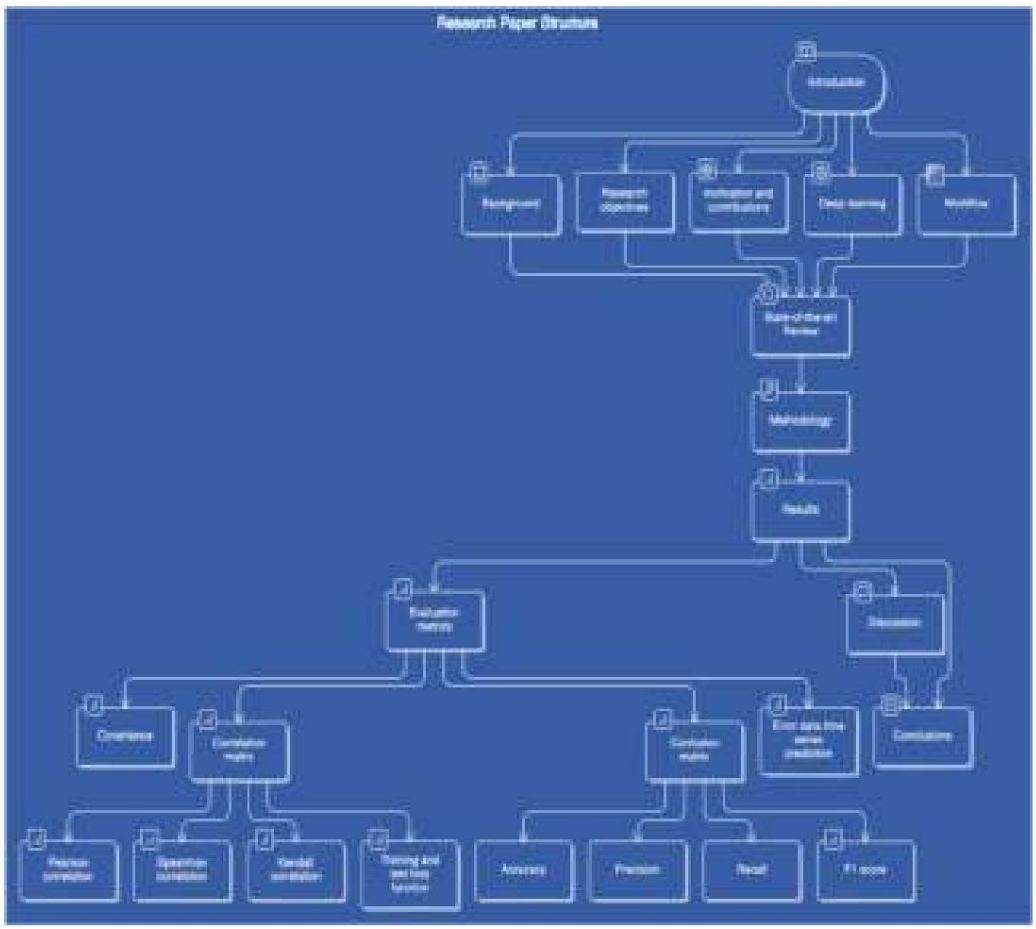


Figure 1. Flux de travail des documents de recherche.

2. Examen de l'état de l'art

Comme souligné précédemment, garantir la sécurité et la fiabilité des opérations des grues est de la plus haute importance. Les données de séries chronologiques provenant de telles opérations de machines lourdes posent des défis uniques en raison de leur grande taille et de leurs modèles complexes. L'identification rapide des erreurs opérationnelles peut réduire les accidents et les temps d'arrêt [23]. Les modèles d'apprentissage automatique et d'apprentissage profond sont devenus des outils puissants pour détecter les anomalies dans ces données. Il existe une littérature abondante sur la détection des défauts et des défaillances en ML [24]. Cependant, très peu de données sont disponibles sur la détection et la prédiction des erreurs. Il est crucial de détecter et de prédire une erreur car une erreur est un précurseur de pannes et de pannes. La figure 2 ci-dessous illustre cette corrélation entre erreur, défaut et échec.

La détection des signaux d'erreur de fonctionnement d'une machine fait référence au processus d'identification de signaux anormaux ou inattendus dans le fonctionnement d'une machine. Ces signaux peuvent indiquer des problèmes, des dysfonctionnements ou des écarts par rapport au comportement normal du système ou de la machine. Cette recherche a utilisé des bases de données réelles de grues STS embarquées et déportées provenant des ports et des grues, telles que la garantie, les alarmes et les données opérationnelles où les signaux d'erreur générés par le PLC de la grue étaient disponibles.

L'importance de l'utilisation de l'apprentissage profond pour la détection des erreurs dans les opérations des grues à conteneurs réside dans sa capacité à capturer efficacement des modèles complexes, des dépendances temporelles et des relations non linéaires inhérentes aux données opérationnelles. Les points essentiels soulignant son importance sont les suivants :

Machine Translated by Google



Figure 2. L'erreur comme précurseur des échecs [25].

1. Complexité des données opérationnelles : les opérations des grues à conteneurs génèrent de grandes quantités de données, notamment les lectures des capteurs, les mouvements des grues, la manutention des marchandises et les conditions environnementales. Les algorithmes d'apprentissage profond, tels que les RNN et les CNN, excellent dans le traitement des données de grande dimension et séquentielles [18], ce qui les rend bien adaptés à l'analyse de la dynamique complexe des opérations de grue.
2. Dépendances temporelles : les opérations de grue présentent des dépendances temporelles, dans lesquelles les actions et les événements à un instant donné influencent les états et les résultats ultérieurs. Les RNN, en particulier les réseaux LSTM, sont conçus pour capturer et apprendre des dépendances temporelles dans les données séquentielles, permettant une modélisation précise de l'évolution temporelle des opérations de grue et la détection de modèles ou d'écarts anormaux [26,27].
3. Apprentissage des fonctionnalités : les modèles d'apprentissage en profondeur peuvent apprendre automatiquement les fonctionnalités pertinentes à partir des données brutes, éliminant ainsi le besoin d'ingénierie manuelle des fonctionnalités. Dans le contexte de la détection des erreurs dans les opérations des grues à conteneurs, les algorithmes d'apprentissage profond peuvent extraire des caractéristiques informatives des données des capteurs, des trajectoires des grues et des journaux opérationnels, permettant une détection plus efficace des anomalies et des écarts indiquant des erreurs.
4. Adaptabilité à des environnements complexes : les opérations des grues à conteneurs sont soumises à divers facteurs environnementaux, tels que le vent, les conditions météorologiques et l'état de la mer, qui peuvent influencer les performances des grues et introduire des incertitudes [28,29]. Les modèles d'apprentissage profond, avec leur capacité à apprendre à partir de sources de données diverses et bruyantes, font preuve de robustesse et d'adaptabilité à des environnements d'exploitation complexes, améliorant ainsi la fiabilité et la précision des systèmes de détection d'erreurs [22].
5. Évolutivité et généralisation : les algorithmes d'apprentissage profond sont intrinsèquement évolutifs et peuvent gérer de grands volumes de données, ce qui les rend adaptés à l'analyse des nombreux ensembles de données générés par les opérations des terminaux à conteneurs. De plus, les modèles d'apprentissage profond peuvent bien généraliser les données invisibles, permettant une détection efficace des erreurs dans différentes configurations de terminaux, types de grues et conditions de fonctionnement [30].
6. Apprentissage et amélioration continus : les modèles d'apprentissage profond soutiennent l'apprentissage et l'amélioration continus grâce à une formation itérative sur de nouvelles données. À mesure que des données opérationnelles supplémentaires deviennent disponibles, les systèmes de détection d'erreurs basés sur l'apprentissage profond peuvent être mis à jour et affinés pour s'adapter à l'évolution des conditions opérationnelles, améliorant ainsi leur efficacité au fil du temps [16,31].

Plusieurs méthodes ont été proposées pour la détection des erreurs dans les opérations des grues à conteneurs STS. Les auteurs de [32] ont introduit une nouvelle méthode pour détecter l'usure des rails de grue, améliorant considérablement la précision, tandis que [33] ont identifié et évalué les erreurs humaines parmi les opérateurs de grues à tour, en se concentrant sur l'importance de surveiller le système anti-collision.

Les auteurs de [34] ont abordé la question de la détection des impacts physiques sur les coins des conteneurs maritimes lors des opérations de manutention en utilisant la méthodologie de détection des impacts. Il a utilisé la méthodologie de détection des impacts pour détecter les impacts récurrents des grues de quai sur les mêmes emplacements de conteneurs lors de l'accrochage. Ces succès contemporains suggèrent de faux

Machine Translated by Google

accrochage, qui pourrait provoquer de fortes déformations du métal sur de courtes durées. L'étude souligne la nécessité de détecter les dommages causés aux conteneurs maritimes pour la sécurité et l'efficacité des ports . Dans [35], les auteurs ont développé un algorithme de détection d'objets couleur à grande vitesse pour les systèmes d'assistance aux opérateurs de grues à quai, augmentant considérablement le taux de détection tout en maintenant une grande précision. Ces études soulignent collectivement l'importance de l'amélioration continue des méthodes de détection d'erreurs pour améliorer la sécurité et l'efficacité des opérations des grues à conteneurs STS. Plusieurs études ont démontré l'efficacité de l'apprentissage profond pour la prédiction des erreurs dans les opérations de grues à conteneurs. La détection rapide des erreurs est essentielle pour optimiser les performances de la grue et garantir la sécurité. Des erreurs ou des problèmes inaperçus peuvent compromettre la production et la sécurité en provoquant des inefficacités, des dommages aux équipements et d'éventuels accidents [36].

L'erreur humaine est un facteur important contribuant aux accidents lors des opérations de grue mobile. Un contrôle efficace de la grue et un positionnement précis dépendent de vitesses préétablies pour chaque direction, déterminées selon les spécifications du système de grue. La détection rapide des problèmes de vitesse, d'accélération ou de décélération peut améliorer les performances et éviter les surcharges ou les accidents. Le développement de systèmes d'assistance intelligents qui détectent les erreurs en temps réel peuvent améliorer la sécurité du fonctionnement des grues en réduisant la probabilité d'erreurs humaines [37].
Une expertise compétente est nécessaire pour le fonctionnement sûr des ponts roulants et pour minimiser l'apparition de dysfonctionnements de l'équipement. Néanmoins, le manque d'opérateurs qualifiés est devenu préoccupant ces derniers temps. L'identification précoce d'erreurs, telles que des problèmes de tension de câble ou des différences de poids de charge, peut contribuer à la prévention des accidents et des dommages matériels, améliorant ainsi la sécurité sans dépendre uniquement des compétences de l'opérateur [38].

Garantir une identification rapide des erreurs est essentiel pour mettre en œuvre avec succès des trajectoires de grue automatisées efficaces et fournir une interface conviviale qui aide les opérateurs à choisir et à exécuter les mouvements. L'identification opportune des erreurs dans la planification ou l'exécution de la trajectoire peut améliorer les performances et prévenir les accidents [36]. Le processus de planification manuelle actuellement utilisé pour la planification du trajet de levage des grues est ardu et sujet aux erreurs, principalement en raison des conditions dynamiques et encombrées des chantiers de construction. La mise en œuvre de systèmes automatisés capables de détecter les erreurs ou les collisions en temps réel peut améliorer l'efficacité et la précision opérationnelles [39].

Les recherches sur la prédiction des erreurs dans les opérations des grues à conteneurs STS sont diverses et couvrent divers aspects des opérations des grues. Les auteurs de [40] se sont concentrés sur l' analyse des dommages causés par la fatigue des structures en acier vieillissantes, un facteur critique dans la prédiction des erreurs. Les auteurs de [41] ont identifié des facteurs de risque spécifiques dans les opérations de manutention, qui peuvent éclairer les stratégies de prédiction des erreurs. Dans [42], les auteurs ont abordé la nécessité de systèmes de contrôle robustes dans les grues à conteneurs offshore, qui peuvent aider à atténuer les erreurs. Les auteurs de [43] ont souligné l' importance de la surveillance de l'état des structures dans la prédiction des erreurs, fournissant ainsi une méthode potentielle de détection précoce. Enfin, [44] ont identifié une diminution du temps moyen entre les pannes dues aux opérations lourdes, à l'augmentation des incidents de maintenance et aux heures de travail. Des indicateurs de performance clés ont été développés pour évaluer les performances des grues navire-terre et fournir des informations sur les améliorations futures du fonctionnement des équipements de grue à conteneurs. Les résultats suggèrent que les indicateurs de performance peuvent guider les futures inspections et améliorations des opérations des grues à conteneurs.

Des progrès prometteurs sont réalisés dans l'utilisation de l'apprentissage profond pour l'identification et la prédiction des défauts dans les données de séries chronologiques. Pour résoudre des problèmes tels que l'étiquetage bruyant, les méthodes d'ensemble pour l'identification des défauts [45] et des modèles innovants tels que le DCNN amélioré par GRU pour un meilleur diagnostic des défauts dans les processus chimiques [46], plusieurs stratégies ont été proposées. En outre, des modèles d'apprentissage profond pour l'identification des anomalies dans les canaux de télémétrie des engins spatiaux ont été étudiés, soulignant la nécessité de méthodes automatisées pour surveiller efficacement les anomalies du système [47]. Ces études démontrent comment les méthodes d'apprentissage en profondeur, telles que les unités récurrentes et les réseaux neuronaux convolutifs, peuvent améliorer l'efficacité, la généralisabilité et la précision de la détection des pannes dans divers domaines industriels.

L'apprentissage profond a montré des capacités remarquables dans l'analyse de modèles de données complexes et la réalisation de prédictions. Les réseaux de neurones profonds (DNN) excellent dans l'apprentissage des représentations

Machine Translated by Google

à partir de données à grande échelle, leur permettant de détecter des signaux subtils indiquant des erreurs imminentes dans les opérations de la grue. Des techniques telles que les RNN et les CNN ont été adaptées pour modéliser les dépendances séquentielles et spatiales des données des capteurs de grue, des journaux d'exploitation et des facteurs environnementaux. Les principales architectures d'apprentissage profond qui peuvent être utilisées pour les données de séries chronologiques booléennes provenant des grues à conteneurs STS sont les suivantes [18,48–50] :

- Réseaux neuronaux récurrents (RNN) : les RNN sont capables de gérer des données séquentielles, de les restituer. adapté à l'analyse de séries chronologiques. Les RNN peuvent capturer efficacement les dépendances temporelles dans les données de fonctionnement des grues [51].
- Mémoire à long terme (LSTM) : les réseaux LSTM, une variante des RNN, sont conçus pour résoudre le problème du gradient de disparition. Les LSTM sont particulièrement aptes à modéliser les dépendances à long terme et ont été utilisés avec succès pour prédire le comportement des grues.
- Gated Recurrent Unit (GRU) : les GRU sont une autre variante du RNN qui conserve la mémoire tout en étant efficace sur le plan informatique. Les GRU se sont révélés prometteurs dans la prévision des valeurs de force d'entrée pour le contrôle des grues.

Les RNN sont une technique d'apprentissage automatique robuste qui excelle dans le traitement des données séquentielles, telles que les séries chronologiques et le traitement du langage naturel. Les RNN contiennent une boucle de rétroaction qui leur permet d'apprendre des corrélations complexes dans les séquences d'entrée et d'exploiter ces structures de manière non linéaire [52]. Les RNN ont été largement utilisés dans divers domaines, tels que l'analyse des sentiments, la prévision de la demande d'énergie et la prévision des flux mensuels [53]. Un avantage essentiel des RNN est leur capacité à gérer des séquences d'entrée de longueurs variables et à produire des séquences de sortie de longueurs variables. Ces caractéristiques les rendent adaptés à des tâches telles que la modélisation du langage, la traduction automatique et la reconnaissance vocale [51]. Des réseaux de neurones profonds, capables d'apprendre des représentations encore plus complexes des données d'entrée, peuvent être créés en empilant des RNN (54).

Des architectures RNN plus sophistiquées, telles que les LSTM et les GRU, ont été créées pour résoudre le problème des dépendances à long terme, qui peuvent être problématiques pour les RNN classiques.
Ces conceptions capturent mieux les dépendances à long terme dans les données d'entrée en employant des techniques de contrôle pour conserver et oublier les informations de manière sélective.

Dans la littérature disponible, les méthodologies utilisées pour la détection, le diagnostic et la prédiction d'erreurs ou de défauts de séries chronologiques booléennes dans les industries manufacturières sur la base de l'apprentissage automatique peuvent être vues dans le tableau 1 ci-dessous, et l'apprentissage profond peut être vu dans le tableau 2 ci-dessous, comme suit :

Tableau 1. Méthodologies d'apprentissage automatique.

Méthodologie	Description Ces	Référence
Détection et diagnostic des défauts (FDD)	méthodologies détectent les irrégularités émergentes et prédisent leurs conséquences sur le comportement du système.	[55]
Diagnostic des défauts basé sur les données	Cette approche utilise des algorithmes d'apprentissage automatique pour la détection des défauts mécaniques ou le pronostic des défauts dans les équipements de fabrication.	[24]
Un tableau de ML, DL et Algorithmes DHL	Ces algorithmes peuvent potentiellement effectuer une détection précoce des défauts pouvant entraîner une panne future de la machine.	[56]
Approche d'apprentissage automatique automatisée	Cette approche est utilisée pour la détection et le diagnostic des défauts en temps réel dans les machines de fabrication discrète.	[57]

Les grues à conteneurs jouent un rôle crucial dans les terminaux à conteneurs et sont fondamentales pour l'économie mondiale car la plupart des marchandises mondiales y transitent. Cependant, leur niveau d'automatisation est inférieur à celui des grues de chantier et des véhicules à guidage automatique (AGV) [21]. Diverses méthodologies ont été explorées ici, mais la lacune réside dans l'affinement et l'expansion de l'application des techniques d'apprentissage en profondeur pour relever des défis spécifiques en matière de détection, de diagnostic et de prédiction des erreurs dans l'analyse des données des grues à conteneurs à quai. Cela améliorerait à terme la sécurité, l'efficacité et l'automatisation des opérations portuaires. La partie essentielle est le diagnostic précoce des pannes, qui accompagne l'étape initiale d'identification et de prédiction des erreurs. La méthodologie proposée vise justement à faire cela.

Machine Translated by Google

Tableau 2. Méthodologies d'apprentissage profond.

Méthodologie	Description Cette	Référence
Apprentissage profond pour les séries chronologiques Détection des anomalies	approche utilise des modèles d'apprentissage profond spécialisés pour détecter des modèles anormaux dans les données de séries chronologiques.	[58]
Modèles génératifs profonds	La reconstruction d'un auto-encodeur variationnel le long de la voie de projection (VAE-RaPP) et un réseau contradictoire génératif de clôture (Fence GAN) sont utilisés pour traiter les données de séries chronologiques déséquilibrées obtenues à partir d'applications de fabrication.	[59]
Modèle d'apprentissage métrique approfondi	Ce modèle prend en compte les données de défaut déséquilibrées et une manière de concevoir des paires de données quadruples.	[60]
Modèle ATT-1D CNN-GRU	Ce modèle combine un CNN unidimensionnel, un GRU, un mécanisme d'attention et un KG (graphique de connaissances) pour le diagnostic des pannes.	[61]
Adaptatif basé sur l'apprentissage profond Prédiction des défauts	Cette approche intègre un module complet d'extraction de caractéristiques convolutives, un module de déclenchement personnalisé et un module d'extraction progressive multicouche.	[62]

3. Méthodologie

L'ensemble de données utilisé ici pour la formation et l'évaluation du modèle d'apprentissage en profondeur comprenait les données réelles des grues à conteneurs dans les terminaux à conteneurs. Il était dans un format propriétaire et ne pouvait pas être lu directement. Il a été converti dans un format lisible à l'aide de la programmation Python. Il s'agissait des données opérationnelles du PLC, des capteurs et d'autres calculs. Les données d'erreur extraites étaient booléennes, ce qui signifie que la sortie était 0 ou 1.

Les systèmes logiciels sont utilisés pour diagnostiquer et conserver les données de l'automate pendant 30 jours, après quoi elles sont supprimées. Il s'agit d'un outil essentiel pour les opérations de grue et l'évaluation quotidienne des performances. Il fournit une source de données fiable sous forme de séries chronologiques, de données brutes provenant directement des capteurs et de leurs calculs. Il détecte et localise les erreurs périodiques au sein de la grue. Il crée environ 680 à 730 fichiers mensuels, stockant des données par heure. Il contient environ 23 000 variables API dans une grue moyenne, mais peut aller beaucoup plus loin dans quelques grues (par exemple, 31 521 variables API dans STS CC2124). Cela représente 518,4 milliards d'échantillons par mois (20 000 données à 100 ms). Les caractéristiques des données ont été sélectionnées (dans ce cas, les erreurs et certains autres signaux d'identification) et les données de ces variables ont été converties en fichiers CSV dans Excel à l'aide du code Python. Les données d'erreur contenaient environ 3 000 variables. Seules deux colonnes n'étaient pas booléennes, dont la dernière pour les horodatages/séries temporelles (de format incorrect). Il y avait environ 3 000 colonnes et plus de 16 100 000 (16 millions) de lignes.

L'analyse de l'ensemble de données opérationnelles a évalué huit grues de différents ports ; Enfin, l'analyse détaillée de trois grues a révélé que la plupart présentaient entre 2 852 et 2 878 signaux d'erreur. Cependant, seuls 3 à 6 signaux s'allumaient, indiquant qu'une véritable erreur s'était produite sur les grues. Cela indiquait la rareté des événements d'erreur et le manque de données pertinentes à analyser plus en profondeur, ce qui représentait un défi important pour le projet.

Un code personnalisé en Python a été développé pour extraire les données du prolétariat acquises grâce aux grues. Les fichiers Excel résultants ont ensuite été utilisés pour un traitement ultérieur des données.

L'algorithme personnalisé a été développé à l'aide des algorithmes SMOTE et LSTM dans la programmation Python, ainsi que de procédures de prétraitement personnalisées, traitant des données déséquilibrées et intégrant des métriques d'évaluation.

Prétraitement des données : un prétraitement personnalisé a été utilisé pour extraire les données booléennes. Initialement, les données non traitées, constituées de séries chronologiques et de relevés de capteurs, ont été chargées. Garantir la qualité des données était primordial afin que toutes les valeurs manquantes et les valeurs aberrantes soient corrigées. Par la suite, des caractéristiques booléennes, constituées d'indicateurs binaires ou de valeurs vraies/fausses, ont été extraites.

Remédier aux données déséquilibrées : dans le cas d'un ensemble de données déséquilibré, lorsqu'une classe est considérablement sous-représentée, SMOTE peut être utilisé. Cette méthode produit des échantillons artificiels pour la catégorie sous-représentée en créant de nouvelles instances grâce à l'interpolation de points de données existants. L'objectif est de parvenir à une répartition équilibrée des classes, garantissant une représentation équitable des deux classes. Ensuite, l'ensemble de données initial est fusionn

Machine Translated by Google

échantillons générés artificiellement. Par la suite, l'ensemble de données est divisé en sous-ensembles distincts à des fins de formation et de validation. Les données ont été divisées dans un rapport de 7 : 3 pour tester le modèle, 70 % étant des données d'entraînement et les 30 % restants étant des données de test. Ces données ont été remodelées dans un format tridimensionnel, les dimensions étant les échantillons, les pas de temps et les caractéristiques.

Construire le modèle LSTM : Le développement de la structure LSTM était crucial. Un réseau neuronal LSTM est construit à l'aide de couches d'entrée représentant des données séquentielles, de couches LSTM capturant les dépendances temporelles et d'une couche de sortie effectuant une classification binaire Le modèle a été compilé en sélectionnant une fonction de perte, telle que l'entropie croisée binaire, un optimiseur, tel qu'Adam, et une métrique d'évaluation, telle que la précision. Le modèle avait une taille de lot de 50 et a été formé pendant 40 époques, composé de 2 couches LSTM et d'une couche dense pour le résultat final. Une valeur de départ a été fournie au début, ce qui a permis de générer des résultats reproductibles pour le modèle afin que des résultats aléatoires différents ne soient pas générés à chaque exécution du modèle. Une fois le modèle entraîné, ses performances ont été évaluées sur l' ensemble de validation et tous les ajustements des hyperparamètres ont été apportés, si nécessaire.

Évaluation du modèle : le modèle a été évalué à l'aide d'une gamme de mesures, telles que l'exactitude, la précision, le rappel et le score F1. La matrice de confusion offre une répartition complète des prédictions, y compris les vrais positifs, les vrais négatifs, les faux positifs et les faux négatifs. Les métriques d'évaluation ont analysé le comportement du modèle et apporté les modifications nécessaires.

L'organigramme de l'algorithme personnalisé est présenté ci-dessous dans la figure 3.

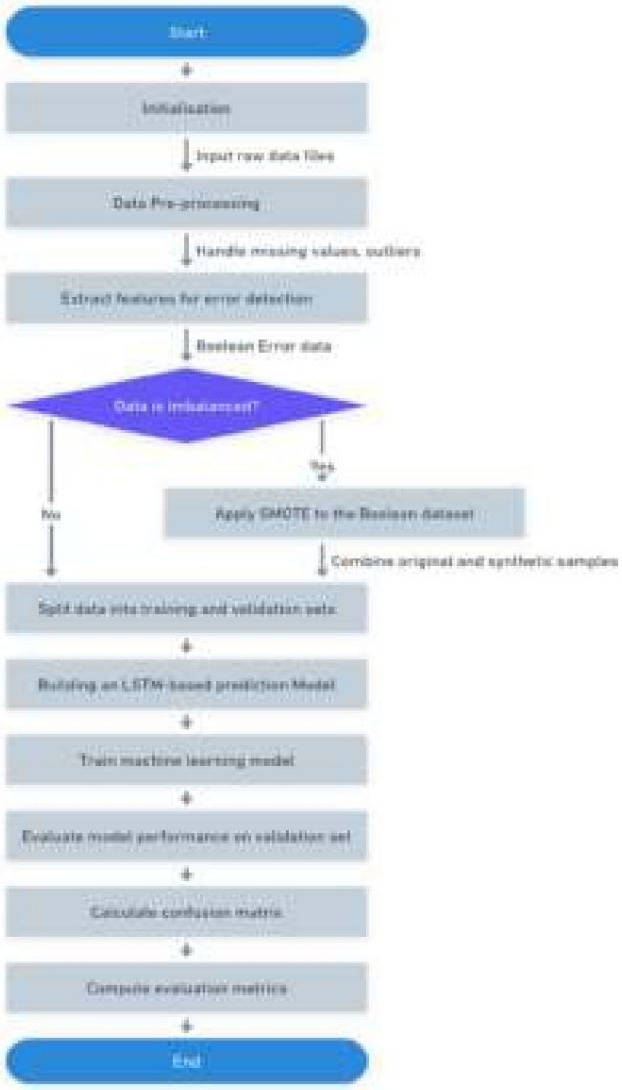


Figure 3. Organigramme d'algorithme personnalisé basé sur SMOTE et LSTM.

Machine Translated by Google

4. Résultats

Cette section représente les résultats expérimentaux de l'application de l'apprentissage profond proposé.
ing algorithme à la tâche de détection et de prédiction des erreurs.

4.1. Paramètres d'évaluation

Les métriques d'évaluation sont des mesures quantitatives utilisées pour évaluer les performances d'un modèle ou d'un algorithme, notamment en matière d'apprentissage automatique, de statistiques et d'analyse de données. Ces métriques aident à déterminer les performances d'un modèle et à guider les améliorations et les optimisations. Les variables affichées sous la forme var_xxxx dans les métriques suivantes représentent des signaux d'erreur spécifiques générés par les grues à conteneurs.

4.1.1. Covariance

La covariance est une relation constamment additive entre les deux échantillons de données. Cette relation peut être résumée entre deux variables, appelée covariance. Il est calculé comme la moyenne du produit entre les valeurs de chaque échantillon, où les valeurs ont été centrées (leur moyenne étant soustraite) [63]. La covariance n'implique pas de causalité : elle mesure uniquement la relation statistique. Comprendre la covariance peut aider à identifier les modèles et les dépendances lors de l'analyse des données. Une matrice de covariance représente les relations entre les paires de variables dans un ensemble de données. Il quantifie la mesure dans laquelle deux variables varient ensemble.

$$\mathrm{Cov}(X,Y) = \mathrm{somme}(x - \mathrm{moyenne}(X)) \times (y - \mathrm{moyenne}(Y))) \times 1/(n-1) \quad (1)$$

Dans cette heatmap, chaque cellule correspond à la covariance entre deux variables. Les cellules diagonales (en haut à gauche en bas à droite) représentent la covariance d'une variable avec elle-même. Ces valeurs sont toujours positives. Les cellules hors diagonale (en haut à droite et en bas à gauche) représentent la covariance entre différentes variables. Ces valeurs peuvent être positives ou négatives. La carte thermique utilise l'intensité des couleurs pour représenter les valeurs de covariance. Le rouge foncé indique une forte covariance positive (proche de 1), ce qui signifie que les variables ont tendance à augmenter ou diminuer ensemble. Le bleu foncé indique une covariance négative (proche de −1), suggérant une relation inverse. Les nuances plus claires représentent une corrélation plus faible (plus proche de 0).

La carte thermique ci-dessus dans la figure 4 montre que les étiquettes des axes X et Y correspondent à des variables spécifiques (par exemple, « var_2663 » et « var_831 »). var_2663 représente .tINTF_General_Visu.tLaneProtection.tCover3.tMotor4.tMotor.tPT100.tAlarm.bStatus et var_831 —.tINTF_General_Visu.tLaneProtection.tCover1.tMotor4.tCB.tSignalWithError.tError.bHasError .

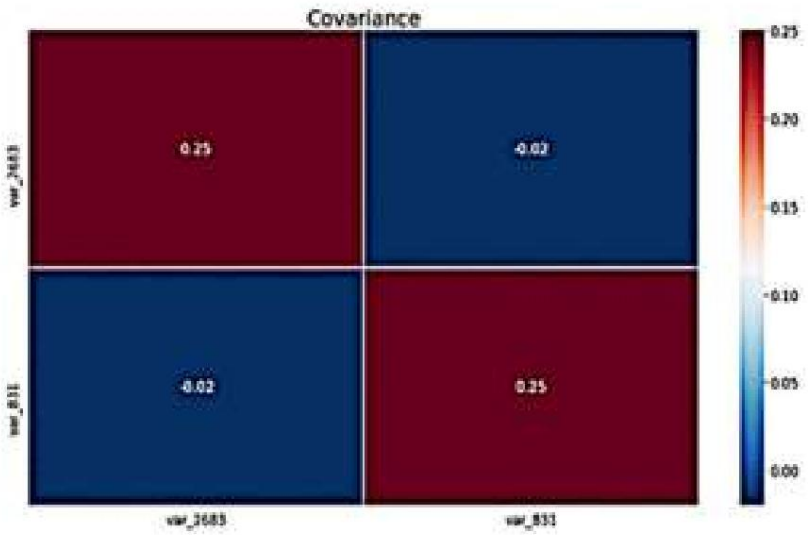


Figure 4. Covariance.

Machine Translated by Google

Le quadrant supérieur gauche a une teinte rouge foncé, indiquant une covariance positive robuste (environ 0,25) entre var_2663 et lui-même. Le quadrant inférieur droit montre également une solide covariance positive pour var_831. Les quadrants bleu clair hors diagonale suggèrent une faible relation inverse (environ -0,02) entre var_2663 et var_831, indiquant que le moteur 4 a une covariance positive avec l'erreur du capteur de température dans le bloc de commande du moteur quatre.

4.1.2. Matrice de corrélation

La matrice de corrélation est une technique multivariée fréquemment utilisée pour déterminer des variables fortement connectées. Cet aspect du projet s'est concentré sur la corrélation de toutes les variables des signaux d'erreur et de certains signaux opérationnels considérés comme des variables dépendantes, telles que CraneOn, CraneOff et le verrouillage par rotation. Les variables d'erreur avec une seule valeur unique, 0, ont été supprimées car les variables homogènes ne contiennent qu'une seule valeur unique et n'apportent pas d'informations supplémentaires au modèle. La corrélation entre les variables a été analysée et visualisée pour donner plus d'informations sur les données. La corrélation a été utilisée pour déterminer la relation entre les variables. Il peut être positif, indiquant que les deux variables évoluent dans la même direction, ou négatif, indiquant que lorsque la valeur d'une variable augmente, les valeurs des autres variables diminuent. La corrélation peut également être neutre ou nulle, indiquant que les variables n'ont aucune relation ou ne sont pas liées.

Corrélation de Pearson

Le coefficient de corrélation de Pearson est calculé en divisant la covariance de deux variables par le produit de leurs écarts types. C'est la normalisation de la covariance entre les deux variables continues pour donner un score interprétable. Il quantifie le degré auquel un changement dans une variable est associé à un changement dans une autre variable [63].

$$\text{Corrélation de Pearson (r)} = \text{covariance (X, Y)} / (\text{stdv (X)} \times \text{stdv (Y)}) \quad (2)$$

Cette carte thermique présentée dans la figure 5 ci-dessus compare les deux variables var_2683, c'est-à-dire .tINTF_General_Visu.tLaneProtection.tCover3.tMotor4.tMotor.tPT100_Group.tPT100_ 2.tError.bHasError et « var_831 », c'est-à-dire .tINTF_General_Visu.tLaneProtection.tCover1. tMotor4.tCB.tSignalWithError.tError.bHasError. La carte thermique affiche le coefficient de corrélation entre ces variables. Les quadrants diagonaux (où une variable est comparée à elle-même) affichent un coefficient de corrélation 1 (comme prévu). Les quadrants hors diagonale révèlent une faible corrélation négative d'environ −0,079. L'échelle de couleurs va de −1 à 1. Le rouge foncé représente des corrélations négatives plus fortes, tandis que le bleu foncé indique des corrélations positives plus fortes.

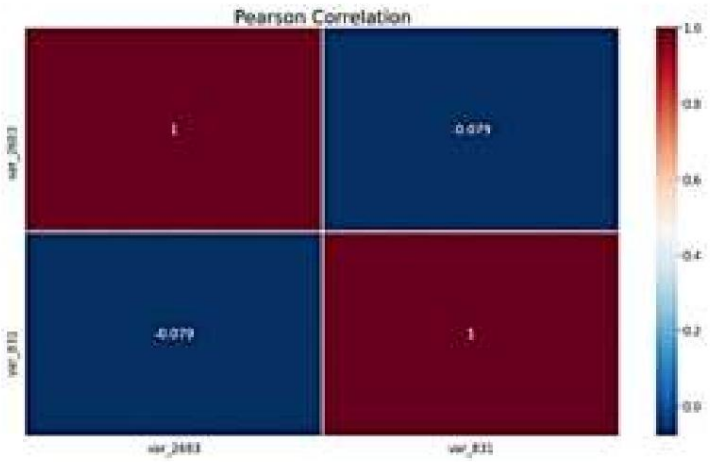


Figure 5. Corrélation de Pearson.

Machine Translated by Google

Corrélation de Spearman

La corrélation de Spearman a été calculée à partir du rang relatif des valeurs de chaque échantillon. Cette approche est couramment utilisée dans les statistiques non paramétriques, telles que les méthodes statistiques où aucune distribution spécifique des données, comme la distribution gaussienne, n'est supposée [64]. Ce coefficient mesure la force et la direction de la relation monotone entre deux variables classées. Il est utile pour identifier et évaluer les relations non linéaires et les données ordinales.

SpearmanCorrelation = covariance(rang(X),rang(Y))/std(rang(X)) ×std(rang(Y)) (3)

Cette carte thermique présentée ci-dessus dans la figure 6 est divisée en quatre quadrants pour montrer le coefficient de corrélation de Spearman entre deux variables étiquetées var_2683 —.tINTF_General_Visu.tLaneProtection.tCover3.tMotor4.tMotor.tPT100_Group.tPT100_2.tError.bHasError sur l' axe y. et var_831—.tINTF_General_Visu.tLaneProtection.tCover1.tMotor4.tCB. tSignalWithError.tError.bHasError sur l' axe des x. Les deux quadrants diagonaux montrent une corrélation positive parfaite (une valeur de 1) pour chaque variable avec elle-même (indiquée par la couleur rouge et le chiffre ff 1). Les quadrants hors diagonale révèlent une très faible corrélation négative de −0,079 entre les deux variables (indiquée en bleu foncé).

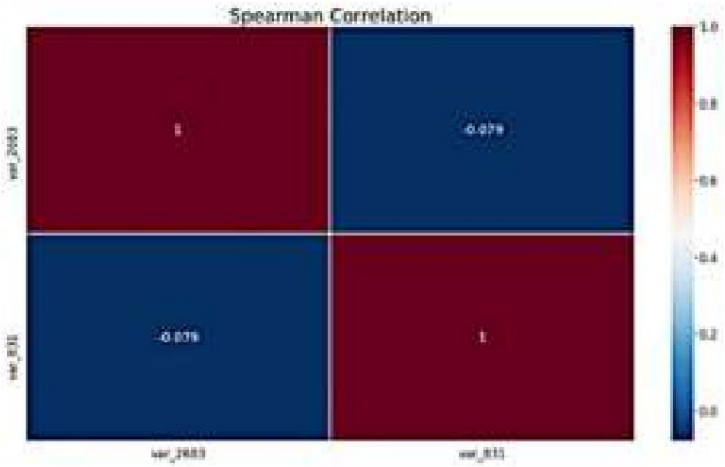


Figure 6. Corrélation de Spearman.

Corrélation de Kendall

Le coefficient de corrélation des rangs de Kendall, Tau mesure le degré de similarité entre deux ensembles rangs attribués au même ensemble d'objets. Contrairement au coefficient de Spearman, le Kendall Tau mesure uniquement l'accord directionnel, et non les différences de rang.

KendallCorrelation = (Nombre de paires concordantes − Nombre de paires discordantes)/n(n − 1)/2 (4)

La carte thermique de corrélation de Kendall présentée dans la figure 7 réaffirme les résultats des autres corrélations. Il se concentre sur les deux mêmes variables que les autres cartes thermiques de corrélation, à savoir var_2683 —.tINTF_General_Visu.tLaneProtection.tCover3.tMotor4.tMotor.tPT100_Group.tPT100_2.tError.bHasError sur l' axe y et var_831—.tINTF_General_ Visu.tLaneProtection. tCover
Les quadrants supérieur gauche et inférieur droit sont rouge foncé, indiquant une corrélation parfaite de 1. Ceci est attendu puisque chaque variable est en corrélation avec elle-même. Les quadrants supérieur droit et inférieur gauche sont bleus, ce qui suggère une faible corrélation négative (environ −0,079) entre var_2683 et var_831.

Machine Translated by Google

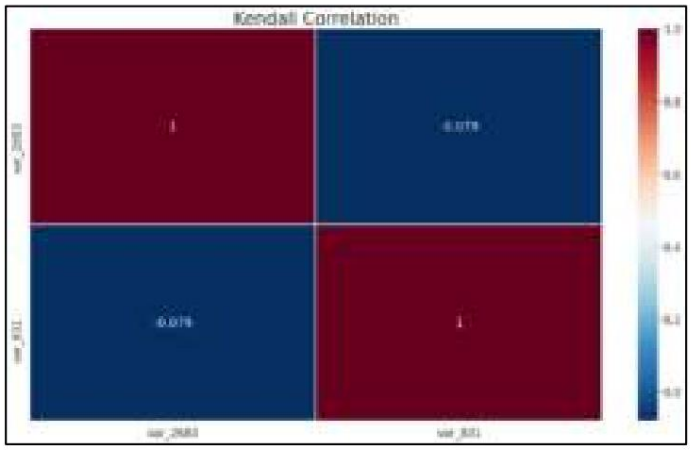


Figure 7. Corrélation de Kendall.

4.1.3. Fonction de perte de formation et de test

Dans l'apprentissage automatique, la perte de formation et la perte de test (ou perte de validation) sont des mesures critiques utilisées pour évaluer les performances d'un modèle pendant le processus de formation. Ces fonctions de perte aident à mesurer les performances du modèle sur les données d'entraînement et invisibles (données de test ou de validation). La perte de formation est l'erreur ou l'écart entre les sorties prévues et réelles sur l'ensemble de données de formation. Il fournit une mesure de l'adéquation du modèle aux données d'entraînement. La perte de test (ou perte de validation) est l'erreur entre les sorties prévues et réelles de l'ensemble de données de test ou de validation. Cet ensemble de données n'est pas utilisé pendant la formation mais est conservé de côté pour évaluer les performances du modèle sur des données invisibles.

La valeur de l'erreur quadratique moyenne (RMSE) est calculée. La racine carrée de la moyenne l'erreur quadratique fournit une mesure dans les mêmes unités que la variable cible.

Le graphique de la figure 8 montre deux lignes : train (en bleu) et test (en orange). L' axe horizontal représente le nombre d'époques, allant de 0 à 40, tandis que l' axe vertical représente la perte, allant de 0,0 à 0,7. Les deux lignes montrent initialement une forte diminution des pertes au cours des premières époques, ce qui indique un apprentissage efficace à partir des données de formation. Au fur et à mesure que les époques avancent, les deux lignes s'aplatissent progressivement, suggérant une diminution des rendements des investissements. l'apprentissage et le risque de surapprentissage si la perte de test augmente. La ligne de train représente la perte des données d'entraînement et la ligne de test représente la perte de données de test invisibles.
La surveillance des deux lignes permet d'évaluer les performances du modèle et sa capacité de généralisation, ainsi que de diagnostiquer des problèmes tels que le surajustement, le sous-ajustement et la convergence. Ce graphique est pertinent sur le plan académique pour évaluer les modèles ML et est utilisé pour optimiser les stratégies de formation et améliorer les performances du modèle.

4.1.4. Matrice de confusion

Une matrice de confusion est un outil fondamental pour évaluer les performances des modèles de classification. Il fournit une vue complète de la manière dont un modèle prédit différentes classes. La matrice de confusion est basée sur un problème de classification binaire avec deux classes : 0 et 1. Ces classes peuvent représenter différents résultats, tels que réussite/échec, oui/non ou vrai/faux. La matrice est organisée en quatre quadrants : • Vrais négatifs (TN) : instances correctement prédites

comme classe 0. Il s'agit de prédictions souhaitables : identifiant correctement les instances comme négatives (classe 0) ;

• Faux positifs (FP) : instances prédites à tort comme étant de classe 1 alors qu'elles sont en réalité de classe 0. Ce sont des erreurs : des instances classées à tort comme positives (classe 1) alors qu'elles ne le sont pas ;

• Faux négatifs (FN) : instances prédites à tort comme étant de classe 0 alors qu'elles sont en réalité de classe 1. Ce sont également des erreurs : des instances classées à tort comme négatives alors qu'elles sont positives ;

Machine Translated by Google

• Vrais positifs (TP) : instances correctement prédites comme classe 1. Il s'agit de prédictions souhaitables, identifiant correctement les instances comme positives.

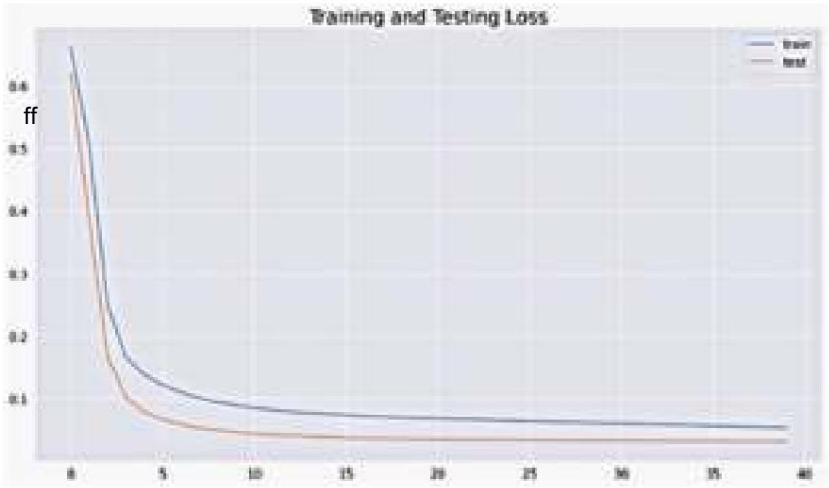


Figure 8. La fonction de perte de formation et de test.

La matrice de confusion présentée à la figure 9 résume les performances du modèle de classification. Il donne un aperçu de la mesure dans laquelle les prédictions du modèle correspondent aux résultats réels.

1. Vrai négatif (TN) : le modèle a correctement prédit les instances négatives (le résultat réel était négatif). Il y a 997 vrais négatifs ; 2. Faux positif (FP) : le modèle a prédit de manière incorrecte les instances positives (le résultat négatif réel). Ici, il y a neuf faux positifs ; 3. Faux Négatifs (FN) : Le modèle a prédit de manière incorrecte les instances négatives (le nombre réel résultat positif). La matrice montre quatre faux négatifs ;
4. Vrai positif (TP) : le modèle a correctement prédit les instances positives (le résultat réel était positif). Dans cette matrice, il y a quatre vrais positifs.

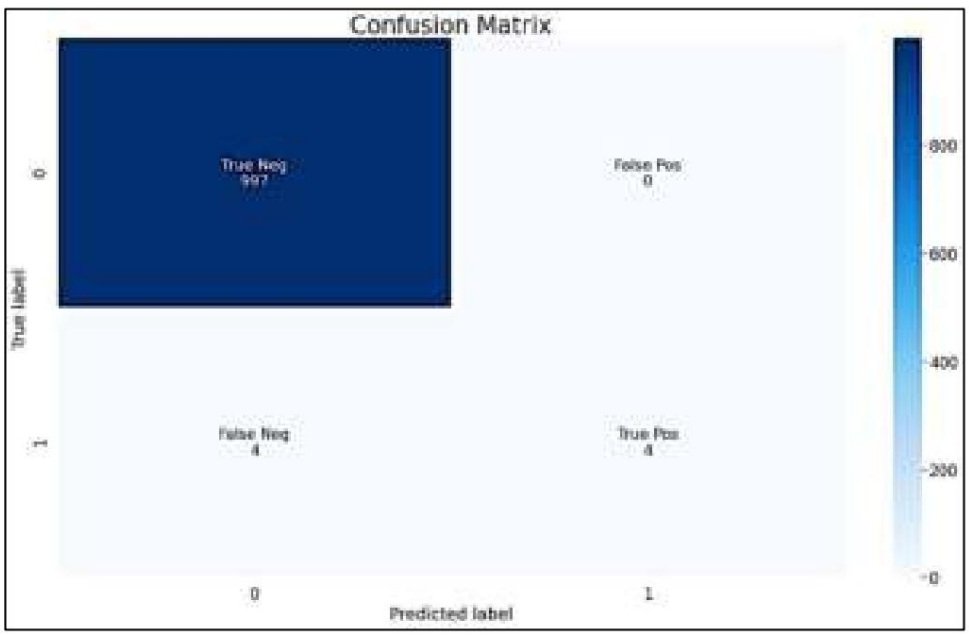


Figure 9. La matrice de confusion.

Machine Translated by Google

4.1.5. Exactitude, précision, rappel, score F1

L'exactitude, la précision, le rappel et le score F1 présentés dans la figure 10 ci-dessous ont également été calculé et imprimé pour les métriques d'évaluation des données prévues. Ces performances modèles ont ensuite été enregistrés dans un fichier texte.

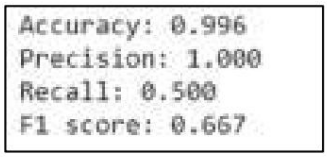


Figure 10. Résultats des mesures d'évaluation.

Précision : le nombre de prédictions correctes est divisé par le nombre total de prédictions.

$$\text{Précision} = (\text{vrais positifs} + \text{vrais négatifs})/\text{nombre total d'instances} \quad (5)$$

Précision : cela mesure l'exactitude des prédictions de classe positives en déterminant combien d'entre eux appartiennent réellement à la classe positive.

$$\text{Précision} = \text{vrais positifs}/(\text{vrais positifs} + \text{faux positifs}) \quad (6)$$

Rappel (sensibilité) : cela calcule la proportion de prédictions positives correctes de tous les exemples positifs de l'ensemble de données.

$$\text{Rappel} = \text{Vrais Positifs}/(\text{Vrais Positifs} + \text{Faux Négatifs}) \quad (7)$$

Score F1 : ceci fournit un score unique qui équilibre la précision et le rappel en un seul chiffre. C'est un moyen harmonique de précision et de rappel.

$$\text{Score F1} = 2 \times((\text{Précision} \times \text{Rappel})/(\text{Précision} + \text{Rappel})) \quad (8)$$

Précision : Atteindre une précision de 0,996 indique une performance globale solide.

Précision : un score de précision de 1,00 suggère que le modèle a rarement été mal classé comme vrai. erreurs comme non-erreurs.

Rappel : le score de rappel de 0,500 indique que le modèle n'a identifié que la moitié des données réelles. erreurs. Ce compromis entre précision et rappel est courant dans les ensembles de données déséquilibrés.

Score F1 : Le score F1 (0,667) équilibre précision et rappel, offrant une évaluation complète vue de l'efficacité du modèle.

4.2. Prédiction des séries chronologiques de données d'erreur

Le résultat final est la prédiction des n secondes de données suivantes où l'utilisateur fournit les n secondes comme entrée au début, et les données correspondantes sont enregistrées dans un fichier CSV. Pour cela, les prochaines 1 s de données sont prédites pour les 60 prochaines s à l'aide de données historiques disponibles pour analyse. La sortie du fichier CSV prédit pour l'un des Les variables d'erreur sont présentées ci-dessous dans le tableau 3 :

Tableau 3. Données d'erreur prévues.

Horodatage	var_831
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0

Machine Translated by Google

Tableau 3. Suite

Horodatage	var_831
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	1
11/12/2020 20:35	1
11/12/2020 20:35	1
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	0
11/12/2020 20:35	1
11/12/2020 20:35	1
11/12/2020 20:35	1
11/12/2020 20:36	1
11/12/2020 20:36	1
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	1
11/12/2020 20:36	1
11/12/2020 20:36	1
11/12/2020 20:36	1
11/12/2020 20:36	1
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0

Machine Translated by Google

Tableau 3. Suite

Horodatage	var_831
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0
11/12/2020 20:36	0

5. Discussion

La détection et la prévision rapides des erreurs sont essentielles pour optimiser les performances des grues. Ils sont essentiels pour garantir la sécurité en prévenant les accidents, en protégeant les équipements contre dommages et réduire les erreurs humaines. En conséquence, ces pratiques améliorent la productivité et des dépenses d'exploitation moindres.

Comprendre les limites de l'analyse prédictive est essentiel. Avant tout, le progrès est impossible sans un ensemble de données de formation complet et de haute qualité. Deuxièmement, avoir un Une description claire du concept à prévoir et des événements passés est essentielle. L'un des Le défi de l'analyse prédictive consiste à identifier une aiguille dans une botte de foin et à prédire son évolution. occurrence. Cependant, les méthodes traditionnelles présentent certaines limites. Ces limites dans les opérations de grue à conteneurs sont les suivantes :

1. Erreurs d'origine humaine : les opérateurs humains jouent toujours un rôle important dans les grues de quai opérations malgré les progrès de l'automatisation. Leurs décisions et actions peuvent conduire à erreurs, affectant la productivité et la sécurité [64].
2. Dynamique complexe : les grues à conteneurs fonctionnent dans un environnement dynamique avec des poids des conteneurs, conditions de vent et mouvements des navires. Prédire les erreurs dans de tels un système complexe est un défi [65-67].
3. Efficacité et temps d'arrêt : il est essentiel de minimiser les temps d'arrêt dus aux erreurs. Erreur opportune la prédiction peut éviter des retards coûteux et optimiser l'utilisation des grues [68,69].
4. Données déséquilibrées : un ensemble de données est dit déséquilibré lorsqu'il existe des disparités substantielles dans le nombre d'exemples de chaque classe, ce qui fait qu'une ou plusieurs classes sont sous-représentés dans l'ensemble de données [70]. Ce phénomène a été constaté dans plusieurs ensembles de données du monde réel rassemblés en milieu industriel.

Par conséquent, cette recherche a développé un nouvel algorithme personnalisé pour répondre à ces problèmes. limites et défis en se concentrant sur la détection et la prévision des erreurs avant qu'elles ne surviennent. se transformer en fautes et en échecs.

Les points suivants abordent les défis et les limites rencontrés au cours la mise en œuvre :

1. Comme il y avait une différence de temps significative entre les fichiers du tableau de données, les données les fichiers de tableau n'ont pas pu être fusionnés pour analyse ; au lieu de cela, une analyse a été effectuée sur chaque fichier de tableau de données séparément.

Machine Translated by Google

2. Pour la prévision des séries chronologiques, les données ont été réduites de millisecondes à secondes, car les données en millisecondes n'étaient pas continues, c'est-à-dire (1, 2, 3, . . ., 999).
3. Il peut y avoir de bons risques d'erreur car la taille de l'ensemble de données était petite, de sorte que le modèle LSTM ne pouvait pas prédire correctement et correctement les nouvelles données. En ajoutant plus de données, ce problème peut être résolu.

Malgré les résultats prometteurs, la mise en œuvre de modèles d'apprentissage profond pour la prédiction des erreurs dans les opérations réelles des grues reste un défi. Ceux-ci incluent les problèmes de qualité des données, l' interprétabilité de modèles complexes et l'intégration avec les systèmes existants. Les futures orientations de recherche pourraient relever ces défis grâce à des architectures de modèles avancées, des techniques de fusion de données et des efforts de collaboration entre le monde universitaire et les parties prenantes de l'industrie. Les pistes de recherche future peuvent progresser grâce à l'amélioration de la qualité et de la quantité des données, qui sont actuellement à la traîne.

6. Conclusions

La recherche sur les grues à conteneurs englobe divers aspects liés à leur conception, leur fonctionnement et leur impact sur la logistique portuaire. Cependant, il y a une absence flagrante d'études axées sur l'identification précoce des erreurs pour prévenir les pannes et les défaillances. La prédiction rapide des erreurs grâce à des algorithmes d'apprentissage profond a le potentiel de révolutionner les performances des grues et de garantir une efficacité optimale dans les opérations des grues à conteneurs.

Un algorithme d'apprentissage profond personnalisé basé sur LSTM a été développé, atteignant une précision de 99,6 % dans la prédiction des signaux d'erreur de grue à l'aide de LSTM et SMOTE pour la prédiction de classification, avec une précision de 1,00, un rappel de 0,500 et un score f1 de 0,667. Cependant, des données supplémentaires sont nécessaires pour améliorer les résultats du modèle.

La disponibilité actuelle des données est limitée à un mois avec environ 30 000 variables, dont moins de 3 000 variables d'erreur. Les données utilisables étaient un petit échantillon, avec seulement 3 à 6 variables d'erreur qui étaient occasionnellement activées. L'étude fournit également d'autres résultats, tels que des matrices de corrélation, de covariance et de confusion. Les expériences ont démontré la capacité de la technique à identifier des éléments cruciaux, ce qui a conduit à des efforts visant à améliorer la structure du modèle afin d'améliorer la prédiction des erreurs dans les mégadonnées industrielles.

Même si le modèle de sortie repose sur des données préexistantes, sa capacité à gérer des données déséquilibrées, un problème courant dans la plupart des secteurs, présente un avantage significatif. Avec des données supplémentaires, le modèle existant peut être amélioré. Des données continues sur plusieurs mois fourniront un ensemble d'échantillons de données d'erreur plus complet, contribuant ainsi à améliorer les résultats. Une fois qu'un ensemble de données plus important sera obtenu à l'avenir, l'exploration d'autres algorithmes tels que la forêt aléatoire, le Bayes naïf et le boost XG pourra être utilisée.

La poursuite de la recherche et de l'innovation dans ce domaine est essentielle pour tirer pleinement parti des avantages de l'analyse prédictive dans les opérations portuaires.

Contributions des auteurs : Conceptualisation, AA ; conservation des données, AA ; analyse formelle, AA; méthodologie , AA ; logiciels, AA; supervision, LK et JW ; validation, AA; visualisation, AA; rédaction – ébauche originale, AA ; rédaction – révision et édition, AA, LK et JW Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Ce projet a été soutenu, en partie, par la subvention 13/RC/2094_P2 de la Science Foundation Ireland et cofinancé dans le cadre du Fonds européen de développement régional par le biais du programme opérationnel régional du Sud et de l'Est de Lero, le Centre de recherche sur les logiciels de la Science Foundation Ireland (www.lero.ie, consulté le 20 juillet 2024) au cours des deux premières années de ce projet.

Déclaration de disponibilité des données : les ensembles de données présentés dans cet article ne sont pas facilement disponibles en raison de problèmes de confidentialité ou de sécurité des données du port à conteneurs. Les demandes d'accès aux ensembles de données doivent être adressées à l'auteur correspondant de cet article.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Machine Translated by Google

Références

1. Renseignements du Mordor. Marché du transport maritime par conteneurs – Analyse de la taille, de la part et de la croissance. Rapports de l'industrie du renseignement du Mordor. [En ligne]. Disponible en ligne : https://www.mordorintelligence.com/industry-reports/global-container-shipping-market (consulté le 17 juillet 2024).
2. Lindeberg, E. ; Améliorer les performances opérationnelles des grues STS. Västerås, Suède, février 2011. Disponible en ligne : www.porttechnology.org (consulté le 18 juin 2021).
3. Taille du marché des grues à conteneurs Ship-to-Shore (STS), part, prévisions de l'industrie 2028. Disponible en ligne : https://www. fortunebusinessinsights.com/ship-to-shore-sts-container-cranes-market-102880 (consulté le 15 juin 2021).
4. Zrniü, N. ; Hoffmann, K. Développement de la conception de grues à conteneurs navire-terre : 1959–2004. Au Colloque international sur Histoire des machines et des mécanismes ; Springer : Dordrecht, Pays-Bas, 2007. [CrossRef]
5. Khoukhi, A. ; Khalid, MH Techniques informatiques hybrides pour la détection et l'isolation des défauts, une revue. Calculer. Electr. Ing. 2015, 43, 17-32. [Référence croisée]
6. Saltzer, JH; Frans Kaashoek, M. 2.2 : Défauts, pannes et conception tolérante aux pannes 2.2.1 : Défauts, pannes et modules. Dans Principes de conception de systèmes informatiques ; Massachusetts Institute of Technology LibreTextes : Davis, CA, USA, 2024 ; Chapitre 2.2, pages 1 à 461. Disponible en ligne : https://eng.libretexts.org/@go/page/58498 (consulté le 19 juillet 2024).
7. Safaï, MM ; Héjazian, S. ; Pedrammehr, S. ; Pakzad, M. ; Ettefagh, M. ; Fotouhi, M. Détection des dommages d'une grue à portique avec une masse en mouvement à l'aide d'un réseau neuronal artificiel. Bâtiments 2024, 14, 458. [CrossRef]
8. Wang, Z. ; Cheng, J. ; Hu, H. Une approche proactive-réactive pour l'attribution continue des postes d'amarrage et l'affectation des grues de quai Problèmes avec l'incertitude hybride. J. Mar. Sci. Ing. 2024, 12, 182. [Réf. croisée]
9. Gattuso, D. ; Pratico, FG; Longo, R. ; Cassone, G. ; Vigna, M. ; Sceni, R. Opérations ferroviaires dans les terminaux de fret : enjeux de sécurité et méthodologie proposée. Lors de la 5e conférence internationale de l'IEEE sur les modèles et technologies pour les systèmes de transport intelligents, MT-ITS 2017 — Actes, Naples, Italie, 26-28 juin 2017 ; Institute of Electrical and Electronics Engineers Inc. : Piscataway, NJ, États-Unis, 2017 ; pp. 164-169. [Référence croisée]
10. Huang, M. ; Liu, Z. ; Tao, Y. Diagnostic et prédiction des défauts mécaniques dans l'IoT basés sur la fusion de données de détection multi-sources. Simul. Modèle. Entraînez-vous. Théorie 2020, 102, 101981. [CrossRef]
11. Meddaoui, A. ; Hain, M. ; Hachmoud, A. Les avantages de la maintenance prédictive pour l'excellence manufacturière : une étude de cas pour établir des méthodes fiables de prévision des pannes. Int. J.Adv. Fab. Technologie. 2023, 128, 3685-3690. [Référence croisée]
12. Fernandes, M. ; Canito, A. ; Bolón-Canedo, V. ; Conceição, L. ; Praça, I. ; Marreiros, G. Analyse des données et sélection de fonctionnalités pour la maintenance prédictive : une étude de cas dans l'industrie métallurgique. Int. J. Inf. Gérer. 2019, 46, 252-262. [Référence croisée]
13. Ouais, JKW ; Wong, JH ; Peng, L. Intégration de modèles d'informations de grue dans le BIM pour vérifier la conformité des exigences du plan de levage. Dans Actes du 33e Symposium international sur l'automatisation et la robotique dans la construction (ISARC 2016), Auburn, AL, États-Unis, 18-21 juillet 2016 ; Disponible en ligne : https://www.iaarc.org/publications/fulltext/ISARC2016-Paper192.pdf (consulté le 20 juillet 2024).
14. Chu, F. ; Gaïlus, S. ; Liu, L. ; Ni, L. L'avenir de l'automatisation portuaire | McKinsey. 2018. Disponible en ligne : https://www.mckinsey.com/industries/voyage-logistique-et-infrastructure/nos-perspectives/l'avenir-des-ports-automatisés (consulté le 18 mai 2024).
15. Lepenioti, K. ; Pertselakis, M. ; Bousdekis, A. ; Louca, A. ; Lampathaki, F. ; Apostolou, D. ; Mentzas, G. ; Anastasiou, S. Apprentissage automatique pour l'analyse prédictive et prescriptive des données opérationnelles dans la fabrication intelligente. Dans les ateliers avancés d'ingénierie des systèmes d'information ; Springer : Cham, Suisse, 2020. [CrossRef]
16. Qu'est-ce que le Deep Learning ?|Un guide du débutant. Disponible en ligne : https://www.scribbr.com/ai-tools/deep-learning/ (consulté sur 19 mai 2024).
17. Zhao, R. ; Yan, R. ; Chen, Z. ; Mao, K. ; Wang, P. ; Gao, RX Deep learning et ses applications à la surveillance de l'état des machines. Mécanique. Système. Processus de signal. 2019, 115, 213-237. [Référence croisée]
18. Kumar, A. ; Gaur, N. ; Chakravarty, S. ; Alsharif, MH; Uthansakul, P. ; Uthansakul, M. Analyse de la détection spectrale à l'aide de profondeurs algorithmes d'apprentissage : CNN et RNN. Aïn Shams ing. J. 2024, 15, 102505. [CrossRef]
19. Ubal, C. ; Di-Giorgi, G. ; Contreras-Reyes, JE; Salas, R. Prédire les dépendances à long terme dans les séries chronologiques à l'aide de données récurrentes Réseaux de neurones artificiels. Mach. Apprendre. Connaître. Extr. 2023, 5, 1340-1358. [Référence croisée]
20. Heilig, L. ; Stahlbock, R. ; Voß, S. De la numérisation à la prise de décision basée sur les données dans les terminaux à conteneurs. arXiv 2019, arXiv : 1904.13251.
21. Shin, HS ; Lee, SP ; Ha, Y.-S. ; Kim, H.-S. Conception d'un modèle d'apprentissage du contrôle des grues à conteneurs à l'aide du deep learning. J.Adv. Mar. Ing. Technologie. 2023, 47, 2023-2234. [Référence croisée]
22. Chalapathy, R. ; Chawla, S. Deep Learning pour la détection des anomalies : une enquête. arXiv 2019, arXiv : 1901.03407.
23. Shagluf, A. ; Bâton long, AP ; Fletcher, S. (PDF) Stratégies de maintenance pour réduire les temps d'arrêt dus aux erreurs de positionnement des machines, mesure et gestion des performances de maintenance (MPMM). Disponible en ligne : https://www.researchgate.net/publication/267864966_Maintenance_Strategies_to_Reduce_Downtime_Due_to_Machine_Positional_Errors (consulté le 18 juillet 2024).
24. Fernandes, M. ; Corchado, JM; Marreiros, G. Techniques d'apprentissage automatique appliquées au diagnostic et au pronostic des défauts mécaniques dans le contexte de cas d'utilisation réels de fabrication industrielle : une revue systématique de la littérature. Appl. Intell. 2022, 52, 14246-14280. [Référence croisée] [Pub Med]

Machine Translated by Google

25. Tests de logiciels : bug, défaut, erreur, défaut ou échec – GeeksforGeeks. Disponible en ligne : https://www.geeksforgeeks. org/software-testing-bug-vs-defect-vs-error-vs-fault-vs-failure/ (consulté le 22 avril 2024).
26. Jalayer, M. ; Orsenigo, C. ; Vercellis, C. Détection et diagnostic des défauts des machines tournantes : un modèle basé sur la convolution LSTM, transformations de Fourier rapide et ondelettes continues. Calculer. Ind.2021 , 125, 103378. [CrossRef]
27. Zhao, R. ; Yan, R. ; Wang, J. ; Mao, K. Apprendre à surveiller la santé des machines avec des réseaux LSTM bidirectionnels convolutifs. Capteurs 2017, 17, 273. [CrossRef] [Pub Med]
28. Zrnic, N. ; Petkovic, Z. ; Bošnjak, SM Automatisation des grues à conteneurs navire-terre : une revue de l'état de l'art. Transactions FME 2005, 33, 111-121. Disponible en ligne : https://www.researchgate.net/publication/228936324_Automation_of_ship-to-shore_conteneur_cranes_A_review_of_state-of-the-art (consulté le 24 mars 2024).
29. Tran, QH ; Hein, J. ; Nguyen, VB; Kang, C. ; Ahn, JH; Park, IJ Analyse de sensibilité pour la conception de grues à conteneurs navire-terre. Appl. Sci. 2018, 8, 1667. [Réf. croisée]
30. Li, Z. Approches basées sur l'apprentissage profond pour la maintenance prédictive Un cadre de diagnostic et de pronostic intelligents des pannes à l'ère de l'industrie 4.0. doctorat Thèse, Université norvégienne des sciences et technologies, Trondheim, Norvège, 2018.
31. Goodfellow, I. ; Bengio, Y. ; Courville, A. 2015. Disponible en ligne : http://www.deeplearningbook.org/ (consulté le 15 juin 2024).
32. Liu, W. ; Xiao, J. Une nouvelle méthode de détection de l'usure des rails de grue et son analyse des erreurs. J. Mech. Force 2014, 36, 878-883. Disponible en ligne : https://www.researchgate.net/publication/293095272_A_new_method_of_detection_the_wear_of_crane_ rails_and_its_error_analysis (consulté le 19 mars 2024).
33. Mohammedfam, I. ; Borgheipour, H. ; Téhérani, G. ; Madadi, S. Identification et évaluation des erreurs humaines chez les opérateurs de grues à tour à l'aide des techniques SHERPA et CREAM. J. Santé Saf. Travail. 2020, 10, 5-8.
34. Jakovlev, S. ; Églynas, T. ; Jusis, M. ; Voznak, M. ; Partila, P. ; Tovarek, J. Détection des impacts physiques sur les coins des conteneurs maritimes lors des opérations de manutention effectuées par les grues de quai. J. Mar. Sci. Ing. 2023, 11, 794. [Réf. croisée]
35. Gao, X. ; Ouais, HG ; Marayong, P. Un algorithme de détection d'objets couleur à grande vitesse pour le système d'assistance aux opérateurs de grues à quai. Dans Actes de la 11e conférence annuelle de l'IEEE International Systems, SysCon 2017 — Actes, Montréal, QC, Canada, 24-27 avril 2017. [CrossRef]
36. Westerberg, S. ; Shiriaev, A. Téléopération de machines forestières basée sur un environnement virtuel : conception de futures méthodes d'interaction. J. Hum.-Robot. Interagir. 2013, 2, 84-110. [Référence croisée]
37. Ren, W. ; Wu, Z. ; Zhang, L. Planification en temps réel d'un système de levage dans des contrôleurs montés sur une grue mobile. Peut. J.Civ. Ing. 2016, 43, 542-552. [Référence croisée]
38. Watanabe, M. ; Momoi, Y. ; Odai, M. ; Ieshige, K. Estimation simultanée de la tension du câble et de la charge des grues à l'aide d'un moteur signaux. Trans. JSME 2023, 89. [CrossRef]
39. Lei, Z. ; Behzadipour, S. ; Al-Hussein, M. ; Hermann, U. Application de l'évitement robotique d'obstacles dans la planification du trajet de levage des grues. Dans Actes du 28e ISARC, Séoul, République de Corée, 29 juin-2 juillet 2011.
40. Lehner, P. ; Kˇrivý, V. ; Krejsa, M. ; Paˇrenica, P. ; Kozák, J. Prédiction stochastique de la durée de vie d'une structure en acier existante chargée par Ponts roulants. Procédure Struct. Intégré. 2018, 13, 1539-1544. [Référence croisée]
41. Shang, KC ; Tseng, WJ Une analyse des risques liés aux opérations de manutention dans les terminaux à conteneurs des ports maritimes. J. Mar. Sci. Technologie. 2010, 18, 201-210. [Référence croisée]
42. Ismaïl, RMTR ; Ça, ND ; Ha, QP Modélisation et suivi de trajectoire robuste pour les systèmes de grues à conteneurs offshore. Automatique. Constr. 2015, 59, 179-187. [Référence croisée]
43. Liu, J. ; Qin, X. ; Soleil, Y. ; Zhang, Q. Modèle de prévision de la réponse pour les structures des grues à conteneurs à quai basé sur la surveillance Données. J. Effectuer. Constr. Facil. 2021, 35. [Réf. croisée]
44. Jo, JH ; Kim, S. Développement d'indicateurs de performance clés pour l'évaluation des performances des grues navire-terre dans les terminaux à conteneurs Opérations. J. Mar. Sci. Ing. 2020, 8, 6. [Réf. croisée]
45. Cheng, C. ; Liu, X. ; Zhou, B. ; Yuan, Y. Diagnostic intelligent des défauts avec des étiquettes bruyantes via un apprentissage semi-supervisé sur l'industrie Série chronologique. IEEETrans. Ind. Informer. 2023, 19, 7724-7732. [Référence croisée]
46. Lee, XY; Kumar, A. ; Vidyaratne, L. ; Rao, AR ; Farahat, A. ; Gupta, C. Un ensemble de méthodes basées sur la convolution pour la détection de défauts à l'aide de signaux de vibration. Dans Actes de la Conférence internationale de l'IEEE 2023 sur les pronostics et la gestion de la santé (ICPHM 2023), Montréal, QC, Canada, 5-7 juin 2023 ; pp. 172-179. [Référence croisée]
47. Zhang, J. ; Zhang, M. ; Feng, Z. ; Ruifang, LV ; Lu, C. ; Dai, Y. ; Dong, L. Réseau neuronal à convolution profonde amélioré par des unités récurrentes pour le diagnostic des défauts des processus industriels en temps réel. Processus Saf. Environ. Prot. 2023, 175, 129-149. [Référence croisée]
48. Comment coder les réseaux de neurones RNN et LSTM en Python. Disponible en ligne : https://www.nbshare.io/notebook/249468051/Comment coder les réseaux neuronaux RNN et LSTM en Python/ (consulté le 28 avril 2022).
49. Quiñones-Grueiro, M. ; Llanes-Santiago, O. ; Silva Neto, AJ Diagnostic des pannes dans les systèmes industriels. Goujon. Système. Décis. Contrôle 2021, 309, 1-14. [Référence croisée]
50. Guide illustré des LSTM et GRU : une explication étape par étape | par Michael Phi | vers la science des données. Disponible en ligne : https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-sa-step-by-step-explanation-44e9eb85bf21 (consulté le 13 avril 2021).
51. Maheswaranathan, N. ; Williams, AH ; Golub, MD; Ganguli, S. ; Sussillo, D. L'ingénierie inverse des réseaux récurrents pour la classification des sentiments révèle la dynamique des attracteurs de lignes. Dans Actes de la 33e Conférence sur les systèmes de traitement de l'information neuronale (NeurIPS 2019), Vancouver, Colombie-Britannique, Canada, 8-14 décembre 2019 ; pages 15696 à 15705.

Machine Translated by Google

52. Alizadeh, Z. ; Yazdi, J. ; Kim, JH ; Al-Shamiri, AK Évaluation des techniques d'apprentissage automatique pour la prévision des flux mensuels. Eau 2018, 10, 1676. [CrossRef]

53. Ahmad, J. ; Tahir, A. ; Larijani, H. ; Ahmed, F. ; Aziz Shah, S. ; Salle, AJ ; Buchanan, WJ Prévision de la demande énergétique des bâtiments à l'aide de réseaux neuronaux aléatoires. J. Intell. Système flou. 2020, 38, 4753-4765. [Référence croisée]

54. Li, S. ; Li, W. ; Membre, S. ; Cuisinier, C. ; Gao, Y. Réseau neuronal récurrent indépendant profond (IndRNN). arXiv 2019, arXiv:1910.06251v3.

55. Abid, A. ; Khan, MT ; Iqbal, J. Une revue des techniques de détection et de diagnostic des pannes : les bases et au-delà. Artif. Intell. Rév. 2021, 54, 3639-3664. [Référence croisée]

56. Shahin, M. ; Chen, FF; Hosseinzadeh, A. ; Zand, N. Utilisation d'algorithmes d'apprentissage automatique et d'apprentissage profond pour minimiser les temps d'arrêt dans les systèmes de fabrication : un service de diagnostic de détection précoce des pannes. Int. J.Adv. Fab. Technologie. 2023, 128, 3857-3883. [Référence croisée]

57. Leite, D. ; Martins, A. ; Rativa, D. ; De Oliveira, JFL; Maciel, AMA Une approche d'apprentissage automatique automatisée en temps réel Détection et diagnostic des pannes. Capteurs 2022, 22, 6138. [CrossRef]

58. Darban, ZZ; Webb, GI ; Poêle, S. ; Aggarwal, CC; Salehi, M. Deep Learning pour la détection d'anomalies de séries chronologiques : une enquête. arXiv 2022, arXiv :2211.05244v2.

59. Chanson, J. ; Lee, YC ; Lee, J. Modèle génératif profond avec codage d'images de séries chronologiques pour la détection des défauts de fabrication dans le processus de moulage sous pression. J. Intell. Fab. 2023, 34, 3001-3014. [Référence croisée]

60. Gui, X. ; Zhang, J. Modèle d'apprentissage métrique profond pour le diagnostic des défauts déséquilibrés. arXiv 2021, arXiv :2107.03786.

61. Liu, H. ; Mar.; Li, D. ; Yan, L. ; Ma, Z. Diagnostic des pannes de machines basé sur l'apprentissage profond pour l'analyse des séries chronologiques et les graphiques de connaissances. J. Processus de signal. Système. 2021, 93, 1433-1455. [Référence croisée]

62. Pang, JL Prédiction adaptative des pannes et maintenance dans les lignes de production grâce à l'apprentissage profond. Int. J. Simul. Modèle. 2023, 22, 734-745. [Référence croisée]

63. Brownlee, J. ; Comment calculer la corrélation entre les variables en Python. Maîtrise de l'apprentissage automatique. Disponible en ligne : https : //machinelearningmastery.com/how-to-use-correlation-to-understand-the-relationship-between-variables/ (consulté le 14 avril 2022).

64. Li, A. Priorisation des risques d'erreur humaine dans les opérations de grue sur la base du CPT et de l'ICWGT. PLoS ONE 2024, 19, e0297120. [Référence croisée] [Pub Med]

65. Eglynas, T. ; Jusis, M. ; Jakovlev, S. ; Sénulis, A. ; Andziulis, A. ; Gudas, S. Analyse de l'efficacité des conteneurs maritimes méthodes et procédures de contrôle de la manutention/chargement. Av. Mécanique. Ing. 2019, 11. [Réf. croisée]

66. Maboudi, M. ; Alamouri, A. ; De Arriba López, V. ; Bajauri, MS ; Berger, C. ; Gerke, M. Inspection des grues à conteneurs par drone : concept, défis et résultats préliminaires. ISPRS Ann. Photogramme. Télédétection. Inf. spatiale. Sci. 2021, V-1-2021, 121-128. [Référence croisée]

67. Shan, J. ; Zhang, M. ; Zhang, D. ; Zhou, X. ; Li, H. ; Wang, B. Grues à conteneurs. Dans Manuel des machines portuaires ; Springer : Singapour, 2024 ; pp. 451-678. [Référence croisée]

68. Viellechner, A. ; Spinler, S. L'analyse des nouvelles données rencontre l'expédition conventionnelle de conteneurs : prévoir les retards en comparant divers algorithmes d'apprentissage automatique. Disponible en ligne : https://hdl.handle.net/10125/63897 (consulté le 12 juin 2024).

69. Tian, X. ; Yan, R. ; Liu, Y. ; Wang, S. Une méthode intelligente de prévision puis d'optimisation pour un transport maritime ciblé et rentable. Transp. Rés. Méthode de la partie B. 2023, 172, 32-52. [Référence croisée]

70. Gestion des données déséquilibrées pour la classification — GeeksforGeeks. Disponible en ligne : https://www.geeksforgeeks.org/handling-données-déséquilibrées-pour-classification/ (consulté le 12 juin 2024).

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Une optimisation efficace de la recherche arborescente de Monte Carlo Algorithme pour les Amazones

Lijun Zhang [1,2], Han Zou 1,2 et Yungang Zhu 1,2,*

1 Collège d'informatique et de technologie, Université de Jilin, Changchun 130012, Chine ; zhanglj2121@mails.jlu.edu.cn (LZ); zouhan1221@mails.jlu.edu.cn (HZ)
2 Laboratoire clé de calcul symbolique et d'ingénierie des connaissances du ministère de l'Éducation, Université de Jilin, Changchun 130012, Chine
* Correspondance : zhuyungang@jlu.edu.cn

Résumé : Amazons est un jeu de société informatisé comportant des positions complexes très difficiles pour les humains. Dans cet article, nous proposons une optimisation efficace de l'algorithme de recherche arborescente de Monte Carlo (MCTS) pour les Amazones, fusionnant la stratégie « Déplacer les groupes » et la stratégie d'optimisation « Évaluation parallèle » (MG-PEO). Plus précisément, nous expliquons la grande efficacité de la stratégie Move Groups en définissant un nouveau critère : la distance de convergence gagnante. Nous soulignons également le problème potentiel de la stratégie de tomber dans un optimum local et proposons que le mécanisme d'évaluation parallèle puisse compenser cette lacune. De plus, nous avons mené des analyses de performances et des expériences rigoureuses. Les résultats de l'analyse des performances indiquent que l'algorithme MCTS avec la stratégie Move Groups peut améliorer la capacité de jeu du jeu Amazons de 20 à 30 fois par rapport à l'algorithme MCTS traditionnel. L'optimisation de l'évaluation parallèle améliore encore la capacité de jeu du jeu Amazons de 2 à 3 fois. Les résultats expérimentaux montrent que l' algorithme MCTS avec la stratégie MG-PEO atteint en moyenne un taux de victoire de 23 % plus élevé que l'algorithme MCTS traditionnel. De plus, le programme MG-PEO Amazons proposé dans cet article a remporté le premier prix du concours Amazons lors du championnat chinois des jeux informatiques collégiaux et du tournoi national des jeux informatiques 2023.

Mots-clés : Amazones ; optimisation d'algorithmes ; Déplacer des groupes ; Évaluation parallèle

check for updates

Référence : Zhang, L. ; Zou, H. ; Zhu, Y. Une optimisation efficace de l'algorithme de recherche d'arbres de Monte Carlo pour les Amazones. Algorithmes 2024, 17, 334. https://doi.org/10.3390/a17080334

Rédacteurs académiques : Wenxin Li et Haifeng Zhang

Reçu : 22 avril 2024
Révisé : 16 juillet 2024
Accepté : 27 juillet 2024
Publié : 1 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Amazons est un jeu de société à deux joueurs introduit par l'Argentin Walter Zamkauskas en 1988 [1,2]. Il s'agit du jeu désigné pour le concours de programmation de jeux informatiques Olympia. En raison de son gameplay complexe, Amazons est très difficile pour les humains [3], ce qui en fait un sujet populaire pour les compétitions et la recherche sur les jeux vidéo. Les règles du jeu sont les suivantes : Le jeu se joue sur un plateau carré de 10 ×10. Chaque joueur contrôle quatre pièces identiques appelées « Amazones ». Ces pièces et flèches suivent les mêmes règles qu'une reine aux échecs : elles peuvent se déplacer de n'importe quel nombre de cases dans n'importe quelle direction (horizontale, verticale ou diagonale) mais ne peuvent pas sauter par-dessus d'autres pièces ou flèches. L'objectif est de restreindre les mouvements de l'adversaire en tirant des flèches jusqu'à ce qu'aucune des quatre pièces adverses ne puisse bouger. Si, au tour d'un joueur, ses quatre pièces sont immobilisées, ce joueur perd et l'autre joueur gagne. Notamment, il n'y a aucun lien chez les Amazones.

Le principal défi du jeu des Amazones réside dans sa complexité : chaque coup offre un grand nombre de choix possibles, généralement plus de 500 [4,5], et le jeu peut s'étendre sur plus de 80 coups [6,7]. Ce haut niveau de complexité fait d'Amazon un excellent échantillon expérimental pour tester les algorithmes de stratégie de recherche [8,9]. Depuis son invention, il a suscité une attention considérable de la part des informaticiens [10] et des étudiants en informatique [11].

La méthode de Monte Carlo [12] est une technique informatique qui produit des résultats numériques par échantillonnage aléatoire répété. Lorsqu'il est intégré à des algorithmes de recherche arborescente

Machine Translated by Google

et des fonctions d'évaluation appropriées, l'algorithme de recherche arborescente de Monte Carlo (MCTS) a donné des résultats impressionnants dans les jeux de société avec un facteur de branchement élevé (comme Go) [13-15], et dans les jeux de stratégie en temps réel (comme StarCraft [13-15]). 16]). Par exemple, en 2016, le programme AlphaGo de DeepMind [17] a battu le champion du monde coréen de Go, Lee Sedol, avec un score de 4 : 1. En 2018, le programme OpenAI Five [18] d'OpenAI a battu l'équipe championne du monde de Dota 2 avec un avantage écrasant ; les deux réalisations sont basées sur l' algorithme de recherche arborescente de Monte Carlo.

La stratégie d'optimisation « Move Groups-Parallel Evaluation » (MG-PEO) s'adapte aux règles du jeu de société Amazons en subdivisant un mouvement complet en deux parties : Amazon Movement et Arrow Shot. Amazon Movement et Arrow Shot sont traités comme des nœuds d'arbre, plutôt que de considérer un mouvement complet comme un seul nœud d'arbre.
Cette approche est connue sous le nom de stratégie Move Groups dans la littérature [19]. Bien que cette stratégie ait été appliquée au jeu des Amazones [20], il lui manque encore une analyse objective de son efficacité. De plus, il existe une croyance largement répandue selon laquelle cette méthode ne présente que des avantages sans aucun inconvénient. Dans cet article, nous expliquons la grande efficacité de la stratégie Move Groups en Amazonie en définissant un nouveau critère : la distance de convergence gagnante . Nous identifions également certaines lacunes de cette stratégie. Pour résoudre ces problèmes, nous proposons d'utiliser un mécanisme d'évaluation parallèle pour compenser les limites de la stratégie Move Groups. Nos résultats expérimentaux démontrent que MG-PEO est une stratégie efficace d'optimisation de la fusion.

Les principales contributions de cet article sont les suivantes :

- Nous proposons une optimisation efficace de l'algorithme de recherche arborescente de Monte Carlo pour le jeu Amazons : l'optimisation de l'algorithme MG-PEO, qui combine la stratégie Move Groups avec la stratégie Parallel Evaluation.
- Nous fournissons une analyse des performances de la stratégie Move Groups sur Amazon. En définissant un nouveau critère, la distance de convergence gagnante, nous avons expliqué la grande efficacité de la stratégie. Par ailleurs, nous soulignons également que cette stratégie présente certaines lacunes. La stratégie Move Groups consiste à affiner les règles du jeu Amazons et à utiliser des simulations pour attribuer des scores de valeur correspondants à des nœuds de couche supplémentaires. En utilisant ces scores de valeur (par exemple, la stratégie Upper Confidence Bound Apply to Tree (UCT) [15]), nous pouvons réaliser un élagage efficace pour gagner du temps. Cependant, l'ajout de nœuds de couche supplémentaires signifie que davantage de nœuds doivent être parcourus et évalués dans le pire des cas, où l'élagage n'a pas eu lieu, ce qui pourrait entraîner une augmentation du temps système. Cette surcharge est générée lorsque l' algorithme s'efforce de trouver la solution optimale locale.
- Nous proposons d'utiliser un mécanisme d'évaluation parallèle pour remédier aux lacunes potentielles de la stratégie Move Groups. En parallélisant la fonction d'évaluation pour le jeu Amazons, nous visons à accélérer le processus d'expansion de la recherche arborescente de Monte Carlo et à minimiser la probabilité que la stratégie Move Groups tombe dans un optimal local.

De plus, nous avons effectué une analyse rigoureuse des performances et des expériences de jeu. L'analyse des performances indique que, par rapport à l'algorithme MCTS traditionnel, l'algorithme MCTS avec la stratégie Move Groups peut améliorer la capacité de jeu des Amazones de 20 à 30 fois, tandis que le composant d'optimisation de l'évaluation parallèle peut améliorer la capacité de jeu des Amazones de 2 à 30 fois. 3 fois. Les résultats expérimentaux démontrent que, par rapport à l'algorithme MCTS traditionnel, l'algorithme MCTS avec la stratégie MG-PEO permet d'obtenir une augmentation moyenne de 23 % du taux de victoire et présente une capacité de jeu plus forte.

Le reste de cet article est structuré comme suit : Dans la section 2, nous présentons des informations préliminaires sur le jeu Amazons et discutons de sa pertinence pour l' algorithme MCTS. Dans la section 3, nous fournissons une description détaillée de l'algorithme MCTS avec la stratégie MG-PEO. Dans la section 4, nous proposons respectivement des analyses des performances de la stratégie Move Groups et de la stratégie Parallel Evaluation. La section 5 présente une série d'expériences

Machine Translated by Google

menée pour vérifier l'efficacité de l'algorithme MCTS avec la stratégie MG-PEO. Enfin, les conclusions et les pistes de recherches futures sont résumées dans la section 6.

2. Travaux connexes

Dans cette section, nous fournissons un aperçu des recherches pertinentes qui constituent la base de notre travail. Plus précisément, nous discuterons de trois modules clés de la littérature qui sont pertinents pour les Amazones : la recherche historique sur les algorithmes des Amazones, l'application de l' algorithme MCTS aux Amazones et les connaissances sur l'évaluation des situations de jeu.

2.1. Algorithme de jeu des Amazones

Le jeu des Amazones a fait l'objet d'études approfondies ; le développement de son algorithme peut être grossièrement classé en trois étapes. Initialement, les méthodes traditionnelles ancrées dans la théorie mathématique, telles que l'algorithme de recherche minimax [21], dominaient le domaine. La deuxième phase a vu l'émergence des algorithmes de Monte Carlo, y compris diverses variantes de la recherche arborescente de Monte Carlo [19, 20, 22-24]. Dans la troisième étape, les techniques d'apprentissage profond exploitant les bases de données de jeux d'Amazon sont apparues au premier plan [25,26]. Notamment, le programme Amazon le plus représentatif , Invader [27-29], s'appuie sur MCTS pour l'exploration prospective, complété par une formation de réseaux neuronaux sur des bases de données d'échecs, ce qui se traduit par une formidable capacité de jeu dans le jeu des Amazones. On peut dire qu'une optimisation plus poussée des algorithmes de Monte Carlo reste cruciale pour faire progresser la recherche sur les Amazones.

2.2. Algorithme de recherche arborescente de

Monte Carlo Le développement de l'algorithme de recherche arborescente de Monte Carlo dans Amazons peut être grossièrement divisé en deux axes de recherche : l'optimisation de l'algorithme en termes de recherche arborescente et l'optimisation de l'algorithme en termes de fonctions d'évaluation. En termes d' optimisation des algorithmes de recherche arborescente, les chercheurs ont travaillé sur l'amélioration des algorithmes de recherche pour améliorer les performances d'Amazon. Par exemple, Guo et al. a divisé le jeu en étapes précoces, intermédiaires et tardives, en employant différentes stratégies à chaque étape [30]. Li et coll. appliqué l'algorithme UCT aux Amazones [31] ; Quan et coll. ont démontré la supériorité de l'algorithme UCT chez Amazons [32].
D'autre part, l'optimisation de la fonction d'évaluation est également l'un des axes de recherche. En optimisant la fonction d'évaluation, le jeu en cours peut être évalué avec plus de précision et l'algorithme de recherche peut être guidé pour prendre des décisions plus éclairées. Lieberum et coll. consacrent leurs recherches à cet aspect [33,34], en se concentrant principalement sur l'amélioration du réglage des hyperparamètres et la conception de nouvelles fonctions d'évaluation. Dans l'ensemble, l'algorithme de recherche arborescente de Monte Carlo ne peut être amélioré sans aucun aspect, et dans cet article, notre proposition de fusion de la stratégie Move Groups et de la stratégie Parallel Evaluation représente une considération complète des deux axes de recherche historiques.
Cette fusion de l'algorithme de recherche arborescente de Monte Carlo est avant-gardiste et efficace.

2.3. Fonction d'évaluation

À l'heure actuelle, presque tous les algorithmes de recherche sont confrontés à un problème : l'évaluation de la situation. Une bonne évaluation de la situation est très importante dans le jeu des Amazones et peut même déterminer la qualité du programme de jeu [33].

Semblable à de nombreux jeux de société, les Amazones peuvent être grossièrement divisées en trois étapes : ouverture, milieu et fin, avec des stratégies distinctes pour chaque étape. Les principaux indicateurs de la fonction d'évaluation des Amazones comprennent le territoire, la position et la mobilité, et l' importance de ces trois caractéristiques varie selon les différentes étapes. Sur la base de ces idées, Guo et al. [30] ont proposé une fonction d'évaluation par étapes, comme le montre l'équation (1) :

$$\text{Valeur} = k1 \quad t1 + k2 \quad t2 + k3 \quad p1 + k4 \quad p2 + k5 \quad \text{mobilité} \tag{1}$$

où t1 et t2 sont des valeurs métriques de territoire, qui représentent respectivement l'évaluation du contrôle de chaque camp sur l'espace, basée sur la méthode QueenMove et la méthode KingMove. p1 et p2 sont des valeurs métriques de position, reflétant la position stratégique des pièces de chaque camp, également basées sur les méthodes QueenMove et KingMove par rapport à l'espace. mobilité

Machine Translated by Google

est la valeur d'évaluation de la flexibilité des deux pièces Amazon. k1, k2, k3, k4 et k5 sont les valeurs de poids correspondantes. La manière dont ces valeurs sont calculées est décrite ci-dessous.

Selon la littérature [2], le calcul des valeurs de territoire et de position implique deux types de méthodes : QueenMove et KingMove. QueenMove signifie qu'un mouvement peut être effectué tant qu'il n'y a aucun obstacle à franchir dans l'une des huit directions (les obstacles incluent des pièces et des flèches), tandis que KingMove signifie que la distance d'un seul mouvement est une, c'est-à-dire une seule. un déplacement peut être effectué vers une case voisine. La valeur QueenMove représente le nombre minimum de mouvements dont les quatre pièces d'un joueur ont besoin pour atteindre un espace donné (c'est-à-dire sans pièces ni flèches) ; la valeur KingMove représente le nombre minimum de mouvements que les quatre pièces d'un joueur doivent effectuer pour atteindre un espace donné. Comme le montre la figure 1 sur le côté gauche, la valeur 2 dans le coin supérieur gauche du carré supérieur gauche (A, 1) indique que la pièce blanche a besoin d'au moins deux mouvements pour atteindre cet espace via la méthode QueenMove, et la valeur 4 dans le coin inférieur droit indique que la pièce noire a besoin de quatre mouvements. De même, les valeurs 2 et 7 dans le même carré de la figure 1 à droite indiquent la situation dans laquelle la méthode KingMove est utilisée.

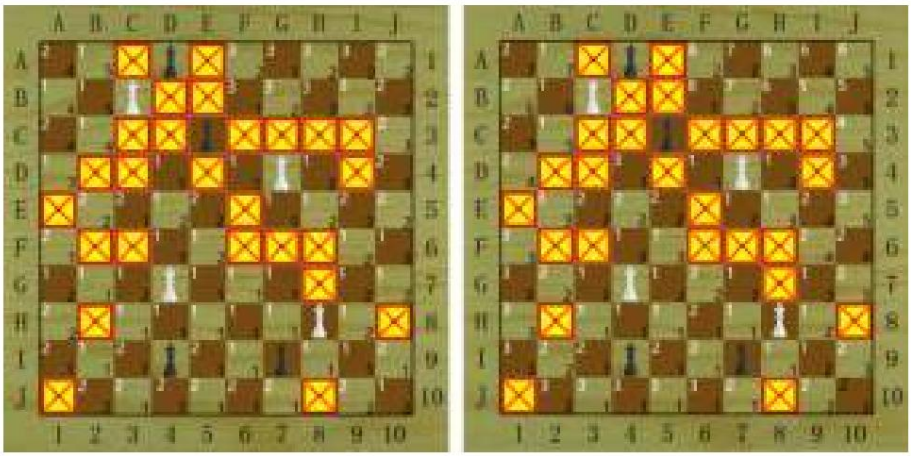


Figure 1. Résultats des calculs des méthodes QueenMove et KingMove. La gauche est le résultat du calcul KingMove et la droite est le résultat du calcul QueenMove.

Ainsi, on peut utiliser la formule de calcul suivante pour tous les indicateurs : t1 représente la différence de contrôle du territoire selon la méthode QueenMove. t2 représente la différence de contrôle du territoire selon la méthode KingMove.
Comme le montrent les équations (2) et (3) :

$$ti = \sum v\,(Bi\ ,\ Wi) \tag{2}$$

$$v\,(Bi\ ,\ Wi) = \begin{cases} 0, & Bi(A) = Wi(A) = \infty \\ k, & Bi(UNE) = Wi(UNE) = \infty \\ 1, & Bi(A) < Wi(A) \\ -1, & Bi(UNE) > Wi(UNE) \end{cases} \tag{3}$$

où B représente les pièces noires, W représente les pièces blanches, l'indice 1 indiquant la méthode QueenMove et 2 indiquant la méthode KingMove. A représente des carrés vides. k représente la prépondérance du carré principal et la plage de valeurs est supérieure à −1 et inférieure à 1. Lorsque le joueur avec le premier coup est blanc, k est supérieur à 0, sinon il est inférieur à 0. p1 représente le positions des deux pièces sur l'espace de contrôle A en utilisant la méthode Queen-

Move. p2 représente les positions des deux pièces sur l'espace de contrôle A en utilisant la méthode KingMove. Comme le montrent les équations (4) et (5) :

Machine Translated by Google

$$p1 = 2\Sigma\ 2\ -B1(A) - 2\ -W1(A) \tag{4}$$

$$p2 = \sum_{UN} \min(1, \max(-1, B2(A) - W2(A)/6)) \tag{5}$$

la mobilité signifie la flexibilité des pièces. Comme le montrent les équations (6) et (7) :

$$\text{mobilité} = \sum_{\text{avec pièce blanche}} F(w) - \sum_{b \quad \text{pièce noire}} F(b) \tag{6}$$

$$F(a) = \begin{cases} \text{espaces environnants, a est l'espace} \\ 0, & \text{a est l'obstacle} \\ \sum s \quad Sa \ \dfrac{Fespace(s)}{D(a,s)}, & \text{a est une pièce Amazon} \end{cases} \tag{7}$$

où F désigne la flexibilité, Sa désigne l'ensemble des espaces pouvant être atteints par une pièce en utilisant la méthode QueenMove, et D(a,s) est le nombre minimum de mouvements de a vers s en utilisant la méthode Méthode KingMove.

Nous avons retenu les paramètres d'hyper-paramétrisation selon la littérature [2] ; détaillé les informations sur chaque indicateur d'évaluation sont présentées dans le tableau 1.

Tableau 1. Tableau des coefficients de pondération.

	Territoire des mesures d'évaluation		Position		Mobilité
	Facteur d'évaluation t1	t2	p1	p2	mobilité
	Coefficient de poids k1	k2	k3	k4	k5
Étapes	Ouverture	0,14	0,37	0,13	0,13	0,20
	Milieu	0,30	0,25	0,20	0,20	0,05
	Fin	0,80	0,10	0,05	0,05	0,00

Le tableau 1 montre que t1 gagne progressivement en importance à mesure que le jeu progresse, commençant plus petit dans l'ouverture, devenant plus grand au milieu et atteignant son apogée en fin de partie. En effet, à mesure que le jeu progresse, le contrôle du plateau devient de plus en plus important. crucial pour la victoire ou la défaite. t2 est extrêmement crucial dans la phase d'ouverture pour assurer la sécurité des régions voisines et des frontières territoriales, mais son rôle diminue à mesure que le jeu progresse et peut être presque ignoré en fin de partie. Cela reflète l'urgence de la défense du tracé initial et la concurrence territoriale. p1 et p2 désignent les positions des pièces, qui jouent principalement un rôle dans la phase d'ouverture lorsqu'une répartition équilibrée des pièces est crucial pour le déploiement stratégique ultérieur ; en fin de partie, le schéma du jeu est déjà déterminé, et la répartition des pièces n'est plus un point clé ; la mobilité dénote la flexibilité, qui est la plus importante au début du jeu, car bloquer les pièces peut être sérieusement désavantageux ; à mesure que le jeu progresse et que l'agencement initial est établi, la répartition des pièces devient de plus en plus critique. Comme le le jeu progresse et la disposition est terminée, l'impact de la mobilité des pièces sur le jeu La situation est affaiblie, son poids est donc réduit à négligeable en milieu et en fin de partie. Cela reflète la nécessité d'une flexibilité stratégique dès les premiers stades de la liberté de mouvement.

3. Méthodologie de recherche

Dans cette section, nous décrivons d'abord le cadre global de l'algorithme MCTS amélioré avec la stratégie MG-PEO. Nous développons ensuite les détails de l'optimisation, en nous concentrant sur deux aspects : Déplacer les groupes et l'évaluation parallèle.

Machine Translated by Google

3.1. Cadre global de l'algorithme Nous

proposons une stratégie d'optimisation de l'algorithme MG-PEO basée sur la recherche arborescente de Monte Carlo [22]. Comme le montre la figure 2, l'algorithme de recherche arborescente de Monte Carlo optimisé implique la construction itérative d'un arbre de recherche, en créant un nouvel arbre à chaque tour jusqu'à ce que la contrainte de temps soit atteinte ou que le jeu se termine. À ce stade, la recherche s'arrête et renvoie l' opération racine la plus performante (une combinaison d'actions de certains nœuds des deuxième et troisième couches).

Plus précisément, les quatre étapes suivantes sont effectuées pour chaque itération de recherche :

(1) Sélection : à partir du nœud racine, une certaine stratégie est appliquée pour descendre de manière récursive dans l'arborescence pour la sélection de nœuds extensible. Un nœud est extensible s'il représente un état non terminal et a des enfants non visités.
(2) Expansion : développez l'arborescence en ajoutant un (ou plusieurs) nœud(s) enfant(s), en fonction des opérations disponibles.
(3) Simulation : exécutez une simulation à partir d'un nouveau nœud pour produire des résultats selon la stratégie par défaut.
(4) Rétropropagation : sauvegardez les résultats de la simulation via les nœuds sélectionnés pour mettre à jour leurs statistiques.

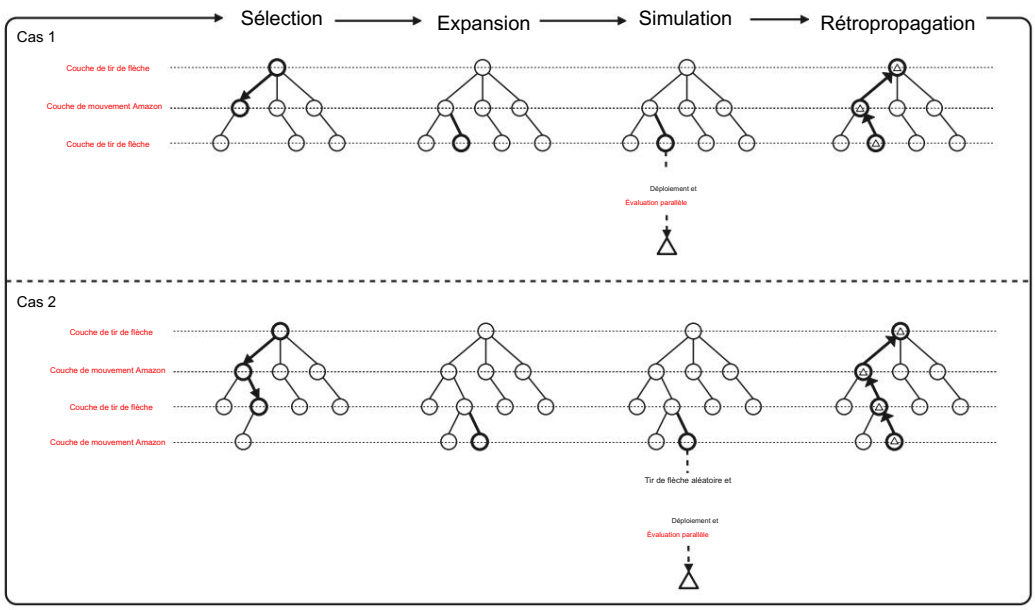


Figure 2. Cadre général MCTS plus MG-PEO.

Comme le montre la figure 2, en raison de l'utilisation de la stratégie Move Groups, l'algorithme fonctionne selon deux scénarios pendant la phase de simulation : si le nœud actuel est un nœud de couche Arrow Shot, il exécute le déploiement et l'évaluation parallèle ; si le nœud actuel est un nœud de couche Amazon Movement, il effectue un tir de flèche aléatoire, un déploiement et une évaluation parallèle.

Le pseudo-code du framework global est décrit dans l'algorithme 1.

Algorithme 1 : cadre d'algorithme général

Entrée : s0 (l'état actuel du jeu)
Sortie : Décision optimale 1 :
fonction MCTSSEARCH (s0) 2 : 3 : 4 :
créer le nœud racine v0 avec
l'état s0 dans le budget de calcul faire v1 ←
Sélection (v0) v2
← Expansion (v1, s0) Δ
5 : 6 : ←Simulation (v2, s0)

Machine Translated by Google


```
                La sauvegarde
7 : 8 :         (v2,Δ) renvoie BESTCHILD(v0)
```

3.2. Stratégie de déplacement des groupes dans l'algorithme

L'idée centrale de la stratégie Move Groups est la suivante : selon les règles, un nœud dans l'arbre de recherche d'origine (Amazon Movement et Arrow Shot combinés) est décomposé en deux nœuds (représentant respectivement Amazon Movement et Arrow Shot). En utilisant la phase de simulation de MCTS pour attribuer des scores de valeur à une couche supplémentaire (la couche nouvellement ajoutée), nous pouvons atteindre les objectifs d'élagage et de gain de temps, en fonction des scores de valeur de la couche supplémentaire (par exemple, en utilisant la stratégie UCT). .

Plus précisément, lorsqu'un nœud de couche supplémentaire n'a pas encore développé ses nœuds enfants, son score de valeur est initialement infini. Par conséquent, chaque nœud de couche supplémentaire développera d'abord ses nœuds enfants une fois et emploiera une stratégie de déploiement superficiel pour obtenir un score de valeur initial. Par la suite, nous pouvons sélectionner le nœud de couche supplémentaire avec le score de valeur le plus élevé pour une expansion et un déploiement ultérieurs, jusqu'à ce que les nœuds enfants d'un nœud de couche supplémentaire particulier soient entièrement générés et que le nœud soit à nouveau sélectionné. À ce stade, nous pouvons alors déduire la décision optimale à partir des nœuds enfants, qui constituent le cœur de l'élagage précoce. Il convient de noter qu'en raison du mécanisme de rétropropagation dynamique du MCTS, les scores de valeur des nœuds de couche supplémentaires changeront dynamiquement à mesure que leurs nœuds enfants se développent. Par conséquent, un élagage précoce peut également se produire au niveau des deuxième, troisième ou autres nœuds de couche supplémentaire. Pour plus de détails sur le calcul et la mise à jour des scores de valeur, veuillez vous référer à la section 3.3. Étant donné que l'étape complète est une combinaison de deux étapes, à savoir Amazon Movement et Arrow Shot, et que la couche supplémentaire ne représente qu'une seule étape, le nombre total de nœuds dans la couche supplémentaire doit être inférieur au nombre total de nœuds, améliorant ainsi le performances de l'élagage précoce dans l'algorithme

Cependant, la stratégie Déplacer les groupes augmente également le nombre total de nœuds d'arborescence (pour une opération complète en une seule étape) en raison de l'introduction de la couche supplémentaire, et si tous les nœuds sont développés sans élagage à l'avance, la surcharge supplémentaire (expansion et évaluation) encourus par les nœuds de la couche supplémentaire affecteront négativement l'algorithme, ce qui conduira également au fait que, dans le pire des cas, l'ajout de la stratégie Move Groups au MCTS fonctionnera légèrement moins bien que l'algorithme MCTS sans la politique de déplacement des groupes. Une analyse théorique complète de la politique de déplacement des groupes est détaillée dans la section 4.1.

3.3. Stratégie d'évaluation parallèle dans l'algorithme

L'idée centrale de la stratégie d'évaluation parallèle est que la partie d'évaluation chronophage de la simulation MCTS est traitée dans le cadre d'une optimisation parallèle. Il a été montré dans la littérature [35] qu'accélérer cette fonction d'évaluation à forte intensité de main-d'œuvre permet d'itérer plus fréquemment l'algorithme de recherche arborescente de Monte Carlo. Cette métrique est très bénéfique pour compenser le fait que la stratégie Move Groups peut tomber dans un optimal local (pire des cas) ; Plus précisément, l'accélération de la fonction d'évaluation permet à Monte Carlo Tree Search d'évoluer le plus rapidement possible, minimisant ainsi l'impact du pire des cas associé à la stratégie Move Groups. Le pseudo-code de la partie évaluation parallèle est tel qu'indiqué dans l'algorithme 2.

Algorithme 2 : Évaluation de la situation

```
Entrée : s0 (l'état actuel du jeu)
Sortie : Valeur d'évaluation de situation 1 :
fonction SituationEvaluation(s0) 2 : 3 :
        t1, t2, p1, p2, mobilité ← ParallelEvaluationMetrics (s0)
        Valeur = k1  t1 + k2  t2 + k3  p1 + k4  p2 + k5  retour de mobilité Valeur
4 :
```

Les nœuds sont soumis à une évaluation et les scores de valeur qui en résultent guident la sélection des nœuds. Les algorithmes traditionnels de recherche arborescente de Monte Carlo choisissent généralement le nœud enfant avec

Machine Translated by Google

score de nœud (taux de victoire) basé sur l'état actuel du jeu [36]. Cependant, cette méthode de sélection n'est pas idéale, car elle tend à donner la priorité aux gains immédiats et peut conduire à des optima locaux.

Pour résoudre ce problème, cet article intègre la « stratégie de déploiement superficiel » [37] et l' algorithme « intervalle de confiance supérieur basé sur un arbre » [19] pour affiner la sélection des nœuds. Plus précisément, la valeur d'utilisation dérivée uniquement des nœuds de niveau décision peut être inexacte. Pour améliorer la précision, une stratégie Shallow Rollout est utilisée, simulant des scénarios plus proches de la fin du jeu. De plus, l'algorithme prend en compte l'ampleur de la valeur exploratoire d'un nœud , qui est élevée pour les nœuds présentant une grande incertitude et de nombreuses possibilités futures . En résumé, l'approche évalue à la fois la valeur d'utilisation à l'équilibre partiel et la valeur d'exploration pour faciliter une prise de décision plus éclairée.

1. « Stratégie de déploiement superficiel » : cela signifie que la valeur d'utilité n'est pas calculée, non seulement en fonction du prochain coup mais également en fonction de la position du jeu après avoir simulé le coup plusieurs fois de manière aléatoire. Dans cet algorithme, nous définissons le nombre de déploiements sur deux et les profondeurs sur 4 et 5. Plus précisément, après avoir développé le nœud de décision complet (Amazon Movement plus Arrow Shot), nous effectuons deux déploiements pour la situation actuelle. Dans chaque déploiement, nous simulons deux étapes : la prochaine décision de l'adversaire (Amazon Movement plus Arrow Shot) et notre décision ultérieure (Amazon Movement plus Arrow Shot), pour obtenir une valeur d'utilité plus prospective. Voir l'algorithme de pseudo-code 3 pour plus de détails. Pour l'analyse des paramètres pertinents, voir l'annexe A.

2. « Intervalle de confiance supérieur basé sur un arbre » : comme le montre l'équation (8), la formule de calcul de l'UCT se compose de deux composants. Le premier terme, Xi , représente la valeur utilitaire des informations dans l'arbre de recherche construit, et le deuxième terme représente la valeur d'exploration des nœuds non visités. La constante C est utilisée pour ajuster la proportion du compromis de l'algorithme entre les deux valeurs. Notez que la valeur de la formule racine est infinie en raison de la présence de nœuds enfants inaccessibles, ce qui garantit que les nœuds enfants inaccessibles sous le nœud actuel seront visités au moins une fois.

$$UCT = Xi + C \sqrt{\frac{\ln N}{ni}} \quad (8)$$

où Xi est le taux de victoire du nœud enfant. N est le nombre de fois où le nœud parent est accédé. ni est le nombre de fois où le nœud enfant est accédé. C est une constante ($\sqrt{2}$ et Szepesvári [36] ont montré que la valeur satisfait l'inégalité de Hoeffding avec des récompenses comprises dans la plage [0, 1]. $\frac{1}{2}$). Kocsis

Ici, nous pouvons fournir le pseudo-code complet comme indiqué dans l'algorithme 3.

Algorithme 3 : MCTS avec la stratégie MG-PEO 1 : fonction

```
Selection(v) 2 : alors que v est
nonTerminal do
                si v n'est pas complètement développé alors
3 : 4 :             revenir v
                sinon
5 : 6 :             v ← MEILLEUR ENFANT(v)
            revenir v

7 : 8 : fonction Expansion(v,&s0) 9 : si v
est un nœud de mouvement Amazon alors
                ch ← créer aléatoirement un nœud de tir de flèche sinon si
10h11 :     v est un nœud de tir de flèche ch ← créer
                aléatoirement le nœud de mouvement Amazon s0 ← Mettre à jour l'état de
12h13 :     la carte retourner ch 14 : 15 : fonction Simulation
            (v, s0)
```

Machine Translated by Google


```
            S ← s0
16 : 17 :   //* profondeur = 4 et 5 (Ordre indépendant deux fois)
18 :        S ← Déploiement (S, profondeur)
19 : Δ ← SituationÉvaluation(S)
20 : retour Δ
21 : fonction Déploiement(S,profondeur)
22 : répéter
23 : si est le terminal (S) alors
24 : passer
                si v est le nœud de mouvement Amazon alors
25 : 26 :           flèche ← Flèche aléatoire (S)
                    S ← faireAction(v,flèche)
27 : 28 :           profondeur ← profondeur - 1
                sinon si v est le nœud Arrow Shot alors
29h30 :             déplacer ← parent de v
                    S ← faireAction(déplacer,v)
31 : 32 :           profondeur ← profondeur - 1
33 :        jusqu'à profondeur = 0 ou gameover
            retour S
34 : 35 : fonction Sauvegarde(v,Δ)
36 :        N(v) ← N(v) + 1
37 :        Q(v) ← Q(v) + Δ(v, parent de v)
38 : v ← parent de v
39 : fonction MEILLEUR ENFANT(v)
40 :        return argmaxv ′  enfants de v  Q(v ′)/N(v ′) + c √(2 ln N(v)/N(v ′))
```

4. Analyse des performances MG-PEO

Dans cette section, nous concevons deux modules pour analyser la stratégie MG-PEO. Premièrement, nous analyser les performances de la stratégie Move Groups dans Amazons. Deuxièmement, nous effectuons une discussion approfondie sur l'amélioration des performances du mécanisme d'évaluation parallèle chez les Amazones.

4.1. Performance de la stratégie Move Groups dans Amazons

L'efficacité d'un algorithme de recherche arborescente de Monte Carlo basé sur la loi des grands les chiffres dépendent en grande partie du nombre de situations qu'il peut rechercher. Plus il y a de situations l'algorithme peut rechercher, plus il sera facile de trouver des « astuces » qui sont généralement difficiles à trouver trouver et ainsi battre l'adversaire. En d'autres termes, plus la puissance de calcul est grande, plus l'algorithme de Monte Carlo est efficace. Ainsi, afin de montrer objectivement influence de la stratégie Move Groups sur les performances réelles de l'algorithme MCTS, nous proposons une nouvelle définition (distance de convergence gagnante) pour évaluer la performance de l'algorithme.

Définition 1. Distance de convergence gagnante. Ceci calcule le nombre de nœuds d'arborescence développés par l'arbre de recherche, à la fois en profondeur et en largeur, depuis l'état actuel (nœud racine) jusqu'à l'état gagnant (jeu terminé). En général, moins l'arbre de Monte Carlo a besoin de s'étendre pour remporter la victoire. le plus tôt possible, plus la distance de convergence gagnante est courte.

Pour visualiser la distance de convergence gagnante, nous utilisons la première étape d'ouverture comme exemple. Puisque le nombre de nœuds pouvant être générés de la première étape au gagnant l'état est incommensurable, nous utilisons la solution optimale locale de la première étape au lieu de la état gagnant, comme le montre la figure 3.

Sur cette base, nous pouvons tracer l'arbre de recherche de Monte Carlo avec la stratégie Move Groups ajoutée et l'arbre de recherche Monte Carlo sans la stratégie Move Groups ajoutée,

Machine Translated by Google

et calculez les numéros de nœuds développés dans le meilleur et le pire des cas, comme le montre la figure 4. et tableau 2.



Figure 3. La solution optimale locale supposée dans la première étape de l'ouverture. De gauche à droite : initiale état, après Amazon Movement et après Arrow Shot.

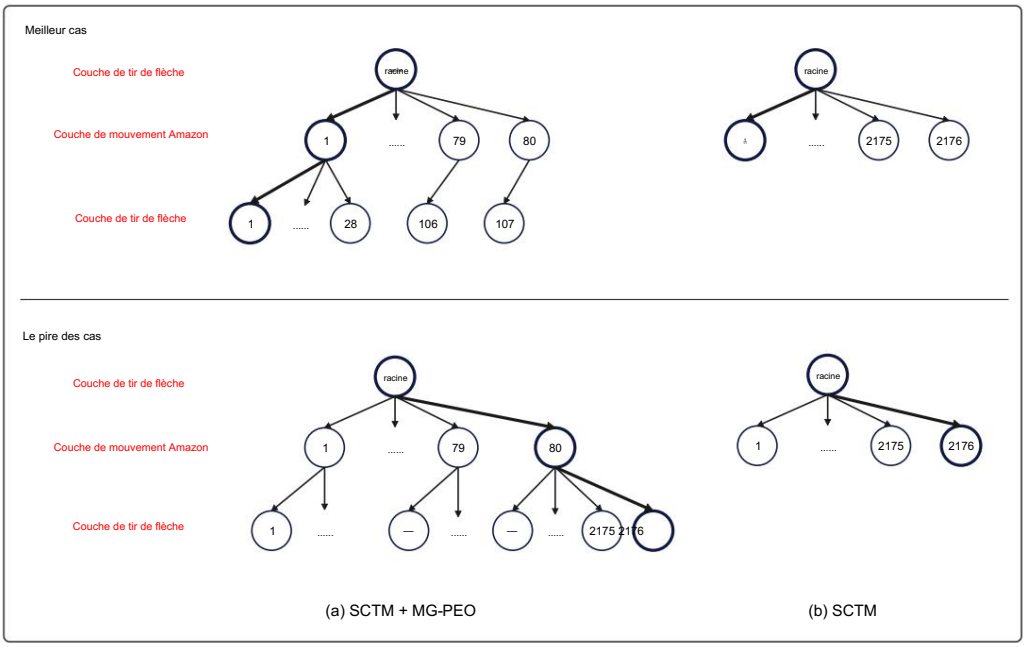


Figure 4. Tracé du calcul gagnant de la distance de convergence. Le numéro sur la figure fait référence au ordre dans lequel le nœud est développé à une certaine couche.

Tableau 2. Distance de convergence gagnante de deux algorithmes (unité : nœuds).

Algorithme	Meilleur cas	Le pire des cas
SCTM + MG-PEO	187	2256
SCTM	2176	2176

Les résultats rouges indiquent le meilleur résultat dans la même colonne.

Dans la figure 4 et le tableau 2, pour la première étape, l'algorithme de recherche arborescente de Monte Carlo avec la stratégie Move Groups a ajouté qu'il faut d'abord étendre 80 nœuds de couche Amazon Movement. Étant donné que le score de valeur d'un nœud de couche Amazon Movement est infini lorsqu'il n'a pas encore été a étendu ses nœuds enfants, chaque nœud de couche Amazon Movement doit se développer davantage un nœud de couche Arrow Shot pour obtenir un score de valeur initiale. Par la suite, nous pouvons sélectionner le nœud de couche Amazon Movement avec le score de valeur le plus élevé pour le nœud enfant suivant expansion. Dans le meilleur des cas, le premier nœud de couche Amazon Movement a la valeur la plus élevée. score. Lors de l'expansion de ses nœuds de couche Arrow Shot suivants, le score de valeur de le nœud parent reste au maximum observé sur toute la couche Amazon Movement nœuds ; les autres nœuds de la couche Amazon Movement n'effectuent pas l'expansion de

Machine Translated by Google

les nœuds de la couche suivante, ce qui reflète un élagage précoce. Enfin, le nombre de nœuds étendu est 187. Dans le pire des cas, tous les nœuds possibles sont entièrement développés et parcourus, ce qui donne lieu à l'extension d'un total de 2 256 nœuds.

L'algorithme de recherche arborescente de Monte Carlo sans l'inclusion des groupes de déplacement la stratégie doit sélectionner le nœud avec le score de valeur le plus élevé comme décision finale après générer tous les nœuds possibles, quels que soient les meilleurs ou les pires scénarios ; résultant en un un total de 2 176 nœuds en cours d'extension.

En résumé, nous pouvons conclure ce qui suit : dans le meilleur des cas, la convergence gagnante la distance des SCTM avec la stratégie Move Groups est nettement plus courte que celle de le SCTM traditionnel, tandis que dans le pire des cas, la distance de convergence gagnante du Le MCTS avec la stratégie Move Groups est légèrement plus faible que celui du MCTS traditionnel. Cela indique que la stratégie Move Groups a de fortes capacités d'élagage et de gain de temps, mais en même temps, il y a quelques lacunes.

4.2. Performance de la stratégie d'évaluation parallèle

Dans cette sous-section, nous examinons le nombre de threads requis pour le mécanisme d'évaluation parallèle, ainsi que le taux d'accélération que le multithreading peut apporter. Depuis le temps la consommation de la fonction d'évaluation provient principalement du temps de calcul de chaque métrique d'évaluation, la partie évaluation parallèle est essentiellement un calcul multithread de paramètres d'évaluation. Le pseudo-code de la fonction métrique d'évaluation parallèle est affiché dans l'algorithme 4.

Algorithme 4 : métriques d'évaluation parallèles

```
Entrée : s0 (l'état actuel du jeu)
Sortie : t1,t2,p1,p2,mobilité
1 : fonction ParallelEvaluationMetrics(s0)
2 :         //* Le calcul des métriques Territory(t1, t2) et Position(p1, p2) utilise la même
                fonction indépendante quatre fois, basée sur le King Move et le QueenMove
                méthodes, qui sont calculées une fois pour le noir et une fois pour le blanc.
3 : //* On optimise ses appels externes en parallèle.
4 : Module1 = lambda (plage)
5 : pour l' index dans la plage :
                  couleur ← indice/2
6 : 7 :           type ← indice%2
                  CalTP(couleur,type)
8 : 9 :     parallèle_pour(block_range(0,2),Module1)

10 :        //* La métrique Mobilité(mobilité) est calculée une fonction à la fois, et la
                la logique interne consiste à traiter chaque position du tableau (GridSize = 100) indépendamment.
11 :        //* Nous optimisons en parallèle sa logique de traitement de boucle interne.
12 : Module2 = lambda (plage)
13 :            pour l' index dans la plage :
14 : couleur ← indice/2
                  type ← indice%2
                  CalEveryBoardPosition()
15 : 16 : 17 : parallèle_pour(block_range(0,GridSize),Module2)
```

A. Discussion sur le nombre de threads : Depuis les principales métriques de l'évaluation la fonction comprend des mesures de territoire, de position et de manœuvrabilité, ainsi que les valeurs du territoire et les métriques de position peuvent être calculées à l'aide d'une seule fonction ; cependant, comme les calculs de ces deux métriques sont liées aux couleurs des pièces et à la méthode de calcul (méthode KingMove et méthode QueenMove), quatre calculs indépendants sont nécessaires pour les valeurs des métriques de territoire et de position. L'indicateur de mobilité est calculé par une fonction qui nécessite un calcul indépendant. L'algorithme est optimisé dans

Machine Translated by Google

parallèle pour les indicateurs de territoire et de position, qui nécessitent théoriquement un total de quatre fils. Pour l'indicateur de maniabilité, dix filetages sont théoriquement nécessaires pour les calculs liés à la taille du conseil d'administration.

B. Exploration du taux d'accélération que peut apporter le multi-threading : Afin de explorer avec précision le taux d'accélération, nous avons compté le temps consommé par le calcul fonction de chaque métrique d'évaluation, et les données sont présentées dans le tableau 3.

Tableau 3. Temps d'exécution du module et taux d'accélération.

Module	Temps d'exécution en série	Temps d'exécution en parallèle	Taux d'accélération
Module1	69 nous	29 nous	2.379
Mudule2	45 nous	23 nous	1.956
Fonction d'évaluation complète 114 us		54 nous	2.111

Comme le montre le tableau 3, le temps de calcul pour le territoire et la position optimisés les métriques ont atteint une accélération de 2,379 fois, tandis que le temps de calcul pour la mobilité optimisée la métrique a atteint une accélération de 1,956 fois. Le taux d'accélération global pour la partie évaluation était de 2,111. Dans la section A, nous avons discuté du nombre de threads et noté que la valeur théorique l'accélération de la parallélisation doit correspondre au nombre de threads. Cependant, le temps la consommation de chaque module avant et après optimisation différait quelque peu de la discussion dans la section A. Nous considérons que la surcharge des bibliothèques parallèles tierces[38] a consommé certaines performances des ressources informatiques. Voir l'annexe B pour plus de détails sur Analyse de l'utilisation du processeur. Néanmoins, le mécanisme d'optimisation parallèle a réduit le temps requis pour traiter les mêmes tâches 2,111 fois, ce qui entraîne une performance significative amélioration des applications pratiques.

5. Expérience et résultats

Dans cette section, afin de vérifier l'efficacité de la stratégie d'optimisation de l'algorithme MG-PEO proposée dans cet article, trois expériences ont été conçues : la première compare les métriques de performance pertinentes de l'algorithme MCTS avec la stratégie MG-PEO et le algorithme MCTS traditionnel, et un test de signification est effectué sur les résultats de la comparaison. La deuxième expérience se concentre sur l'évaluation de l'ampleur de l'effet de chaque module d'optimisation de MG-PEO sur le MCTS, c'est-à-dire l'expérience d'ablation. La troisième expérience est une analyse expérimentale généralisée du MG-PEO.

De plus, il convient de mentionner que le programme MG-PEO Amazons proposé dans cet article, a remporté le premier prix du concours Amazons au China Collegiate 2023 Championnat de jeux informatiques et tournoi national de jeux informatiques.

5.1. Informations sur l'environnement expérimental

Afin de garantir la fiabilité des résultats expérimentaux, nous avons effectué toutes les expériences dans un environnement unifié et informations détaillées sur les expériences Les paramètres d'environnement sont présentés dans le tableau 4.

Tableau 4. Informations sur l'environnement expérimental.

Matériel et contraintes	Détails
GPU	GeForce RTX 3070
Processeur	Intel Core i7-11800H
Temps de prise de décision	5 s
Nombre de matchs joués par le même adversaire	2
Nombre de sets par match	3
Plateforme de correspondance expérimentale	Plateforme de jeu SAU
Mode d'exécution de Visual Studio	[Version x64]

Contre le même adversaire, un total de six sets seront joués, dont trois sets pour le premier joueur et trois pour le premier joueur. deuxième joueur.

Machine Translated by Google

5.2. Expériences quantitatives et qualitatives

5.2.1. Statistiques et analyses du pourcentage de victoires

Afin d'évaluer objectivement et précisément l'efficacité de la stratégie d'optimisation de l'algorithme MG-PEO, nous avons sélectionné l'algorithme MCTS sans optimisation Move-Groups et sans optimisation Parallel Evaluation à des fins de comparaison, et nous avons joué un total de 100 matchs dans les mêmes conditions, avec chaque match divisé en 3 jeux chacun (première main et seconde main). Plus précisément, nous avons sélectionné 33 joueurs ayant participé à une compétition de jeux vidéo et 17 joueurs qui ont publié en open source leur classement (ladder list) sur la plateforme Botzone (https://www.botzone.org.cn/, consulté le 2 mars 2024), totalisant 50 opposan

En raison de la nécessité de bibliothèques tierces, nous avons exporté localement les programmes exe pour tous les joueurs, puis avons mené les jeux les uns contre les autres sur la plateforme SAU (http://computergames.caai.cn/platform/SAU_Game_Platform_2.1.0_r3.rar , consulté le 20 mars 2024), et les résultats des jeux sont présentés dans le tableau 5. Des informations détaillées sur tous les jeux peuvent être trouvées à l'annexe C.

Tableau 5. Le taux de réussite des deux algorithmes.

Algorithme	Pourcentage gagnant
SCTM	74%
SCTM + MG-PEO	97%

Les résultats rouges indiquent le meilleur résultat dans la même colonne.

Dans le tableau 5, l'algorithme MCTS a un taux de victoire de 74 % dans Amazons Duel sur la plateforme de mise en relation SAU. Dans le même environnement de test, l'algorithme MCTS avec la stratégie MG-PEO montre une plus grande stabilité et son taux de victoire dans l'Amazons Match sur la plateforme de matchmaking SAU atteint 97 %. En résumé, l'algorithme MCTS avec la stratégie MG-PEO présente une optimisation plus complète et plus approfondie lors de l'examen et de l'utilisation des stratégies que l'algorithme MCTS et montre donc un taux de victoire plus élevé dans ce scénario de test.

Il convient de noter qu'en observant les correspondances des deux algorithmes, nous avons constaté que l'algorithme MCTS optimisé prend souvent des décisions plus précises que l' algorithme MCTS non optimisé lors de certains tours critiques. Ces tours clés impliquent souvent des positions cruciales pour des stratégies « encercler le territoire » ou « bloquer l'adversaire ». De plus, lors des matchs, nous avons remarqué que l'algorithme MCTS optimisé présente un net avantage sur certains algorithmes traditionnels (tels que Minimax et MCTS traditionnels). Cependant, ses performances laissent légèrement à désirer face à des programmes de réseaux neuronaux entraînés avec une grande quantité de données de jeu. Néanmoins, cela fournit une direction pour améliorer encore les performances de l'algorithme dans l'étape suivante.

5.2.2. Visualisation et analyse gagnantes de la distance de convergence

Dans cette sous-sous-section, nous comptons le nombre de tours de points de match au cours desquels le pourcentage de victoire de l'algorithme atteint 90 % ou plus (le seuil d'une victoire solide) sur 100 parties jouées par les deux algorithmes sur la plateforme SAU, afin de visualiser la définition de la distance de convergence d'un gain, ainsi que comparer les forces et les faiblesses des capacités de jeu des deux algorithmes du jeu des Amazones. Ici, nous considérons généralement que le résultat d'un jeu est clair lorsque l'algorithme évalue un taux de victoire de 90 % ou plus (sauf erreurs de bas niveau de la part du camp dominant) ; de tels moments sont considérés comme des points de match et nous pensons que plus cela se produit tôt, plus les algorithmes sont puissants. Les données moyennes résultantes pour le nombre de tours de points de match pour les deux algorithmes sont présentées dans le tableau 6. Voir l'annexe C pour les données complètes :

Machine Translated by Google

Tableau 6. Nombre moyen de tours de points de match dans 100 matchs où les valeurs d'évaluation du taux de victoire de deux algorithmes dépassent 90 %.

Algorithme	Valeur moyenne
SCTM	42,57
SCTM + MG-PEO	39,68

Si le joueur perd, les tours de points de match sont considérés comme le nombre maximum de tours, soit 47. De plus, les résultats rouges indiquent le meilleur résultat dans la même colonne.

Comme le montre le tableau 6, le nombre moyen de tours de points de match requis pour que l'algorithme MCTS atteigne le seuil de victoire stable (valeur d'évaluation supérieure à 90 %) est de 42,57 tours. D'autre part, l'algorithme MCTS avec la stratégie MG-PEO surpasse MCTS en termes d'atteinte du même seuil de victoire stable avec une moyenne de 39,68 tours.

Cela indique que l'algorithme MCTS avec la stratégie MG-PEO converge plus rapidement que MCTS. Il nécessite moins de tours moyens que l'algorithme MCTS traditionnel. L' algorithme MCTS avec la stratégie MG-PEO utilise la stratégie Move Groups et la stratégie d'évaluation parallèle pour optimiser le processus de sélection de stratégie de l'algorithme MCTS afin qu'il puisse trouver les meilleurs chemins de recherche dans un laps de temps plus court et atteindre un état gagnant stable. avec un taux de gain de plus de 90% à l'avance.

En résumé, l'algorithme MCTS avec la stratégie MG-PEO a une distance de convergence gagnante plus courte que l'algorithme MCTS traditionnel. Il faut moins de tours pour obtenir un taux de victoire plus élevé et démontre ainsi une capacité de jeu supérieure.

5.2.3. Test de signification

Dans cette sous-section, nous effectuons un test de signification sur les données de décompte des points de match, telles que comptées dans la section 5.2.2, pour déterminer s'il existe une différence significative entre elles. Nous effectuons les tests avec un niveau de signification de 0,05. Les données complètes sur le nombre de tours de points de correspondance sont détaillées à l'annexe C. Tout d'abord, nous utilisons le test de Shapiro – Wilk pour supposer que les deux ensembles de données obéissent à une distribution normale. Cependant, aucun des deux ensembles de données n'est conforme à une distribution normale. Sur la base de ce résultat, nous utilisons le test U de Mann – Whitney pour supposer qu'il n'y a pas de différence significative entre les deux groupes de données. Nous comparons ensuite les caractéristiques de distribution des données des tours de match des deux algorithmes. Les résultats détaillés sont présentés dans le tableau 7.

Tableau 7. Signification des résultats des tests.

Données des tours de points de match	Valeur p pour le test de Shapiro-Wilk (P1)	Valeur p pour Mann – Whitney U Essai (P2)
SCTM	0,0002	1,16 ×10−8
SCTM + MG-PEO	0,0013	

À la suite du test de Shapiro-Wilk, les deux algorithmes testés ont montré une valeur P1 inférieure à 0,05 pour les données sur le nombre de tours de points de match, ce qui nous a amené à rejeter l'hypothèse nulle selon laquelle les données pour MCTS+MG-PEO et le les données pour MCTS sont toutes deux conformes à une distribution normale. Les résultats du test U de Mann-Whitney indiquent que la valeur P2 pour les données sur le nombre de tours de points de match entre les deux algorithmes est nettement inférieure à 0,05, nous rejetons donc l'hypothèse initiale. Autrement dit, il existe une différence significative dans le nombre moyen d'étapes gagnantes entre les deux algorithmes.
Plus précisément, le nombre moyen de coups gagnants requis par les MCTS avec la stratégie MG-PEO est nettement inférieur à celui requis par les MCTS traditionnels contre 100 joueurs, ce qui suggère que les MCTS avec la stratégie MG-PEO surpassent les MCTS traditionnels en termes de capacité de jeu chez les Amazones.

Machine Translated by Google

5.3. Expérience d'ablation

Dans cette sous-section, pour évaluer l'efficacité de la stratégie Move Groups et de la stratégie d'évaluation parallèle chez Amazons, nous concevons deux expériences : une qui utilise uniquement la stratégie Move Groups sans utiliser la stratégie Parallel Evaluation, et une autre qui utilise uniquement la stratégie Parallel Evaluation. Stratégie d'évaluation sans utiliser la stratégie Move Groups. Les résultats expérimentaux sont présentés dans le tableau 8. D'une part, la stratégie Move Groups se concentre sur les itérations MCTS dans un temps limité sur les sous-nœuds des nœuds de couche supplémentaire les plus précieux en exploitant leurs scores de valeur, évitant ainsi l'expansion de nœuds sans importance. nœuds. Bien que la stratégie Move Groups présente de fortes capacités d'élagage, les scores de valeur des nœuds de couche supplémentaire changent dynamiquement avec l'expansion de leurs sous-nœuds en raison du mécanisme de mise à jour de rétropropagation dynamique du MCTS. Par conséquent, tous les nœuds de la couche supplémentaire peuvent être entièrement développés et parcourus plusieurs fois (l'élagage n'a pas eu lieu), ce qui pourrait entraîner une augmentation du temps passé par rapport aux algorithmes MCTS sans la stratégie Move Groups. D'un autre côté, étant donné que la proportion d'évaluation de la situation est limitée, l'utilisation uniquement de l'optimisation de l'évaluation parallèle sans adopter la stratégie Move Groups offre une amélioration limitée des performances. Par conséquent, l'optimisation combinée de l'utilisation de la stratégie Move Groups et de l'évaluation parallèle est efficace et nécessaire.

Tableau 8. Analyse de l'expérience d'ablation de la stratégie Move Groups et de la stratégie d'évaluation parallèle.

Algorithme	Pourcentage gagnant	Nombre d'itérations MCTS en 1 s
SCTM	74%	2176
SCTM + PEO	89%	10 026
SCTM + MG	83%	5 663
SCTM + MG-PEO	97%	45 536

Le « Nombre d'itérations MCTS en 1 s » fait référence au nombre d'itérations de recherche arborescente Monte Carlo effectuées. en une seconde. Dans le cadre de la stratégie Move Groups, chaque itération MCTS aboutit soit à un nœud Amazon Movement, soit à un nœud Arrow Shot. Sans utiliser la stratégie Move Groups, chaque itération MCTS aboutit à un nœud de décision complet (Amazon Movement plus Arrow Shot). De plus, les résultats rouges indiquent le meilleur résultat dans la même colonne.

5.4. Analyse de généralisation

Dans cette sous-section, nous avons effectué une analyse générale de la stratégie Move Groups et de la stratégie Parallel Evaluation, discutant en profondeur de leur transférabilité à d'autres jeux de société.

1. Stratégie de déplacement des groupes : dans certains jeux, le facteur de branchement peut être très important, mais de nombreux mouvements sont similaires. La stratégie Move Groups crée une couche de décision supplémentaire où toutes les actions possibles sont regroupées, ce qui nous permet de gérer plus efficacement le grand facteur de branchement du jeu. Cette stratégie est bénéfique pour d'autres jeux, tels que Go [39] et le jeu de bandit multi-armé avec Move Groups [40]. Nous pensons que cette stratégie mérite encore d'être explorée et étudiée.
2. Stratégie d'évaluation parallèle : La stratégie d'évaluation parallèle peut fournir une accélération efficace, que ce soit pour des jeux simples basés sur des règles comme Gomoku, des jeux plus complexes comme Chess and Go, ou même des jeux de société innovants modernes. Son principal avantage réside dans sa capacité à explorer efficacement l'arbre de décision, ce qui est une exigence courante dans divers jeux de société.

6. Conclusions et travaux futurs

Dans cet article, nous proposons une optimisation efficace de l'algorithme Monte Carlo Tree Search pour le jeu Amazons, qui affine la capacité de Monte Carlo Tree Search grâce à la stratégie Move Groups. Le mécanisme d'évaluation parallèle aborde le risque que la stratégie Move Groups tombe dans un optimal local (ces composants sont

Machine Translated by Google

complémentaires les uns des autres). Les résultats expérimentaux valident que la méthode proposée dans cet article améliore considérablement la capacité de jeu du jeu Amazons.

De futurs efforts de recherche sont également inclus dans nos plans. Premièrement, pour maximiser le gain de temps potentiel de la stratégie d'élagage des Move Groups, nous espérons envisager d'utiliser des outils plus intelligents. approches (telles que les fonctions heuristiques) pour guider le comportement d'élagage des groupes de déplacement stratégie. Deuxièmement, la série d'expériences présentées dans cet article démontre la puissance capacité d'amélioration des performances de l'optimisation parallèle, nous prévoyons donc d'envisager davantage des possibilités parallèles et des stratégies parallèles plus efficaces à l'avenir. Enfin, nous aussi J'espère que cet algorithme pourra être étendu davantage pour résoudre davantage de problèmes de gameplay.

Contributions des auteurs : Tous les auteurs ont contribué à la conception et au design de l'étude. Méthodologie, logiciels, validation, enquête, analyse des données, vérification des modèles, ressources, conservation des données, visualisation, rédaction, édition et révision : LZ et HZ Supervision, administration du projet, financement acquisition : YZ Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche n'a reçu aucun financement externe.

Déclaration de disponibilité des données : Une partie du code de la concurrence utilisé dans cette étude est accessible au public. code open source, consultable sur le site suivant : https://www.botzone.org.cn/, consulté le 2 mars 2024.

Conflits d'intérêts : Les auteurs déclarent qu'ils n'ont aucun intérêt financier ni relation personnelle pertinent par rapport au contenu de cet article.

Annexe A

Dans l'Annexe A, nous discutons et analysons l'effet de la profondeur du déploiement dans la phase de simulation sur la capacité de l'algorithme à jouer à Amazons et à utiliser la convergence gagnante. métrique de distance (voir la section 4.2 pour plus de détails) pour évaluer la capacité de jeu. Déploiement superficiel Les stratégies peuvent considérer des positions simulées plus proches de la fin de partie et des jeux plus prospectifs pour calculer des valeurs d'évaluation de position plus précises, qui peuvent être utilisées pour mieux guider la sélection des nœuds pour les SCTM. Il est donc crucial d'équilibrer la profondeur de Dérouler; une profondeur trop faible rendra la valeur d'utilisation (taux gagnant) dans la position la formule d'évaluation est moins crédible, tandis qu'une profondeur trop profonde conduira à une seule simulation processus d'itération trop long et le nombre de nœuds étendu sur une période finie de temps deviendra plus petit, réduisant la précision de la décision de recherche dans l'arborescence de Monte Carlo. Nous avons ajusté la valeur de la profondeur de déploiement de 2 à 10 et avons constaté que la valeur optimale de La profondeur de déploiement est de 4 ou 5, comme indiqué dans le tableau A1.

Tableau A1. Analyse paramétrique de la profondeur de déploiement.

Profondeur de déploiement	1	2	3	4	5	6	7-10	4 et 5
Tours de points de match 40.25		40.14	40,0	39,9	39,84	40.04	40.1	39,68

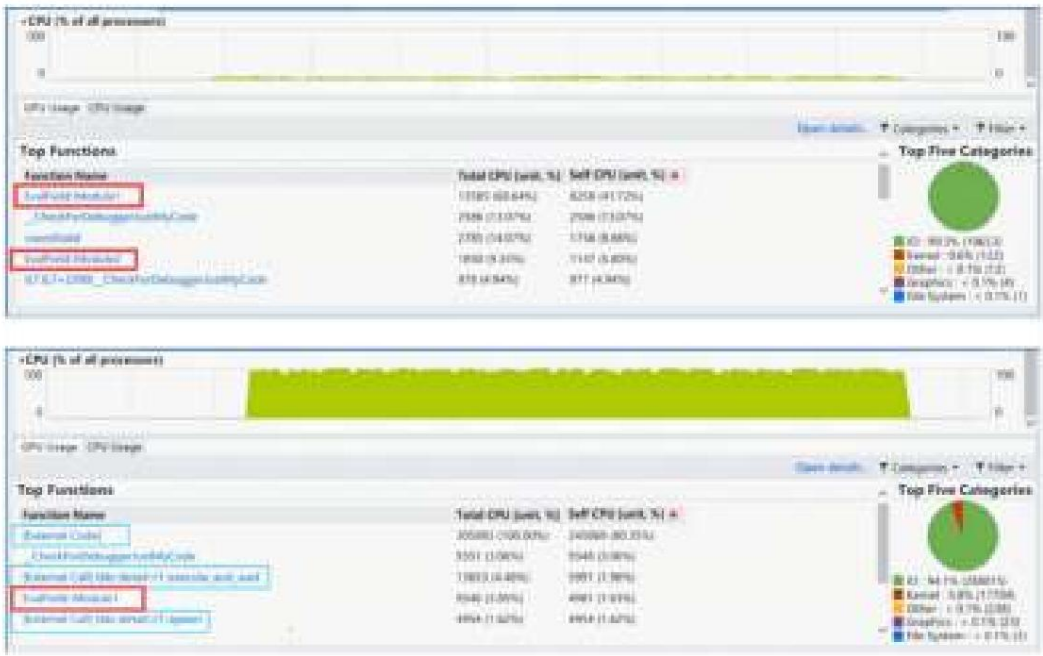
Les résultats rouges indiquent le meilleur résultat dans la même colonne.

De plus, nous avons constaté dans nos expériences que le déploiement deux fois (en choisissant 4 et 5 en profondeur) fonctionne mieux que le déploiement une fois (en choisissant seulement 4 ou 5), c'est-à-dire le nombre de tours aux points de match est plus petite, la distance de convergence pour gagner est plus courte et le l'algorithme est plus capable de jouer aux Amazones.

Annexe B

L'annexe B montre l'utilisation détaillée du processeur de l'algorithme avec et sans le Stratégie d'évaluation parallèle. L'environnement de surveillance est le profileur de performances dans Microsoft Visual Studio 2019, l'algorithme ne prenant qu'une seule décision, qui prend 20 s. Comme le montre la figure A1, la partie supérieure de l'image représente l'utilisation du processeur, et la partie fluctuante des valeurs d'utilisation du processeur correspond aux valeurs de l'algorithme. phase de prise de décision, tandis que la partie restante correspond à l'entrée et à la sortie phases de repos.

Machine Translated by Google



Graphique A1. Utilisation du processeur et 5 principaux modules occupant les ressources du processeur dans la condition [Debug ×64]. La partie supérieure montre les résultats de surveillance de l'algorithme exécuté séquentiellement, tandis que la partie inférieure La partie montre les résultats de surveillance de l'algorithme exécuté en parallèle.

D'après la figure A1, il est évident que l'utilisation du processeur par le programme algorithmique dans l'exécution parallèle est nettement plus élevée que l'exécution séquentielle. Quand l'algorithme Le programme s'exécute séquentiellement, les deux modules encadrés en rouge de la fonction d'évaluation sont parmi les 5 premiers modules occupant le plus de ressources CPU. Cela indique que parallèlement l'optimisation de la partie évaluation est très nécessaire. Lors de l'exécution parallèle du programme d'algorithme, la bibliothèque parallèle tierce encadrée en bleu occupe une place substantielle quantité de ressources CPU. Cela indique indirectement que l'amélioration des performances après la parallélisation, l'accélération théorique n'est pas atteinte en raison d'une ressource importante consommation par code externe et appels externes.

Annexe C

L'Annexe C présente les informations complètes sur le jeu, ainsi que le tour de points de match. nombre, comme le montrent les tableaux A2 et A3.

Tableau A2. Informations complètes sur le jeu (MCTS plus MG-PEO).

Identifier le premier joueur		Deuxième joueur	Résultat	Correspondre Indiquer Tours	Gagnant
1	Amazones basées sur Qt Cadre QMainWindow	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
2	Mouvement divin (le nôtre)	Amazones basées sur Qt Cadre QMainWindow	3:0	39	Mouvement divin (le nôtre)
3	Gomoku	Mouvement divin (le nôtre)	0:3	39	Mouvement divin (le nôtre)
4	Mouvement divin (le nôtre)	Gomoku	3:0	36	Mouvement divin (le nôtre)
5	Othello	Mouvement divin (le nôtre)	0:3	35	Mouvement divin (le nôtre)
6	Mouvement divin (le nôtre)	Othello	3:0	33	Mouvement divin (le nôtre)
7	L'algorithme de Dieu	Mouvement divin (le nôtre)	0:3	43	Mouvement divin (le nôtre)

Machine Translated by Google

Tableau A2. Suite.

Identifier le premier joueur	Deuxième joueur	Résultat	Correspondre Indiquer Tours	Gagnant
8 Mouvement divin (le nôtre)	L'algorithme de Dieu	3:0	42	Mouvement divin (le nôtre)
9 Système de jeu Amazons	Mouvement divin (le nôtre)	0:3	39	Mouvement divin (le nôtre)
10 Mouvement divin (le nôtre)	Système de jeu Amazons	3:0	37	Mouvement divin (le nôtre)
11 Avancez les petites reines	Mouvement divin (le nôtre)	0:3	40	Mouvement divin (le nôtre)
12 Mouvement divin (le nôtre)	Avancez les petites reines	3:0	37	Mouvement divin (le nôtre)
13 Échec et mat	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
14 Mouvement divin (le nôtre)	Échec et mat	3:0	38	Mouvement divin (le nôtre)
15 Amazones à suspense	Mouvement divin (le nôtre)	0:3	40	Mouvement divin (le nôtre)
16 Mouvement divin (le nôtre)	Amazones à suspense	3:0	37	Mouvement divin (le nôtre)
17 La victoire de Qi Kaide	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
18 Mouvement divin (le nôtre)	La victoire de Qi Kaide	3:0	36	Mouvement divin (le nôtre)
19 Super Amazone	Mouvement divin (le nôtre)	0:3	39	Mouvement divin (le nôtre)
20 Mouvement divin (le nôtre)	Super Amazone	3:0	36	Mouvement divin (le nôtre)
21 L'algorithme de Dieu	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
22 Mouvement divin (le nôtre)	L'algorithme de Dieu	3:0	37	Mouvement divin (le nôtre)
23 Amazones voyageuses	Mouvement divin (le nôtre)	0:3	40	Mouvement divin (le nôtre)
24 Mouvement divin (le nôtre)	Amazones voyageuses	3:0	35	Mouvement divin (le nôtre)
25 Échecs suprêmes d'Amazon	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
26 Mouvement Divin (le nôtre)	Échecs suprêmes d'Amazon	3:0	39	Mouvement divin (le nôtre)
27 Troupes d'échecs tombant de Paradis	Mouvement divin (le nôtre)	0:3	43	Mouvement divin (le nôtre)
28 Mouvement divin (le nôtre)	Troupes d'échecs tombant de Paradis	3:0	41	Mouvement divin (le nôtre)
29 Amazones aux clous de girofle	Mouvement divin (le nôtre)	0:3	40	Mouvement divin (le nôtre)
30 Mouvement divin (le nôtre)	Amazones aux clous de girofle	3:0	37	Mouvement divin (le nôtre)
31 Les Amazones de Shao Guang	Mouvement divin (le nôtre)	0:3	43	Mouvement divin (le nôtre)
32 Mouvement divin (le nôtre)	Les Amazones de Shao Guang	3:0	39	Mouvement divin (le nôtre)
33 L'IA joue aux échecs	Mouvement divin (le nôtre)	0:3	43	Mouvement divin (le nôtre)
34 Mouvement divin (le nôtre)	L'IA joue aux échecs	3:0	39	Mouvement divin (le nôtre)
35 Canopus Un	Mouvement divin (le nôtre)	0:3	39	Mouvement divin (le nôtre)
36 Mouvement divin (le nôtre)	Canopus Un	3:0	35	Mouvement divin (le nôtre)
37 Gagnez l'équipe d'ouverture 38	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
Mouvement divin (le nôtre)	Gagnez l'équipe d'ouverture	3:0	39	Mouvement divin (le nôtre)
39 Amazones de Wukong 40	Mouvement divin (le nôtre)	0:3	34	Mouvement divin (le nôtre)
Mouvement divin (le nôtre)	Amazones de Wukong	3:0	30	Mouvement divin (le nôtre)
41 Amazones voyageuses	Mouvement divin (le nôtre)	0:3	44	Mouvement divin (le nôtre)
42 Mouvement divin (le nôtre)	Amazones voyageuses	3:0	40	Mouvement divin (le nôtre)
43 Yue	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
44 Mouvement divin (le nôtre)	Yue	3:0	40	Mouvement divin (le nôtre)
45 Amazonie finale	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
46 Mouvement divin (le nôtre)	Amazonie finale	3:0	38	Mouvement divin (le nôtre)
47 Université de l'information - Monte Carlo	Mouvement divin (le nôtre)	1:2	42	Mouvement divin (le nôtre)
48 Mouvement divin (le nôtre)	Université de l'information - Monte Carlo	3:0	40	Mouvement divin (le nôtre)
49 Haha Salut 50	Mouvement divin (le nôtre)	1:2	42	Mouvement divin (le nôtre)
Mouvement Divin (le nôtre)	Haha Salut	3:0	39	Mouvement divin (le nôtre)
51 Victoire du Dragon	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
52 Mouvement Divin (le nôtre)	Victoire des dragons	3:0	37	Mouvement divin (le nôtre)
53 Super Amazone 54	Mouvement divin (le nôtre)	0:3	43	Mouvement divin (le nôtre)
Mouvement Divin (le nôtre)	Super Amazone	3:0	38	Mouvement divin (le nôtre)
55 Équipe de jumelage de base	Mouvement divin (le nôtre)	0:3	44	Mouvement divin (le nôtre)
56 Mouvement divin (le nôtre)	Équipe de jumelage de base	3:0	41	Mouvement divin (le nôtre)
57 Passer le niveau	Mouvement divin (le nôtre)	3:0	47	Passer le niveau

Machine Translated by Google

Tableau A2. Suite.

Identifier le premier joueur	Deuxième joueur	Résultat	Correspondre Indiquer Tours	Gagnant
58 Mouvement divin (le nôtre)	Passer le niveau	0:3	47	Passer le niveau
59 Université Jiaotong de Dalian Équipe des Amazones 1	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
60 Mouvement divin (le nôtre)	Université Jiaotong de Dalian Équipe des Amazones 1	3:0	39	Mouvement divin (le nôtre)
61 Débarquez	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
62 Mouvement divin (le nôtre)	Débarquez	3:0	40	Mouvement divin (le nôtre)
63 Vide	Mouvement divin (le nôtre)	0:3	44	Mouvement divin (le nôtre)
64 Mouvement divin (le nôtre)	Vide	3:0	41	Mouvement divin (le nôtre)
65 Taureau	Mouvement divin (le nôtre)	0:3	42	Mouvement divin (le nôtre)
66 Mouvement divin (le nôtre)	Taureau	3:0	40	Mouvement divin (le nôtre)
67 Amazones de Wukong	Mouvement divin (le nôtre)	1:2	34	Mouvement divin (le nôtre)
68 Mouvement divin (le nôtre)	Amazones de Wukong	3:0	31	Mouvement divin (le nôtre)
69 Équipe de soutien aux fans de Gangzi	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
70 Mouvement divin (le nôtre)	Équipe de soutien aux fans de Gangzi	3:0	39	Mouvement divin (le nôtre)
71 Pantalon thaïlandais épicé	Mouvement divin (le nôtre)	0:3	44	Mouvement divin (le nôtre)
72 Mouvement divin (le nôtre)	Pantalon thaïlandais épicé	3:0	41	Mouvement divin (le nôtre)
73 Gâteau à l'herbe verte	Mouvement divin (le nôtre)	0:3	34	Mouvement divin (le nôtre)
74 Mouvement divin (le nôtre)	Gâteau à l'herbe verte	3:0	30	Mouvement divin (le nôtre)
75 Prenez l'hypoxie cérébrale végétale Équipe	Mouvement divin (le nôtre)	0:3	40	Mouvement divin (le nôtre)
76 Mouvement divin (le nôtre)	Je n'ai pas réussi à saisir l'air et je ne l'ai pas fait Prenez l'hypoxie cérébrale végétale Équipe	3:0	37	Mouvement divin (le nôtre)
77 DG 78	Mouvement divin (le nôtre)	0:3	44	Mouvement divin (le nôtre)
Mouvement divin (le nôtre)	DG	3:0	39	Mouvement divin (le nôtre)
79 Pourquoi Tang Yang est-il un Dieu 80	Mouvement divin (le nôtre)	0:3	42	Mouvement divin (le nôtre)
Mouvement divin (le nôtre)	Pourquoi Tang Yang est-il un Dieu	3:0	39	Mouvement divin (le nôtre)
81 Équipe de rêve	Mouvement divin (le nôtre)	0:3	43	Mouvement divin (le nôtre)
82 Mouvement divin (le nôtre)	Équipe de rêve	3:0	39	Mouvement divin (le nôtre)
83 Jupe visage de cheval quotidienne et facile porter	Mouvement divin (le nôtre)	0:3	43	Mouvement divin (le nôtre)
84 Mouvement divin (le nôtre)	Jupe visage de cheval quotidienne et facile porter	3:0	40	Mouvement divin (le nôtre)
85 Genshin Impact Expert Divine	Mouvement divin (le nôtre)	0:3	41	Mouvement divin (le nôtre)
86 Move (le nôtre)	Genshin Impact Expert Divine	3:0	37	Mouvement divin (le nôtre)
87 Code Apprentice Mouvement	Move (le nôtre)	0:3	42	Mouvement divin (le nôtre)
divin (le nôtre) 88 89	Code Apprentice Divine	3:0	41	Mouvement divin (le nôtre)
Notes de dérive amazonienne	Move (le nôtre)	0:3	39	Mouvement divin (le nôtre)
90 Mouvement divin (le nôtre)	Amazon Drift Notes Divine	3:0	37	Mouvement divin (le nôtre)
91 L'amour va disparaître, n'est-ce pas ?	Move (le nôtre)	0:3	38	Mouvement divin (le nôtre)
93 Mouvement divin (le nôtre) 92	L'amour va disparaître, n'est-ce pas ?	3:0	33	Mouvement divin (le nôtre)
Petit Su et le chat 94 Mouvement divin (le	Mouvement divin (le nôtre)	0:3	43	Mouvement divin (le nôtre)
nôtre) 95 Je ne veux pas de Marla 96	Little Su et le chat Divine	3:0	42	Mouvement divin (le nôtre)
Mouvement divin (le nôtre) 97 Équipe	Move (le nôtre)	0:3	44	Mouvement divin (le nôtre)
d'ajustement des paramètres 98	Je ne veux pas que Marla	3:0	43	Mouvement divin (le nôtre)
Mouvement divin (le nôtre)	Divine Move (la nôtre)	0:3	44	Mouvement divin (le nôtre)
	Équipe de réglage des paramètres Divine	3:0	41	Mouvement divin (le nôtre)
99 AlphaAmazon 100 Mouvement	Move (la nôtre)	0:3	43	Mouvement divin (le nôtre)
Divin (le nôtre)	AlphaAmazon	3:0	42	Mouvement divin (le nôtre)

Machine Translated by Google

Tableau A3. Informations complètes sur le jeu (MCTS).

Identifier le premier joueur	Deuxième joueur	Résultat	Correspondre Indiquer Tours	Gagnant
1 Amazones basées sur Qt Cadre QMainWindow	SCTM (référence)	1:2	43	SCTM (référence)
2 SCTM (référence)	Amazones basées sur Qt Cadre QMainWindow	3:0	41	SCTM (référence)
3 Gomoku	SCTM (référence)	3:0	47	Gomoku
4 SCTM (référence)	Gomoku	3:0	45	SCTM (référence)
5 Othello	SCTM (référence)	2:1	47	Othello
6 SCTM (référence)	Othello	2:1	44	SCTM (référence)
7 L'algorithme de Dieu	SCTM (référence)	0:3	43	SCTM (référence)
8 SCTM (référence)	L'algorithme de Dieu	3:0	40	SCTM (référence)
9 Système de jeu Amazons	SCTM (référence)	0:3	42	SCTM (référence)
10 SCTM (référence)	Système de jeu Amazons	3:0	40	SCTM (référence)
Avancer les petites reines 11	SCTM (référence)	3:0	47	Avancez les petites reines
12 SCTM (référence)	Avancez les petites reines	3:0	43	SCTM (référence)
13 Échec et mat	SCTM (référence)	2:1	47	Échec et mat
14 SCTM (référence)	Échec et mat	3:0	40	SCTM (référence)
15 Les Amazones à suspense	SCTM (référence)	3:0	47	Échec et mat
16 SCTM (référence) 17	Amazones à suspense	3:0	44	SCTM (référence)
La victoire de Qi Kaide	SCTM (référence)	0:3	44	SCTM (référence)
18 SCTM (référence) 19	La victoire de Qi Kaide	3:0	41	SCTM (référence)
Super Amazone	SCTM (référence)	2:1	47	Super Amazone
20 SCTM (référence) 21	Super Amazone	2:1	43	SCTM (référence)
L'algorithme de Dieu	SCTM (référence)	3:0	47	L'algorithme de Dieu
22 SCTM (référence) 23	L'algorithme de Dieu	2:1	40	SCTM (référence)
Amazones voyageuses	SCTM (référence)	2:1	47	Amazones voyageuses
24 SCTM (référence)	Amazones voyageuses	3:0	42	SCTM (référence)
25 Échecs suprêmes d'Amazon	SCTM (référence)	0:3	40	SCTM (référence)
26 SCTM (référence)	Échecs suprêmes d'Amazon	3:0	41	SCTM (référence)
27 Troupes d'échecs tombant de Paradis	SCTM (référence)	0:3	43	SCTM (référence)
28 SCTM (référence)	Troupes d'échecs tombant de Paradis	3:0	39	SCTM (référence)
29 Amazones aux clous de girofle	SCTM (référence)	0:3	44	SCTM (référence)
30 SCTM (référence) 31	Amazones aux clous de girofle	3:0	40	SCTM (référence)
Les Amazones de Shao Guang	SCTM (référence)	1:2	43	SCTM (référence)
32 SCTM (référence)	Les Amazones de Shao Guang	3:0	40	SCTM (référence)
33 L'IA joue aux échecs	SCTM (référence)	2:1	47	L'IA joue aux échecs
34 SCTM (référence) 35	L'IA joue aux échecs	3:0	42	SCTM (référence)
Canopus Un	SCTM (référence)	0:3	43	SCTM (référence)
36 SCTM (référence)	Canopus Un	2:1	40	SCTM (référence)
37 Gagnez l'équipe d'ouverture	SCTM (référence)	3:0	47	Gagnez l'équipe d'ouverture
38 SCTM (référence)	Gagnez l'équipe d'ouverture	3:0	44	SCTM (référence)
39 Amazones de Wukong	SCTM (référence)	0:3	42	SCTM (référence)
40 SCTM (référence) 41	Amazones de Wukong	3:0	39	SCTM (référence)
Amazones voyageuses	SCTM (référence)	0:3	41	SCTM (référence)
42 SCTM (référence) 43	Amazones voyageuses	3:0	36	SCTM (référence)
Yue	SCTM (référence)	0:3	44	SCTM (référence)
44 SCTM (référence) 45	Yue	3:0	41	SCTM (référence)
Amazonie finale	SCTM (référence)	0:3	42	SCTM (référence)
46 SCTM (référence)	Amazonie finale	3:0	37	SCTM (référence)
47 Université de l'information - Monte Carlo	SCTM (référence)	3:0	47	Information Université-Monte-Carlo
48 SCTM (référence)	Informations Université-Monte Carlo	0:3	47	Information Université-Monte-Carlo

Machine Translated by Google

Tableau A3. Suite.

Identifier le premier joueur	Deuxième joueur	Résultat	Correspondre Indiquer Tours	Gagnant
49 Haha Salut	SCTM (référence)	3:0	47	Haha Salut
50 SCTM (référence) 51	Haha Salut	3:0	41	SCTM (référence)
Victoire des dragons	SCTM (référence)	0:3	43	SCTM (référence)
52 SCTM (référence) 53	Victoire des dragons	3:0	38	SCTM (référence)
Super Amazone	SCTM (référence)	0:3	41	SCTM (référence)
54 SCTM (référence) 55	Super Amazone	3:0	38	SCTM (référence)
Équipe de jumelage de base	SCTM (référence)	3:0	45	Équipe de jumelage de base
56 SCTM (référence) 57	Équipe de jumelage de base	0:3	42	Équipe de jumelage de base
Passer le niveau	SCTM (référence)	3:0	47	Passer le niveau
58 SCTM (référence)	Passer le niveau	0:3	47	Passer le niveau
59 Université Jiaotong de Dalian Équipe des Amazones 1	SCTM (référence)	0:3	44	SCTM (référence)
60 MCTS (référence)	Université Jiaotong de Dalian Équipe des Amazones 1	3:0	42	SCTM (référence)
61 Débarquez	SCTM (référence)	3:0	47	Débarquez
62 SCTM (référence) 63	Débarquez	0:3	47	Débarquez
Vide	SCTM (référence)	0:3	45	SCTM (référence)
64 SCTM (référence)	Vide	3:0	41	SCTM (référence)
65 Taureau	SCTM (référence)	0:3	45	SCTM (référence)
66 SCTM (référence)	Taureau	3:0	40	SCTM (référence)
67 Amazones de Wukong	SCTM (référence)	2:1	47	Amazones de Wukong
68 SCTM (référence)	Amazones de Wukong	3:0	39	SCTM (référence)
69 Équipe de soutien aux fans de Gangzi	SCTM (référence)	3:0	44	Fans de Gangzi Équipe d'assistance
70 MCTS (référence)	Équipe de soutien aux fans de Gangzi	1:2	41	Fans de Gangzi Équipe d'assistance
71 Pantalon thaïlandais épicé	SCTM (référence)	0:3	42	SCTM (référence)
72 SCTM (référence) 73	Pantalon thaïlandais épicé	3:0	40	SCTM (référence)
Gâteau à l'herbe verte	SCTM (référence)	0:3	42	SCTM (référence)
74 SCTM (référence)	Gâteau à l'herbe verte	3:0	37	SCTM (référence)
Je n'ai pas réussi à saisir l'air et je ne l'ai pas fait				
75 Prenez l'hypoxie cérébrale végétale Équipe	SCTM (référence)	0:3	41	SCTM (référence)
	Je n'ai pas réussi à saisir l'air et je ne l'ai pas fait			
76 SCTM (référence)	Prenez l'hypoxie cérébrale végétale Équipe	3:0	40	SCTM (référence)
77 DG	SCTM (référence)	0:3	43	SCTM (référence)
78 SCTM (référence)	DG	3:0	39	SCTM (référence)
79 Pourquoi Tang Yang est-il un Dieu	SCTM (référence)	2:1	47	Pourquoi Tang Yang un Dieu
80 SCTM (référence) 81	Pourquoi Tang Yang est-il un Dieu	3:0	44	SCTM (référence)
Équipe de rêve	SCTM (référence)	0:3	43	SCTM (référence)
82 SCTM (référence)	Équipe de rêve	3:0	44	SCTM (référence)
83 Jupe visage de cheval quotidienne et facile porter	SCTM (référence)	0:3	44	SCTM (référence)
84 SCTM (référence)	Jupe visage de cheval quotidienne et facile porter	3:0	40	SCTM (référence)
85 Expert Genshin Impact	SCTM (référence)	0:3	44	SCTM (référence)
86 SCTM (référence)	Expert Genshin Impact	3:0	39	SCTM (référence)
87 Code Apprenti	SCTM (référence)	0:3	42	SCTM (référence)
88 SCTM (référence)	Apprenti code	3:0	40	SCTM (référence)
89 Notes sur la dérive amazonienne	SCTM (référence)	0:3	44	SCTM (référence)
90 SCTM (référence) 91	Notes sur la dérive amazonienne	3:0	42	SCTM (référence)
L'amour va disparaître, n'est-ce pas ?	SCTM (référence)	0:3	40	SCTM (référence)
92 SCTM (référence)	L'amour va disparaître, n'est-ce pas ?	3:0	38	SCTM (référence)

Machine Translated by Google

Tableau A3. Suite.

Identifier le premier joueur	Deuxième joueur	Résultat	Correspondre Indiquer Tours	Gagnant
93 Petit Su et le chat	SCTM (référence)	0:3	44	SCTM (référence)
94 SCTM (référence)	95 Petit Su et le chat	3:0	40	SCTM (référence)
Je ne veux pas de Marla	SCTM (référence)	0:3	41	SCTM (référence)
96 SCTM (référence)	97 Je ne veux pas de Marla	3:0	37	SCTM (référence)
Équipe de réglage des paramètres	SCTM (référence)	0:3	42	SCTM (référence)
98 SCTM (référence)	Équipe de réglage des paramètres	3:0	39	SCTM (référence)
99 AlphaAmazone	SCTM (référence)	0:3	41	SCTM (référence)
100 MCTS (référence)	AlphaAmazon	0:3	47	AlphaAmazon

Références

1. Li, R. ; Gao, M. Conception de l'algorithme de recherche d'Amazon basé sur le modèle CNN. Chiffre. Technologie. Appl. 2022, 40, 164-166.
2. Guo, Q. ; Li, S. ; Bao, H. Recherche sur la fonction d'évaluation du jeu informatique d'Amazon. Calculer. Ing. Appl. 2012, 48, 50-54.
3. Guo, T. ; Qiu, H. ; Tong, B. ; Wang, Y. Optimisation et comparaison de plusieurs algorithmes de jeu chez Amazons. Dans les actes de la Conférence chinoise sur le contrôle et la décision (CCDC) 2019, Nanchang, Chine, du 3 au 5 juin 2019 ; pp. 6299-6304.
4. Quan, J. ; Qiu, H. ; Wang, Y. ; Li, F. ; Qiu, S. Application des technologies UCT pour les jeux informatiques d'Amazon. Dans les actes du Conférence chinoise sur le contrôle et la décision (CCDC) 2016, Yinchuan, Chine, 28-30 mai 2016 ; pages 6896 à 6899.
5. Ju, J. ; Qiu, H. ; Wang, F. ; Wang, X. ; Wang, Y. Recherche sur l'optimisation des threads et l'ouverture d'une bibliothèque basée sur PVS parallèle Algorithme chez les Amazones. Dans les actes de la 33e Conférence chinoise sur le contrôle et la décision (CCDC) 2021, Kunming, Chine, 28-30 mai 2021 ; pages 2207 à 2212.
6. Li, Z. ; Ning, C. ; Cao, J. ; Li, Z. Amazones Basé sur l'algorithme hybride UCT-PVS. Dans les actes du 33e contrôle chinois 2021 et décision (CCDC), Kunming, Chine, 28-30 mai 2021 ; pages 2179 à 2183.
7. Ding, M. ; Bo, J. ; Qi, Y. ; Fu, Y. ; Li, S. Conception du système de jeu Amazons basé sur l'apprentissage par renforcement. Dans les actes du Conférence chinoise sur le contrôle et la décision (CCDC) 2019, Nanchang, Chine, 3-5 juin 2019 ; pages 6337 à 6342.
8. Tong, B. ; Qiu, H. ; Guo, T. ; Wang, Y. Recherche et application du calcul parallèle de l'algorithme PVS basé sur Amazon Jeu Homme-Machine. Dans les actes de la Conférence chinoise sur le contrôle et la décision (CCDC) 2019, Nanchang, Chine, 3-5 juin 2019 ; pp. 6293-6298.
9. Chen, X. ; Yang, L. Recherche sur la fonction d'évaluation chez les Amazones. Calculer. Connaître. Technologie. 2019, 15, 224-226.
10. Wang, C. ; Ding, M. Conception d'interface et mise en œuvre d'Amazones personnalisées. Intell. Calculer. Appl. 2017, 7, 78-80.
11. Zhang, L. Recherche sur le système de jeu Amazon basé sur l'algorithme de recherche Minimax. Mémoire de maîtrise, Université du Nord-Est, Shenyang, Chine, 2010.
12. Métropole, N. ; Ulam, S. La méthode de Monte Carlo. Confiture. Stat. Assoc. 1949, 44, 335-341. [Référence croisée] [Pub Med]
13. Coulom, R. Opérateurs de sélectivité et de sauvegarde efficaces dans la recherche arborescente de Monte-Carlo. Dans Actes des ordinateurs et des jeux, Turin, Italie, 1er-5 avril 2007 ; pp. 72-83.
14. Gelly, S. ; Kocsis, L. ; Schönauer, M. ; Sébag, M. ; Argent, D. ; Szepesvari, C. ; Teytaud, O. Le grand défi de l'ordinateur Go : Monte Recherche et extensions de l'arborescence Carlo. Commun. ACM 2012, 55, 106-113. [Référence croisée]
15. Gelly, S. ; Silver, D. Recherche arborescente de Monte-Carlo et estimation rapide de la valeur d'action dans l'ordinateur Go. Artif. Intell. 2011, 175, 1856-1875. [Référence croisée]
16. Vinyals, O. ; Babuchkin, I. ; Czarnecki, WM; Mathieu, M. ; Dudzik, A. ; Chung, J. ; Choi, DH ; Powell, R. ; Ewalds, T. ; Georgiev, P. ; et coll. Niveau Grand Maître dans StarCraft II utilisant l'apprentissage par renforcement multi-agents. Nature 2019, 575, 350-354. [Référence croisée]
17. Argent, D. ; Huang, A. ; Maddison, CJ ; Guez, A. ; Sifré, L. ; van den Driessche, G. ; Schrittwieser, J. ; Antonoglou, I. ; Panneershelvam, V. ; Lanctot, M. ; et coll. Maîtriser le jeu de Go avec les réseaux de neurones profonds et la recherche arborescente. Nature 2016, 529, 484-489. [Référence croisée]
18. Berner, C. ; Brockman, G. ; Chan, B. ; Cheung, V. ; D'ebiak, P. ; Dennison, C. ; Farhi, D. ; Fischer, Q. ; Hashme, S. ; Hesse, C. ; et coll. Dota 2 avec un apprentissage par renforcement profond à grande échelle. arXiv 2019, arXiv : 1912.06680.
19. Browne, CB; Powley, E. ; Maison Blanche, D. ; Lucas, SM ; Cowling, PI ; Rohlfshagen, P. ; Tavener, S. ; Pérez, D. ; Samothrakis, S. ; Colton, S. Une enquête sur les méthodes de recherche d'arbres de Monte Carlo. IEEETrans. Calculer. Intell. Jeux AI 2012, 4, 1–43. [Référence croisée]
20. Kloetzer, J. ; Iida, H. ; Bouzy, B. L'approche Monte-Carlo chez les Amazones. Dans les actes de l'atelier sur les jeux informatiques, Amsterdam, Pays-Bas, 15-17 juin 2007 ; pp. 185-192.
21. Shannon, CE Programmation d'un ordinateur pour jouer aux échecs. Londres. Edinb. Dublin Philos. Mag. J. Sci. 1950, 41, 256-275. [Référence croisée]
22. Chaslot, G. ; Bakkes, S. ; Szita, moi.; Spronck, P. Monte-Carlo Tree Search : un nouveau cadre pour l'IA de jeu. Dans les actes du Conférence AAAI sur l'intelligence artificielle et le divertissement numérique interactif, Californie, Californie, États-Unis, 22-24 octobre 2008 ; pp. 216-217.
23. Świechowski, M. ; Godlewski, K. ; Sawicki, B. ; Ma 'ndziuk, J. Monte Carlo Tree Search : Un examen des modifications et applications récentes. Artif. Intell. Rév.2023 , 56, 2497-2562. [Référence croisée]

Machine Translated by Google

24. Kloetzer, J. Monte-Carlo Livres d'ouverture pour les Amazones. Dans Ordinateurs et jeux : 7e Conférence internationale (CG 2010), Kanazawa, Japon, 24-26 septembre 2010, articles sélectionnés révisés 7 ; Springer : Berlin/Heidelberg, Allemagne, 2011 ; pp. 124-135.
25. Chanson, J. ; Müller, M. Un solveur amélioré pour le jeu des Amazones. IEEETrans. Calculer. Intell. Jeux AI 2014, 7, 16-27. [Référence croisée]
26. Zhang, G. ; Chen, X. ; Chang, R. ; Zhang, Y. ; Wang, C. ; Caution.; Wang, J. ; Xu, C. Maîtriser rapidement le jeu des Amazones en découplant l'apprentissage en réseau. Dans Actes de la Conférence internationale conjointe 2021 sur les réseaux de neurones (IJCNN), Shenzhen, Chine, 18-22 juillet 2021 ; p. 1 à 8.
27. de Koning, J. INVADER prolonge le titre d'Amazon. ICGA J. 2011, 34, 96. [CrossRef]
28. Kloetzer, J. INVADER remporte le tournoi des Amazones. ICGA J. 2009, 32, 112-113. [Référence croisée]
29. Lorentz, R. Invader remporte la huitième médaille d'or des Amazones. ICGA J. 2017, 39, 228-229. [Référence croisée]
30. Guo, Q. ; Li, S. ; Bao, H. La recherche d'un algorithme de recherche dans le jeu Amazons. Dans Actes de la Conférence chinoise sur le contrôle et la décision (CCDC) de 2011, Mianyang, Chine, 23-25 mai 2011 ; pages 1859-1862.
31. Li, X. ; Hou, L. ; Wu, L. Algorithme UCT dans les jeux homme-machine d'Amazon. Dans les actes du 26e contrôle chinois et Conférence de décision (CCDC), Changsha, Chine, 31 mai-2 juin 2014 ; pp. 3358-3361.
32. Quan, J. ; Qiu, H. ; Wang, Y. ; Li, Y. ; Wang, X. Étudiez les performances des algorithmes de recherche dans Amazons. Dans les actes du 27 Conférence chinoise sur le contrôle et la décision (CCDC), Qingdao, Chine, 23-25 mai 2015 ; pages 5811 à 5813.
33. Lieberum, J. Une fonction d'évaluation pour le jeu des Amazones. Théorique. Calculer. Sci. 2005, 349, 230-244. [Référence croisée]
34. Chai, Z. ; Croc, Z. ; Zhu, J. Stratégie d'optimisation de l'évaluation d'Amazon basée sur l'algorithme PSO. Dans les actes de l'année 2019 Conférence chinoise sur le contrôle et la décision (CCDC), Nanchang, Chine, 3-5 juin 2019 ; pages 6334 à 6336.
35. Soleil, Y. ; Yuan, D. ; Gao, M. ; Zhu, P. Accélération GPU de l'algorithme de recherche d'arborescence de Monte Carlo pour les Amazones et sa fonction d'évaluation. Dans les actes de la Conférence internationale 2022 sur l'intelligence artificielle, le traitement de l'information et le cloud computing (AIIPCC), Kunming, Chine, 21-23 juin 2022 ; pp. 434-440.
36. Kocsis, L. ; Szepesvari, C. ; Willemson, J. Recherche Monte-Carlo améliorée. Univ. Tartu Est. Technologie. Rapport 2006, 1, 1-22.
37. Finnsson, H. ; Björnsson, Y. Propriétés de l'arbre de jeu et performances du MCTS. IJCAI 2011, 11, 23-30.
38. Phatt, C. Blocs de construction de thread Intel®. J. Informatique. Sci. Coll. 2008, 23, 164-166.
39. Childs, ÊTRE ; Brodeur, JH; Kocsis, L. Transpositions et groupes de déplacement dans la recherche arborescente de Monte Carlo. Dans Actes du Symposium IEEE 2008 sur l'intelligence informatique et les jeux, Perth, Australie, 15-18 décembre 2008 ; pp. 389-395.
40. Van Eyck, G. ; Müller, M. Revisiter les groupes de mouvements dans la recherche arborescente de Monte-Carlo. Dans Advances in Computer Games : 13th International Conference, ACG 2011, Tilburg, Pays-Bas, 20-22 novembre 2011, articles sélectionnés révisés 13 ; Springer : Berlin/Heidelberg, Allemagne, 2012 ; p. 13-23.

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Standardisation des couleurs des images de solutions chimiques à l'aide Correspondance d'histogramme basée sur un modèle en profondeur Régression d'apprentissage

Patrycja Kwiek et Mał gorzata Jakubowska *

Faculté des Sciences des Matériaux et de la Céramique, Université AGH de Cracovie, al. Mickiewicza 30, 30-059 Cracovie, Pologne ;
pakwiek@agh.edu.pl * Correspondance : jakubows@agh.edu.pl

Résumé : La distorsion des couleurs dans une image présente un défi pour la classification et la régression par apprentissage automatique lorsque les données d'entrée sont constituées d'images. En conséquence, un nouvel algorithme de standardisation des couleurs des photos est proposé, constituant la base d'un modèle de régression de réseau neuronal profond. Cette approche utilise un modèle de couleur auto-conçu qui a été développé sur la base d'une première série d'études et d'imagerie numérique. En utilisant l'histogramme égalisé des canaux R, V, B du modèle numérique et de sa photo, une stratégie de mappage des couleurs a été calculée. En appliquant cette approche, les histogrammes ont été ajustés et les couleurs des photos prises avec un smartphone ont été standardisées.

L'algorithme proposé a été développé pour une série d'images dans lesquelles la surface entière conservait à peu près une couleur uniforme et les différences de couleur entre les photographies d'objets individuels étaient mineures. Cette approche optimisée a été validée dans la procédure de détermination colorimétrique de la vitamine C. L'ensemble de données pour le réseau neuronal profond dans la variante de régression a été constitué à partir de photos d'échantillons dans deux conditions d'éclairage distinctes. Pour la plage de concentrations de vitamine C de 0 à 87,72 µg·mL−1 , le RMSE pour l'ensemble de tests variait entre 0,75 et 1,95 µg·mL−1 , par rapport à la variante non standardisée, où cet indicateur était au niveau de 1,48–2,29 µg·mL−1 . La cohérence des résultats de concentration prévus avec les données réelles, exprimées en R2 , variait entre 0,9956 et 0,9999 pour chacune des variantes standardisées. Cette approche permet de supprimer les reflets lumineux sur les surfaces brillantes des solutions, problème courant dans les échantillons liquides. Cet algorithme de correspondance des couleurs a un caractère universel et son champ d'application n'est pas limité.

check for updates

Référence : Kwiek, P. ; Jakubowska, M. Standardisation des couleurs des produits chimiques Images de solutions utilisant la correspondance d'histogramme basée sur un modèle dans la régression d'apprentissage profond. Algorithmes 2024, 17, 335. https://doi.org/10.3390/a17080335

Rédacteur académique : Marcin Iwanowski

Reçu : 11 juillet 2024
Révisé : 29 juillet 2024
Accepté : 30 juillet 2024
Publié : 1 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

Mots-clés : standardisation des couleurs ; correspondance de modèles de couleurs ; Ajustement de l'histogramme RVB ; apprentissage profond ; détermination colorimétrique de la vitamine C

1. Introduction

Le développement de méthodes de mesure et de stratégies de recherche dans divers domaines scientifiques et technologiques est étroitement lié au soutien potentiel offert par les techniques d'apprentissage automatique (ML) et d'apprentissage profond (DL). Si les informations relatives au problème à résoudre sont communiquées au moyen d'images, de dessins, de croquis ou de photographies, l'utilisation de DL dans la modélisation des données constitue souvent l'approche optimale. L'application des réseaux de neurones dans l'apprentissage supervisé typique nécessite l'accès à de grands ensembles de données, ce qui fait de la méthode d'obtention et d'étiquetage des données une limitation importante. La technologie actuelle propose de nombreux appareils électroniques, tels que des caméras ou des scanners, qui peuvent être utilisés pour obtenir des données permettant de construire des modèles rapidement et sans qualifications particulières. Les informations peuvent souvent être codées simplement sous la forme d'objets, de relations entre les objets dans la zone analysée, de textures, de quantité de petits composants et de couleurs.

Plusieurs espaces colorimétriques sont largement utilisés dans différentes industries pour diverses applications [1]. L'espace RVB, basé sur les couleurs primaires de la lumière (rouge, vert et bleu), est largement utilisé dans les affichages et l'imagerie numériques. Le CMJN (cyan, magenta, jaune et noir) est utilisé dans l'impression couleur et s'appuie sur le modèle de couleur soustractif pour produire une large gamme de coul

Machine Translated by Google

sur papier. HSI/HSV (teinte, saturation, intensité/valeur) est souvent utilisé dans l'analyse d'images et la vision par ordinateur en raison de son alignement sur la perception humaine des couleurs. YUV et YCbCr sont utilisés dans la compression et la diffusion vidéo, séparant l'image en composants de luminance et de chrominance pour optimiser le stockage et la transmission des données. L'espace colorimétrique CIE Lab , développé par la Commission internationale de l'éclairage, fournit une représentation des couleurs perceptuellement uniforme, ce qui le rend idéal pour les tâches de différenciation et de transformation des couleurs . RAL CLASSIC est un système de correspondance des couleurs largement utilisé en Europe, composé de 213 couleurs standardisées, tandis que RAL DESIGN propose une palette plus étendue de 1 625 couleurs pour les designers et les architectes. RAL DIGITAL facilite une communication précise des couleurs dans les applications numériques grâce aux valeurs RVB et CMJN. Le PANTONE Matching System (PMS) est mondialement reconnu dans le domaine de la conception graphique, de l'impression et de la mode, fournissant des codes de couleurs standardisés pour garantir la cohérence entre les différents supports et matériaux. Chacun de ces espaces colorimétriques présente ses propres avantages et applications, répondant aux besoins spécifiques de diverses industries.

Si la couleur est le critère principal et décisif dans l'analyse d'images qui code les informations que nous recherchons, il est crucial de prendre des photos de manière à maintenir la qualité appropriée de ce paramètre. La vraie couleur de l'image permet l'interprétation correcte des diverses caractéristiques des objets photographiés. Une variabilité indésirable peut résulter de l'utilisation de plusieurs appareils pour capturer des images ou de la méthode d'acquisition d'images avec un seul appareil dans des conditions incontrôlées ou changeantes. Les facteurs importants incluent les changements dans les conditions d'éclairage, les reflets sur les surfaces brillantes et les variations telles que l' heure de la journée. Ces éléments peuvent ne pas avoir d'impact direct sur la qualité de la photo mais peuvent influencer sa valeur informative. Par conséquent, l'importance de la normalisation des couleurs en tant qu'étape cruciale du prétraitement des données doit être prise en compte. Diverses approches informatiques permettant de préserver ou de restaurer la vraie couleur des images au sein d'un ensemble de données pour l'apprentissage en profondeur sont décrites dans de nombreux ouvrages, proposant des solutions à ce problème.

Cette étude présente un algorithme original permettant de normaliser les couleurs des photos capturées avec un seul appareil photo de smartphone dans diverses conditions d'éclairage. Il prend également en compte la variabilité résultant de facteurs fortuits associés au processus de photographie et au manque inhérent de répétabilité parfaite de l'appareil photo. La recherche s'est concentrée sur un liquide homogène présentant une variation de couleur minimale entre les objets testés. L'intégralité de la photo utilisée dans le processus de modélisation partageait la même couleur, mais avec un certain niveau de variabilité aléatoire qui n'a pas pu être entièrement éliminée. Dans ce contexte, même de légers changements de couleur avaient une valeur informative significative, soulignant l'importance d'une interprétation précise des couleurs pour améliorer la qualité du modèle de régression d'apprentissage profond et minimiser les erreurs de prédiction.

Un algorithme de standardisation a été proposé, utilisant l'histogramme d'un modèle spécialement conçu, photographié dans des conditions contrôlées, ainsi que l'objet testé en relation avec l'histogramme du modèle préparé à l'aide d'une application graphique et sa photographie. Cette étude présente également la vérification de l'approche proposée pour la détermination colorimétrique de la vitamine C à l'aide d'un modèle de régression défini par DNN (réseaux de neurones profonds).

2. Travaux connexes

2.1. Standardisation des couleurs

Il existe de nombreuses catégories de problèmes d'étalonnage des couleurs, avec deux groupes fondamentaux liés soit à l'utilisation de différents appareils au sein d'une série de photos, soit aux variations des conditions d'éclairage affectant les sujets photographiés. La compensation matérielle des couleurs peut être obtenue en utilisant des capteurs de caméra qui ajustent automatiquement les paramètres en fonction de l'éclairage de la scène capturée. L'approche alternative implique le traitement numérique des photos à l'aide de diverses stratégies visant à obtenir des tons de couleurs cohérents sur plusieurs images. Pour ce faire, des tableaux d'étalonnage sont initialement utilisés, qui corrèlent et mappent les pixels individuels des photos capturées avec les valeurs standardisées des zones de couleur correspondantes sur le tableau [2,3]. Une couleur standard

Machine Translated by Google

La carte d'étalonnage sert d'outil de référence en photographie, création de films, enregistrement vidéo ou conception graphique, garantissant une reproduction précise et cohérente des couleurs dans les images. Typiquement, il comprend une grille de plages colorées de couleurs connues et standardisées. Ces graphiques sont appliqués lors de l'étalonnage des équipements et des ajustements pendant le post-traitement, garantissant ainsi la cohérence de la représentation des couleurs sur divers appareils et conditions. Dans le commerce, plusieurs cartes d'étalonnage sont disponibles, allant de celles comportant six à 100 pla Cependant, la mire d'étalonnage des patchs à 24 couleurs, souvent appelée Macbeth ColorChecker, a été largement utilisée dans de nombreuses applications [4]. De plus , le développement de tableaux d'étalonnage personnalisés joue un rôle important dans la standardisation des couleurs des images.

La normalisation des couleurs joue un rôle important dans de nombreux domaines scientifiques et technologiques, notamment le diagnostic médical, la bioanalyse, l'agriculture, l'analyse des preuves en sciences médico-légales, la chimie analytique, etc. Dans le traitement des images agricoles, l'accent est mis sur la liaison des modifications de la couleur des images , dans des zones d'intérêt désignées, avec des attributs de qualité particuliers tels que la santé des plantes, le stress des cultures et le niveau de maturité. Dans [5], une technique a été proposée pour calibrer les images numériques, garantissant leur uniformité et améliorant les comparaisons phénologiques en agriculture. La transformation des couleurs des images a été réalisée à l'aide d'une mire d'étalonnage standard de 24 couleurs. L'étalonnage fournit des résultats typiques et attendus en ajustant les valeurs RVB des taches de couleur dans l'image d'entrée aux valeurs RVB standard. Dans [6], un dispositif numérique de correspondance de teinte a été développé pour déterminer la couleur des dents, car une évaluation précise de la teinte est cruciale dans les procédures de restauration dentaire. Les enregistrements d'images réalisés avec une caméra intra-orale ont servi de base aux calculs et aux transformations.

L'algorithme de la machine à vecteurs de support (SVM) et l'espace colorimétrique CIEL*a*b* ont été utilisés pour évaluer la couleur des dents. La correction des couleurs était basée sur le calcul du coefficient euclidien distance entre les stores. Les auteurs ont créé un modèle dédié, distinct du modèle classique, avec un large spectre de couleurs, spécifiquement adapté à la gamme de variations de couleur des dents abordées dans l'étude. La prochaine proposition concerne l'amélioration de la précision de la reproduction des couleurs dans les appareils photo numériques et les smartphones. La première méthode utilise une correction de la balance des blancs suivie d'une transformation de l'espace colorimétrique. La deuxième approche, applicable aux images traitées hors ligne, est basée sur la balance complète des couleurs au lieu de la balance des blancs et nécessite l'utilisation de l'apprentissage automatique pour prédire la matrice de balance des couleurs [7]. L' article de synthèse [8] discute de l'utilisation de divers algorithmes de reconnaissance d'images histopathologiques. Les différences de couleur en histopathologie résultent de l'utilisation de divers équipements ou techniques de coloration des tissus. En règle générale, la normalisation des couleurs implique le transfert de la couleur moyenne de l'image cible vers l'image source et la séparation de la tache présente d L'article décrit un ensemble complet de méthodes de normalisation des couleurs qui ont été vérifiées à l'aide d'images histopathologiques, notamment celles obtenues lors du diagnostic du cancer du sein et du côlon.

En raison du développement de la technologie d'imagerie numérique et, par conséquent, de l'émergence de nombreuses applications entièrement nouvelles, des solutions basées sur les images des smartphones sont proposées dans divers domaines. Le tableau 1 contient une liste d'articles sélectionnés dans lesquels les images couleur servent de supports d'informations importants et où la correction et la standardisation des couleurs ont un impact significatif sur les résultats de modélisation et la qualité des solutions propo

Tableau 1. Comparaison des travaux connexes.

Réf.	Année	Problème résolu	Ensemble de données utilisé	Stratégie de correction des couleurs
[9]	2022	Le concept général de constance relative des couleurs (RCC), c'est-à-dire la capacité d'aligner les couleurs des mêmes objets entre des images indépendamment de l'éclairage et de la caméra.	Trois ensembles de données d'images collectées par plusieurs caméras sous diverses conditions d'éclairage et d'exposition (Belfast, Middlebury, Gehler-Shi)	Trois étapes du cadre d'alignement des couleurs : étalonnage de la réponse de la caméra avec des patchs de couleur, linéarisation de la réponse, intensité des couleurs par pixel et correspondance de chromaticité

Machine Translated by Google

Tableau 1. Suite

Réf.	Année	Problème résolu	Ensemble de données utilisé	Stratégie de correction des couleurs
[10]	2022	Variations de couleur au sein des cliniques images d'échantillons de prothèses maxillo-faciales en silicone (approche in vitro)	Images des échantillons de silicone prothétique pigmentés dans différentes ambiances conditions d'éclairage	Étalonnage informatisé de la balance des blancs après traitement (PPWBC) basé sur un logiciel utilisant une carte grise et un nuancier Macbeth
[11]	2023	Conservation non invasive du patrimoine culturel dans un contexte de dégradation matérielle	Quatre séries d'images dans différentes conditions d'éclairage, acquises avec une caméra de téléphone portable classique dans l'espace colorimétrique sRGB	Une fonction de transfert adaptatif, fonctionnant avec une charte d'étalonnage des couleurs créée par vous-même ; réglage final des fonctions d'étalonnage basé sur un réseau neuronal
[12]	2023	Colorimétrie urinaire sur smartphone	Échantillons d'urine de 58 patients photographiés dans une boîte photo personnalisée, dans cinq conditions d'éclairage simulées, à l'aide de cinq smartphones dans l'espace colorimétrique RVB	Calibrage des couleurs avec Nuancier SpyderCHEKRTM et algorithmes propriétaires d'Adobe Photoshop pour mettre à l'échelle les couleurs non corrigées. Valeurs RVB
[13]	2023	Une solution basée sur smartphone pour des mesures colorimétriques précises, avec un système de guidage en réalité augmentée	Images simulées et réelles, images de kits de bandes colorées pour la lecture du pH	Un algorithme de correction des couleurs avec le modèle de régression spatiale variable de premier ordre qui exploite à la fois l'ampleur et l'échelle absolues des couleurs ; le système comprend un tableau de référence de couleurs
[14]	2023	Évaluation du processus de cicatrisation des plaies chroniques	Images de tissus de plaie prises à différents moments, dans différentes conditions d'éclairage, distances, objectifs et caméras de smartphone dans l'espace colorimétrique RVB	Images de tissus de plaie prises avec un Tableau Macbeth composé de 24 carrés de couleur et de quatre marqueurs ArUco aux coins du tableau ; calcul de la matrice de transformation à l'aide du Matrice inverse de Moore – Penrose et la matrice cible
[15]	2024	Un algorithme à usage général	Photos de paysages naturels et urbains, de végétation et d'objets fabriqués par l'homme dans l'espace colorimétrique RVB	Une transformation matricielle utilisation du produit tensoriel B-splines avec optimisation du paramètre de lissage λ et du nombre de fonctions de base spline
[16]	2024	Reproduction de la couleur de référence des vêtements textiles en sélectionnant des colorants dispersés appropriés	Résultats de mesures colorimétriques à l'aide d'un spectrophotomètre de type Spectraflash 600 Plus (Datacolor International, USA)	Un algorithme génétique pour optimiser la sélection et la concentration des colorants pour une correspondance exacte des couleurs dans les filaments bicomposants (PET/PTT)
[17]	2024	Optimisation de l'apparence des aliments d'origine végétale étant donné que les produits d'origine animale peuvent contenir une large gamme de pigments et d'éléments structurels différents qui diffusent la lumière	Valeurs L*, a*, b* des produits d'origine animale et des émulsions à base de plantes produites, mesurées par colorimètre (ColorFlex EZ 45/0-LAV, Hunter Associates Laboratory Inc., Reston, VA, USA)	Le modèle de correspondance des couleurs basé sur Théorie de Kubelka-Munk, utilisée pour calculer la réflectance spectrale des émulsions mixtes à partir du réflectance spectrale des émulsions colorées individuelles
[18]	2024	Conception de couleurs pour la mode : sélection des couleurs et correspondance de différentes combinaisons de couleurs	Objets graphiques conçus en CorelDraw	Correspondance des couleurs basée sur un algorithme génétique interactif utilisant des images de référence

Machine Translated by Google

2.2. Détermination de la vitamine C

Les techniques de standardisation des couleurs sont cruciales en chimie analytique, notamment dans la détermination colorimétrique de différentes substances. Cela est particulièrement évident dans les méthodes dans lesquelles les solutions des composés testés sont photographiées et les informations sur les concentrations sont dérivées de ces photos à l'aide de divers outils statistiques et d'une analyse multivariée. Une illustration d'une telle méthodologie est la détermination colorimétrique de la vitamine C.

L'acide ascorbique (AA), c'est-à-dire la vitamine C, est un composé naturel hautement soluble dans l'eau, qui joue un rôle dominant dans la santé humaine et est essentiel pour une variété de réactions métaboliques et de processus biologiques, avec une forte activité antioxydante, limitant les radicaux libres. dommage. Une concentration insuffisante de vitamine C dans l'organisme, entre autres, provoque le scorbut et l'anémie et est associée à certaines anomalies psychologiques, par exemple la dépression.
Une surconsommation de vitamine C peut avoir des conséquences négatives, comme empêcher l'absorption de la vitamine B12 et potentiellement causer des problèmes gastro-intestinaux ou des calculs rénaux [19,20]. Comme les humains sont incapables de produire de la vitamine C en interne, il est essentiel de compléter avec une quantité régulée d'AA, généralement provenant de fruits et légumes frais [21,22].

La chimie analytique propose de nombreuses stratégies pour déterminer les AA dans les produits alimentaires et les suppléments de vitamine C. Traditionnellement, des techniques d'analyse instrumentale ont été utilisées pour mesurer les niveaux d'acide ascorbique, offrant une large plage de concentrations avec une faible limite de détection. Ces méthodes comprennent le titrage direct [23], la chromatographie liquide ultraperformance (UPLC), la chromatographie liquide haute performance (HPLC) [24,25], la spectrométrie infrarouge à transformée de Fourier (FT-IR) [26], la chimiluminescence [27], Spectroscopie Raman [28], spectroscopie proche infrarouge (NIR), spectroscopie FT-Raman [29] et électrophorèse capillaire [30]. De plus , diverses méthodes électrochimiques ont été utilisées pour déterminer l'AA, offrant une sensibilité, une sélectivité, une reproductibilité et une fiabilité extrêmement élevées [31-33]. Cependant, les méthodes d'analyse instrumentale chimique mentionnées nécessitent des salles de laboratoire spécialisées, des instruments coûteux, de nombreux réactifs auxiliaires, des étapes d'extraction fastidieuses et des procédures de mesure compliquées qui nécessitent un personnel hautement qualifié pour des tâches telles que la préparation de réactifs et d'échantillons, la conduite de recherches ou l'interprétation. résultats.

Le développement de nouvelles méthodologies qui remplacent ou soutiennent les stratégies de laboratoire traditionnelles avec des instruments rentables répond au besoin d' approches hautement efficaces et adaptables en matière d'analyse alimentaire, pharmaceutique et environnementale. Les smartphones populaires et largement disponibles offrent une plate-forme analytique prometteuse caractérisée par une automatisation complète, une grande capacité de mémoire locale, un accès au stockage cloud externe et des caméras particulièrement de haute qualité, ce qui les rend adaptés à une large utilisation comme appareils de mesure po
Par rapport aux méthodes conventionnelles, les mesures basées sur smartphone utilisent l'imagerie numérique, offrant de nombreuses applications dans divers domaines et des avantages significatifs. Ceux-ci incluent un faible coût, un fonctionnement simple, la possibilité d'installer divers capteurs dédiés, des mesures précises en temps réel, un caractère non destructif des tests dans de nombreux cas, une analyse rapide et la portabilité vers d'autres appareils similaires. En chimie analytique, des appareils couramment disponibles, tels que des caméras, des webcams, des scanners et des smartphones, sont utilisés pour enregistrer des images de réactions colorimétriques dans lesquelles l'intensité de la couleur résultante est directement corrélée à la concentration de l'analyte. La littérature fournit diverses descriptions de l'utilisation de méthodes d'imagerie numérique pour déterminer l'AA. Le travail [34] présente une technique simple, rapide et rentable utilisant l'analyse d'images sur smartphone pour quantifier l'AA dans une solution aqueuse. L'étude était basée sur l'évaluation de l'effet perturbateur de l'AA dans la détection colorimétrique enzymatique du glucose. Une autre approche a été suggérée, impliquant la colorimétrie d'images numériques pour déterminer la teneur en vitamine C des jus de fruits naturels. Cette méthode consiste à mesurer la couleur du complexe fer(II)-o-phénanthroline [35]. Dans l'article suivant, les auteurs ont développé un nouveau spectromètre à dichroïsme circulaire basé sur des smartphones pour une analyse colorimétrique hautement sensible et ultraportable, offrant les avantages de rentabilité et de simplicité. Ce spectromètre utilisait un modèle de couleur HSV et déterminait l'absorbance en additionnant la valeur V des positions adjacentes pour calculer l'intensité globale. La réaction nanofeuillet TMB-MnO2 a également été utilisée et, en effet, un système très sensible et spécifique a été proposé pour la détection des vitamines.

Machine Translated by Google

C [36]. Dans [37], un capteur au point d'intervention a été développé, composé d'une puce de papier fluorescente, de pièces imprimées en 3D et d'un smartphone. Ce système a permis une détection visuelle en temps réel quantitative et très sensible des AA. La conception de cet appareil impliquait de placer un capteur fluorescent sur une plateforme de détection portable connectée à un smartphone.

La méthode de mesure analytique utilisée dans ce travail était basée sur les propriétés réductrices de la vitamine C. AA réduit le Fe(III) en Fe(II) puis forme un complexe avec l'o-phénanthroline. Le complexe rouge intense de Fe(II) et d'o-phénanthroline est largement utilisé pour la détermination colorimétrique et comme indicateur [38]. Cette réaction se caractérise par une vitesse, une stabilité et une reproductibilité élevées. L'utilisation de cette approche basée sur une caméra associée à l'imagerie numérique des composés colorés améliore la précision et réduit le coût de mesure de la teneur en vitamine C.

2.3. Algorithmes de régression multivariée

Les techniques de régression sont largement utilisées comme méthode d'apprentissage supervisé, impliquant la modélisation de la relation entre les caractéristiques et la variable cible pour aborder des tâches visant à prédire des valeurs continues. Les algorithmes typiques utilisés à cette fin sont PCR (régression en composantes principales) [39], PLSR (régression des moindres carrés partiels) [40, 41] et SVR (régression vectorielle de support) [42]. DNN, avec diverses architectures qui cartographient la structure non linéaire des données multivariées, peut également être appliqué aux problèmes de régression [43,44]. Les réseaux de neurones convolutionnels (CNN) sont un type d'algorithme d'apprentissage profond dont le fonctionnement est basé sur une opération mathématique appelée convolution. Ils se composent de plusieurs couches de filtres convolutifs qui apprennent automatiquement les représentations hiérarchiques des caractéristiques directement à partir des données brutes de pixels, ce qui les rend très efficaces pour des tâches telles que la classification ou la segmentation d'images et la détection d'objets. Généralement, la qualité des modèles de régression est évaluée à l'aide des carrés moyens des erreurs entre les valeurs réelles et les prédictions (MSE) ou de l'erreur absolue moyenne (MAE).

MSE et MAE sont des mesures d'erreur différentes utilisées pour évaluer les performances des modèles prédictifs. MSE est plus sensible aux erreurs importantes car elles sont au carré, tandis que MAE fournit une compréhension plus intuitive de l'erreur moyenne et est exprimée dans l'unité de la variable cible. Ces scores utilisés comme fonctions de perte lors de l'entraînement du réseau neuronal minimisent la différence moyenne entre les valeurs prédites et réelles. Les performances d'entraînement et de prédiction du modèle d'étalonnage sont également évaluées sur la base des scores suivants : l'erreur quadratique moyenne d'étalonnage (RMSEC) et de prédiction (RMSEP), et le coefficient de détermination R2 pour l'étalonnage et la prédiction [45]. Puisque le résultat de la prédiction devrait correspondre à la valeur réelle, les points avec des coordonnées (réelles, prédites) devraient idéalement se trouver sur la ligne y = x. Cependant, en raison de la variabilité aléatoire et de l'ajustement imparfait du modèle aux données, deux conditions doivent être remplies : la valeur un doit se situer dans l'intervalle de confiance de la pente, et une valeur de zéro doit se situer dans l' intervalle de confiance de l'ordonnée à l'origine [46].

3. Méthode proposée 3.1.

Algorithme de correspondance des couleurs

Dans ce travail, l'objectif principal était de développer une approche de détermination de la vitamine C basée sur le changement de couleur de la solution influencé par sa concentration. Par conséquent, une étape cruciale pour obtenir un modèle de régression de haute qualité consistait à effectuer la procédure de standardisation des couleurs sur les échantillons de vitamine C photographiés. L'algorithme de standardisation des couleurs proposé a été divisé en deux parties. Dans la première étape, une correspondance d'histogramme standard a été réalisée entre un modèle de couleur numérique et une photographie d'un modèle imprimé obtenu dans des conditions d'éclairage données. Dans la deuxième étape, une correspondance d'histogramme a été réalisée entre une photo de la solution (échantillon testé) prise sous le même éclairage que le motif et le modèle de référence. En fin de compte, 512 carrés aléatoires plus petits ont été découpés dans l'image reconstruite afin d'obtenir un ensemble de données suffisamment grand pour l' entraînement ultérieur du réseau neuronal de régression. Le nombre de petits carrés sélectionnés au hasard a été optimisé, car les calculs préliminaires ont montré que les données étaient insu

Machine Translated by Google

aboutissait à une qualité de modèle inférieure. Il a également fallu franchir une étape supplémentaire consistant à adapter le modèle de référence auto-conçu à la plage de valeurs appropriée. Cette opération a abouti au développement d'un algorithme d'appariement d'histogrammes adapté.

La correspondance d'histogramme est une opération dans laquelle les pixels d'une image d'entrée sont manipulés de manière à permettre à son histogramme de correspondre à l'histogramme de l'image de référence. Dans ce travail, l'espace colorimétrique RVB a été choisi pour les calculs. Il est largement disponible et populaire, car la plupart des caméras et appareils de capture d'images enregistrent les données directement au format RVB, éliminant ainsi le besoin de transformations supplémentaires de l'espace colorimétrique. L'espace RVB est également simple à mettre en œuvre, étant intuitif et directement lié aux propriétés physiques de la lumière, ce qui simplifie le développement d'algorithmes. De plus, dans notre application spécifique, le RVB s'est avéré donner des résultats satisfaisants. Par conséquent, malgré les avantages des autres espaces colorimétriques, le RVB était le plus approprié pour des raisons pratiques et techniques. Dans le cas des images couleur, chaque canal (RVB) a été considéré indépendamment. Dans le contexte du traitement d'images, il s'agissait d'une tâche de normalisation, associée par exemple à un changement d



Figure 1. Exemple de correspondance d'histogramme à des fins d'illustration.

En termes mathématiques, la correspondance d'histogramme peut être représentée comme une transformation d'image telle que la fonction de distribution cumulative (CDF) des valeurs dans chaque bande correspond à la CDF des bandes dans une autre image.

Pour faire correspondre les histogrammes de deux images, il est nécessaire d'égaliser leurs histogrammes. Après cela, chaque pixel de la première image est modifié et mappé sur la deuxième image. Cette opération a été réalisée pour chacune des voies (R, V, B) d'un gabarit photographié dans une condition d'éclairage donnée (lumière naturelle et lumière artificielle) et pour les voies d'un gabarit numérique. L'ensemble du processus de transformation des couleurs de l'image est représenté dans la figure 2.

Afin de normaliser les images des solutions chimiques testées, la plage correcte de valeurs RVB a été sélectionnée à partir des histogrammes de la photo de l'échantillon testé. Si la valeur de l'histogramme d'un échantillon était supérieure ou égale à 1% de la valeur maximale de l'histogramme de la solution chimique, la valeur de l'histogramme n'était pas modifiée, sinon elle était fixée à 0. L'écart du nombre de pixels était neutralisé par leur distribution aléatoire uniforme sur toute la plage de valeurs. La dernière étape de l'algorithme de standardisation consistait à faire correspondre les histogrammes d'une image d'une solution avec les histogrammes modèles adaptés.

Cela a été réalisé de la même manière que décrit sur la figure 3, suivi de la reconstruction d' une image à partir des canaux R, V et B.

Machine Translated by Google

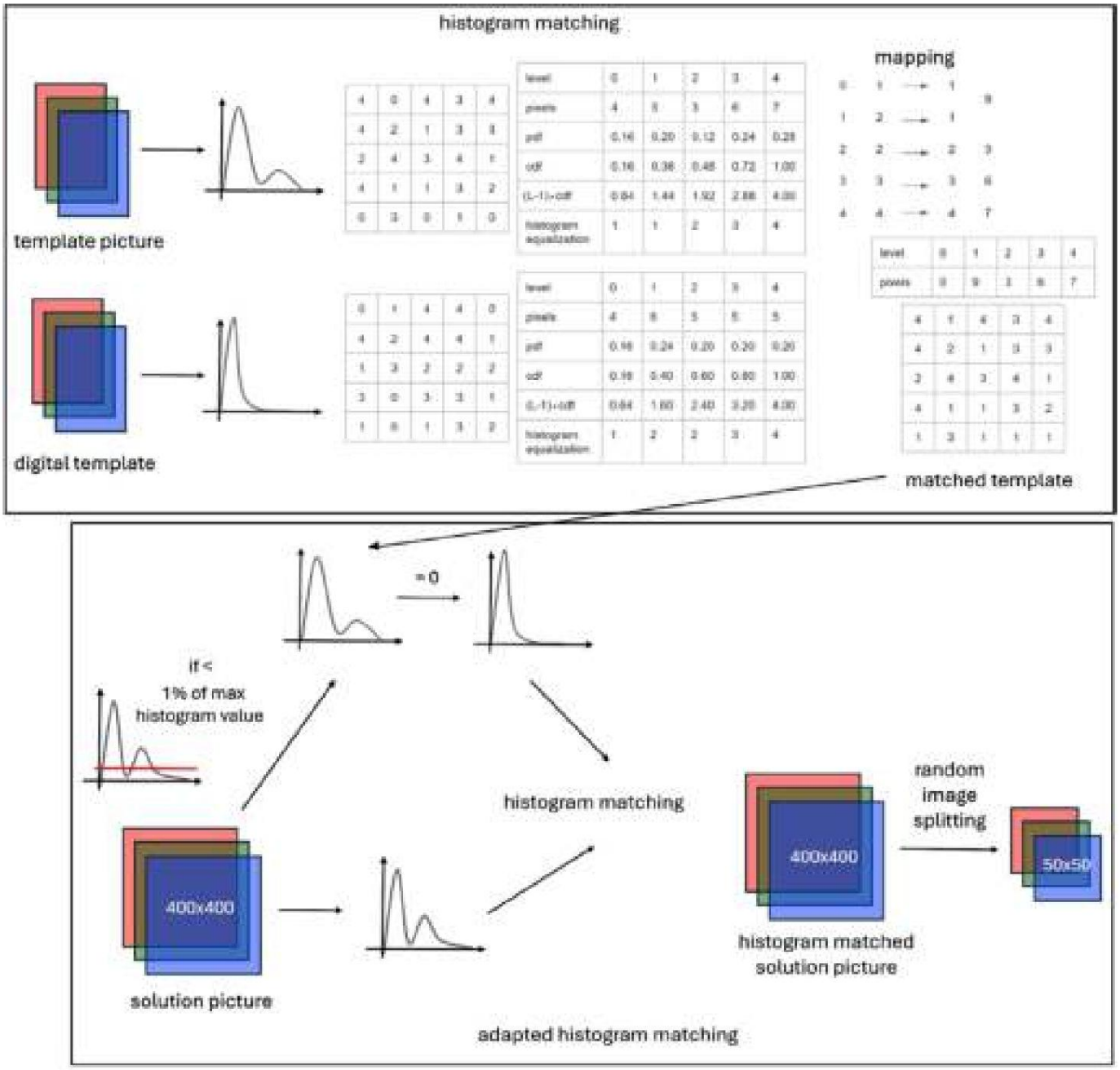


Figure 2. Schéma de l'algorithme de correspondance des couleurs.



l'histogramme de la solution chimique, la valeur de l'histogramme n'a pas été modifiée, sinon elle a été

répartition sur toute la plage de valeurs. La dernière étape de la normalisation

Figure 3. Solutions de vitamine C (0, 4,54, 11,31, 18,05, 22,52, 33,63, 44,64, 66,37 et 87,72 µg·mL−1) en présence d'eau bidistillée, de tampon acétate, de sulfate d'ammonium et de fer(III), et orthophénanthroline.

Machine Translated by Google

3.2. Logiciel

Tous les calculs ont été effectués en Python 3.10.0. Les bibliothèques et packages suivants ont été utilisés : opencv 4.8.1, scikit-image 0.22.0, numpy 1.26.1, matplotlib 3.5.3, scikit-learn 1.0.2, tensorflow 2.9.1 et keras 2.10.0.

4. Résultats et discussion 4.1.

Réactifs, équipement de laboratoire et procédure de mesure Dans ce travail,

la réaction colorimétrique utilisée pour déterminer l'acide ascorbique (AA) était basée sur la réduction de Fe(III) en Fe(II) par AA, qui forme un complexe de couleur rouge avec l'orthophénanthroline. Dans cette expérience, acide L(+)-ascorbique pa pur, CAS : 50-81-7 (POCH, Gliwice, Pologne), 1,10 phénanthroline pa, CAS : 66-71-7 (POCH, Pologne) et ammonium du sulfate de fer (III) dodécahydraté pur pa, CAS : 7783-83-7 (POCH, Pologne) a été utilisé. Le tampon acétate 1 M a été préparé dans notre laboratoire en mélangeant 1,0 mol·L (CH3COOH) avec 1,0 mol·L d'acide acétique (CH3COONa) (les deux réactifs achetés chez Avantor $^{-1}$ acétate de sodium Performance Materials, Gliwice, Pologne^{-1}) et en ajustant au pH souhaité à l'aide de 10N. HCl. Toutes les solutions, y compris AA à une concentration de 2,5 gL−1 , la phénanthroline à une concentration de 4,0 gL−1 , le sulfate d'ammonium et de fer (III) à une concentration de 2,5 gL−1 et un tampon acétate 1 M (pH 4,6), ont été préparé avec de l'eau bidistillée. De petits équipements de laboratoire ont été utilisés dans les expériences, notamment une balance analytique précise (RADWAG, modèle AS 60/220.XS, Radom, Pologne), un agitateur magnétique (WIGO, Toszek, Pologne), des pipettes automatiques de différents volumes (Eppendorf , Allemagne) et la verrerie, c'est-à-dire les béchers et les boîtes de Pétri. Pour ajuster les valeurs de pH de la solution tampon, nous avons utilisé le pH-mètre de laboratoire SevenCompact S210 (Mettler Toledo, Greifensee, Suisse). Toutes les expériences ont été réalisées à température ambiante.

Des solutions avec différentes concentrations de vitamine C pour une évaluation directe de la couleur ont été préparées en ajoutant d'abord 9 ml d'eau distillée dans les béchers, puis 1 ml de tampon acétate, puis le volume approprié d'acide ascorbique et 0,5 ml de sulfate d'ammonium et de fer (III). .
Après 3 minutes d'attente, 0,5 ml de phénanthroline ont été ajoutés. Les ajouts d'AA étaient de 0, 20, 50, 80, 100, 150, 200, 300 et 400 µL. La figure 3 montre la photo des échantillons préparés pour des tests ultérieurs. Les paramètres étudiés pour l'optimisation étaient la concentration du réactif, le volume et le temps entre les différentes étapes (influence de la cinétique chim
Après avoir attendu environ 20 minutes, 10 ml de chaque solution colorée ont été versés dans des boîtes de Pétri à fond plat et des photos ont été prises avec l'appareil photo d'un smartphone.

4.2. Préparation d'un modèle de couleur et acquisition d'images

L'expérience a été divisée en plusieurs étapes et la procédure est illustrée à la figure 4.

La première étape a été la préparation du modèle pour la normalisation des couleurs. Pour déterminer quelles couleurs doivent être utilisées comme référence, une série de solutions de vitamine C ont été préparées couvrant la plage de concentrations considérée dans cette étude. Cela a été fait pour estimer la plage de valeurs RVB significatives pour ces calculs. Des photos des solutions testées ont été prises et, à l'aide d'un logiciel maison dédié, 12 configurations de valeurs RVB (couleurs) ont été choisies pour être utilisées sur le modèle. Les valeurs RVB moyennes ont servi de base pour déterminer la gamme de couleurs dans le modèle dédié au problème à résoudre. Pour élargir légèrement la palette de couleurs, des valeurs supérieures et inférieures supplémentaires ont été estimées et le modèle a été créé pour contenir 12 couleurs. Le modèle, chaque couleur étant représentée par un carré, a été imprimé en plusieurs exemplaires avec une seule imprimante. La taille de chaque carré était de 180 ×180 pixels et la taille totale du modèle de 600 ×790 pixels.

Machine Translated by Google

tt

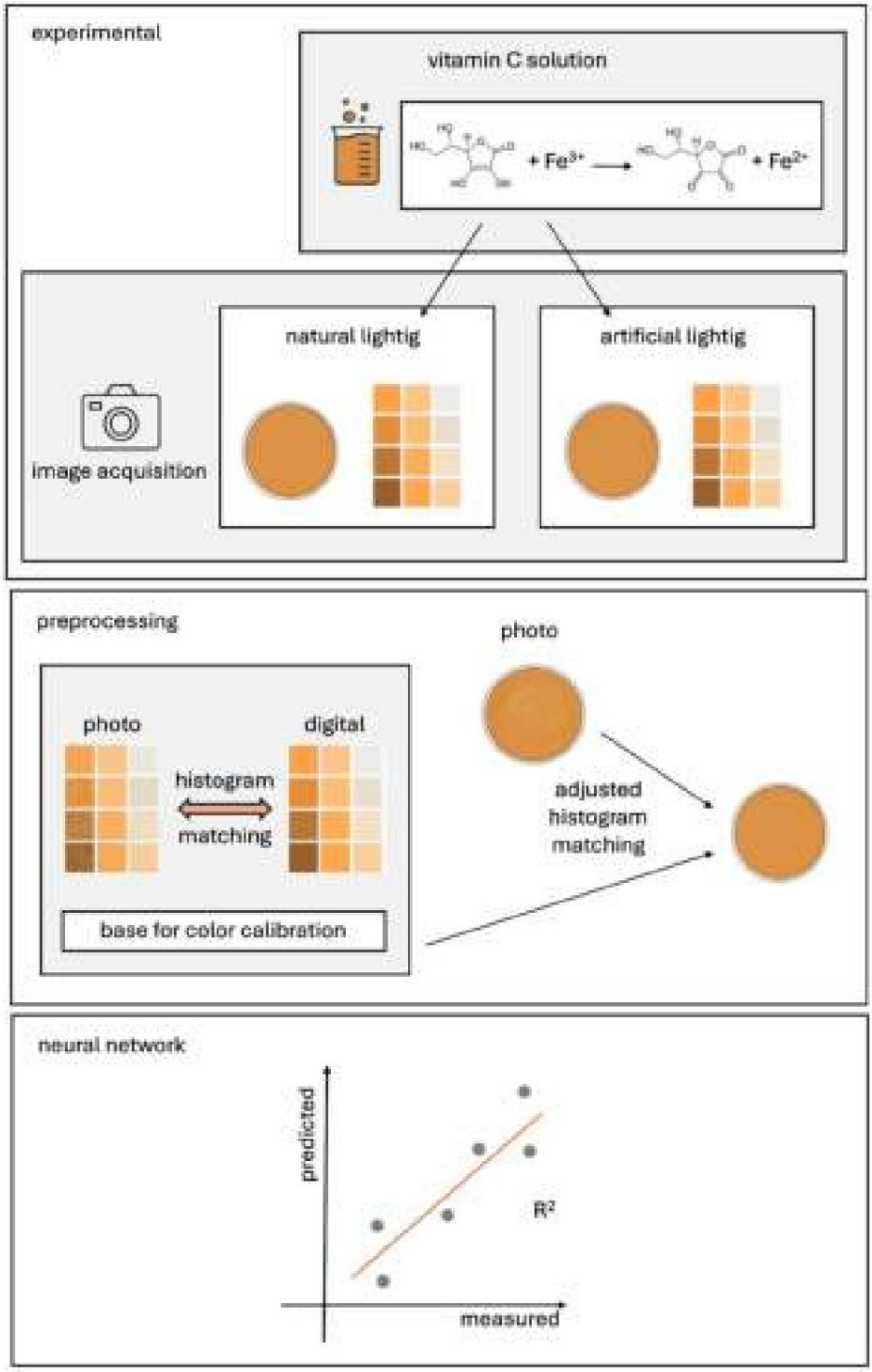


Figure 4. Schéma des expériences et calculs menés : préparation de solutions AA à concentrations variables ; application de solutions sur boîtes de Pétri ; capture d'images dans différentes conditions d'éclairage (solutions AA avec modèles de couleurs) ; prétraitement de l'image : normalisation des couleurs ; formation et évaluation du réseau de régression.

ff

Le processus de prise de photos impliquait de placer 10 ml de la solution dans une boîte de Pétri sur fond blanc (en soulignant le contraste entre le fond et la boîte de Pétri contenant la solution), à côté du modèle (qui servait de jeu de contrôle des couleurs) sur une surface noire. Les photos ont été prises dans deux conditions : l'éclairage naturel et la lumière de la torche d'un smartphone. Dans la deuxième variante, une boîte entièrement

Machine Translated by Google

isolant les objets photographiés de l'environnement a été placé sur le dispositif expérimental pour couper l'accès à la lumière naturelle. Des différences de couleurs des solutions et des nuanciers photographiés ont été observées sous ces deux illuminations (Figure 5). Le processus a été répété 11 fois pour chaque échantillon avec une concentration différente de vitamine C Toutes les photos ont été prises avec un smartphone Xiaomi (Xiaomi Redmi Note 9, Xiaomi, Pékin, Chine), doté d'un appareil photo de 48 MP avec une ouverture f/1,79. La résolution des images était de 3984 ×1840 px.

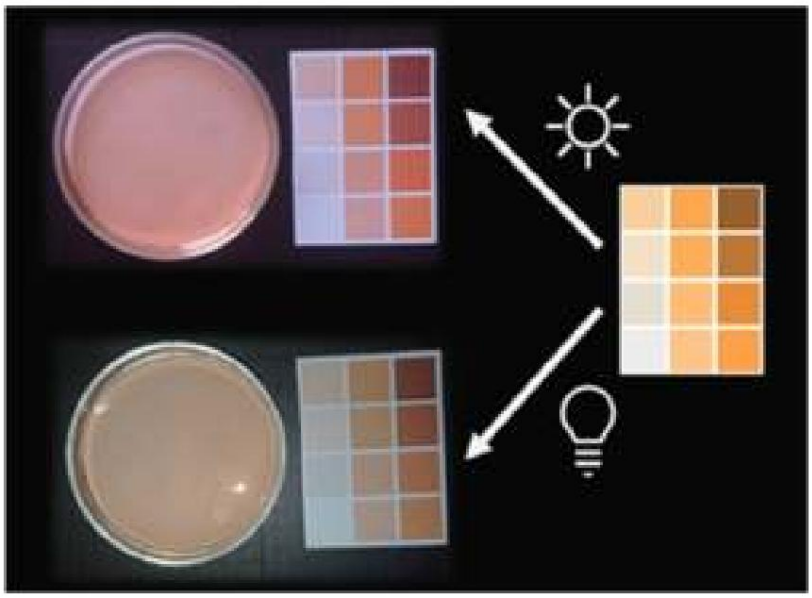


Figure 5. Contrôle des conditions d'éclairage : deux approches : lumière naturelle (en haut) et éclairage par lampe de smartphone (en bas). Côté droit d'une image : modèle numérique utilisé pour la correspondance d'histogramme ; les flèches indiquent les conditions d'éclairage dans lesquelles les images ont été prises.

Pour faire correspondre l'histogramme aux modèles utilisés dans la normalisation, des photos supplémentaires ont été prises du modèle dans les deux conditions. Les modèles ont été découpés à partir des images et une correspondance d'histogramme a été effectuée entre les modèles résultants et celui d'origine. Le processus avec les résultats est visible sur la figure 6. Chaque modèle est présenté sous forme d' image RVB, avec ses histogrammes pour chacun des canaux R, V et B. Cette figure illustre clairement le fonctionnement de notre algorithme. La première colonne montre l'histogramme idéal du modèle de couleur numérique conçu pour cette tâche. Les histogrammes de couleurs des photos modèles doivent correspondre aux caractéristiques idéales. Cependant, en conditions réelles, il existe des différences significatives (deuxième et quatrième colonnes) entre les histogrammes des photos et ceux de la version numérique. Cependant, après avoir appliqué l'algorithme proposé dans ce travail, la forme attendue de l'histogramme des couleurs pour les trois composantes R, V et B peut être reproduite. Les graphiques des troisième et cinquième colonnes correspondent à ceux de la premièr

Les histogrammes cumulés des images correspondantes sont les mêmes que l'histogramme cumulé du modèle numérique (qui sert de référence) pour chaque canal. Chaque pointe sur un histogramme donné correspond à un composant d'une certaine couleur sur le m L'apparence moins discrète des histogrammes des images modèles d'origine est due à la variance de l'impression et à différentes variables qui affectent la façon dont la caméra perçoit l' objet avec certains changements (par exemple, les ombres et les hautes lumières).

Machine Translated by Google

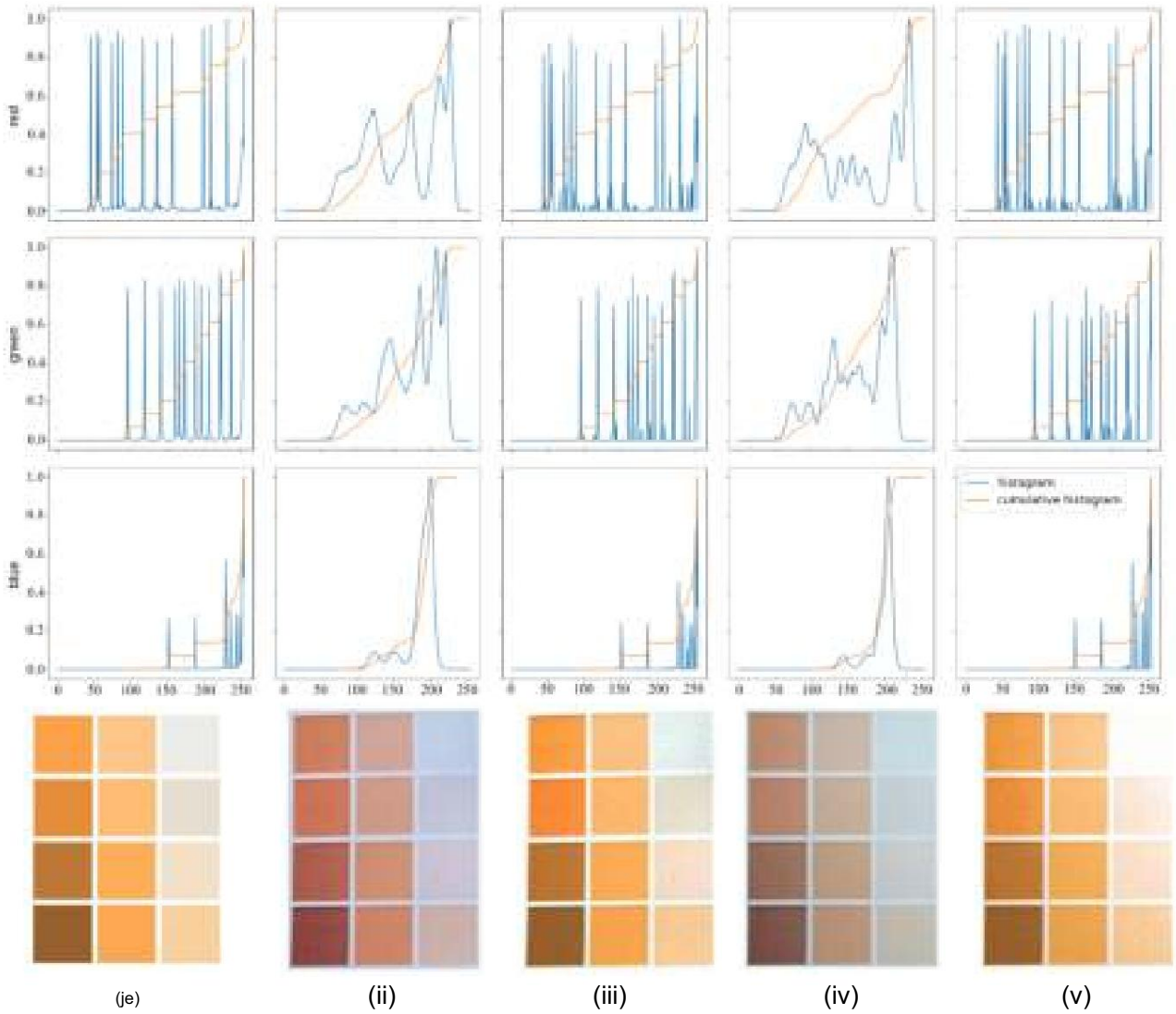


Figure 6. Modèles et leurs histogrammes pour les canaux RVB : (i) modèle numérique original, (ii) image de un modèle sous un éclairage naturel, (iii) image d'un modèle correspondant sous un éclairage naturel, (iv) image d'un modèle sous éclairage artificiel, et (v) image d'un modèle correspondant sous éclairage artificiel.

L'ensemble de données pour les calculs a été préparé comme suit : 198 photos ont été prises pour neuf concentrations de vitamine C dans deux conditions d'éclairage différentes. Cela signifie que pour une solution donnée pour chacune des conditions, 11 photos ont été prises. Chacun des les images ont subi un processus de standardisation, comme décrit à la section 3.1. L'exemplaire L'effet est illustré dans la figure 7. Dans la figure 8, l'effet de l'application de l'algorithme à divers des images de la solution de vitamine C avec une concentration de 33,63 µg/mL sont présentées. Ce a abouti à la génération d'ensembles de données pour deux conditions d'éclairage. La combinaison des deux ont abouti à la préparation d'un ensemble de conditions mixtes, composé de :

- ensemble de formation : 60 416 images
- Ensemble de validation : 20 480 images
- ensemble de tests : 20 480 images.

Les ensembles de données d'images préparés ont servi de données d'entrée pour former un système neuronal de régression. réseau, qui a été utilisé pour obtenir un modèle pour déterminer la concentration d'AA basé sur les photos prises.

Machine Translated by Google

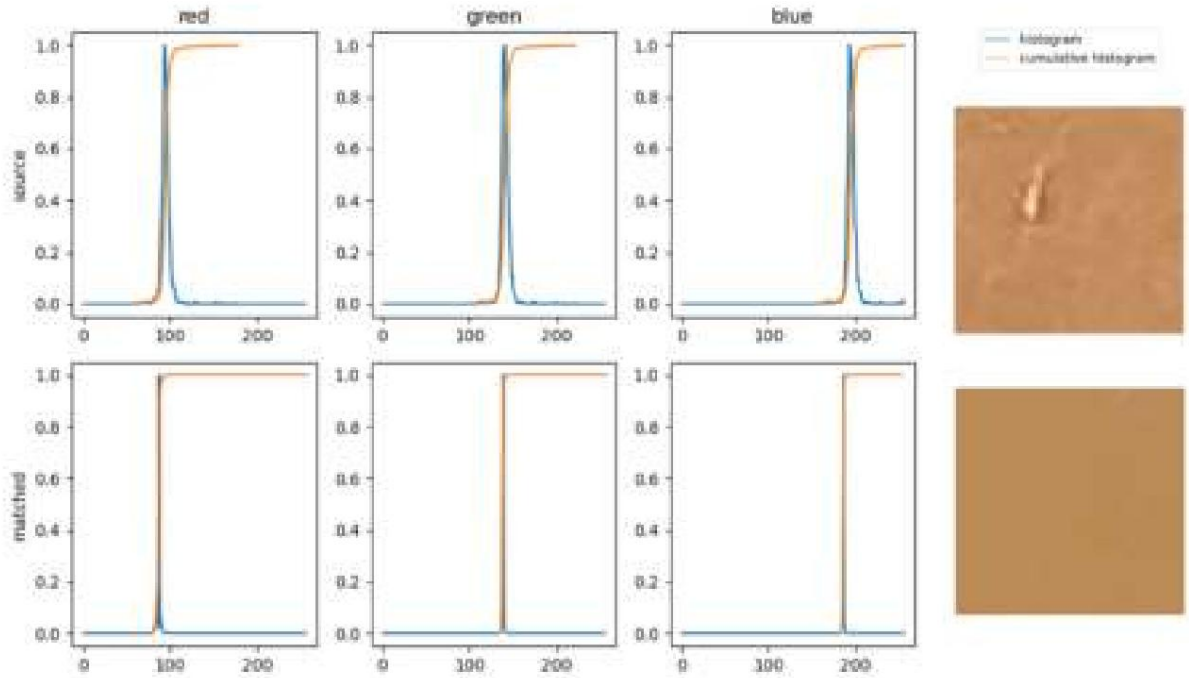


Figure 7. Images de la solution de vitamine C (33,63 µg/mL) dans des conditions d'éclairage artificiel avant (source) et après (correspondance) application de l'algorithme de standardisation et histogrammes correspondants des canaux RVB.

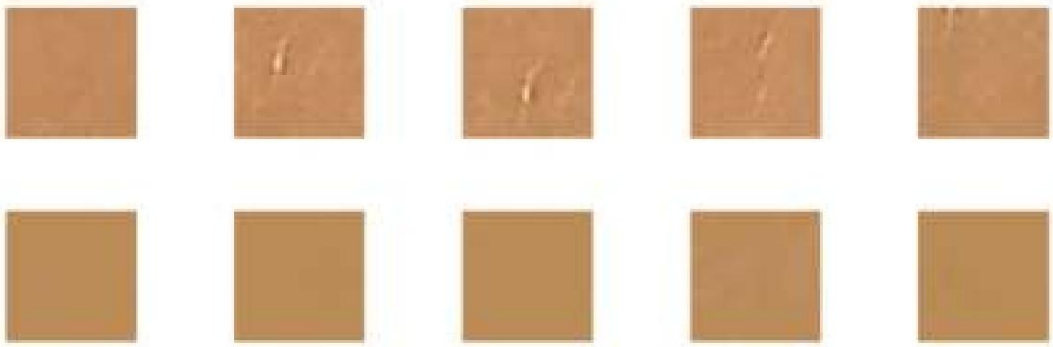


Figure 8. Diverses images exemplaires de la solution de vitamine C (33,63 µg/mL) dans des conditions d'éclairage artificiel avant (en haut) et après (en bas) l'application de l'algorithme de standardisation.

tt

tt

4.3. Architecture réseau

Un modèle de régression multivariée a été défini utilisant des réseaux de neurones profonds avec une architecture non conventionnelle spécifiquement dédiée à cette tâche. La conception du réseau a pris en compte des informations essentielles telles que la modélisation basée sur les couleurs, la similarité des couleurs correspondant aux valeurs successives de la variable cible et la couleur approximativement uniforme de chaque photo.

La couche d'entrée du réseau neuronal acceptait des images de dimensions 50 ×50 et trois canaux de couleur (RVB), sélectionnés au hasard sur chaque photo. Par la suite, des couches convolutives Conv2D avec de petits filtres 1 ×1, une régularisation L2 et la fonction d'activation GELU (unité linéaire d'erreur gaussienne) ont été appliquées après chaque couche convolutive. Une couche de normalisation a également été utilisée, opérant le long de l'axe des caractéristiques pour chaque échantillon afin de stabiliser et d'améliorer le processus d'apprentissage. L'architecture comprenait également des couches de regroupement qui réduisaient la taille spatiale de la carte de fonctionnalités, notamment AveragePooling2D et GlobalMaxPooling2D, qui réduisaient la dimension spatiale à une valeur par carte de fonctionnalités. La combinaison du regroupement global et de l'aplatissement est une technique standard pour la réduction de dimension avant les couches entièrement connectées. À l'étape suivante, les données ont été aplaties en un vecteur 1

tt

tt

Machine Translated by Google

à des couches denses entièrement connectées avec activation GELU et régularisation L2, tandis que le l'ajout d'une couche Dropout a empêché le surajustement. La sortie était une valeur correspondant à la concentration prévue de la solution dans l'image. Une description détaillée du L'architecture du réseau est fournie dans le tableau 2 et une vue d'ensemble est illustrée dans la figure 9. le nombre total de paramètres était de 23 879, et la formation en un seul cycle prenait quelques à plusieurs secondes (selon la variante de calcul).

Tableau 2. Architecture du réseau neuronal et paramètres utilisés dans cette étude.

Calque (Type)	Forme de sortie	Nombre de paramètres
Couche d'entrée	(Aucun, 50, 50, 3)	0
Conv2D	(Aucun, 50, 50,3 2)	128
Activation	(Aucun, 50, 50, 32)	0
Conv2D	(Aucun, 50, 50, 64)	2112
Activation	(Aucun, 50, 50, 64)	0
MoyennePooling2D	(Aucun, 46, 46, 64)	0
LayerNormalisation	(Aucun, 46, 46, 64)	128
Conv2D	(Aucun, 46, 46, 64)	4160
Activation	(Aucun, 46, 46, 64)	0
MoyennePooling2D	(Aucun, 44, 44, 64)	0
GlobalMaxPooling2D	(Aucun, 64)	0
Aplatir	(Aucun, 64)	0
Dense	(Aucun, 150)	9750
Abandonner	(Aucun, 150)	0
Dense	(Aucun, 50)	7550
Dense	(Aucun, 1)	51
	Paramètres totaux : 23 879 (93,28 Ko) Paramètres entraînables : 23 879 (93,28 Ko) Paramètres non entraînables : 0 (0,00 octet)	

Dans une couche convolutive telle que Conv2D, la régularisation L2 appliquée fonctionne sur le filtres convolutifs (noyaux). Ces filtres sont chargés d'extraire les fonctionnalités de les images d'entrée. L'ajout de la régularisation L2 signifie que pendant le processus d'optimisation, la fonction de perte que le modèle tente de minimiser inclut un terme supplémentaire qui pénalise valeurs de poids élevées des filtres. L'ajout d'une régularisation L2 à la couche convolutive aide empêcher le surajustement, stabilise le processus d'optimisation et favorise une plus grande uniformité et des solutions stables, conduisant à une meilleure généralisation du modèle.

La taille du noyau de 1 ×1 dans la couche convolutive possède des propriétés spécifiques et utiles dans réseaux de neurones. Les filtres 1 ×1 sont généralement utilisés pour réduire les dimensions spatiales sans l'agrégation spatiale, qui peut être utile pour réduire le nombre de canaux de fonctionnalités ou introduire la non-linéarité sans changer la taille spatiale. Cette configuration est souvent appelée une « convolution ponctuelle ». Le noyau 1 ×1 modifie le nombre de canaux de sortie (profondeur) sans changer la largeur et la hauteur de l'image d'entrée, permettant la combinaison de informations contenues dans différents canaux d'entrée sans intégrer d'informations spatiales. Chaque point de sortie est une combinaison linéaire des valeurs du canal d'entrée à un espace spatial donné. indiquer. Par exemple, si à un point d'entrée donné (x, y) nous avons des valeurs dans les canaux (w1, w2, w3), le noyau 1 ×1 applique des poids différents à ces valeurs et les combine, créer de nouveaux canaux de sortie. Un tel noyau a moins de paramètres par rapport à un noyau plus grand. noyaux convolutifs, ce qui entraîne moins d'opérations de calcul. Ceci est efficace dans termes de mémoire et de temps de calcul. Dans la modélisation basée sur des composants RVB, en utilisant un noyau 1 ×1 signifie examiner les relations entre les composants de couleur individuels,

Machine Translated by Google

plutôt que le voisinage de chaque pixel. Dans notre tâche, il n'était pas nécessaire de détecter de petits
puis des détails plus grands, comme c'est le cas dans la reconnaissance de formes.

La fonction d'activation GELU [47], appliquée comme alternative à ReLU (Rectified Linear
filtres convolutifs (noyaux). Ces filtres sont responsables de l'extraction des fonctionnalités de l'unité t) et d'autres fonctions
d'activation dans les réseaux de neurones, garantissant un meilleur comportement de gradient
et un apprentissage plus efficace dans les réseaux de neurones profonds (Figure 10). Sa continuité, sa différentiabilité et
sa capacité à préserver les caractéristiques des données non linéaires en font un outil précieux dans le domaine des données.
conception de modèles d'apprentissage profond. Contrairement à ReLU, qui peut souffrir de problèmes de débit graduel
valeurs de poids des filtres. L'ajout d'une régularisation L2 à la couche convolutive aide à p- pendant
l'entraînement (connu sous le nom de problème de gradient de disparition), GELU fournit un meilleur gradien
évacue le surajustement, stabilise le processus d'optimisation et favorise un débit plus uniforme et plus rapide en raison de sa
structure. GELU préserve les caractéristiques non linéaires qui sont cruciales pour
solutions efficaces, conduisant à une meilleure généralisation du modèle.
apprendre les représentations de données, aidant les modèles à mieux capturer les dépendances complexes dans
La taille du noyau de 1 ×1 dans la couche convolutive possède des données de formation de propriétés spécifiques et utiles.

dans les réseaux de neurones. Les filtres 1 ×1 sont généralement utilisés pour réduire les dimensions spatiales sans



ou introduire la non-linéarité sans modifier la taille spatiale. Cette configuration est souvent

combinaison d'informations contenues dans différents canaux d'entrée sans intégrer de naissain

canaux (w1, w2, w3). Le noyau 1 ×1 applique des poids différents à ces valeurs et co-

opérations de calcul. C'est e

, appliqué comme alternative à ReLU (Rectified Li-ear Unit) et à
d'autres fonctions d'activation dans les réseaux de neurones, garantit un meilleur comportement de gradient et un
apprentissage plus efficace dans les réseaux de neurones profonds (Figure 10). Sa continuité, d-

dans la conception de modèles d'apprentissage profond. Contrairement à ReLU, qui peut souffrir de problèmes de gradient fl
pendant l'entraînement (connus sous le nom de problème de gradient de disparition), GELU permet d'être

spécial pour l'apprentissage des représentations de données, aidant les modèles à mieux capturer les dépendances complexes

Figure 9. Architecture de réseau neuronal utilisée dans cette étude.

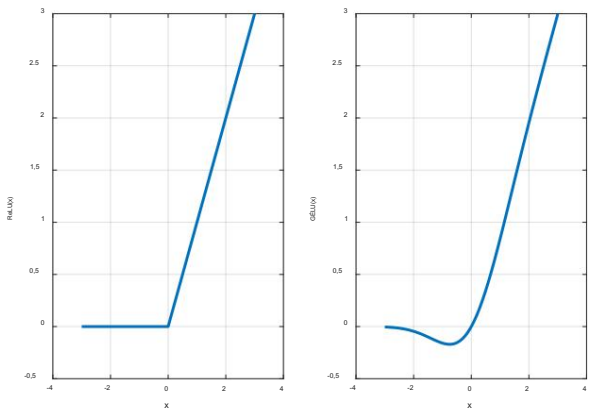


Figure 10. Fonctions d'activation ReLU et GELU.

tt

Machine Translated by Google

Dans la couche LayerNormalization, la normalisation est effectuée le long de l'axe des caractéristiques pour chaque échantillon, ce qui signifie que la normalisation est indépendante des autres échantillons du lot. Chaque vecteur de caractéristiques (par exemple, les canaux d'une image) a été normalisé séparément, en considérant uniquement les valeurs de ce vecteur, et non l'ensemble du lot. Contrairement à BatchNormalization, qui fonctionne en fonction de la taille du lot, LayerNormalization reste stable même avec de petits lots ou une taille de lot d'un seul. La normalisation peut aider à prévenir des problèmes tels que la disparition ou l'explosion des gradients et à accélérer la convergence. La normalisation des activations aide à maintenir les valeurs dans une plage raisonnable, stabilisant ainsi le processus d'apprentissage. Lorsque les activations sont normalisées, l'optimiseur (algorithme d'apprentissage) trouve plus facile d'ajuster les pondérations du réseau, accélérant ainsi la convergence vers des valeurs de pondération optimales.

GlobalMaxPool2D est une couche utilisée dans les réseaux de neurones pour réduire les dimensions des données spatiales après avoir appliqué des couches convolutives ou d'autres couches traitant des données spatiales, telles que des couches de pooling [48,49]. GlobalMaxPool2D fonctionne en sélectionnant la valeur maximale de chaque canal de caractéristiques sur toute la zone spatiale (tous les points) du tenseur de caractéristiques d'entrée. GlobalMax-Pool2D est un outil efficace pour la réduction de la dimensionnalité spatiale, ne conservant que les caractéristiques les plus significatives (valeurs les plus élevées) de chaque canal de caractéristiques. Cela réduit le nombre de paramètres du modèle, ce qui peut contribuer à réduire le surajustement et la com GlobalMaxPool2D n'a aucun paramètre pouvant être entraîné ; son seul objectif est de sélectionner la valeur maximale de chaque canal de fonctionnalités tout au long de la carte des fonctionnalités. Pour chaque canal de fonctionnalités, la couche GlobalMaxPool2D sélectionne la valeur la plus élevée sur toute la zone spatiale (hauteur et largeur). Le résultat de GlobalMaxPool2D est un tenseur aux dimensions réduites, contenant uniquement les valeurs maximales pour chaque canal de fonctionnalités. Après avoir appliqué GlobalMaxPool2D, le tenseur résultant peut être transmis à une couche entièrement connectée ou à d'autres types de couches qui traitent des vecteurs de données unidimensionnels.

Dans cette étude, le modèle a été compilé à l'aide de l'optimiseur Adam avec des paramètres bêta personnalisés et une fonction de perte définie, qui était la somme de MSE et MAE. Lorsque la fonction de perte est une combinaison de MSE et MAE, elle affecte la façon dont le modèle apprend à minimiser les erreurs. MSE pénalise plus sévèrement les grosses erreurs (en raison des carrés des différences), tandis que MAE traite toutes les erreurs de la même manière. La combinaison de ces fonctions de perte signifie que le modèle visera à réduire à la fois les erreurs, grandes et petites. Le gradient de MSE est plus raide pour les erreurs plus importantes, ce qui permet au modèle d'apprendre plus rapidement des erreurs significatives. Le gradient MAE est constant, ce qui conduit à un taux d'apprentissage plus cohérent pour toutes les erreurs. Le MAE est plus résistant aux valeurs aberrantes que le MSE car il n'implique pas de différences au carré. Une fonction de perte combinant MSE et MAE peut conduire à une convergence plus rapide dans certains cas en tirant parti des atouts des deux méthodes. MSE peut accélérer la réduction des erreurs importantes, tandis que MAE peut assurer la stabilité et un rythme d'apprentissage cohérent.

Pendant la formation, une modification du taux d'apprentissage a été appliquée après avoir atteint un niveau de val_loss stable pendant cinq époques. La valeur initiale était de 0,001. Les détails concernant les paramètres d'apprentissage sont fournis dans le tableau 3.

Tableau 3. Paramètres de formation en réseau.

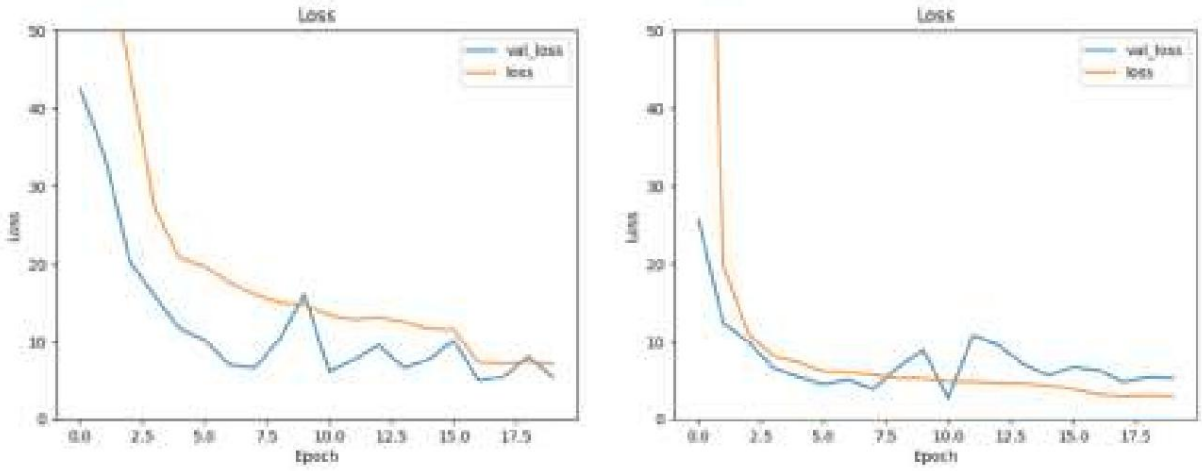
Paramètre	Paramètre
Taille de l'image	50 ×50 ×3
Fonction de perte	MSE + MAE
Optimiseur	Adam
Taux d'apprentissage initial	0,001
Métrique	RMSE
Taille du lot	30
Époque	20
Mélanger	Chaque époque

L'ensemble de données d'entrée était constitué d'images de dimensions 400 ×400 et était divisé en trois parties : ensembles de formation, de validation et de test dans un rapport 60/20/20. Pendant la formation,

Machine Translated by Google

Des carrés de 50 ×50 ont été recadrés dynamiquement à partir des images, et ces carrés ont servi de l'entrée directe à la première couche du réseau.

Il existe six variantes différentes de tests d'apprentissage profond pour entraîner une régression. algorithme pour déterminer la concentration de vitamine C sur la base d'images : images originales pour éclairage artificiel, images assorties pour l'éclairage artificiel, images originales pour l'éclairage naturel, images correspondantes pour l'éclairage artificiel, images originales pour un ensemble de données mixtes et correspondances images pour un ensemble de données mixtes. Pour chacun d'eux, le modèle a été formé (Section 3.2) et des prédictions ont été faites. Pertes de formation et de validation pour les calculs de lumière artificielle Les données, pour les images originales et celles correspondant à l'histogramme, sont présentées dans la figure 11. Les données la portée était similaire dans les deux cas ; cependant, les fluctuations de valeur étaient plus faibles pour les images, et dans ces conditions, les pertes de validation et de formation étaient plus similaires les unes aux autres.



(je) (ii)

Figure 11. Pertes de formation et de validation pour les calculs des conditions de foudre artificielle pour les deux images originales et (i) et correspondant à l'histogramme (ii).

La figure 12 présente les résultats de la prédiction effectuée à l'aide du système neuronal présenté. réseau. Il y avait trois séries de paires de résultats : la foudre artificielle, la foudre naturelle, et des données mixtes. Chaque série a été répétée à la fois pour les images originales et les images traitées. ceux. Il y a eu une amélioration des valeurs de concentration plus élevées pour chacun des conditions (tableau 4). L'écart type des prédictions était inférieur lors de l'application la méthode de normalisation proposée. La plus grande différence peut être observée dans le cas de lumière artificielle, qui pourrait être causée par l'uniformisation des reflets lumineux visibles dans les images originales.

Tableau 4. Évaluation des modèles de régression DL avec les meilleurs résultats pour six variantes différentes de tests.

	Artificiel Éclairage, Original	Artificiel Éclairage, Correspondant	Naturel Éclairage, Original	Naturel Éclairage, Correspondant	Mixte Éclairage, Original	Mixte Éclairage, Correspondant
Pente	0,9799	0,9961	0,9706	0,9452	0,9784	0,9825
Intercepter	−0,0036	−0,2091	1.1888	0,6051	0,5832	0,4086
RMSE/µg·mL−1	1,5769	0,7525	2.2900	1.9542	1,4828	1.3510
R^2 score	0,9979	0,9999	0,9953	0,9956	0,9966	0,9968

Machine Translated by Google

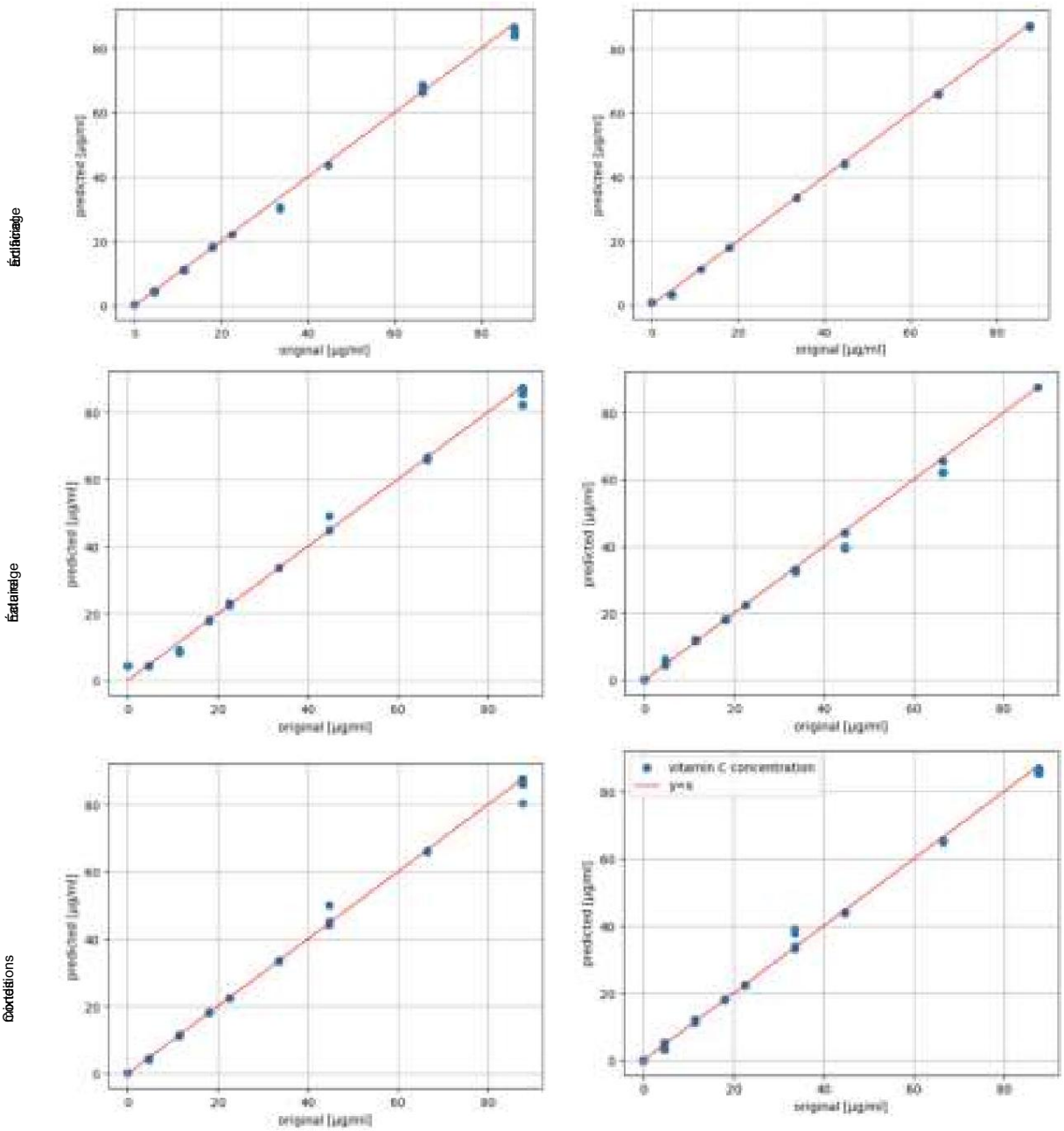


Figure 12. Évaluation des modèles de régression DL pour six variantes différentes de tests, avant (première colonne) et après correspondance des couleurs (deuxième colonne).

Les valeurs des facteurs d'évaluation et d'ajustement du modèle numérique sont présentées dans le tableau 4. Le score R2 était très élevé (supérieur à 0,99) dans tous les cas ; cependant, dans tous les cas, R^2 le score était plus élevé lorsque la standardisation était effectuée. Quant aux valeurs d'erreur quadratique moyenne , elles étaient inférieures pour les images correspondantes par rapport aux images originales. En fonction des conditions d'éclairage, le taux d'apprentissage passe d'une valeur initiale de 10^{-3} à 10^{-5} .

4.4. Étude sur l'ablation

Dans cette étude, nous avons également analysé comment différentes modifications de l'architecture du réseau affectaient la qualité du modèle de régression. Dans un premier temps, nous avons considéré un réseau ff qui ne tenait pas compte des spécificités du problème. Le modèle contenait six couches : une convolution, une mise en commun, une aplatie et trois denses. Il a été entraîné dans les mêmes conditions que précédemment. Les résultats quantitatifs détaillés sont présentés dans le tableau 5. L'utilisation

Machine Translated by Google

d'une architecture plus conventionnelle a détérioré la capacité du modèle à adapter les données et leurs capacités prédictives. Cependant, des améliorations de la qualité de la modélisation ont encore été observées. après avoir appliqué l'algorithme de normalisation des couleurs.

Tableau 5. Évaluation des modèles de régression DL pris en compte dans l'étude d'ablation pour six variations des tests, en fonction des conditions d'éclairage et de l'ensemble de données choisi.

	Artificiel Éclairage, Original	Artificiel Éclairage, Correspondant	Naturel Éclairage, Original	Naturel Éclairage, Correspondant	Mixte Éclairage, Original	Mixte Éclairage, Correspondant
Pente	0,9058	1,0138	0,9538	1.0071	1.0005	1,0112
Intercepter	0,6891	−0,3598	−0,8347	0,2263	0,8747	0,0203
RMSE/µg·mL−1	5.8415	1.7077	3.9504	2,5151	4.3488	2.8211
$R.^2$ score	0,9557	0,9962	0,9798	0,9918	0,9755	0,9897

De plus, plusieurs changements dans l'architecture ont été introduits avant la version finale. version a été choisie. Certains d'entre eux, par comparaison avec la version finale, sont évalués dans le tableau 6.

Tableau 6. Évaluation des modèles de régression DL pris en compte dans l'étude d'ablation pour un éclairage mixte et état d'essai correspondant, en fonction des modifications apportées au modèle.

Changement d'architecture	RMSE/µg·mL−1	$R.^2$ Score	Comparaison
Taille du lot = 120	2,3535	0,9947	Diminution du score R2 et augmentation du RMSE en comparaison avec le modèle final.
Taille de l'image = 16	4.0146	0,9825	Diminution significative du score R2 et augmentation de RMSE par rapport au modèle final. Baisse moindre du tarif initial par rapport à la version finale.
Modification de la fonction de perte	3.8203	0,9856	Diminution du score R2 et augmentation significative de RMSE par rapport au modèle final. Minimisé différences entre la perte de formation et de validation pendant la formation.
Optimiseur = Adagrad	3.1625	0,9877	Diminution du score R2 et augmentation du RMSE en comparaison avec le modèle final.
Optimiseur = SGD	-	-	Il n'y a eu aucun changement dans la valeur de la perte entre le époques. Le modèle n'a pas appris à prédire le concentration de solutions de vitamine C.
Taux d'apprentissage initial = 0,01 ou 0,0001	-	-	Aucun changement significatif dans les résultats n'a été détecté. Les résultats obtenus ne varient pas significativement selon comparaison avec les résultats finaux du modèle.

5. Conclusions

Ce travail présente un nouvel algorithme de standardisation des couleurs, dans lequel les principales étapes consiste à préparer un modèle à partir d'une première série de photos, à élaborer une couleur principe de cartographie utilisant des histogrammes numériques RVB égalisés d'un motif idéal et de sa photo, conversion des couleurs de photographies de l'objet examiné impliquant l'utilisation d'informations sur les histogrammes de modèles de cartographie et la reconstruction d'images basée sur la transformation histogramme. L'algorithme a été développé pour les cas où les objets examinés diffèrent peu en couleur et la couleur de chaque objet photographié est presque uniforme. Le fonctionnement du la procédure proposée a été démontrée et vérifiée dans le cadre d'une tâche de chimie analytique, qui impliquait la détermination colorimétrique de la vitamine C. Photographies de solutions AA prises dans des conditions contrôlées, n'a pas montré beaucoup de différence de couleur au sein de la concentration gamme. La recherche a été réalisée à la lumière du jour et dans des conditions d'éclairage artificiel.

Machine Translated by Google

afin de les combiner en un seul ensemble de données et d'obtenir un modèle permettant de déterminer la vitamine C quelles que soient les conditions d'éclairage utilisées lors de la prise de photos. Des ensembles d'images standardisées ont constitué la base d'un apprentissage automatique approfondi utilisant l' architecture des réseaux neuronaux dans la variante de régression.

Pour la plage de concentrations de vitamine C de 0 à 87,72 µg·mL−1 , le RMSE pour le L'ensemble de tests variait entre 0,75 et 1,95 µg·mL−1 par rapport à la variante non standardisée, où le RMSE était compris entre 1,35 et 2,29 µg·mL−1 . La compatibilité des résultats de prévision de concentration avec les données réelles, exprimée par R2 , était supérieure à 0,99 pour chacune des variantes standardisées. La plus grande amélioration a été constatée dans le cas de la lumière artificielle, ce qui peut s'expliquer par l'élimination des reflets sur les surfaces des solutions dans les images après le processus de standardisation. D'autres études sont nécessaires pour déterminer l'utilité de l'algorithme dans différents scénarios (par exemple, prédiction colorimétrique de la concentration de différents composés). Cette nouvelle approche pourrait contribuer au développement d'un nouveau domaine d'analyse chimique, qui comprend l'évaluation quantitative d'un large spectre d'analytes chimiques à l'aide de téléphones mobiles et de l'analyse d'images. Cette étude présentée est prometteuse pour des applications dans plusieurs domaines scientifiques et technologiques, tels que le génie biomédical, le génie chimique et l'industrie alimentaire.

Contributions des auteurs : PK : Conceptualisation ; Méthodologie; Conservation des données ; Analyse formelle ; Enquête ; Logiciel; Ressources; Visualisation; Rédaction – ébauche originale ; Rédaction – Révision et édition.
MJ : Conceptualisation ; Analyse formelle ; Acquisition de financement ; Enquête; Méthodologie; Gestion de projet ; Surveillance; Validation; Rédaction – ébauche originale ; Rédaction – Révision et édition. Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Projet de recherche soutenu par le programme « Initiative d'excellence – université de recherche » pour l'Université AGH de Cracovie (ID1479).

Déclaration de disponibilité des données : les algorithmes implémentés en Python et l'échantillon de données pour les calculs sont disponibles sur la plateforme Github à l'adresse : https://github.com/pwkwiek/vitaminC (consulté le : 13 juillet 2024).

Remerciements : Les auteurs remercient Inz. Filip Ciepiela (Université AGH de Cracovie) pour ses précieux conseils lors de la mise en œuvre du projet.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Références

1. Fairchild, MD Modèles d'apparence couleur, 3e éd.; John Wiley & Sons, Ltd. : Hoboken, NJ, États-Unis, 2013 ; ISBN9781119967033.
2. Minz, PS ; Saini, CS Évaluation du cadre d'étalonnage des cubes RVB et effet des graphiques d'étalonnage sur la mesure des couleurs de Fromage Mozzarella. J. Mesures alimentaires. Caractère. 2019, 13, 1537-1546. [Référence croisée]
3. Ernst, A. ; Papst, A. ; Ruf, T. ; Garbas, JU Check My Chart : un outil de suivi de nuancier robuste pour l'étalonnage des caméras colorimétriques. Dans Actes de la 6e Conférence internationale sur les techniques et applications de collaboration en vision par ordinateur/infographie , MIRAGE'13, Berlin, Allemagne, 6-7 juin 2013.
4. McCamy, CS ; Marcus, H. ; Davidson, JG Tableau de rendu des couleurs. J.Appl. Photogr. Ing. 1976, 2, 95-99.
5. Sunoj, S. ; Igathinathane, C. ; Saliendra, N. ; Hendrickson, J. ; Archer, D. Calibrage des couleurs des images numériques pour l'agriculture et Autres applications. ISPRS J. Photogramme. Télésens. 2018, 146, 221-234. [Référence croisée]
6. Kim, M. ; Kim, B. ; Parc, B. ; Lee, M. ; Gagné, Y. ; Kim, CY; Lee, S. Un dispositif numérique de correspondance de teintes pour la détermination de la couleur dentaire à l'aide de l'algorithme de la machine à vecteurs de support. Capteurs 2018, 18, 3051. [CrossRef]
7. Karaimer, HC ; Brown, MS Améliore la précision de la reproduction des couleurs sur les appareils photo. Dans les actes de la conférence de l'IEEE Computer Society sur la vision par ordinateur et la reconnaissance de formes, Salt Lake, UT, États-Unis, 18-23 juin 2018 ; pages 6440 à 6449.
8. Roy, S. ; Kumar Jain, A. ; Lal, S. ; Kini, J. Une étude sur les méthodes de normalisation des couleurs pour les images histopathologiques. Micron 2018, 114, 42-61. [Référence croisée]
9. Zhao, Y. ; Ferguson, S. ; Zhou, H. ; Elliott, C. ; Rafferty, K. Alignement des couleurs pour une constance relative des couleurs via des méthodes non standard Références. IEEETrans. Processus d'images. 2022, 31, 6591-6604. [Référence croisée]
10. Rashid, F. ; Jamayet, Nouveau-Brunswick; Farook, TH; AL-Rawas, M. ; Barman, A. ; Johari, Y. ; Noorani, TY; Abdallah, JY ; Eusufzai, SZ; Alam, MK Variations de couleur lors de l'imagerie numérique de prothèses faciales soumises à une lumière ambiante non filtrée et à des techniques d'étalonnage d'image dans les cliniques dentaires : une analyse in vitro. PLoS ONE 2022, 17, e0273029. [Référence croisée]
11. Barbero-Álvarez, MA; Rodrigo, JA; Menéndez, JM Calibrage RVB adaptatif à erreur minimale dans un contexte d' incertitude colorimétrique pour la préservation du patrimoine culturel. Calculer. Vis. Image comprise. 2023, 237, 103835. [Réf. croisée]

Machine Translated by Google

12. Noor Azhar, M. ; Bustam, A. ; Naseem, FS ; Shuin, SS ; Md Yusuf, MH ; Hishamudin, NU ; Poh, K. Améliorer la fiabilité de la colorimétrie urinaire sur smartphone à l'aide d'une méthode d'étalonnage par carte couleur. Chiffre. Santé 2023, 9, 20552076231154684. [Référence croisée]
13. Zhang, G. ; Chanson, S. ; Panescu, J. ; Shapiro, N. ; Dannemiller, KC; Qin, R. Une nouvelle solution système pour une colorimétrie précise Mesure via la réalité augmentée sur smartphone. PLoS ONE 2023, 18, e0287099. [Référence croisée]
14. Chairat, S. ; Chaichulee, S. ; Dissaneewate, T. ; Wangkulangkul, P. ; Kongpanichakul, L. Évaluation assistée par l'IA des tissus de la plaie avec calibrage automatique des couleurs et des mesures sur les images prises avec un smartphone. Santé 2023, 11, 273. [CrossRef]
15. Suominen, J. ; Egiazarian, K. Correction des couleurs de la caméra à l'aide de splines. Dans Actes du Symposium international IS&T sur l'imagerie électronique, Burlingame, Californie, États-Unis, 21-25 janvier 2024 ; pages 165-1-165-6.
16. Souissi, M. ; Chaouch, S. ; Moussa, A. Correspondance des couleurs de filaments bicomposants (PET/PTT) avec de hautes performances à l'aide d' un algorithme génétique. Sci. Rep.2024 , 14, 10949. [CrossRef]
17. Wannasin, D. ; Grossmann, L. ; McClements, DJ optimisant l'apparence des aliments à base de plantes à l'aide de pigments naturels et Théorie de la correspondance des couleurs. Biophysique alimentaire. 2024, 19, 120-130. [Référence croisée]
18. Wu, Y. Conception de correspondance des couleurs assistée par image de référence basée sur un algorithme génétique interactif. J. Electr. Système. 2024, 20, 400-410. [Référence croisée]
19. Organisation pour l'alimentation et l'agriculture ; Organisation Mondiale de la Santé. Besoins en vitamines et minéraux en nutrition humaine, 2e éd. ; FAO/OMS : Genève, Suisse, 1998 ; pp. 1–20, ISBN 9241546123.
20. Organisation pour l'alimentation et l'agriculture ; Organisation Mondiale de la Santé. Besoins humains en vitamines et minéraux ; FAO/OMS : Genève, Suisse, 2001.
21. Lykkesfeldt, J. Sur l'effet de l'apport en vitamine C sur la santé humaine : comment (mauvais) interpréter les preuves cliniques. Rédox Biol. 2020, 34, 101532. [Réf. croisée]
22. Dosé, M. ; Jirkovsk, E. ; Kujovsk, L. ; Javorsk, L. ; Pourov, J. ; Mercolini, L. ; Remi, F. Vitamine C — Sources, rôle physiologique, cinétique, carence, utilisation, toxicité et détermination. Nutriments 2021, 615, 1–34.
23. Suntornsuk, L. ; Gritsanapun, W. ; Nilkamhank, S. ; Paochom, A. Quantification de la teneur en vitamine C dans les jus de plantes par titrage direct. J.Pharm. Bioméde. Anal. 2002, 28, 849-855. [Référence croisée]
24. Klimczak, I. ; Gliszczy ´nska-Swigł o, A. Comparaison des méthodes UPLC et HPLC pour la détermination de la vitamine C. Chimie alimentaire. 2015, 175, 100-105. [Référence croisée]
25. Gazdik, Z. ; Zitka, O. ; Petrova, J. ; Adam, V. ; Zehnalek, J. ; Horna, A. ; Reznicek, V. ; Beklova, M. ; Kizek, R. Détermination de la vitamine C (acide ascorbique) par chromatographie liquide haute performance couplée à une détection électrochimique. Capteurs 2008, 8, 7097-7112. [Référence croisée]
26. Bunaciu, AA; Bacalum, E. ; Aboul-Enein, HY; Udristioiu, GE; Fleschin, ¸S. Analyse spectrophotométrique FT-IR de l'acide ascorbique et de la biotine et de leurs formulations pharmaceutiques. Anal. Lett. 2009, 42, 1321-1327. [Référence croisée]
27. Zhu, Q. ; Dong, D. ; Zheng, X. ; Chanson, H. ; Zhao, X. ; Chen, H. ; Chen, X. Détermination par chimiluminescence de l'acide ascorbique à l'aide de cadres organométalliques à base de cuivre et d'oxyde de graphène comme catalyseur. RSC Adv. 2016, 6, 25047-25055. [Référence croisée]
28. Berg, RW Enquête sur l'acide L (+) -ascorbique avec spectroscopie Raman dans la lumière visible et UV. Appl. Spectrosc. Rév.2015 , 50, 193-239. [Référence croisée]
29. Yang, H. ; Irudayaraj, J. Détermination rapide de la vitamine C par les techniques NIR, MIR et FT-Raman. J.Pharm. Pharmacol. 2010, 54, 1247-1255. [Référence croisée]
30. Zykova, EV; Sandetskaïa, NG ; Ostrovskii, OV ; Verovskii, VE Méthodes d'analyse et de contrôle des processus de détermination de l'acide ascorbique dans les préparations médicinales par électrophorèse de zone capillaire et micellaire. Pharma. Chimique. J. 2010, 44, 463-465. [Référence croisée]
31. Dodevska, T. ; Hadzhiev, D. ; Shterev, I. Une revue des microcapteurs électrochimiques pour la détection de l'acide ascorbique : clinique, Applications pharmaceutiques et de sécurité alimentaire. Micromachines 2023, 14, 41. [CrossRef]
32. Huang, L. ; Tian, S. ; Zhao, W. ; Liu, K. ; Guo, J. Capteurs électrochimiques de vitamines : une revue critique. Talante 2021, 222, 121645. [Référence croisée]
33. Broncová, G. ; Prokopec, V. ; Shishkanova, Langue électronique potentiométrique TV pour l'analyse pharmaceutique : détermination de Acide ascorbique à base de films électropolymérisés. Chemocapteurs 2021, 9, 110. [CrossRef]
34. Coutinho, MS ; Morais, CLM; Neves, ACO; Ménezès, FG ; Lima, Détermination colorimétrique KMG de l'acide ascorbique basée sur son effet interférent dans l'analyse enzymatique du glucose : une approche utilisant l'analyse d'images sur smartphone. J. Braz. Chimique. Soc. 2017, 28, 2500-2505. [Référence croisée]
35.Porto , ISA ; Santos Neto, JH ; dos Santos, LO; Gomes, AA; Ferreira, SLC Détermination de l'acide ascorbique dans les jus de fruits naturels à l'aide de la colorimétrie d'image numérique. Microchimie. J. 2019, 149, 104031. [CrossRef]
36. Kong, L. ; Gan, Y. ; Liang, T. ; Zhong, L. ; Poêle, Y. ; Kirsanov, D. ; Légin, A. ; Wan, H. ; Wang, P. Un nouveau spectromètre CD basé sur smartphone pour une détection colorimétrique très sensible et rentable de l'acide ascorbique. Anal. Chim. Actes 2020, 1093, 150-159. [Référence croisée]
37. Li, C. ; Xu, X. ; Wang, F. ; Zhao, Y. ; Shi, Y. ; Zhao, X. ; Liu, J. Plate-forme pour smartphone portable intégrée à un capteur de fluorescence assisté par bande de papier pour la quantification ultrasensible et visuelle de l'acide ascorbique. Chimie alimentaire. 2023, 402, 134222. [Réf. croisée]
38. Zhaoa, W. ; Caoa, P. ; Zhua, Y. ; Liua, S. ; Gaob, H.-W.; Huang, C. Détection rapide de la teneur en vitamine C dans les fruits et légumes Utilisation d'un appareil photo numérique et de Color Reaction. Quim. Novembre 2020, 43, 1421-1430. [Référence croisée]

Machine Translated by Google

39. Dumancas, GG; Ramasahayam, S. ; Bello, G. ; Hughes, J. ; Kramer, R. Techniques de régression chimiométrique comme émergentes et puissantes Outils dans les études d'associations génétiques. Analyse des tendances TrAC. Chimique. 2015, 74, 79-88. [Référence croisée]

40. Wold, S. ; Sjöström, M. ; Eriksson, L. PLS-Regression : un outil de base en chimiométrie. Chimio. Intell. Laboratoire. Système. 2001, 58, 109-130. [Référence croisée]

41. Li, B. ; Morris, J. ; Martin, Sélection de modèle EB pour la régression des moindres carrés partiels. Chimio. Intell. Laboratoire. Système. 2002, 64, 79-89. [Référence croisée]

42. Smola, AJ; Schokopf, B. Un didacticiel sur la régression des vecteurs de support. Stat. Calculer. 2004, 14, 199-222. [Référence croisée]

43. Goodfellow, I. ; Bengio, Y. ; Courville, A. Apprentissage profond ; MIT Press : Cambridge, MA, États-Unis, 2016 ; ISBN9780262035613.

44. Lathuilière, S. ; Mesejo, P. ; Alameda-Pineda, X. ; Horaud, R. Une analyse complète de la régression profonde. IEEETrans. Modèle Anal. Mach. Intell. 2020, 42, 2065-2081. [Référence croisée]

45. Pascual, L. ; Gras, M. ; Vidal-Brotons, D. ; Alcañiz, M. ; Martínez-Máñez, R. ; Ros-Lis, JV Un outil voltammétrique E-Tongue pour l'émulation de l'analyse sensorielle et la discrimination des laits végétaux. Sens. Actionneurs B Chem. 2018, 270, 231-238. [Référence croisée]

46. Wójcik, S. ; Ciepiela, F. ; Jakubowska, M. Analyse de vision par ordinateur des couleurs d'échantillons par rapport à la langue électronique voltammétrique à quatre disques Iridium-Platine pour la reconnaissance de la falsification naturelle du miel. Mesures. J. Int. Mesures. Confé. 2023, 209, 112514. [Réf. croisée]

47. Lee, M. Analyse mathématique et évaluation des performances de la fonction d'activation GELU dans l'apprentissage profond. J. Math. 2023, 1, 4229924. [Réf. croisée]

48. Sikandar, S. ; Mahum, R. ; Alsalman, AM Une nouvelle approche hybride pour une récupération d'images basée sur le contenu utilisant la fusion de fonctionnalités. Appl. Sci. 2023, 13, 4581. [Réf. croisée]

49. Hasan, MA; Haque, F. ; Sabuj, SR ; Sarker, H. ; Goni, MOF; Rahman, F. ; Rashid, MM Un CNN multi-échelle léger de bout en bout pour la classification du cancer du poumon et du côlon avec intégration XAI. Technologies 2024, 12, 56. [CrossRef]

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Pièges dans la résolution de métaheuristiques basées sur la stœchiométrie Modèles d'optimisation pour les réseaux métaboliques

Mónica Fabiola Briones-Báez 1,*, Luciano Aguilera-Vázquez [1], Nelson Rangel Valdez [1], Cristal Zuñiga [2,*], Ana Lidia Martínez-Salazar [1] 1 et Claudia Gomez-Santillan

[1] División de Estudios de Posgrado e Investigación, Instituto Tecnológico de Ciudad Madero (TECNM), Los Mangos 89440, Mexique ; luciano.av@cdmadero.tecnm.mx (LA-V.); nelson.rv@cdmadero.tecnm.mx (NR-V.); ana.ms@cdmadero.tecnm.mx (ALM-S.); claudia.gs@cdmadero.tecnm.mx (CG-S.)

[2] Département de biologie, Université d'État de San Diego, 5500 Campanile Drive, San Diego, CA 92182, États-Unis

* Correspondance : d00070469@cdmadero.tecnm.mx (MFB-B.) ; czuniga2@sdsu.edu (CZ)

Résumé : L'analyse de l'équilibre des flux (FBA) est une méthode basée sur des contraintes qui est couramment utilisée pour guider les métabolites à travers des voies restrictives qui impliquent souvent des conditions telles que des cycles anaplérotiques comme celui de Calvin, des réactions réversibles ou irréversibles et des nœuds où se ramifient les voies métaboliques. La méthode peut identifier les meilleures conditions pour un cours, mais échoue lorsqu'elle traite les voies de plusieurs métabolites d'intérêt. Des études récentes sur le métabolisme considèrent qu'il est plus naturel d'optimiser simultanément plusieurs métabolites plutôt qu'un seul ; de plus, ils soulignent l'utilisation des métaheuristiques comme une alternative intéressante qui étend FBA pour aborder de multiples objectifs. Cependant, la littérature prévient également que le recours à de telles techniques ne doit pas être sauvage. Au lieu de cela, il doit être soumis à des processus minutieux de mise au point et de sélection pour obtenir les résultats souhaités. Ce travail analyse l'impact sur la qualité des voies construites à l'aide des algorithmes NSGAII et MOEA/D et de plusieurs nouveaux modèles d'optimisation ; il mène une étude sur deux études de cas, la biosynthèse pigmentaire et le nœud du métabolisme du glutamate de la microalgue Chlorella vulgaris, dans trois conditions de culture (autotrophe, hétérotrophe et mixotrophe) tout en optimisant simultanément trois intermédiaires métaboliques comme fonctions objectives indépendantes. Les résultats montrent des performances variables entre NSGAII et MOEA/D, démontrant que la sélection d'un modèle d'optimisation peut grandement affecter les phénotypes prédits.

Mots clés : métabolisme cellulaire ; Expédié par Amazon ; optimisation multi-objectifs ; NSGAII ; MOEA/D

check for updates

Citation : Briones-Baez, MF ; Aguilera-Vazquez, L. ; Rangel-Valdez, N. ; Zuñiga, C. ; Martinez-Salazar, AL; Gomez-Santillan, C. Pièges liés aux métaheuristiques résolvant l'optimisation stœchiométrique Modèles de réseaux métaboliques. Algorithmes 2024, 17, 336. https:// doi.org/10.3390/a17080336

Rédacteurs académiques : Massimiliano Caramia et Frank Werner

Reçu : 30 mars 2024
Révisé : 1er juillet 2024
Accepté : 16 juillet 2024
Publié : 1 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Les microalgues sont des micro-organismes cellulaires photosynthétiques connus depuis la nuit des temps. Ils peuvent être cultivés soit dans les eaux usées, soit dans des eaux propres ou salées ; certaines souches, comme Dunaliella salina, peuvent être cultivées dans des eaux salées, et d'autres souches, comme Chlorella vulgaris, sont cultivées dans de l'eau douce et peuvent survivre à des températures de croissance élevées. Les microalgues ont besoin d'une source de carbone pour réaliser la photosynthèse et produire leur biomasse. Le carbone peut être obtenu sous forme de CO2 à partir de sources polluantes, et celles-ci le transforment en oxygène, contribuant ainsi de manière circulaire à réduire le réchauffement climatique [1].

Les microalgues sont de gros producteurs de biomasse ; à l'intérieur se trouvent des métabolites tels que des lipides qui pourront à l'avenir être utilisés comme biocarburants, des acides aminés et des pigments actuellement utilisés dans les industries pharmaceutiques et cosmétiques, et des protéines qui sont utilisées comme compléments alimentaires [2,3]. De plus, les microalgues présentent des applications biotechnologiques en tant que sources de biorestauration de la qualité de l'eau et ont été utilisées comme alternatives pour l'élimination des métaux lourds car certaines souches, telles que Chlorella vulgaris et Scenedesmus obliquos, sont capables d'absorber des métaux lourds tels que le cadmium (Cd) et Plomb (Pb) [4,5].

Toutes les caractéristiques ci-dessus rendent l'étude du métabolisme des microalgues attractive pour l'ingénierie métabolique. Cette discipline se concentre sur l'étude de la topographie du

Machine Translated by Google

réseau, la régulation d'une voie métabolique, l'identification des goulots d'étranglement, la détermination des flux métaboliques et l'élimination des réactions secondaires par délétion de gène [6].

En particulier, diverses techniques d'ingénierie métabolique ont été utilisées pour analyser les voies et optimiser les flux afin de manipuler le métabolisme et de modifier les flux vers un produit souhaité et ainsi pouvoir ajouter de la valeur commerciale. Cependant, les flux métaboliques ne peuvent pas être mesurés in vivo ; pour cette raison, des approches de modélisation sont nécessaires pour les mesurer ou les prédire [7]. Parmi elles figurent des approches basées sur des contraintes à objectif unique , telles que FBA (Flux Balance Analysis), des approches mathématiques exactes multi-objectifs et basées sur des heuristiques.

Le FBA est l'une des techniques les plus utilisées pour étudier le métabolisme cellulaire. Il s'agit de l' approche FBA à objectif unique basée sur des contraintes. Cette approche est largement utilisée dans l'analyse des flux des réseaux métaboliques puisqu'elle peut être utilisée même si les données cinétiques ne sont pas disponibles, mais elle nécessite des informations sur les données stoechiométriques des réactions présentes dans le réseau, les exigences de croissance et les paramètres spécifiques. des méthodes de mesure du système biologique, en particulier la reconstruction du réseau métabolique à l'échelle du génome [8] qui incluent toutes les réactions connues présentes dans les organismes étudiés et les gènes qui codent pour chaque enzyme [9].

Le métabolisme cellulaire dans la modélisation métabolique est décrit comme l'ensemble des réactions chimiques présentes dans un organisme. Ceci est représenté mathématiquement par une matrice stœchiométrique, S, de taille (m ×n), où n sont les réactions et m sont les métabolites impliqués dans chaque réaction, attribuant un coefficient négatif s'il représente un réactif et un coefficient positif s'il s'agit d'un réactif. un produit, et un coefficient nul signifie que le métabolite n'est pas présent dans la réaction ; chaque réaction aura une borne inférieure et une borne supérieure limitant l'espace des solutions ou la valeur maximale et minimale du flux autorisé. La FBA recherche le

optimisation linéaire d'une fonction objectif ; cette fonction représente la combinaison linéaire des flux qui représentent généralement la production de biomasse [10].

$$\begin{aligned} &\text{FBA max } F(v) = v_{\text{biomasse}} \\ &\text{Sous réserve de} \quad S \cdot v = 0 \\ &\qquad\qquad\quad LB_j \le v_j \le UB_j, \quad j \in \{1, \ldots, n\} \end{aligned} \tag{1}$$

L'équation (1) définit le problème d'optimisation linéaire FBA associé [11], où v est le vecteur de flux à travers les réactions. La matrice stœchiométrique Sm×n représente le réseau métabolique, où il existe un métabolite par ligne et une réaction par colonne. La valeur de la cellule Sij est le coefficient stoechiométrique du métabolite i impliqué dans la réaction j [9], et LBj et UBj sont les limites inférieure et supérieure des flux autorisés dans le système métabolique. L'hypothèse de régime permanent est établie par Sv = 0 [12].

Sa polyvalence a fait que le FBA a été largement utilisé dans différents organismes, notamment les microalgues, les bactéries, les consortiums de micro-organismes, etc. Un exemple en est la prédiction de la croissance cellulaire de la cyanobactérie Synechocystis dans [13] ; il a également été utilisé dans la dégradation du glucose par digestion anaérobie pour prédire la distribution des métabolites et révéler la transformation du carbone afin d'évaluer la conversion de l'éthanol, de l'acide propionique et de l'acide butyrique en acide acétique [14]. FBA a servi d'étude en médecine, où les activités de flux ont été calculées pour étudier les différences dans les voies métaboliques, en comparant les sous-types de cancer du sein (15).

La méthodologie FBA a été utilisée pour évaluer les flux métaboliques dans différentes souches de microalgues en utilisant trois conditions de croissance autotrophe hétérotrophes et mixotrophes. La première microalgue à utiliser le FBA fut la microalgue Chlamydomonas reinhardtii [16], avec laquelle la première carte métabolique a été obtenue. Plus tard, la microalgue Chlorella vulgaris a été utilisée pour l'étude de la production de lipides [17] ; tous deux avaient pour unique fonction objective la production de biomasse.

Bien que cette approche ait été largement utilisée, la distribution des métabolites à travers les voies est conditionnée par les métabolites présents dans l'équation de la fonction objectif et dans les paramètres expérimentaux ; le blocage de certaines voies métaboliques entraîne

Machine Translated by Google

une distribution de flux avec des valeurs nulles. De plus, le FBA a été largement utilisé dans la recherche de maximisation de la production de composés d'intérêt, mais le métabolisme cellulaire dans son état naturel ne guide pas les voies métaboliques vers la production d' un métabolite particulier . Dans la recherche d'une meilleure compréhension du métabolisme des microalgues, de nouvelles techniques d'optimisation ont émergé, comme les métaheuristiques pour l'optimisation multi-objectifs [18].
Ces techniques recherchent une répartition plus uniforme et sont plus proches de la réalité de ce qui se passe au niveau métabolique en optimisant simultanément diverses fonctions aux objectifs contradictoires.

L'optimisation multi-objectifs repose généralement sur la recherche de solutions à différents objectifs contradictoires qui doivent être optimisés simultanément. L'optimisation multi-objectifs est d'une grande importance et a été réalisée au niveau technologique et scientifique.
Quelques exemples dans l'industrie chimique ont été rapportés dans l'optimisation des processus des unités opérationnelles, de la bioraffinerie, de l'ingénierie des réactions, de la prévention et du contrôle, etc. [19]. Il a également été utilisé dans le secteur de la biologie et de la médecine [20], ainsi qu'en ingénierie métabolique [18].

L'optimisation multi-objectifs contraste avec Cobrapy FBA en libre accès, qui ne maximise ou minimise qu'une seule fonction objectif et où une seule solution est obtenue.
En optimisation multi-objectifs, un ensemble de solutions est obtenu. Les solutions obtenues sont dites non dominées car aucune autre solution dans l'espace de recherche n'est meilleure que les autres lorsque tous les objectifs sont considérés simultanément. Cet ensemble de solutions est connu sous le nom de solutions optimales de Pareto [21].

Entre les méthodes utilisées pour l'optimisation multi-objectif des micro-organismes se trouvent des algorithmes évolutifs tels que NSGAII [18], MOMO, basés sur le modèle bio-objectif, et des méthodes mathématiques exactes qui consomment beaucoup de ressources informatiques [22]. .

Les algorithmes métaheuristiques proviennent de l'évolution naturelle des groupes biologiques ; ils font partie de l'intelligence artificielle et sont nés de l'informatique naturelle et de méthodes heuristiques (algorithmes de recherche partielle). Comparées aux méthodes mathématiques qui génèrent une dépense de calcul importante, ces méthodes fournissent des solutions suffisamment bonnes pour un algorithme d'optimisation avec un temps et un espace de calcul acceptables [23].

Les algorithmes évolutifs multi-objectifs (MOEA) sont largement reconnus dans la communauté scientifique comme une approche pour résoudre les problèmes d'optimisation multi-objectifs
En particulier, le NSGAII (EA multi-objectif basé sur une classification non dominée) proposé par [24] s'est avéré assez efficace lors du traitement de deux ou trois objectifs [25,26].
L'algorithme évolutif MOEA/D basé sur la décomposition est testé car il a la particularité de fonctionner correctement lorsqu'il y a plus de trois fonctions objectifs.

Auparavant, des méthodes métaheuristiques avaient été utilisées pour étudier le métabolisme dans [18]. L' algorithme NS-GAII a été développé pour optimiser trois fonctions objectives, les protéines, les glucides et le CO2, dans un réseau métabolique de la microalgue Chlamydomonas reinhardtii en utilisant l'algorithme NSGAII, un schéma de codage basé sur l'analyse du bilan de flux (FBA). Cependant, la solution algorithmique peut être différente pour chaque problème d'optimisation, une différence qui peut augmenter en fonction de la manière dont le modèle mathématique impliqué explique les phénomènes étudié
Les réseaux métaboliques pourraient ne pas être exemptés de ces problèmes, et ces travaux analysent des algorithmes distincts (NSGAII et MOEA/D) et des modèles d'optimisation distincts pour les réseaux métaboliques (quatre approches d'optimisation multiobjectifs) du point de vue de la qualité des solutions qui pourraient être obtenues. par eux, et la commodité des informations fournies.
L'étude est réalisée sur deux études de cas impliquant des conditions complexes impliquant des cycles, des bifurcations et des réactions réversibles et irréversibles. Les principales contributions de ce travail incluent le développement de trois nouveaux modèles d'optimisation multi-objectifs pour les réseaux métaboliques, une nouvelle solution algorithmique basée sur la décomposition et une analyse qui guide l'identification appropriée de métaheuristiques et de modèles pour résoudre le processus d'optimisation derrière les métabolismes.

Le reste du document est structuré comme suit. La section 2 décrit les métaheuristiques utilisées à des fins d'analyse dans cette recherche ; en particulier, il présente la définition générale et les constituants. La section 3 détaille les nouveaux modèles d'optimisation pour le

Machine Translated by Google

réseaux métaboliques proposés dans ce travail. La section 4 discute des fonctionnalités originales incluses dans la conception du NSGAII et du MOEA/D utilisés pour résoudre les modèles d'optimisation proposés. Les sections 5 et 6 décrivent la conception des expériences menées pour tester les modèles d'optimisation proposés et leurs solutions métaheuristiques ; il contient les définitions des cas d'études ainsi que les expériences et les résultats, se terminant par une brève discussion des données observées.

2. Matériels et méthodes

Ce travail considère les algorithmes NSGAII et MOEA/D comme métaheuristiques envisagées pour aborder le problème d'optimisation dans les réseaux métaboliques. La notion générale de conception de tels algorithmes est fournie dans la suite de cette section.

2.1. Algorithme NSGAII

NSGAII est un algorithme évolutif multi-objectif (MOEA) qui utilise le tri non dominé et la distance de regroupement pour exercer une pression sélective vers le front de Pareto (24). La métaheuristique fait évoluer une population initiale de parents P0 en utilisant des opérateurs génétiques calculables courants tels que la mutation, le croisement et la sélection en tournoi qui créent une nouvelle progéniture, Qt , à chaque génération. Pour maintenir l'élitisme, la population actuelle, Pt , est combinée avec la progéniture Qt , et une nouvelle population Pt+1 est choisie en fonction des rangs de non-domination et de la distance de surpopulation pour se diversifier et rompre les liens si nécessaire.

Les trois composants principaux de NSGAII sont son approche de tri rapide non dominée, la procédure d'estimation de distance rapide et encombrée et l' opérateur de comparaison simple encombré. La méthode générale dérivée de NSGAII peut être décrite comme suit :

1. Initialisation d'une population Po de taille N en utilisant une distribution uniforme.
2. Créez une population de progéniture Qt en utilisant une sélection de tournoi binaire basée sur un opérateur de comparaison de foule, un croisement et une mutation effectués sur la population parentale (Pt), où l'indice « t » désigne le nombre de générations. La population descendante et sa population parentale sont combinées pour produire la population entière Rt , Rt = Pt + Qt . La population Rt sera de taille 2N.
3. Effectuer une approche de tri rapide non dominé sur l'ensemble de la population Rt pour identifier différents fronts de fonctions objectives. F = rapide − non dominé − sort(Rt), où F = (F0, F1, F2, . . .) aura dans F0 l'ensemble non dominé de solutions de Rt qui se rapproche le mieux de la frontière de Pareto.
4. Construire une nouvelle population parentale (Pt+1) de taille N à partir des fronts obtenus (Fi). Cette population de taille N est désormais utilisée pour la sélection, le croisement et la mutation afin de créer une nouvelle population (Qt+i) de taille N.
5. Le processus doit être répété jusqu'à ce que le nombre maximum d'itérations soit atteint.

2.2. MOEA/D

MOEA/D est une stratégie basée sur la décomposition du problème d'optimisation multiobjectif MOP (tel que défini dans [27]) en un certain nombre de sous-problèmes d'optimisation scalaire optimisés simultanément. Chaque sous-problème est optimisé à l'aide d'informations provenant exclusivement de ses sous-problèmes voisins, ce qui permet d'obtenir moins de complexité informatique à chaque génération. Il existe plusieurs approches pour transformer un problème multi-objectif en un nombre scalaire de problèmes d'optimisation. L'une des approches les plus populaires est la MOEA/D proposée par [28], où les problèmes d'optimisation scalaire peuvent être formulés comme suit :

$$g^{te}(y \mid wi, z) = \max\{wi \mid fk(x) - zk \mid\}, \quad \text{Sous réserve de : } x \quad \Omega \tag{2}$$

où w = (w1, w2, . . . , wN) est un vecteur de poids et wi ≥ 0 pour tout i = 1, . . . , k, et l'ensemble z = {z1, z2, . . . , zk} est le point de référence, où zi = max{ fi(x) | × ω } pour i = 1, . . . , m.

Machine Translated by Google

Pour chaque point optimal de Pareto il existe un vecteur de poids w, où x * est l'optimal
x solution de 2.

MOEA/D fonctionne sur un certain nombre de générations, et au cours de chaque génération, il exerce également une pression sélective vers le front de Pareto en utilisant des opérateurs génétiques tels que la mutation et le croisement. L'élément clé de cette stratégie est que la progéniture remplace les parents en fonction de la fonction scalaire et de leur proximité à chaque itération jusqu'à ce qu'un front final soit formé.

délivré par l'algorithme.

3. Approches proposées Ce

travail propose trois nouveaux modèles d'optimisation pour les réseaux métaboliques qui étendent FBA à un problème d'optimisation multi-objectifs. Les modèles, appelés MOFBA2, MOFBA3 et MOFBA4, représentent des améliorations par rapport au MOFBA1 proposé dans [18] et représenté dans l'équation (3). MOFBA1 optimise simultanément un ensemble de bioproduits {vb1 , . . . vbm} au lieu d'un seul, maintient dans les limites les flux de réaction et satisfait la condition d'état stationnaire, c'est-à-dire qu'ils garantissent que Sv = 0 (où S est la matrice stoechiométrique et v est le vecteur de flux).

$$\begin{aligned} &\text{MOFBA max } F(v) = \{vb1 , \ldots , vbm \} \\ &\quad \text{Sous réserve de} \\ &\qquad S \cdot v = 0 \\ &\qquad LBj \le vj \le UBj , \quad j \in \{1, \ldots , n\} \end{aligned} \tag{3}$$

MOFBA1 est le problème d'optimisation résultant de la mise en œuvre directe du problème défini dans l'équation (3). Elle considère autant de fonctions objectives que d'ensembles de métabolites d'intérêt. De même, il considère autant de variables de décision que de flux de réaction nécessaires pour définir le système métabolique. Notez que, pour une approche d'optimisation, l'espace de recherche dépend des variables de décision, et sur la base de cette définition, il y en a une pour chaque flux possible, c'est-à-dire qu'une métaheuristique doit rechercher des valeurs de flux appropriées dans les limites fournies de n variables de décision distinctes.

Les conditions décrites dans le paragraphe précédent caractérisent un piège courant dans la conception de stratégies de solution pour des problèmes d'optimisation. La difficulté apparaît parce que les métaheuristiques peuvent nécessiter des temps d'exécution plus longs pour localiser des solutions réalisables lorsque le nombre de variables de décision est grand. Ce travail considère cette situation et propose trois nouveaux modèles d'optimisation pour les réseaux métaboliques qui réduisent l'espace de recherche (c'est-à-dire le nombre de variables de décision qu'une métaheuristique utilise dans la recherche). Ces modèles sont intégrés dans un plan expérimental approprié pour démontrer que, à des fins de qualité, il est important de choisir le modèle choisi pour résoudre certains problèmes.

Bien que des détails sur les expériences soient fournis dans d'autres sections, le reste de cette section contient une description approfondie des trois nouveaux MOFBA proposés dans ce travail, avec un résumé de leur pertinence et de leur impact à la fin.

3.1. MOFBA2

Dans MOFBA2, comme dans MOFBA1, les fonctions objectifs à optimiser sont les ensembles de métabolites d'intérêt ; par conséquent, le nombre d'objectifs, m, est le même. D'autre part, MOFBA2 considère un ensemble réduit de variables de décision constitué uniquement des flux de réaction {vb1 , . . . vbm} associé aux mêmes métabolites d'intérêt présents dans la fonction objectif, et un autre, v, qui indique lequel des métabolites d'intérêt mène la recherche. En d'autres termes, MOFBA2 possède m + 1 variables de décision au lieu de n. L'équation (4) définit formellement MOFBA2.

$$\begin{aligned} &\text{MOFBA max } F(v) = \{vb1 , \ldots , vbm \} \\ &\quad \text{Sous réserve de} \\ &\qquad FBA(vbk, v) \text{est réalisable} \\ &\qquad LBj \le vb\,je \le UBj , \; 1 \le i \le k \le m \in \{1, \ldots , m\} \end{aligned} \tag{4}$$

Machine Translated by Google

MOFBA2 décrit un modèle d'optimisation à deux niveaux dans lequel le modèle interne optimise le principal métabolite d'intérêt, v, à l'aide de FBA et délimite les limites des métabolites d' intérêt à ceux définis dans le modèle externe. Les limites des réactions restantes sont supposées connues et fixées en fonction du réseau métabolique analysé.

3.2. MOFBA3

MOFBA3 propose un modèle de substitution pour optimiser les métabolites d'intérêt. Le modèle de substitution cherche à améliorer deux indicateurs bien connus : l'hypervolume (HV) et la distance générationnelle (GD). Ces indicateurs reflètent dans quelle mesure une solution converge vers le front de Pareto. Alors que le HV doit être maximisé, le GD doit être minimisé.

MOFBA3 a m + 1 variables de décision, les mêmes que MOFBA2, c'est à dire les flux de réaction {vb1 , . . . vbm} associé aux mêmes métabolites d'intérêt présents dans la fonction objectif et au principal flux de métabolites v. D'autre part, le nombre d'objectifs est toujours de 2, quel que soit le nombre de métabolites d'intérêt considérés. La particularité de ce modèle est sa substitution ; au lieu de rechercher directement les valeurs de flux appropriées sur les métabolites d'intérêt, il utilise des indicateurs de performance dans un contexte multi-objectifs (c'est-à-dire les indicateurs HV et GD). L'équation (5) définit formellement MOFBA3.

$$\begin{aligned} &\text{MOFBA min } F(v) = -1HV(\{vb1 , \ldots , vbm \}, R), DG(\{vb1 , \ldots , vbm \}, Zr) \\ &\quad \text{Sous réserve de} \\ &\qquad \text{FBA}(vbk, v)\text{est réalisable} \\ &\qquad LBj \le vb\, i \le UBj , \quad i \quad \{1, \ldots , m\}\ 1 \le k \le m \end{aligned} \tag{5}$$

Ce problème d'optimisation suppose qu'il existe un point de référence, R, et un ensemble de référence, Zr . Étant donné que la gestion de FBA était sous le package COBRApy, compte tenu de ses limites, le point de référence considéré pour ce travail est R = {1000, 1000, . . . , 1000}. De plus, étant donné la disponibilité d'une implémentation FBA grâce au même package, l'ensemble de référence Zr est formé par l'ensemble des solutions optimales formées par celles obtenues lors de la résolution de FBA à l'optimalité, ayant chaque métabolite d'intérêt comme biomasse optimisée.

3.3. MOFBA4

MOFBA4 est le dernier problème d'optimisation proposé combinant les idées de MOFBA2 et MOFBA3. Autrement dit, il propose d'optimiser non seulement les métabolites d'intérêt mais également les indicateurs HV et GD. Le nombre de variables de décision pour ce modèle reste égal à m + 1 et le nombre d'objectifs est m + 3. L'équation (6) définit formellement ce problème d'optimisation.

$$\begin{aligned} &\text{MOFBA min } F(v) = -1HV(\{vb1 , \ldots , vbm \}, R), DG(\{vb1 , \ldots , vbm \}, Zr)(vb1 , \ldots , vbm) \\ &\quad \text{Sous réserve de} \\ &\qquad \text{FBA}(vbk, v)\text{est réalisable} \\ &\qquad LBj \le vb\, je \le UBj , 1 \quad i \quad \{1, \ldots , m\} \\ &\qquad \le k \le m \end{aligned} \tag{6}$$

3.4. Le

tableau d'analyse 1 résume les caractéristiques les plus notables des modèles d'optimisation proposés dans ce travail et les compare au MOFBA1 proposé dans [18]. L'espace de recherche est considérablement réduit dans les nouveaux modèles, et certains d'entre eux utilisent des informations de convergence dans leur définition. Les caractéristiques uniques démontrent la richesse des modèles pouvant être conçus pour résoudre un problème d'optimisation spécifique.

Machine Translated by Google

Tableau 1. Caractéristiques pertinentes des modèles MOFBA pour les réseaux métaboliques.

Modèle	N° Variables de décision	N° Objectifs	Substitut
MOFBA1	n	m	Non
MOFBA2	m + 1	m	Non
MOFBA3	m + 1	2	Oui
MOFBA4	m + 1	m + 2	Oui

Les MOFBA proposés ne peuvent pas être résolus avec des solveurs linéaires traditionnels tels que FBA. Les alternatives consistent à utiliser des schémas énumératifs ou des approches approximatives qui permettent un pour obtenir des solutions appartenant à la frontière optimale de Pareto. En ce sens, cette recherche analysé l'utilisation de métaheuristiques qui intègrent FBA dans leur processus de recherche comme une approche de solution appropriée car elles améliorent leur approximation du front de Pareto dans chaque itération.

4. Conceptions métaheuristiques

Cette section présente les détails particuliers requis dans cette recherche pour la mise en œuvre des métaheuristiques NSGAII et MOEA/D afin de résoudre les quatre problèmes d'optimisation. problèmes MOFBA1, MOFBA2, MOFBA3 et MOFBA4. Ces métaheuristiques sont basées sur les cadres NSGAII et MOEA/D.

Les métaheuristiques considérées nécessitent la définition des caractéristiques suivantes : (1) schémas de codage ; (2) fonction d'évaluation de la condition physique ; (3) les opérateurs génétiques ; et (4) contrainte stratégie de gestion. La méthode d'initialisation de la population pour les deux stratégies (NSGAII et MOEA/D) est aléatoire. La conception proposée pour que le reste de ces composants gère les nouveaux MOFBA sont détaillés dans le reste de cette section. Les adaptations romanes pour les frameworks NSGAII et MOEA/D incluent la représentation calculable intelligente de solutions associées aux schémas de codage.

4.1. Schémas de codage

Ce travail propose l'utilisation d'ensembles de codage de solution distinctifs pour chaque MOFBA (comme défini dans les équations (3) à (6). Les schémas de codage impliquent la définition d'une structure de données cela représente une solution du réseau métabolique. Le script développé pour l'expérimentation se trouve dans ce dépôt Github (https://github.com/multiobjectiveoptimization2 /MOFBA, consulté le 24 juin 2024).

Pour MOFBA1, la structure des données est un vecteur à valeur réelle, V. Le schéma de codage considère un réseau métabolique, MN, constitué d'un ensemble de réactions, V, et de deux sous-ensembles. V^M , V^b $\subseteq$ V, où $V^M \cap V^b = \emptyset$, qui représentent les réactions des métabolites de intérêt pour un décideur. De plus, soit v = (v1, . . . , vn) le vecteur flux pour V et supposons qu'il existe des limites inférieure et supérieure initiales, LBi , UBi , pour chaque vi , 1 ≤ i ≤ n. Ensuite, le schéma de codage W propose de redéfinir les limites de chaque vi associé avec une réaction en V^M V^b en utilisant deux valeurs (li , Δi). Les nouvelles limites sont calculées comme $LB_{nouveau}$ = li et $UB_{nouveau}$ = (UBi − li)Δi + li . Tous les flux restants garderont leurs limites inchangé. En d'autres termes, la solution code les changements de limites pour que FBA résout MN en utilisant un bioproduit prédéfini, qui dans ce travail est supposé être v^b_1 . Le résultat le vecteur codant W est de taille O(n), asymptotique en nombre de réactions.

Pour MOFBA2, MOFBA3 et MOFBA4, un vecteur de taille m + 1 est considéré comme le schéma de codage, où m est le nombre de fonctions objectives. Les m premiers éléments du vecteur sont des variables réelles dont la valeur représente la limite supérieure du flux pour chaque métabolite d'intérêt dans les réactions associées. L'élément supplémentaire est un objectif unique sélecteur d'optimisation qui prend des valeurs comprises entre 1 et m, indiquant quel métabolite est va être optimisé à son tour. Lorsqu'il s'agit d'une réaction réversible, la valeur dans les valeurs indiquées La variable sera la même pour la borne inférieure (mais négative).

Machine Translated by Google

4.2. Fonction d'évaluation de la condition physique

Parmi tous les problèmes d'optimisation MOFBA, les fonctions d'évaluation de la condition physique (ou FEA) sont considérées comme un sous-ensemble dérivé de l'ensemble composé des valeurs des flux de métabolites d'intérêt, tels que les métriques Hypervolume et Distance générationnelle. Les informations requises sur les bioproduits étant associées à des réactions spécifiques, la pertinence d'une solution obtenue par métaheuristique sur les MOFBA est évaluée en considérant leurs valeurs de flux. Dans MOFBA1 et MOFBA2, les critères ou objectifs à optimiser seront b et de-être les flux de réaction correspondant aux fonctions bioproduits $v_{1,\ldots,}^{b}{}^{b}$ m). Dans MOFBA3, l'Hypervolume et la Distance de Génération obtenus d'intérêt choisis dans V notés comme (v d'une solution, le point de définis comme suit, sont référence R, et l'ensemble de référence Zr , optimisé. Le point R est la pire valeur extrême possible des flux, qui est de 1000 pour tout métabolite d'intérêt, compte tenu de sa définition pour FBA, dans des plateformes largement utilisées, telles que CobraPy. L'ensemble Zr est composé de trois points, dont. les flux optimaux obtenus en résolvant le cas en question à l'aide de la méthode FBA en optimisant individuellement chaque métabolite d'intérêt donc, s'il y a n fonctions objectives, Zr aura une cardinalité de n. dans MOFBA2 à MOFBA4, celui-ci est choisi dérivé de la valeur d'une des variables de décision considérées, comme décrit précédemment.

4.3. Opérateurs génétiques

Ces opérateurs créent de nouvelles solutions en faisant varier dynamiquement et aléatoirement les valeurs des variables de décision dans les solutions existantes. Cette sélection était due à son succès dans la résolution de problèmes impliquant des variables de décision avec des valeurs réelles [29]. Les opérateurs choisis pour NSGAII sont respectivement la mutation, le croisement et une sélection aléatoire simple mais fiable . Les valeurs spécifiques de ces paramètres sont tirées de la littérature et sont présentées dans le tableau 2.

Tableau 2. Paramètres spécifiques à la mise en œuvre de NSGAII.

Paramètre	Valeur
Mutation polynomiale	Probabilité = 1,0/d, où d est le nombre de variables de décision. Indice de distribution = 20
SBXCrossover	Probabilité = 100 % Indice de distribution = 20
Critère d'arrêt	jusqu'à atteindre 100 000 évaluations
Taille de la population	100

Les opérateurs choisis pour MOEA/D pour la mutation et le crossover étaient la mutation polynomiale [30], avec crossover par évolution différentielle. La stratégie de sélection est simple mais fiable, et la fonction d'agrégation utilisée est la distance de Tshebycheff. Pour une référence plus complète sur les opérateurs, voir [28]. Les valeurs spécifiques de ces paramètres sont tirées de la littérature et sont présentées dans le tableau 3.

Tableau 3. Paramètres spécifiques à la mise en œuvre de MOEA/D.

Paramètre	Valeurs
Mutation polynomiale	Probabilité = 1,0/d, où d est le nombre de variables de décision. Indice de distribution = 20
Evolution différentielle	CR = 1 F = 0,5 K = 0,5
Critère d'arrêt	À la fin de 100 000 évaluations
Taille de la population	100

Machine Translated by Google

4.4. Stratégie de gestion des contraintes

Ce travail utilise la méthode de gestion des contraintes proposée dans [31] pour générer une pression sélective vers des solutions réalisables. À mesure que les générations évoluent dans les deux métaheuristiques, la compétition entre les solutions préférera toujours la solution réalisable malgré l' état de non-domination. À long terme, une telle stratégie tend à éradiquer les solutions irréalisables dans le rapport final sur l'algorithme. L'optimisation multi-objectifs est utilisée lorsqu'il y a plusieurs objectifs à optimiser simultanément. Il existe plusieurs algorithmes évolutifs multi-objectifs (MOEA), tels que NSGAII et MOEA/D. Bien qu'ils soient utilisés pour optimiser des problèmes multi-objectifs, ils sont très différents. NSGAII est un algorithme de classification non dominé , tandis que l'algorithme MOEA/D est basé sur la décomposition.

5. Plan d'expériences Cette

sous-section présente l'ensemble des expériences réalisées afin de valider l' application des modèles d'optimisation proposés comme outils pour améliorer la compréhension des métabolismes des microalgues. Dans le domaine de la recherche de solutions efficaces à des problèmes multi-objectifs, des expériences ont été menées sur deux réseaux de microalgues Chlorella vulgaris pour évaluer les performances de différents algorithmes et des MOFBA respectifs. Les sous- sections présentent le cas des études, la conception expérimentale et les détails du logiciel utilisé dans la mise en œuvre, dans le but de vérifier les hypothèses suivantes.

Hypothèses 0 (H0). Cela n'est pas pertinent pour la sélection d'un modèle et/ou d'un algorithme de solution pour optimiser les flux dans un réseau métabolique.

5.1. Cas d'étude Par

rapport à une enquête précédente [18], deux réseaux, le métabolisme du glutamate et la distribution du flux pigmentaire de la microalgue Chlorella vulgaris, ont été inclus dans les deux études de cas ; les réactions réversibles et irréversibles ont été ajoutées, la représentation d'une réaction réversible utilisant les intervalles de flux de limite inférieure de −1 000 et de limite supérieure de 1 000, et les réactions irréversibles avec des intervalles de limite inférieure de 0 à la limite supérieure de 1 000. De plus, des nœuds ont été inclus où les métabolites bifurquent vers différentes routes et cycles fréquemment présentés dans le métabolisme des cellules.

5.1.1. Cas d'étude 1 : Réseau métabolique Chlorella vulgaris

Le réseau métabolique de la microalgue Chlorella vulgaris [17] a été étudié à l'aide des algorithmes NS-GAII et MOEA/D pour trois conditions de culture différentes : photoautotrophie (lumière + composants), hétérotrophie (composant) et mixotrophie (CO2 + lumière + composant) .). Parmi les composés utilisés comme nutriments pour la culture figuraient l'ajout de sources d'azote, telles que NO3 et NH4, ainsi que de sulfates, tels que SO4, Fe2 et magnésium. Les différentes sources de cultures affectent la production de métabolites. La figure suivante montre la répartition des pigments dans la microalgue C. vulgaris.

Dans ce cas, une partie du réseau de distribution des pigments sera étudiée puisque la distribution dans les microalgues telles que C. vulgaris est d'une grande importance dans l'étude de la synthèse des pigments. Les réactions impliquées dans le métabolisme sont présentées à l'Annexe A.1, Tableau A1. Ce réseau comprend la complexité des réactions, des nœuds et des cycles réversibles et irréversibles (FRDPth, GRDPth), et sont illustrés dans la figure 1.

Machine Translated by Google

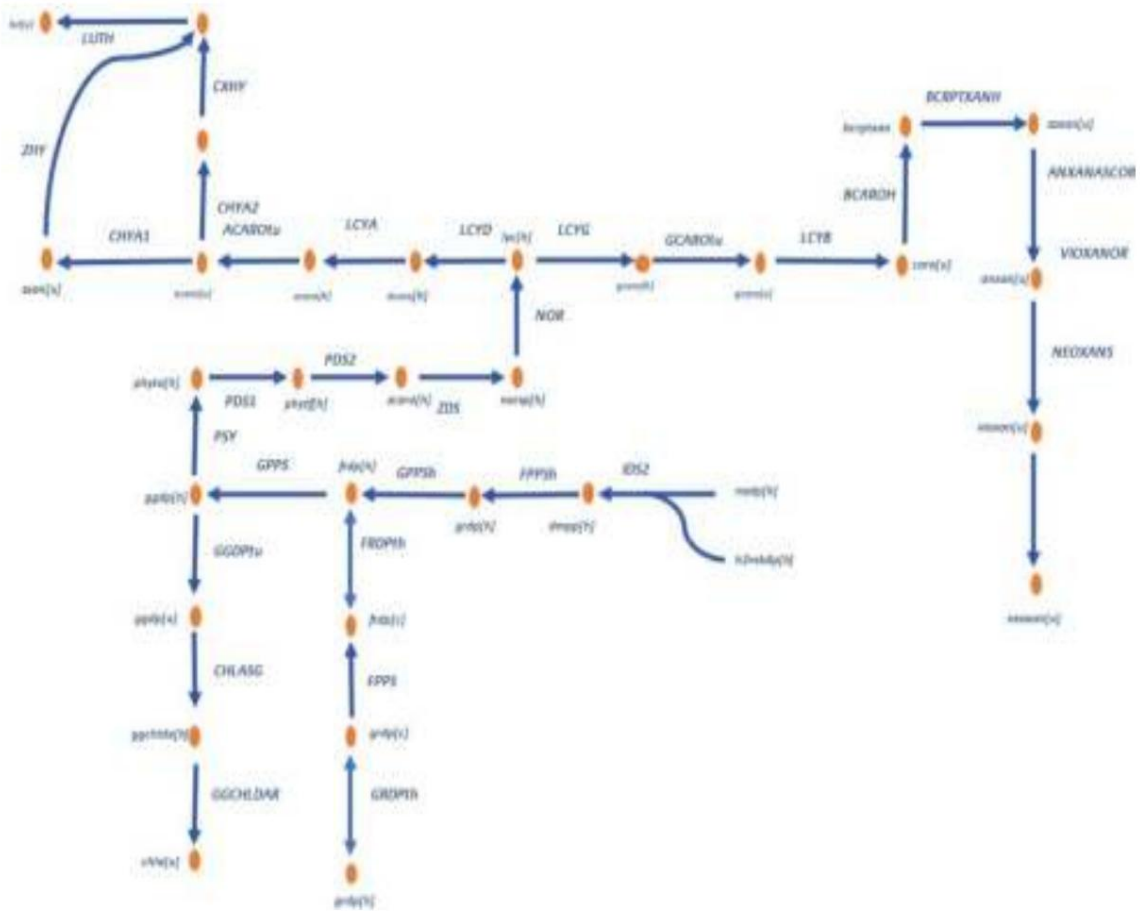


Figure 1. Distribution du flux des voies de byosinthèse pigmentaire [17].

5.1.2. Cas d'étude 2 : Optimisation multiobjectif du réseau métabolique du métabolisme Glutamate de la microalgue Chlorella vulgaris Le

métabolisme est constitué de différentes voies métaboliques qui s'entrelacent pour en former une plus complexe. L'une de ces voies est la distribution des flux du métabolisme du glutamate, qui remplit différentes fonctions, notamment la synthèse des acides aminés.

Ce réseau métabolique représente une grande complexité en raison du nombre de métabolites qui se ramifient au niveau du nœud central et de la présence de réactions réversibles telles que ASPATh et ASPNA1Th. Ce réseau a été évalué dans trois conditions de croissance différentes, autotrophie, hétérotrophie et mixotrophie, en utilisant les algorithmes NSGAII et MOEA/D avec leurs quatre MOFBA. Ce métabolisme est d'une grande importance car les voies de production de différents produits d'intérêt, des acides aminés comme la tyrosine, la valine, la leucine, etc., sont impliqués, qui peuvent ensuite être utilisés pour produire des protéines. Les réactions impliquées dans le métabolisme sont présentées à l'Annexe A.1, Tableau A2. La figure 2 représente la répartition des flux associés au métabolisme du glutamate dans le chloroplaste et le cytoplasme.

Machine Translated by Google

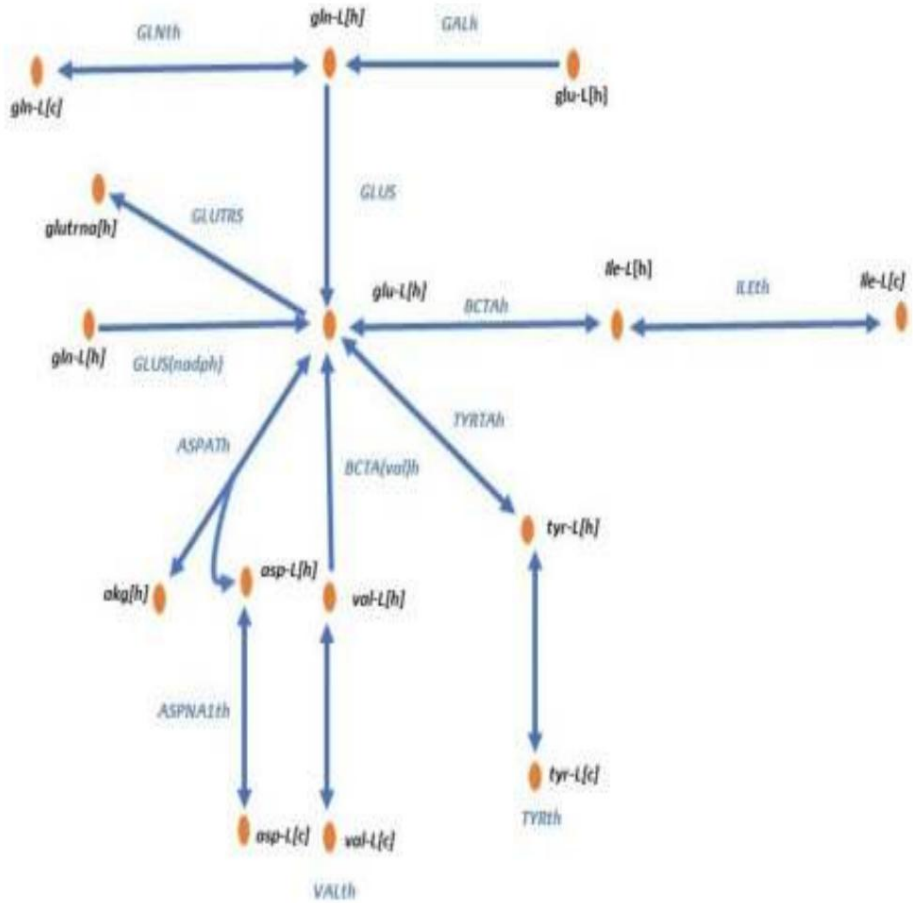


Figure 2. Distribution des flux associés au métabolisme du glutamate dans le chloroplaste et le cytoplasme [17].

5.2. Définition des expériences

Définir une méthodologie basée sur des métaheuristiques qui améliorent les informations sur les flux du réseau métabolique fournies par la méthode FBA en redéfinissant l'analyse de l'équilibre des flux comme un problème d'optimisation multi-objectifs est possible ; quatre expériences ont été proposées, résumées dans le tableau 4 suivant .

Tableau 4. Conception d'expériences pour démontrer le soutien métaheuristique à la compréhension du métabolisme des microalgues.

Objectif de l'expérience		Variables impliquées
Expérience1	L'utilisation de différents modèles d'optimisation produit des résultats différents.	C. vulgaris, NSGAII, MOFA1 , MOFBA2 , MOFBA3 , MOFBA4 .
Expérience2	Démontrer que l'utilisation de métaheuristiques soutient la compréhension du métabolisme des microalgues.	C. vulgaris, Expédié par Amazon, NSGAII, MOFBA4
Expérience3	Il existe des algorithmes d'optimisation plus adaptés à la résolution de problèmes spécifiques	C. vulgaris, NSGAII, MOEA/D, MOFBA4 .
Expérience 4 Valider qu'une sélection aléatoire n'est pas suffisante.		C. vulgaris, NSGAII, FBA, aléatoire.

Machine Translated by Google

L'expérience 1 a démontré que différents modèles d'optimisation multi-objectifs offrent des résultats différents. L'algorithme a été réglé sur NSGAII, les microalgues ont été modifiées pour démontrer la polyvalence de l'approche et enfin, les quatre modèles d'optimisation proposés ont été analysés. Pour les cas étudiés, il a été observé que le meilleur modèle était MOFBA4, qui a été utilisé dans les expériences ultérieures.

Dans l'expérience 2, l'algorithme, les microalgues et le modèle FBA mono-objectif ont été comparés au modèle multi-objectif défini par MOFBA4. Il a été observé que le métabolisme d'une microalgue peut être décrit avec différents flux, et non avec un seul, et que ceux-ci peuvent être contrôlés pour obtenir des informations sur différents métabolites d'intérêt. Ceci confirme la facilité d'adaptation des méthodes proposées à différents types de réseaux métaboliques, en considérant différentes configurations.

L'expérience 3 a évalué les performances des algorithmes NSGAII et MOEA/D sur le même problème d'optimisation. Il a été observé que, pour les trois objectifs (c'est-à-dire les trois métabolites d'intérêt) considérés, NSGAII était le meilleur. Ce résultat est cohérent avec la littérature, étant donné que pour deux ou trois objectifs, NSGAII montre de meilleures performances que MOEA/D. Cela laisse ouverte la question de savoir si MOEA/D s'améliorera pour un plus grand nombre d'objectifs, ce qui constitue une piste d'investigation ouverte pour son application dans l'étude des réseaux métaboliques de microalgues.

L'expérience 4 montre que le simple échantillonnage aléatoire n'est pas suffisant pour obtenir une meilleure distribution de solutions, ce qui est possible grâce à l'utilisation de métaheuristiques.

Les métaheuristiques NSGAII et MOEA/D ont été implémentées à l'aide du framework jMetalPy [32]. Les modèles d'optimisation ont été développés en Python et utilisés dans le cadre de l'implémentation FBA fournie par le package COBRApy [33]. Les graphiques ont été recréés à l'aide de l'interface pyplot de matplotlib [34]. L'ordinateur utilisé pour exécuter les expériences dispose d'un processeur 64 bits 2,6 GHz avec 32 mémoire RAM.

6. Résultats

Cette section résume les données obtenues à la suite de la mise en œuvre des expériences décrites à la section 3.2. À la fin, il propose une discussion sur les objectifs atteints dans la recherche. Pour visualiser les résultats des figures 3 à 6, la bibliothèque matplotlib a été importée.

6.1. L'expérience 1

En [18] a montré que NSGAII présente des solutions de meilleure qualité que le FBA classique pour trois fonctions objectives. La figure 3 montre les résultats de l'expérience où sont testées les quatre variantes décrites ci-dessus, MOFBA1, MOFBA2, MOFBA3 et MOFBA4, pour l'algorithme NSGAII, avec l'optimisation de trois fonctions objectives de la distribution du flux pigmentaire de la microalgue Chlorella vulgaris. On peut observer que les différents MOFBA proposent des comportements différents les uns des autres ; MOFBA1, MOFBA3 et MOFBA4 améliorent FBA. Cependant, le meilleur comportement de solution était MOFBA4 sur la figure 3d, car il fournit davantage de solutions non dominées et maintient une bonne diversité de population pour un environnement de même taille de population.

La figure 3 montre la distribution du flux pigmentaire dans la microalgue C. vulgaris ; on peut voir que chaque variante de MOFBA offre des comportements différents et améliore toutes le FBA sur les figures 3a à c, mais le meilleur comportement des solutions peut être observé dans la variante de MOFBA4 sur la figure 3d.

Machine Translated by Google

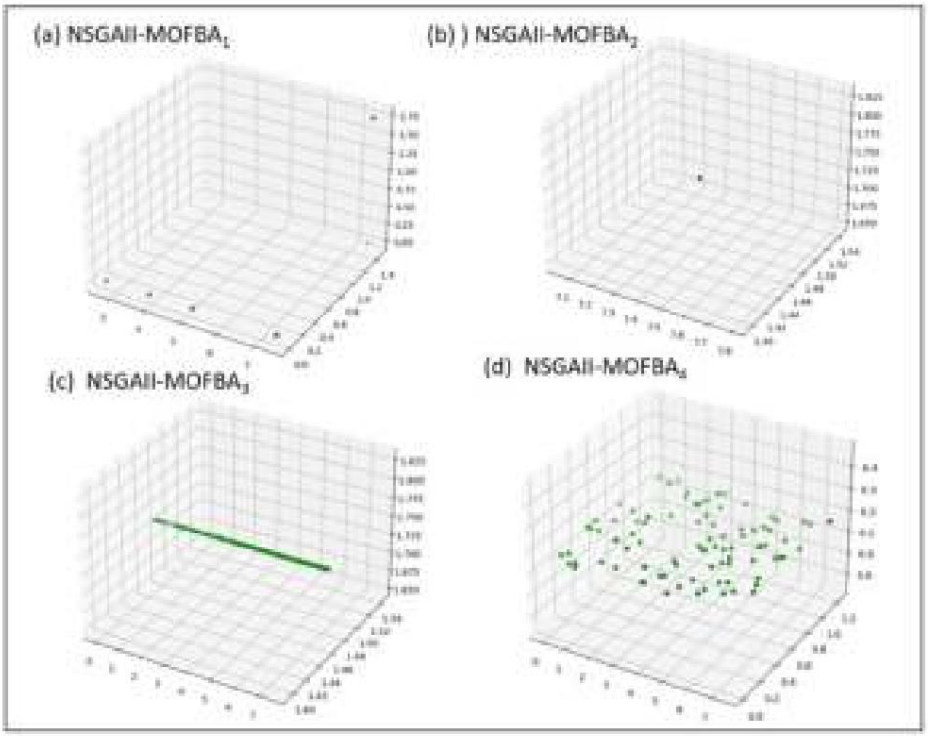


Figure 3. Comparaison entre l'algorithme NSGAII et les variantes (a) MOFBA1 , (b) MOFBA2 , (c) MOFBA3 et (d) MOFBA4 dans la distribution des flux pigmentaires.

6.2. Expérience

2 En comparant les performances de l'algorithme NSGAII avec la variante MOFBA4 , figure 4a, et l'optimisation classique FBA à objectif unique de la figure 4b, on peut observer que l' algorithme NSGAII-MOFBA4 présente une supériorité en démontrant que les informations fournies sont amélioré en fournissant plus de solutions et, surtout, une distribution considérablement améliorée. Cette distribution améliorée est particulièrement évidente sur les figures 3d et 4a avec deux réseaux métaboliques différents étudiés, tels que le métabolisme du glutamate et la distribution des flux pigmentaires dans la microalgue C. vulga

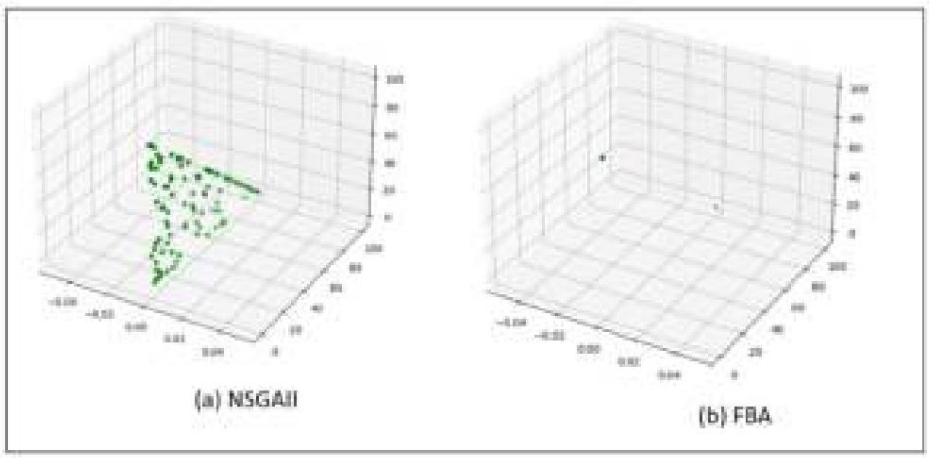


Figure 4. Comparaison entre (a) NSGAII-MOFBA4 et (b) FBA dans la distribution des flux associés au métabolisme du glutamate de la microalgue Chlorella vulgaris.

6.3. Expérience

3 La figure 5 présente l'évaluation entre les algorithmes NSGAII, Figure 5a, c, et MOEA/D, Figure 5b, d, dans la distribution des flux associés au métabolisme du glutamate.

Machine Translated by Google

et le réseau de distribution et les flux pigmentaires avec NSGAII, Figure 5c, et MOEA/D, Figure 5d, de la microalgue C. vulgaris avec la variante de l' algorithme MOFBA4. Il a été démontré que l'algorithme NSGAII offre plus de solutions et une meilleure diversité de population que MOEA/D. Les métaheuristiques sont importantes ; dans ce cas, NSGAII est le meilleur, ce qui est cohérent avec la littérature [35], car cet algorithme fonctionne bien avec 2 et 3 o

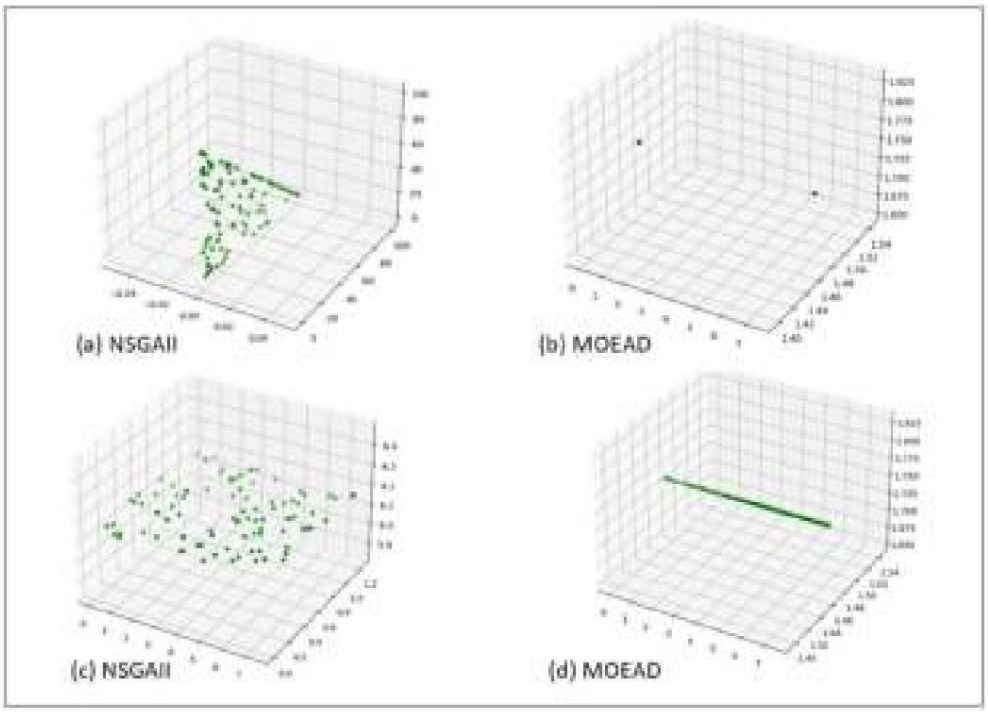


Figure 5. Comparaison entre les algorithmes (a,c) NSGAII et (b,d) MOEA/D dans le réseau de distribution des pigments et dans la distribution des flux associés au métabolisme du glutamate de la microalgue Chlorella vulgaris.

6.4. Expérience

4 En plus de tester les différentes études de cas avec les approches FBA, NSGAII et MOEA/D, une expérimentation a été réalisée selon une approche aléatoire rapide, que nous appelons aléatoire, dans la répartition du flux pigmentaire chez la microalgue C. vulgaris. . La figure 6 montre la comparaison entre FBA, figure 6a, aléatoire, figure 6c, et NSGAII, figure 6b. La méthode aléatoire, bien que rapide, ne pouvait pas offrir de meilleurs résultats, figure 6b, par rapport à ce qui est présenté dans la figure 6c, NSGAII.

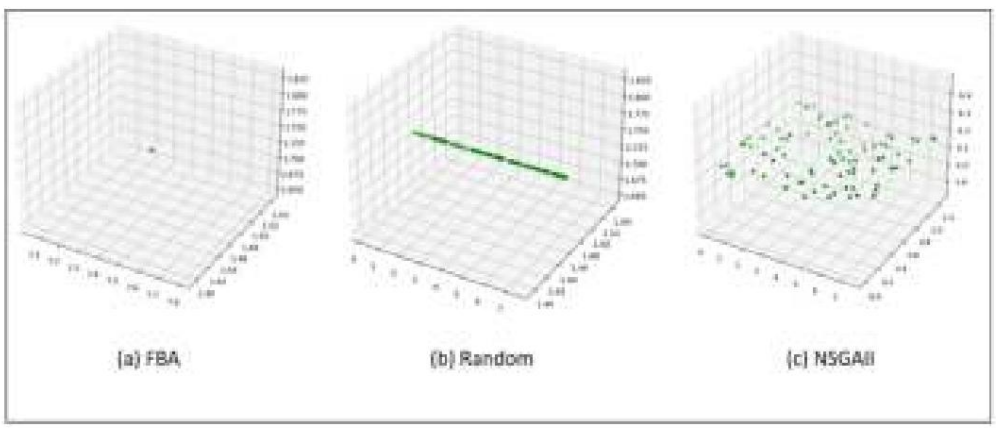


Figure 6. Comparaison entre les variantes (a) FBA, (b) aléatoire et (c) NSGAII chez la microalgue C. vulgaris.

Machine Translated by Google

6.5. Analyse statistique

Les tableaux 5 et 6 montrent que les méthodes proposées obtiennent des solutions réalisables ; il est montré que ils satisfont aux conditions identifiées dans [17] et FBA, démontrant ainsi la corrélation de les solutions in silico et sa capacité à émuler les résultats dans différentes conditions de culture. Dans le tableau 6 on peut voir une comparaison entre les flux en mmol h−1 obtenus avec FBA et NSGAII-MOFBA4 et la distance euclidienne présentée entre eux. Le tableau 5 démontre que NSGAII présente une grande polyvalence pour limiter les paramètres dans les conditions de croissance grâce à les valeurs des limites inférieure et supérieure, en plus de pouvoir simuler des cycles et bifurcations entre réseaux métaboliques. De même, quelques solutions réalisables correspondant à NSGAII avec la variante MOFBA4 et les solutions obtenues à partir du FBA classique sont présentées.

Tableau 5. Répartition des flux en mmol h−1 obtenus par NSGA II, associés à la synthèse de pigments dans le métabolisme de la microalgue C. vulgaris.

Réaction	KG	UB	S1	S2	S3	S4
NADPH	0,000487	0,000487	0,000487	0,000487	0,000487	0,000487
IDS2	0	0,000487	7,42 ×10³	4,72 8	2,58 ×10−5	4,33 ×10−7
FPPSh	0	0,000487	7,42 ×10−5	4,72 ×10−5	2,58 ×10−5	4,33 ×10−7
GPPSh	0	0,000487	7,42 ×10−5	4,72 ×10−5	2,58 ×10−5	4,33 ×10−7
FRDPème	0	0	0	0	0	0
FPS	0	0	0	0	0	0
GRDPème	0	0	0	0	0	0
GRDPH	0	0	0	0	0	0
GGPPS	0	0,000487	7,42 ×10−5	4,72 ×10−5	2,58 ×10−5	4,33 ×10−7
v1	0	0,000402078	0	0	0	0
GGDPtu	−0,0000849	0,0000849	7,42 ×10−5	4,72 ×10−5	2,58 ×10−5	4,33 ×10−7
CHLASG	0	0,0000742	7,42 ×10−5	4,72 ×10−5	2,58 ×10−5	4,33 ×10−7
GGCHLDAR 0		0,0000742	7,42 ×10−5	4,72 ×10−5	2,58 ×10−5	4,33 ×10−7
CHLAU	0	0,0000742	7,42 ×10−5	4,72 ×10−5	2,58 ×10−5	4,33 ×10−7
PSY	0	2,24 ×10−8	0	0	0	0
PDS1	−2,24 ×10−8	2,24 ×10−8	0	0	0	0
PDS2	−2,24 ×10−8	2,24 ×10−8	0	0	0	0
ZDS	−2,24 ×10−8	2,24 ×10−8	1,59 ×10−8	1,54 ×10−8	1,06 ×10−8	3,58 ×10−9
NI	−2,24 ×10−8	2,24 ×10−8	1,59 ×10−8	1,54 ×10−8	1,06 ×10−8	3,58 ×10−9
v2	0,00	8h00 ×10−11	0	0	0	0
LCYG	−1,59 ×10−8	1,59 ×10−8	1,59 ×10−8	1,54 ×10−8	1,06 ×10−8	3,58 ×10−9
GCAROtu	−1,59 ×10−8	1,59 ×10−8	1,59 ×10−8	1,54 ×10−8	1,06 ×10−8	3,58 ×10−9
LCYB	−1,59 ×10−8	1,59 ×10−8	1,59 ×10−8	1,54 ×10−8	1,06 ×10−8	3,58 ×10−9
v3	0,00	1,24 ×10−8	1,24 ×10−8	1,20 ×10−8	7,10 ×10−9	1,00 ×10−10
BCAROH	−3,48 ×10−9	3,48 ×10−9	3,48 ×10−9	3,48 ×10−9	3,48 ×10−9	3,48 ×10−9
BCRPTXANH −3,48 ×10−9		3,48 ×10−9	3,48 ×10−9	3,48 ×10−9	3,48 ×10−9	3,48 ×10−9
v4	0,00	6,16 ×10−9	6,16 ×10−9	6,16 ×10−9	6,16 ×10−9	6,16 ×10−9
ANXANAS	−2,68 ×10−9	2,68 ×10−9	−2,68 ×10−9	−2,68 ×10−9	−2,68 ×10−9	−2,68 ×10−9
v5	0,00	3,88 ×10−9	3,88 ×10−9	3,88 ×10−9	3,88 ×10−9	3,88 ×10−9
VIOXANOR −2,41 ×10−9		2,41 ×10−9	−2,41 ×10−9	−2,41 ×10−9	−2,41 ×10−9	−2,41 ×10−9
v6	0,00	2,70 ×10−10	2,70 ×10−10	2,70 ×10−10	2,70 ×10−10	2,70 ×10−10
NÉOXANS	−1,47 ×10−9	1,47 ×10−9	1,47 ×10−9	1,47 ×10−9	1,47 ×10−9	1,47 ×10−9
NÉOXANU 0		1,47 ×10−9	1,47 ×10−9	1,47 ×10−9	1,47 ×10−9	1,47 ×10−9
LCYD	−6,42 ×10−9	6,42 ×10−9	0	0	0	0
LCYA	−6,42 ×10−9	6,42 ×10−9	0	0	0	0

Machine Translated by Google

Tableau 5. Suite

Réaction	KG	UB	S1	S2	S3	S4
v7	0,00	1,33 ×10−9	0	0	0	0
ACAROTu	−5,09 ×10−9	5,09 ×10−9	0	0	0	0
CHYA1	0	0	0	0	0	0
ZHY	0	0	0	0	0	0
CHYA2	0	5,09 ×10−9	0	0	0	0
CXHY	0	5,09 ×10−9	0	0	0	0
v8	0	3,35 ×10−9	−1,74 ×10−9	−1,74 ×10−9	−1,74 ×10−9	−1,74 ×10−9
LUTH	−1,74 ×10−9	1,74 ×10−9	1,74 ×10−9	1,74 ×10−9	1,74 ×10−9	1,74 ×10−9
LOROXANU 0		1,74 ×10−9	1,74 ×10−9	1,74 ×10−9	1,74 ×10−9	1,74 ×10−9
FRDPème	0	1	0	0	0	0

Tableau 6. Répartition des flux en mmol h−1 obtenus via FBA et l'algorithme NSGAII-MOFBA4 .

	Distance euclidienne entre Expédié par Amazon et NSGAII-MOFBA4			
Réaction.	Expédié par Amazon	MOFBA4	MOFBA4	MOFBA4
ACAROTu'	5,09 ×10−9	0	0	0
ANXANASCOR'	−2,68 ×10−9	−2,68 ×10−9	−2,68 ×10−9	−2,68 ×10−9
BCAROH'	3,48 ×10−9	3,48 ×10−9	3,48 ×10−9	3,48 ×10−9
BCRPTXANH'	3,48 ×10−9	3,48 ×10−9	3,48 ×10−9	3,48 ×10−9
CHLASG'	0	1,36106 ×10−9	2,35109 ×10−5	0,0000742
CHYA1'	0	0	0	0
CHYA2'	5,09 ×10−9	0	0	0
CXHY'	5,09 ×10−9	0	0	0
FPS'	0	0	0	0
FRDPth'	0	0	0	0
GCAROtu'	1,59 ×10−8	7,94367 ×10−9	1,35944 ×10−8	1,59 ×10−8
GGCHLDAR'	0	1,36106 ×10−9	2,35109 ×10−5	0,0000742
GGDPtu'	0,000148102	1,36106 ×10−9	2,35109 ×10−5	0,0000742
GGPS'	0.000486596	1,36106 ×10−9	2,35109 ×10−5	0,0000742
GRDPème'	0	0	0	0
IDS1'	0,001459788	1,36106 ×10−9	2,35109 ×10−5	0,0000742
LCYA'	6,42 ×10−9	0	0	0
LCYB'	1,59 ×10−8	7,94367 ×10−9	1,35944 ×10−8	1,59 ×10−8
LCYD'	6,42 ×10−9	0	0	0
LCYG'	1,59 ×10−8	7,94367 ×10−9	1,35944 ×10−8	1,59 ×10−8
LUTH'	1,74 ×10−9	1,74 ×10−9	1,74 ×10−9	1,74 ×10−9
LES NÉOXANS	1,47 ×10−9	1,47 ×10−9	1,47 ×10−9	1,47 ×10−9
NI'	2,24 ×10−8	7,94367 ×10−9	1,35944 ×10−8	1,59 ×10−8
PDS1'	2,24 ×10−8	0	0	0
PDS2'	2,24 ×10−8	0	0	0
PSY'	2,24 ×10−8	0	0	0
VIOXANOR'	−2,41 ×10−9	−2,41 ×10−9	−2,41 ×10−9	−2,41 ×10−9
ZDS'	2,24 ×10−8	7,94367 ×10−9	1,35944 ×10−8	1,59 ×10−8
ZHY'	0	0	0	0
Distance euclidienne		0.001545861	0,001514586	0,001451344

Machine Translated by Google

Cette section valide statistiquement qu'il existe une différence lors de l'utilisation de différents
des problèmes d'optimisation ou des algorithmes afin de montrer que le choix est pertinent. Faire
Ceci, il résume les résultats en comparant par Hypervolume (l'indicateur de proximité à
le front Pareto optimal) pour voir s'il existe ou non une différence significative entre le
les modèles d'optimisation MOFBA3 et MOFBA4 et les algorithmes NSGAII et MOEA/D.

La première analyse considère les modèles MOFBA3 et MOFBA4, définit la solution
algorithme vers NSGAII et évalue tous les réseaux. Pour l'analyse, l'Hypervolume a été
obtenu à partir de chacune des 30 exécutions de l'algorithme par problème. Utilisation du non paramétrique
Wilcoxon a signé le test de classement, avec un niveau de confiance de 95%, le H0 a été validé, ce qui
précise qu'il est impossible de définir une méthodologie basée sur des métaheuristiques qui améliore
l'information des flux du réseau métabolique fournie par la méthode FBA en redéfinissant
l'analyse de l'équilibre des flux en tant que problème d'optimisation multi-objectifs. Le tableau 7 résume
les résultats, montrant la valeur Hypervolume sur une échelle logarithmique par exécution pour chaque réseau
et l'état d'acceptation du H0 dans la dernière ligne. On voit que l'hypothèse est
rejeté dans presque tous les réseaux métaboliques analysés. À l'exception de /textitChlorella, il peut
Il convient de noter que le meilleur modèle d'optimisation est MOFBA4.

Tableau 7. Les données ont été analysées statistiquement pour valider les différences entre MOFBA3 et MOFBA4 .
L'hypothèse nulle, H0, a été acceptée lorsque la valeur p obtenue était inférieure à 0,05.

Métabolisme du glutamate chez C. vulgaris		Réseau de pigments chez C. vulgaris	
MOFBA4	MOFBA3	MOFBA4	MOFBA3
34.539	20.723	20h65	20.647
34.539	20.723	20h65	20.648
34.539	20.723	20h65	20.649
34.539	20.723	20h65	20.649
34.539	20.723	20h65	20.651
34.539	20.723	20h65	20.651
34.539	20.723	20h65	20.651
34.539	20.723	20h65	20.651
34.539	20.723	20h65	20.652
34.539	20.723	20h65	20.648
34.539	20.723	20h65	20.650
34.539	20.723	20h65	20.651
34.539	20.723	20h65	20.648
34.539	20.723	20h65	20.653
34.539	20.723	20h65	20.654
34.539	20.723	20h65	20.652
34.539	20.723	20h65	20.651
34.539	20.723	20h65	20.649
34.539	20.723	20h65	20.651
34.539	20.723	20h65	20.651
34.539	20.723	20h65	20.650
34.539	20.723	20h65	20.650
34.539	20.723	20h65	20.651
34.539	20.723	20h65	20.650
34.539	20.723	20h65	20.650
34.539	20.723	20h65	20.647
34.539	20.723	20h65	20.651
34.539	20.723	20h65	20.651
34.539	20.723	20h65	20.652
34.539	20.723	20.66	20.655
H0	ACCEPTÉ	H0	REJETÉ

Machine Translated by Google

La deuxième analyse considère les algorithmes NSGAII et MOEA/D, définit le modèle d'optimisation sur MOFBA4 et évalue tous les réseaux. Pour l'analyse, l'Hypervolume a été obtenu à partir de chacune des 30 exécutions de chaque algorithme sur le problème résolu. En utilisant le test non paramétrique de rang signé de Wilcoxon, avec un niveau de confiance de 95 %, le H0 était validé, ce qui précise que la différence de moyennes entre les échantillons est la même.
Le tableau 8 résume les résultats et montre la valeur Hypervolume de chaque réseau par exécution à l'échelle logarithmique et le statut d'acceptation de H0 dans la dernière ligne. On peut voir que l'hypothèse est rejetée dans tous les réseaux métaboliques analysés. Ces résultats des deux les analyses confirment ce qui était attendu, qu'il est pertinent de considérer quel modèle d'optimisation utiliser, et quel algorithme, car leurs performances lors de l'obtention d'ensembles de solutions peuvent être différent.

Tableau 8. Les données ont été analysées statistiquement pour établir les différences entre l'utilisation de NSGAII et MOEA/D. L'hypothèse nulle, H0, a été acceptée lorsque la valeur p obtenue était inférieure à 0,05.

Métabolisme du glutamate chez C. vulgaris		Réseau de pigments chez C. vulgaris	
NSGAII	MOEA/D	NSGAII	MOEA/D
20.723	20.723	20.647	20.723
20.723	20.723	20.648	20.723
20.723	20.723	20.649	20.723
20.723	20.723	20.649	20.723
20.723	20.723	20.651	20.723
20.723	20.723	20.651	20.723
20.723	20.723	20.651	20.723
20.723	20.723	20.651	20.723
20.723	20.723	20.652	20.723
20.723	20.723	20.648	20.723
20.723	20.723	20.650	20.723
20.723	20.723	20.651	20.723
20.723	20.723	20.648	20.723
20.723	20.723	20.653	20.723
20.723	20.723	20.654	20.723
20.723	20.723	20.652	20.723
20.723	20.723	20.651	20.723
20.723	20.723	20.649	20.723
20.723	20.723	20.651	20.723
20.723	20.723	20.651	20.723
20.723	20.723	20.650	20.723
20.723	20.723	20.650	20.723
20.723	20.723	20.651	20.723
20.723	20.723	20.650	20.723
20.723	20.723	20.650	20.723
20.723	20.723	20.647	20.723
20.723	20.723	20.651	20.723
20.723	20.723	20.651	20.723
20.723	20.723	20.652	20.723
20.723	20.723	20.655	20.723
H0	REJETÉ	H0	REJETÉ

A travers les résultats graphiques observés, principalement dus au volume et à la dispersion des des solutions obtenues dans tous les réseaux métaboliques d'algues considérés, il est démontré que la méthode proposée basée sur l'optimisation multi-objectifs résolue par métaheuristique offre un meilleur support à l'analyse. En revanche, l'analyse statistique
présenté dans cette section démontre qu'il est pertinent de considérer l'optimisation
modèle et l'algorithme puisque ceux-ci peuvent contribuer à différents types d'améliorations.

Machine Translated by Google

Les analyses statistiques présentées ici démontrent qu'il peut y avoir une différence significative entre les modèles d'optimisation et entre les algorithmes métaheuristiques.

6.6. Discussion

Presque aucune expérience n'a été réalisée auparavant sur le métabolisme des microalgues, à l'exception de [18]. Bien qu'il existe des méthodologies exactes, les approches évolutionnistes nécessitent moins de ressources informatiques dans le domaine des objectifs multiples ; par exemple, cela donne l' avantage d'utiliser moins de temps et de mémoire. Des approches telles que NSGAII et MOEA/D permettent un plus grand pouvoir de choix dans le processus de prise de décision en raison de la variété et du nombre de solutions et de la possibilité d'une reconnaissance plus facile des flux les plus importants dans un réseau ainsi que de leur influence et de leur impact, au lieu de ne pas le faire. avoir une méthodologie.

Certaines informations supplémentaires ressortent des résultats ci-dessus. L'expérience 2 démontre la polyvalence de NSGAII pour s'adapter à différentes circonstances et sa capacité à améliorer l' analyse du réseau métabolique étant donné le plus grand nombre de solutions qu'il produit pour chacune d'elles. Comme démontré dans les expériences 1 à 4, la capacité d'analyse d'un réseau métabolique est améliorée par l'introduction de l'algorithme NSGAII.

Les problèmes d'optimisation multi-objectifs présents dans la littérature considèrent actuellement différentes métriques de solution. L'expérience 1 compare l'utilisation de l'algorithme NSGAII avec les quatre variantes de problèmes d'optimisation, le problème d'optimisation MOFBA4 étant le plus prometteur, en introduisant différentes stratégies d'optimisation, telles que l'optimisation non seulement des métabolites d'intérêt mais également de l'hypervolume et de la distance générationnelle. Par rapport à MOFBA3, qui minimise l'hypervolume, dans MOFBA2, les variables de décision ne sont que les flux des réactions d'intérêt.

Bien que, dans les études de cas, NSGAII ait de meilleures performances graphiquement observables que MOEA/D, comme cela s'est produit dans l'expérience 3, car les études de cas avaient trois fonctions objectives et, selon la littérature, NSGAII est meilleur que MOEA/D lorsqu'il y a objectif à trois fonctions, s'ouvre la possibilité de pouvoir utiliser MOEA/D dans des réseaux où plus de trois fonctions objectifs doivent être optimisées. On peut également observer, à travers l'expérience 4, que la simple utilisation d'un échantillon aléatoire de solutions ne suffit pas pour obtenir un bon ensemble de solutions comme l'utilisation de métaheuristiques. Toutefois, des considérations particulières doivent être prises pour permettre le respect des restrictions ou informations de contrôle souhaitées par une personne intéressée.

Les algorithmes ont été testés sur différentes souches de microalgues, comme vu dans [18] avec C. reinhardti, et dans cette recherche utilisant C. vulgaris, dans des réseaux métaboliques complexes contenant des cycles, des bifurcations et des réactions réversibles, on vérifie leur viabilité dans différents processus métaboliques. réseaux, confirmant qu'ils ne peuvent pas être utilisés uniquement dans une seule microalgue. Cela laisse ouverte la possibilité de l'utiliser dans d'autres types d'espèces où il est nécessaire d'optimiser plus d'une fonction objective.

7. Conclusions

Le présent travail de recherche a réalisé une étude des flux métaboliques chez les microalgues vertes. L'objectif était de vérifier l'adéquation des méthodes in silico comme stratégies d'accompagnement pour améliorer l'analyse du métabolisme des microalgues. Grâce aux expériences développées, des preuves ont été obtenues qui soutiennent les conclusions suivantes :

L'étude des flux métaboliques chez les microalgues s'améliore en augmentant le nombre de solutions satisfaisant les conditions d'une microalgue pour qu'elle puisse vivre. Ceci est observable car, contrairement aux méthodes traditionnelles comme FBA qui n'offrent qu'une solution, élargie de manière limitée par l'analyse de sensibilité, elle est grandement favorisée par son intégration dans une méthodologie basée sur des métaheuristiques et des problèmes d'optimisation multi-objectifs ; il augmente à la fois le nombre de flux qui satisfont aux conditions recherchées dans le réseau métabolique, permettant également simultanément l'optimisation de plusieurs métabolites d'intérêt.

Il existe plusieurs alternatives à l'analyse d'un réseau métabolique en optimisant plusieurs métabolites d'intérêt, la méthode FBA étant au cœur du processus d'optimisation.

Machine Translated by Google

Le présent travail a proposé quatre modèles d'optimisation démontrant ce résultat, chacun offrant des angles d'analyse différents de ceux proposés par FBA.

Il est possible de résoudre les problèmes d'optimisation supportant l'étude métabolique en considérant différentes métaheuristiques évolutives et en obtenant des résultats significatifs pour l'analyse. Ceci est démontré en utilisant NSGAII et MOEA/D pour résoudre les problèmes d'optimisation proposés. Dans cette étude, NSGAII a montré les meilleures performances en général, ce qui est cohérent avec la littérature, en abordant exclusivement l'optimisation simultanée de trois objectifs. Cela montre que, pour les travaux futurs, l'analyse de la meilleure métaheuristique doit être réalisée avant l'étude.

Les paramètres de recherche de solutions peuvent être contrôlés lors de l'analyse d'un réseau métabolique en ajustant les limites de la réaction. Cela contribue à améliorer encore l'étude des microalgues puisque la définition d'environnements contrôlés est possible. Ceci est observé dans le processus de validation, où les paramètres permettant de générer des solutions étaient limités aux valeurs trouvées dans les travaux sur des échantillons in vivo.

La sélection d'un algorithme ou d'un modèle pour optimiser un réseau métabolique spécifique peut être difficile et nécessite un réglage fin pour identifier la configuration qui correspond le mieux aux intérêts de recherche de l'ingénierie métabolique réalisée. Cela est évident compte tenu de la variation des performances entre les algorithmes et les modèles d'optimisation, ou des différentes combinaisons testées dans ce travail de recherche.

Un décideur, par exemple un chercheur en ingénierie métabolique, améliore sa capacité de prise de décision en visualisant un ensemble de flux métaboliques qui satisfont aux conditions spécifiées pour le réseau métabolique qu'il étudie.

Contributions de l'auteur : Conceptualisation, MFB-B. et LA-V.; méthodologie, MFB-B. et NR-V.; logiciel, MFB-B. et NR-V.; validation, LA-V., ALM-S. et CZ ; analyse formelle, MFB-B., LA-V. et NR-V.; enquête, MFB-B.; ressources, MFB-B. et LA-V.; conservation des données, LA-V., ALM-S. et CZ ; rédaction – préparation du projet original , MFB-B., NR-V. et CG-S.; rédaction – révision et édition, LA-V., ALM-S. et CZ ; visualisation, MFB-B.; supervision, LA-V. et ALM-S.; administration de projet, MFB-B. Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : CZ a été soutenu par la National Science Foundation (subvention n° 2313313). Toutes les opinions, constatations et conclusions ou recommandations exprimées dans ce document sont celles du ou des auteurs et ne reflètent pas nécessairement les points de vue de la National Science Foundation.

Déclaration de disponibilité des données : le script des algorithmes métaheuristiques modifiés NSGAII et MOEA/D avec leurs MOFBA respectifs ont été stockés dans le référentiel GitHub : https://github. com/multiobjectiveoptimization2/MOFBA (consulté le 24 juin 2024).

Remerciements : Les auteurs remercient le soutien des projets CONAHCYT no. 3058, A1-S-11012 et le Laboratorio Nacional de Tecnologías de la Información (LANTI). Ils remercient également le soutien de l'ITCM et du projet TECNM 14612.22-P. Mónica Fabiola Briones Báez remercie la bourse n° 784358 du CONAHCYT pour poursuivre ses études de troisième cycle.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Annexe A

Annexe A.1

Comme mentionné précédemment, un réseau métabolique est l'ensemble des réactions présentes dans le micro-organisme. La liste de réactions suivante représente le réseau métabolique pour la synthèse des pigments dans la microalgue Chlorella vulgaris .

Machine Translated by Google

Tableau A1. Réactions issues de la synthèse des pigments dans le métabolisme des microalgues C. vulgaris [17].

Nom	Formule
FRDPème	frdp[h] frdp[c]
GRDPème	grdp[h] grdp[c]
ACAROTu	acaro[h] → acaro[u]
GCAROtu	gcaro[h] → gcaro[u]
GGDPtu	ggdp[h] → ggdp[u]
FPS	grdp[c] + ipdp[c]→ frdp[c] + ppi[c]
FPPSh	grdp[h] + ipdp[h]→ frdp[h] + h[h] + ppi[h]
GGP	frdp[h] + ipdp[h]→ ggdp[h] + h[h] + ppi[h]
GPPSh	dmpp[h] + ipdp[h]→ grdp[h] + h[h] + ppi[h]
IDS2	h[h] + h2mb4p[h] + nadph[h]→ dmpp[h] + h2o[h] + nadp[h]
ANXANASCOR	anxan[u] + ascb-L[u] → dhdascb[u] + h2o[u] + zaxan[u]
BCAROH	caro[u] + h[u] + nadph[u] + o2[u] → bcrptxan[u] + h2o[u] + nadp[u]
BCRPTXANH	bcrptxan[u] + h[u] + nadph[u] + o2[u] → h2o[u] + nadp[u] + zaxan[u]
CHYA1	acaro[u] + h[u] + nadph[u] + o2[u]→ h2o[u] + nadp[u] + zxan[u]
CHYA2	acaro[u] + h[u] + nadph[u] + o2[u]→ crpxan[u] + h2o[u] + nadp[u]
CXHY	crpxan[u] + h[u] + nadph[u] + o2[u]→ h2o[u] + lut[u] + nadp[u]
LCYB	gcaro[u] → caro[u]
LCYA	lyc[h] → décaro[h]
LCYG	lyc[h] → gcaro[h]
NÉOXANS	vioxan[u] → néoxan[u]
NI	norsp[h] + o2[h] + pqh2[h] → 2 h2o[h] + lyc[h] + pq[h]
PDS1	phyto[h] + pq[h] → phytfl[h] + pqh2[h]
PDS2	phytfl[h] + pq[h] → pqh2[h] + zcaro[h]
PSY	2 ggdp[h]→ 2 h[h] + phyto[h] + 2 ppi[h]
VIOXANOR	ascb-L[u] + vioxan[u] → anxan[u] + dhdascb[u] + h2o[u]
ZDS	o2[h] + pqh2[h] + zcaro[h] → 2 h2o[h] + norsp[h] + pq[h]
ZHY	h[u] + nadph[u] + o2[u] + zxan[u]→ h2o[u] + lut[u] + nadp[u]
CHLASG	chlda[u] + ggdp[u]→ ggchlda[u] + h[u] + ppi[u]
GGCHLDAR	ggchlda[u] + 3 h[u] + 3 nadph[u]→ chla[u] + 3 nadp[u]
GGDR	ggdp[h] + 3 h[h] + 3 nadph[h] → 3 nadp[h] + pdp[h]
CHLBSG	chldb[u] + ggdp[u]→ ggchldb[u] + h[u] + ppi[u]

Tableau A2. Réactions dérivées de la distribution des flux associées au métabolisme du glutamate [17].

Nom	Formule
GLNième	gln-L[c] + h[c] gln-L[h] + h[h]
GALh	atp[h] + glu-L[h] + nh4[h] → adp[h] + gln-L[h] + h[h] + pi[h]
GLUS	glu-L[h] + h2o[h] + nad[h] → akg[h] + h[h] + nadh[h] + nh4[h]
GLUTRES	atp[h] + glu-L[h] + trnaglu[h] → amp[h] + glutrna[h] + h[h] + ppi[h]
GLUS(nadph)	akg[h] + gln-L[h] + h[h] + nadph[h] → 2 glu-L[h] + nadp[h]
ASPAth	akg[h] + asp-L[h] glu-L[h] + oaa[h]
ASPNA1ème	asp-L[c] + na1[c] asp-L[h] + na1[h]
VALth	h[c] + val-L[c] h[h] + val-L[h]
BCTA(val)h	akg[h] + val-L[h] → 3mob[h] + glu-L[h]
TYRTAh	34hpp[h] + glu-L[h] akg[h] + tyr-L[h]
TYRth	h[c] + tyr-L[c] h[h] + tyr-L[h]
BCTAh	3mop[h] + glu-L[h] akg[h] + ile-L[h]
ILEth	h[c] + ile-L[c] h[h] + ile-L[h]

Références

1. Estèves, AF; Soares, OS ; Vilar, VJ; Pires, JC ; Gonçalves, AL L'effet de la longueur d'onde de la lumière sur la capture du CO2, la biomasse production et absorption de nutriments par les microalgues vertes : Un pas en avant dans l'intégration et l'optimisation des procédés. Énergies 2020, 13, 333. [Réf. croisée]
2. Wang, Y. ; Tibbetts, SM; McGinn, PJ Microalgues comme sources de protéines de haute qualité pour l'alimentation humaine et les suppléments protéiques. Aliments 2021, 10, 3002. [CrossRef] [Pub Med]
3. Brun, M. ; Jeffrey, S. Composition biochimique des microalgues des classes d'algues vertes Chlorophyceae et Prasinophyceae. 1. Acides aminés, sucres et pigments. J. Exp. Mars Biol. Écol. 1992, 161, 91-113. [Référence croisée]
4. Kholssi, R. ; Ramos, PV ; Marques, EA ; Montero, O. ; Rad, C. 2Utilisations biotechnologiques des microalgues : état de l'art et les défis de l'économie circulaire. Biocatal. Agricole. Biotechnologie. 2021, 36, 102114. [Réf. croisée]
5. Romero, DCV ; Cardozo, AP; Montes, VD Utilisation de microalgues comme alternative pour l'élimination des métaux lourds. RIAA2022 , 13, 10.
6. Woolston, BM ; Edgar, S. ; Stephanopoulos, G. Ingénierie métabolique : passé et futur. Ann. Révérend Chem. Biomol. Ing. 2013, 4, 259-288. [Référence croisée] [Pub Med]
7. Kaste, JA; Shachar-Hill, Y. Validation et sélection de modèles dans l'analyse des flux métaboliques et l'analyse du bilan des flux. Biotechnologie. Programme. 2024, 40, e3413. [Référence croisée] [Pub Med]
8. Anand, S. ; Mukherjee, K. ; Padmanabhan, P. Un aperçu de l'analyse du bilan de flux pour les réseaux biochimiques. Biotechnologie. Genet. Ing. Rév.2020 , 36, 32-55. [Référence croisée] [Pub Med]
9. Orth, JD; Thiele, I. ; Palsson, B.Ø. Qu'est-ce que l'analyse du bilan de flux ? Nat. Biotechnologie. 2010, 28, 245-248. [Référence croisée]
10. Liu, Y. ; Westerhoff, HV Analyse de l'équilibre des flux compétitive, multi-objectif et compartimentée pour traiter les problèmes spécifiques aux tissus erreurs innées du métabolisme. J. Hériter. Métab. Dis. 2023, 46, 573-585. [Référence croisée]

Machine Translated by Google

11. Raman, K. ; Chandra, N. Analyse de l'équilibre des flux des systèmes biologiques : applications et défis. Bref. Bioinformer. 2009, 10, 435-449. [Référence croisée]
12. Lewis, N. ; Nagarajan, H. ; Palsson, B. Contraindre la relation métabolique génotype-phénotype à l'aide d'une phylogénie in silico méthodes. Nat. Révérend Microbiol. 2012, 10, 291-305. [Référence croisée]
13. Toyoshima, M. ; Toya, Y. ; Shimizu, H. L'analyse de l'équilibre de flux des cyanobactéries révèle une utilisation sélective des composants photosynthétiques de transport d'électrons dans différentes conditions de lumière spectrale. Photosynthèse. Rés. 2020, 143, 31-43. [Référence croisée] [Pub Med]
14. Huang, J. ; Hou, J. ; Li, L. ; Wang, Y. Analyse du bilan de flux de la dégradation du glucose par digestion anaérobie en pression négative. Int. J. Hydrogène Énergie 2020, 45, 26822-26830. [Référence croisée]
15. Trilla-Fuertes, L. ; Gámez-Pozo, A. ; Diaz-Almirón, M. ; Prado-Vázquez, G. ; Zapater-Moros, A. ; López-Vacas, R. ; Nanni, P. ; Zamora, P. ; Espinosa, E. ; Fresno Vara, JA La modélisation informatique du métabolisme prédit le risque de survie sans rechute à distance chez les patientes atteintes d'un cancer du sein. Futur Oncol. 2019, 15, 3483-3490. [Référence croisée]
16. Boyle, NR; Morgan, JA Analyse de l'équilibre du flux du métabolisme primaire chez Chlamydomonas reinhardtii. Système BMC. Biol. 2009, 3, 4. [Référence croisée] [Pub Med]
17. Zuñiga, C. ; Li, CT ; Huelsman, T. ; Levering, J. ; Zielinski, DC; McConnell, BO; Longue, CP ; Knoshaug, EP; Guarnieri, MT; Antoniewicz, MR ; et coll. Le modèle métabolique à l'échelle du génome de l'algue verte Chlorella vulgaris UTEX 395 prédit avec précision les phénotypes dans des conditions de croissance autotrophes, hétérotrophes et mixotrophes. Physiol végétal. 2016, 172, 589-602. [Référence croisée] [Pub Med]
18. Briones-Baez, MF; Aguilera-Vazquez, L. ; Rangel-Valdez, N. ; Martinez-Salazar, AL; Zuñiga, C. Optimisation multi-objectifs du métabolisme des microalgues : un algorithme évolutif basé sur FBA. Métabolites 2022, 12, 603. [CrossRef] [Pub Med]
19. Rangaiah, médecin généraliste ; Petriciolet, A. Optimisation multi-objectifs en génie chimique : développements et applications ; Rangaiah, généraliste, Bonilla-Petriciolet, A., éd. ; Wiley : New York, NY, États-Unis, 2013.
20. Liu, X. ; Tian, J. ; Duan, P. ; Yu, Q. ; Wang, G. ; Wang, Y. GrMoNAS : Un cadre NAS multi-objectifs basé sur la granularité pour un diagnostic médical. Calculer. Biol. Méd. 2024, 171, 108118. [Réf. croisée]
21. PDE : Une approche d'évolution différentielle de Pareto-frontière pour les problèmes d'optimisation multi-objectifs. Dans Actes du Congrès 2001 sur le calcul évolutif (IEEE Cat. No. 01TH8546), Séoul, République de Corée, 27-30 mai 2001 ; Volume 2, pp. 971-978.
22. Andrade, R. ; Doostmohammadi, M. ; Santos, J. ; Sagot, MF; Mira, NP; Vinga, S. MOMO—métabolisme multi-objectifs mixte optimisation entière : Application à l'ingénierie des souches de levure. BMC Informer. 2020, 21, 69. [Réf. croisée]
23. Wang, GG; Zhao, X. ; Li, K. Algorithmes métaheuristiques : théorie et pratique ; CRC Press : Boca Raton, FL, États-Unis, 2024.
24. Déb, K. ; Pratap, A. ; Agarwal, S. ; Meyarivan, T. Un algorithme génétique multiobjectif rapide et élitiste : NSGA-II. IEEETrans. Évol. Calculer. 2002, 6, 182-197. [Référence croisée]
25. Cruz, L. ; Fernández, E. ; Gomez, C. ; Rivera, G. ; Perez, F. Optimisation de portefeuille à objectifs multiples de projets interdépendants avec intégration « a priori » des préférences des décideurs. Appl. Mathématiques. Inf. Sci. 2014, 8, 1517-1531. [Référence croisée]
26. Rivera, G. ; Florence, R. ; Guerrero, M. ; Porras, R. ; Sanchez-Solis, J. Analyse de portefeuille multicritère en ligne grâce à des modèles de programmation de compromis construits sur les principes sous-jacents du surclassement flou. Inf. Sci. 2021, 580, 734-755. [Référence croisée]
27. Chang, KH Optimisation multiobjectif et sujets avancés. En théorie de la conception et méthodes utilisant la CAO/CAE ; Elsevier: Amsterdam, Pays-Bas, 2015 ; pp. 325-406.
28. Zhang, Q. ; Li, H. MOEA/D : Un algorithme évolutif multiobjectif basé sur la décomposition. IEEETrans. Évol. Calculer. 2007, 11, 712-731. [Référence croisée]
29. Cruz-Reyes, L. ; Fernández, E. ; Rangel-Valdez, N. Une élicitation indirecte des paramètres de préférence basée sur une optimisation métaheuristique pour résoudre des problèmes à objectifs multiples. Int. J. Informatique. Intell. Système. 2017, 10, 56-77. [Référence croisée]
30. Déb, K. ; Goyal, M. Une recherche génétique adaptative combinée (GeneAS) pour la conception technique. Calculer. Sci. Informer. 1996, 26, 30-45.
31. Deb, K. Une méthode efficace de gestion des contraintes pour les algorithmes génétiques. Calculer. Méthodes Appl. Mécanique. Ing. 2000, 186, 311-338. [Référence croisée]
32. Benítez-Hidalgo, A. ; Nébro, AJ ; García-Nieto, J. ; Oregi, moi.; Del Ser, J. jMetalPy : Un framework Python pour les projets multi-objectifs optimisation avec métaheuristiques. Essaim Evol. Calculer. 2017, 12, e0171744. [Référence croisée]
33. Ebrahim, A. ; Lerman, JA; Palsson, BO; Hyduke, DR COBRApy : Reconstruction et analyse basées sur les contraintes pour Python. Système BMC. Biol. 2013, 7, 74. [Réf. croisée]
34. Hunter, JD Matplotlib : Un environnement graphique 2D. Calculer. Sci. Ing. 2007, 9, 90-95. [Référence croisée]
35. Liu, D. ; Huang, Q. ; Yang, Y. ; Liu, D. ; Wei, X. Algorithme bi-objectif basé sur le cadre NSGA-II pour optimiser le fonctionnement des réservoirs. J. Hydrol. 2020, 585, 124830. [Réf. croisée]

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Python hyperspectral : HypPy

Wim Bakker *, Frank van Ruitenbeek, Harald van der Werff*, Christoph Hecker, Arjan Dijkstra et Freek van der Meer

Faculté ITC, Université de Twente, 7522 NH Enschede, Pays-Bas ; fjavanruitenbeek@utwente.nl (FvR) ; cahecker@utwente.nl (CH); ahdijkstra@utwente.nl (AD); fdvandermeer@utwente.nl (FvdM)
* Correspondance : whbakker@utwente.nl (WB) ; hmavanderwerff@utwente.nl (HvdW)

Résumé : Cet article décrit la conception, la mise en œuvre et l'utilisation d'un package Python appelé Hyperspectral Python (HypPy). Les logiciels propriétaires de traitement des images hyperspectrales sont coûteux et les outils développés à l'aide de ces packages ne peuvent pas être distribués librement. L'idée de HypPy est de pouvoir traiter des images hyperspectrales à l'aide d'un logiciel gratuit et open source. HypPy a été développé en utilisant Python et s'appuie sur les capacités de traitement de tableaux de packages tels que NumPy et SciPy. HypPy a été conçu en pensant à la spectrométrie d'imagerie pratique et a mis en œuvre un certain nombre d'idées novatrices. Pour ne citer que quelques-unes de ces idées, HypPy dispose des outils BandMath et SpectralMath pour traiter les images et les spectres à l'aide d'instructions Python, peut traiter les bibliothèques spectrales comme s'il s'agissait d'images et peut adresser les bandes par longueur d'onde plutôt que par numéro de bande. Nous espérons qu'HypPy sera bénéfique pour la recherche, l'éducation et les projets utilisant des données hyperspectrales, car il est flexible et polyvalent.

Mots clés : imagerie hyperspectrale ; traitement d'images; Python; minéraux

1. Introduction

Les logiciels de traitement d'images existent depuis la création des premiers ordinateurs. Cependant, au début de ce siècle, peu de logiciels étaient capables de gérer des données d'images hyperspectrales. Les images hyperspectrales ont tendance à être volumineuses et les opérations standard telles que la classification des images ne sont pas adaptées à la gestion de la haute dimensionnalité des données hyperspectrale Traditionnellement, les images hyperspectrales sont traitées à l'aide de packages propriétaires tels que ENVI (ENVI et IDL sont des marques déposées de NV5 Geospatial Solutions, Inc.) (Environnement pour la visualisation des images) [1] et d'autres. Cependant, avec l'avènement des logiciels libres et open source [2], le besoin s'est fait sentir de pouvoir utiliser et traiter des images hyperspectrales indépendamment des progiciels propriétaires.

Avec des packages comme numpy [3], scipy [4] et matplotlib [5], le langage Python [6] est devenu un environnement de traitement de données et de prototypage, rivalisant avec les systèmes (propriétaires et gratuits) tels que MATLAB (MatLab est un marque commerciale de The MathWorks, Inc.), IDL (ENVI et IDL sont des marques commerciales de NV5 Geospatial Solutions, Inc.), Octave (Octave est une marque commerciale d'Octave Technologies, Inc.) et R.

Python, en plus des packages numpy et scipy susmentionnés, est gratuit et open source. Les utilisateurs peuvent accéder au code source, apporter des modifications et distribuer leurs propres versions. En principe, tout code écrit en Python peut être distribué librement.

Nous utilisons Python pour l'implémentation du package Hyperspectral Python, HypPy (prononcé [hip-ee], comme la sous-culture dans les années 1960), qui est un progiciel permettant de lire et d'écrire, de visualiser, de traiter et d'analyser des données hyperspectrales.

HypPy a commencé avec un certain nombre de classes d'images et de fonctions d'usine permettant la lecture et l'écriture d'images au format ENVI. La création de nouveaux outils dans HypPy a pris de l'ampleur lorsque notre groupe de recherche s'est impliqué dans le projet pilote de cartographie planétaire de Mars [7] en 2009 et 2010. Par la suite, divers outils généraux et spécifiques pour gérer les données hyperspectrales de Mars ont été écrits. Parallèlement, HypPy proposait

check for updates

Référence : Bakker, W. ; Van Ruitenbeek, F. ; van der Werff, H. ; Hecker, C. ; Dijkstra, A. ; van der Meer, F. Python hyperspectral : HypPy. Algorithmes 2024, 17, 337. https://doi.org/10.3390/a17080337

Rédactrices académiques : Laura Antonelli et Lucie Maddalena

Reçu : 27 juin 2024
Révisé : 22 juillet 2024
Accepté : 30 juillet 2024
Publié : 1 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

Machine Translated by Google

plateforme de mise en œuvre d'idées expérimentales. Progressivement, le paquet a été étendu jusqu'à devenir ce qu'il est aujourd'hui. Cet article décrit les considérations de conception et la mise en œuvre de HypPy en fournissant un certain nombre d'exemples.

2. Conception et mise en œuvre

Dans cette section, nous visons à montrer qu'HypPy possède un ensemble de fonctionnalités qui le distinguent des autres packages et en font un outil polyvalent et flexible pour le traitement des données hyperspectrales.

HypPy possède les fonctionnalités suivantes : • HypPy prend en charge trois niveaux d'interface : une interface utilisateur graphique (GUI), une interface de ligne de commande (CLI) et une interface de programmation (API). Les interfaces utilisateur de HypPy sont accessibles à partir d'un menu de niveau supérieur. • Le menu de niveau supérieur est configurable et peut être adapté à différents objectifs. • HypPy dispose de visionneuses d'images hyperspectrales et de bibliothèques spectrales. Ces visualiseurs ont des interfaces simples mais offrent des fonctionnalités suffisantes pour explorer la qualité et le contenu informatif des données.

• HypPy offre une capacité de lecture et d'écriture pour les fichiers de données ENVI. • Les images hyperspectrales ont tendance à être énormes, mais HypPy est respectueux de la mémoire.

• HypPy prend en charge les trois formats d'image natifs d'ENVI : BSQ, BIL et BIP. En interne, HypPy les traduit au format BIP, évitant ainsi au programmeur d' avoir à gérer trois formats différents.

• HypPy offre des fonctionnalités permettant de travailler avec des longueurs d'onde plutôt qu'avec des numéros de bande. Les numéros de bande n'ont aucune signification physique, contrairement aux longueurs d'onde. • Les données d'image brutes peuvent ne pas avoir de bandes dans l'ordre des longueurs d'onde croissantes. Donc, HypPy est capable de trier les longueurs d'onde dans l'image à la volée.

• HypPy est capable d'utiliser la liste des bandes incorrectes (bbl) et d'ignorer les bandes marquées comme mauvaises.

• HypPy peut traiter les bibliothèques spectrales ENVI comme s'il s'agissait d'images ENVI.

• Si nécessaire, les spectres sont rééchantillonnés à la volée. Cela élimine le besoin de rééchantillonner spectralement avant de combiner les spectres de différentes sources.

• HypPy est un logiciel gratuit et open source, ce qui en fait un outil idéal pour l'éducation. Les étudiants peuvent récupérer le package gratuitement et l'utiliser également à la maison.

Dans les sections suivantes, les points mentionnés ci-dessus seront abordés plus en détail.

2.1. Les niveaux d'utilisation : GUI, CLI et API

Les packages de traitement d'images prennent en charge différents types d'interfaces avec le logiciel. Par exemple, le package Khoros [8] prend en charge une interface graphique, une CLI et une API. En plus de cela, Khoros propose une interface de programmation visuelle, appelée Cantata [9]. Construire un environnement de programmation visuel est complexe et n'entre pas dans le cadre de HypPy. Cependant, comme Khoros, HypPy propose une interface graphique, une CLI et une API.

L'avantage d'avoir ces trois niveaux est que cela offre de la flexibilité. Les interfaces graphiques facilitent l'utilisation des outils, même pour les débutants. Depuis l'interface, il est clair de quelles entrées et sorties un certain outil peut avoir besoin. En plus de cela, une interface graphique peut être utilisée pour appeler plusieurs tâches à la fois. De cette manière, deux programmes ou plus peuvent être exécutés en parallèle, permettant ainsi l' utilisation des capacités multi-traitements de l'ordinateur sans avoir besoin d'un codage parallèle.

La CLI peut être utilisée pour exécuter des outils sur la ligne de commande. Cependant, le véritable pouvoir de la CLI réside dans le fait qu'elle peut être utilisée pour créer des scripts exécutant une séquence d'outils afin d'exécuter une chaîne de traitement complète. De plus, les scripts peuvent également être utilisés pour des tâches fastidieuses qui s'exécutent sur une longue période ou pour automatiser le traitement de grands ensembles d'images.

La plupart des fonctionnalités de HypPy peuvent également être importées dans d'autres programmes Python à l'aide de l'API.

L'interface graphique d'un outil HypPy se compose d'un module séparé, qui importe la fonction HypPy, crée l'interface graphique et transmet les paramètres à la fonction HypPy. Un module contient l'interface graphique, tandis qu'un autre module contient l'API et la CLI. Par exemple, le module edgy.py contient les fonctions de filtrage des bords hyperspectraux ainsi que la commande-

Machine Translated by Google

interface de ligne. L'interface graphique est invoquée à l'aide du module tkEdgy.py. Dans ce cas, les trois niveaux sont les suivants :

• Programme GUI : tkEdgy.py ; •
Programme CLI : edgy.pyl • Fonction
API : edgy().

Le menu de niveau supérieur HypPy peut être utilisé pour démarrer le programme GUI tkEdgy.py. L'énervé Le filtre peut être trouvé en tant que « Filtre de bord spatial-spectral » dans le menu « Filtre ».

2.1.1. Interface graphique

La plupart des outils HypPy sont divisés en deux programmes, un pour l'interface graphique et un pour l'API et la CLI. La plupart des interfaces graphiques sont de simples interfaces permettant de définir les paramètres d'entrée et de sortie, qui sont transmis à l'API. Un exemple d'interface graphique peut être trouvé dans la section 2.4.1, où l' outil Band Math est discuté.

Les interfaces utilisateur des outils de traitement GUI sont créées à l'aide de Tkinter, le package GUI (interface utilisateur graphique) qui fait partie de la distribution officielle de Python. Tkinter est une fine couche orientée objet au-dessus de Tcl/Tk [10]. Il existe de nombreux packages qui peuvent être utilisés pour développer des interfaces graphiques en Python. Par exemple, wxPython est un package plus complet (par exemple, un ensemble de widgets) que Tkinter. Cependant, le module Tkinter (interface Tk) est l'interface Python standard de la boîte à outils Tk GUI [11] et a été choisi pour HypPy pour sa simplicité et sa disponibilité. Tous les noms des outils HypPy commencent par « tk » pour indiquer qu'ils disposent d'une interface graphique.

2.1.2. CLI et scripts L'interface

de ligne de commande est utile pour exécuter des processus à l'invite de commande.
Cependant, la raison la plus importante d'avoir une CLI est de pouvoir exécuter une chaîne de traitement à partir d'un script. Les scripts sont pratiques pour répéter un ensemble de commandes de manière itérative.
De cette manière, un ensemble potentiellement important d'images peut être traité exactement de la même manière.
L'exécution de scripts réduit les erreurs et offre la possibilité d'exécuter des commandes nécessitant beaucoup de temps de traitement pendant la nuit.

HypPy adhère aux conventions utilisées pour fournir des options de ligne de commande ainsi que des entrées et sorties sur la ligne de commande. Dans ce but. il utilise le module standard Python appelé argparse, qui offre un moyen pratique de créer des interfaces de ligne de commande pouvant être utilisées dans les shells Unix ainsi que Windows.

2.1.3. L'API

Une interface de programmation d'application (API) est une interface logicielle par laquelle un logiciel offre des fonctionnalités à un autre logiciel. En Python, les fonctionnalités d'un programme peuvent être importées dans un autre programme en utilisant l'instruction import.
Une partie importante de l'API est la spécification de l'API, qui décrit les fonctions et les structures de données nécessaires pour utiliser les fonctionnalités disponibles.

Actuellement, aucun manuel d'API dédié n'est présent dans HypPy, mais des classes essentielles telles que Image et Spectrum sont couvertes respectivement dans le manuel des programmeurs et le manuel de mathématiques spectrales. De plus, de nombreuses fonctions HypPy sont décrites dans le manuel de script HypPy (CLI) et le manuel d'utilisation HypPy (GUI), car la CLI et l'interface graphique agissent souvent comme une fine coque autour de ces fonctions.

Étant donné que les objets de données constituent une partie importante de l'API, l'en-tête et l'image de HypPy les objets sont discutés ci-dessous ; voir la section 2.2.

2.1.4. Le menu de niveau supérieur

Le menu de niveau supérieur de HypPy peut être configuré à l'aide du fichier de configuration de menu appelé hyppymenu.cfg. Les fonctions peuvent être des scripts Python (généralement les programmes HypPy GUI), des fichiers PDF ou HTML, ou des URL complètes. De cette manière, le menu supérieur peut être utilisé pour démarrer des scripts ainsi que pour présenter des documents tels que des manuels ou des pages Web.

Machine Translated by Google

Les outils sont exécutés en tant que processus distincts ; par conséquent, plusieurs outils HypPy peuvent être démarrés et exécutés en parallèle en même temps. La figure 1 montre le menu supérieur de HypPy.

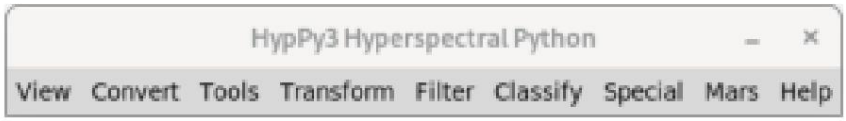


Figure 1. Le menu de niveau supérieur de HypPy.

Les scripts Python sont exécutés par Python en tant que sous-processus. Le PDF, HTML ou URL est délégué au lecteur par défaut du système, qui démarrera le lecteur PDF ou le navigateur Web sur le système.

2.2. Formats d'image

HypPy a adopté les formats ENVI comme formats natifs. Les formats d'image ENVI et de bibliothèque spectrale ENVI sont des fichiers binaires accompagnés d'un fichier d'en-tête de texte. ENVI prend en charge trois types différents d'entrelacement dans son format natif : bande séquentielle (BSQ), bande entrelacée par pixel (BIP) et bande entrelacée par ligne (BIL).

HypPy prend en charge les trois formats ENVI : BSQ, BIP et BIL. De plus, HypPy peut mapper une bibliothèque spectrale ENVI sur une image. La bibliothèque spectrale est alors représentée sous la forme d'une image y par x par z , dans laquelle y est le nombre de spectres stockés dans la bibliothèque spectrale, x vaut toujours 1 (la bibliothèque spectrale est visible sous forme d'image avec une seule colonne de pixels), et z est le nombre de bandes dans la bibliothèque spectrale. L'avantage est que les bibliothèques spectrales peuvent être visualisées comme des images spectrales. De plus, une chaîne de traitement destinée aux images spectrales peut d'abord être testée sur une bibliothèque spectrale pour voir quel effet la chaîne de traitement aura sur les différents minéraux présents dans la bibliothèque spectrale.

Pour des raisons de cohérence dans le codage, les images à bande unique sont également mappées sur trois dimensions , auquel cas la dimension de bande est toujours 1. Il existe une exception à cette règle : les images de classification sont mappées sous forme d'images à bande unique bidimensionnelles. Les images de classification ont une utilisation et une sémantique différentes de celles des données hyperspectrales, et pour cette raison, les images de classification sont traitées d'une manière différente dans HypPy.

En interne, HypPy mappe les trois formats, BSQ, BIP et BIL, sur un tableau avec un BIP structure, évitant ainsi au programmeur d'avoir à gérer trois formats différents.

2.2.1. Objets d'en-tête et d'image

Le module d'en-tête de HypPy est utilisé pour lire et écrire des fichiers d'en-tête ENVI. Les images ENVI et les bibliothèques spectrales se composent généralement de deux fichiers, les données et l'en-tête. Les fichiers de données contiennent des données binaires brutes. Les fichiers d'en-tête contiennent du texte lisible par l'homme.

Dans HypPy, l'en-tête ENVI est représenté comme un objet Python de la classe Header. Les objets de cette classe contiennent des attributs qui reflètent les attributs tels qu'ils se trouvent dans les fichiers d'en-tête ENVI. Dans le code Python, il existe quelques différences. Premièrement, les noms des attributs dans l'en-tête ENVI peuvent contenir des espaces, ce qui n'est pas pratique dans un langage de programmation comme Python car les noms de variables ne peuvent pas contenir d'espaces. Un choix a été fait de remplacer les espaces dans les noms d'attributs ENVI par un trait de soulignement. Par exemple, l'attribut « data-type » dans l'en-tête du fichier ENVI sera renommé « data-type » en Python. Deuxièmement, certaines valeurs contenues dans l'en-tête ENVI peuvent changer d'apparence en Python. Par exemple, les enregistrements IDL, tels qu'ils apparaissent dans les en-têtes ENVI, tels que les valeurs de l' attribut de longueur d'onde, sont remplacés par le type de données de liste Python standard. La traduction entre les enregistrements IDL et les listes Python est gérée en interne par HypPy.

HypPy prend en charge tous les types de données binaires brutes IDL/ENVI, des entiers non signés de 8 bits aux entiers signés de 64 bits et des nombres à virgule flottante simple de 32 bits aux nombres complexes à double précision de 64 bits avec une paire réelle-imaginaire de double précision. De plus, grâce à l'utilisation du nom du champ « ordre des octets », HypPy prend en charge les fichiers d'ordre des octets big-endian et small-endian.

Machine Translated by Google

HypPy a deux fonctions d'usine dans son module image, la fonction Open() pour ouvrir les images existantes et la fonction New() pour créer de nouvelles images (vides).

La fonction d'usine Open() renvoie l'un des objets image de la classe ImageBIP, Im-ageBIL, ImageBSQ, Classification ou ImageSL (une bibliothèque spectrale mappée sous forme d'image). Les images accessibles à l'aide de la fonction Open() peuvent uniquement être lues, pas écrites. Pour protéger les images existantes, celles-ci sont toujours ouvertes en lecture seule.

La fonction d'usine New() renvoie l'un des objets image de la classe ImageBIP, Image- BIL, ImageBSQ ou Classification. En plus de ces classes, la classe ImageSL est également prise en charge, qui est une représentation image d'une bibliothèque spectrale ENVI.

2.2.2. Carte mémoire

HypPy utilise la carte mémoire de Numpy, « numpy.memmap », pour mapper des fichiers de données binaires sur des tableaux. Les fichiers mappés en mémoire sont utilisés pour accéder à des fichiers volumineux sur un disque sans avoir à lire au préalable l'intégralité du fichier en mémoire. Les objets memmap NumPy sont des objets de type tableau ; ils peuvent être manipulés de la même manière que les tableaux numpy. Parce que dans HypPy le choix a été fait d'avoir la structure BIP pour le tableau interne, les axes des deux autres formats, BSQ et BIL, doivent être intervertis pour ressembler à BIP dans la carte mémoire renvoyée. Ceci est réalisé en appelant la fonction de transposition de la carte mémoire. BSQ est mappé à l'aide d'une transpose(1, 2, 0). BIL est mappé à l'aide d'une transpose (0, 2, 1). Le format BIP n'a pas besoin d'être transposé.

On pourrait penser que l'utilisation de la fonction de transposition créerait une copie des données avec les axes des données échangés d'une manière différente. Cependant, ce n'est pas le cas car, pour certaines opérations, numpy prend en charge ce qu'on appelle une « vue » sur les données. Une vue crée une référence à un tableau existant, avec la réorganisation requise des données ajoutée aux métadonnées. Cela présente deux avantages : d'une part, la vue ne crée pas de copie des données en mémoire, et d'autre part, lorsque la vue est ouverte en lecture-écriture, toute modification apportée à la vue sera immédiatement répercutée dans les données d'origine.

De plus, le boutisme et l'ordre des octets peuvent être modifiés en créant une vue. Les trois formats (BSQ, BIP et BIL), l'endianité (little endian et big endian) et le type de données (des entiers non signés de 8 bits aux virgules flottantes de 64 bits, etc.) peuvent être gérés en créant une vue du données au dossier. Pour les données d'image d'entrée et de sortie, aucune copie n'a besoin de résider dans la mémoire de l'ordinateur. De cette manière, HypPy peut gérer des fichiers volumineux qui ne tiendraient normalement pas en mémoire.

Les fonctions suivantes sont appliquées pour créer des vues : • Création d'une carte mémoire : numpy.memmap(fname, mode = 'r', shape = shape, dtype = data_type); • Changer l'ordre des octets : méthode newbyteorder(); • Permutation des axes de BSQ à BIP : méthode transpose(1, 2, 0) ; • Permutation des axes de BIL à BIP : méthode transpose(0, 2, 1).

Malheureusement, changer le type de données d'un fichier d'entrée crée une copie en mémoire, et non une vue :

• Changez le type de données de l'entrée : méthode astype(dtype).

Cependant, il existe quelques mises en garde concernant l'utilisation des cartes mémoire numpy. Toutes les opérations et fonctions pouvant être utilisées sur les tableaux numpy ne peuvent pas être utilisées sur des cartes mémoire, et seul un nombre limité d'opérations ou de fonctions sur une carte mémoire sont capables de renvoyer une vue.

De plus, l'utilisation d'une carte mémoire elle-même présente également des inconvénients. La configuration d'une carte mémoire utilise un peu de mémoire supplémentaire et toute la réorganisation d'une vue doit être exécutée à la volée à chaque accès aux données du fichier. De cette façon, la cartographie mémoire est moins efficace que la copie d'un tableau en mémoire.

Les avantages d'une carte mémoire sont évidents : elle permet de gérer de grands ensembles de données qui ne peuvent pas tenir dans la mémoire physique de l'ordinateur. De plus, en interne, les données sont toujours mappées sur la même structure BIP ; dans un programme, il n'est pas nécessaire

Machine Translated by Google

pour garder une trace du format d'entrelacement réel du fichier de données. Tout cela est géré de manière transparente par la carte mémoire et ses vues.

2.2.3. Formats de bibliothèque spectrale

HypPy prend en charge deux types de base de bibliothèques spectrales : les bibliothèques spectrales ENVI et les bibliothèques spectrales ASCII ou textuelles. Dans les bibliothèques spectrales ENVI, un inconvénient majeur est que la liste des longueurs d'onde doit être la même pour tous les spectres de la bibliothèque spectrale. Par conséquent, le nombre de bandes et les longueurs d'onde centrales de tous les spectres doivent être les mêmes pour tous les spectres inclus. Un avantage de ce format est qu'il s'agit d'un format binaire et qu'il peut être adressé comme une image, ce qui offre la possibilité de traiter des bibliothèques spectrales de la même manière que des images spectrales. Cela peut être utilisé pour tester des chaînes de traitement sur des bibliothèques spectrales avant de les exécuter sur des images. Un procédé de détection de certains minéraux peut par exemple être testé sur la bibliothèque spectrale de l'USGS [12] avant de l'appliquer à une image. Les problèmes liés aux faux positifs de certains minéraux peuvent être détectés dès les premières étapes du développement d'une chaîne de traitement.

En revanche, les bibliothèques spectrales ASCII offrent plus de flexibilité. Les fichiers texte peuvent être importés dans d'autres programmes et les plages spectrales des spectres individuels n'ont pas besoin d'être toutes identiques. Une complication est que les spectres doivent être rééchantillonnés avant utilisation.

HypPy rééchantillonne les spectres à la volée, en utilisant la fonction interp1d() du module scipy.interpolate pour unir les deux ensembles de longueurs d'onde. Une étape intermédiaire de rééchantillonnage des spectres n'est pas nécessaire.

2.2.4. Listes de mauvais groupes

Le format d'image ENVI prend en charge la notion de ce que l'on appelle une liste de bandes incorrectes (BBL), indiquée dans un fichier d'en-tête ENVI par le mot-clé BBL. Il doit y avoir autant d'éléments dans le BBL qu'il y a de bandes dans l'image. Séquentiellement, les bandes sont marquées comme étant mauvaises ou bonnes. Un zéro (0) pour une bande particulière signifie que les données de cette bande ne doivent pas être fiables en raison d'une défaillance du capteur ou d'un bruit excessif. Un un (1) pour une bande signifie des données valides. HypPy peut prendre en compte la liste des mauvaises bandes, de sorte que les mauvaises bandes n'apparaissent pas dans les données du tableau. Dans la plupart des outils HypPy, cela peut être indiqué en définissant l'option use_bbl sur true.

Si cette option est définie sur false, alors toutes les données sont visibles dans le tableau, y compris les bandes défectueuses. Une telle image virtuelle, sur laquelle use_bbl est défini sur true, peut avoir moins de bandes que le fichier de données d'origine.

En interne, cela est géré par un tableau d'index à index, qui est utilisé comme tableau d'indexation pour accéder à plusieurs éléments à la fois. L'inconvénient est que la gestion de ces images virtuelles est plus lente que l'utilisation de l'image complète. Un autre inconvénient est que les numéros de bande dans l'image virtuelle ne sont pas les mêmes que ceux de l'image originale. Cependant, l'une des idées de conception de HypPy est d'utiliser la longueur d'onde plutôt que le numéro de bande pour adresser une bande spectrale d'une image (voir section suivante).

2.2.5. Tri des longueurs d'onde Souvent,

les capteurs sont constitués d'un certain nombre de bancs de détecteurs, chacun couvrant une plage de longueurs d'onde spécifique. Les données de ces détecteurs peuvent avoir des plages de longueurs d'onde qui se chevauchent ou même apparaître dans un ordre inattendu. Par exemple, les données brutes de l'instrument OMEGA de l'européen Mars Express comportent les données des banques de détecteurs SWIR1 (0,93 µm à 2,69 µm) et SWIR2 (2,53 µm à 5,09 µm) avant les données du détecteur VNIR (0,36 µm à 5,09 µm). µm à 1,07 µm) [13]. Un autre instrument hyperspectral bien connu, AVIRIS, se compose de quatre spectromètres dont les plages se chevauchent [14]. Notez que les bandes redondantes dans le chevauchement spectral des spectromètres AVIRIS peuvent également être supprimées ou placées dans la liste des bandes incorrectes ; voir, par exemple, [15].

Évidemment, l'ordre commuté et les plages qui se chevauchent peuvent entraîner des problèmes lors du tracé ou du traitement de ces données. Par conséquent, l'un des critères de conception d'HypPy était de pouvoir trier les longueurs d'onde d'un cube de données hyperspectrales à la volée. Dans un compromis entre l'utilisation de la mémoire et l'utilisation du processeur, HypPy ne convertit pas l'image d'entrée en mémoire ; donc,

Machine Translated by Google

il doit convertir l'image à la volée. De nombreux outils HypPy proposent l'option « sort_wavelengths » lors de l'ouverture d'une image.

HypPy trie les longueurs d'onde en conservant une table d'index à index comme attribut des objets images. Ce tableau traduit les numéros de bande d'une image d'entrée en bandes triées par longueur d'onde. L'avantage est que, en interne, les données semblent avoir leurs bandes triées par longueur d'onde. L'inconvénient évident est que les numéros de bande de l'image d'entrée peuvent ne pas correspondre aux numéros de bande de l'objet image d'origine.

Cependant, si les bandes sont abordées par longueurs d'onde plutôt que par numéros de bande, alors ce dernier point n'est pas pertinent. Les longueurs d'onde ont une signification physique et devraient constituer le moyen privilégié pour traiter les bandes d'images. Les numéros de bande peuvent changer et même avoir une signification différente d'un capteur distant à l'autre. Malheureusement, toutes les images n'ont pas de longueurs d'onde, soit parce que les longueurs d'onde sont manquantes, soit parce que les images ne contiennent pas de données spectrales, c'est pourquoi les numéros de bande peuvent toujours être utilisés dans HypPy.

2.3.

Téléspectateurs 2.3.1.

Visionneuse d'images La visionneuse d'images de HypPy offre un affichage simple pour visualiser des images hyperspectrales et récupérer des spectres. La visionneuse d'images a été conçue pour une évaluation interactive rapide de la qualité de l'image et la récupération des spectres des membres finaux. Sa simplicité le rend facile à utiliser pour les utilisateurs novices de données hyperspectrales. Des curseurs peuvent être utilisés pour changer l'affichage d'une bande à une autre. En mode couleur, trois curseurs sont disponibles, un pour chacun des canaux RVB de l'écran. Une barre d'état en bas de la fenêtre présente des informations sur les longueurs d'onde des bandes, la position du curseur, etc. À l'aide des curseurs et du bouton gauche de la souris, les bandes et les spectres de l'image d'entrée peuvent être évalués en termes de qualité et de contenu informatif. La figure 2 montre l'image hyperspectrale d'un échantillon de roche. Cet échantillon de roche particulier 15-91-FE fait partie du complexe Fen en Norvège, qui comprend une intrusion de carbonatite [16].

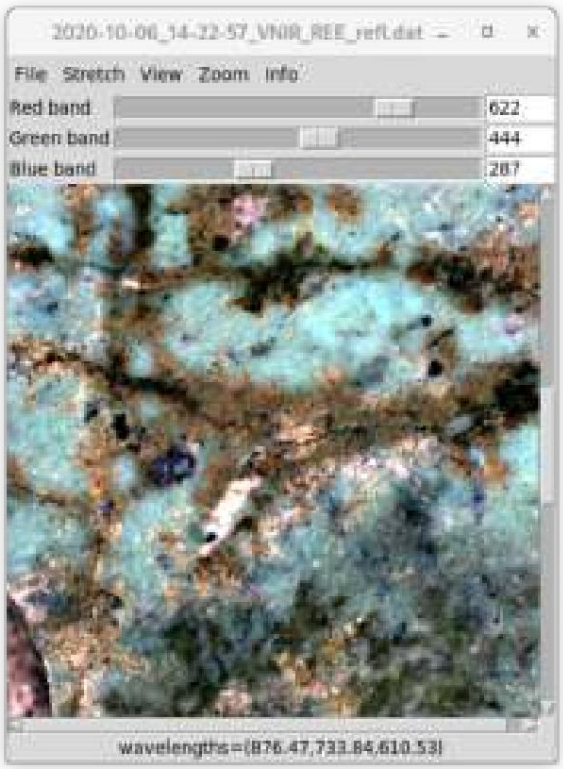


Figure 2. Visionneuse d'images montrant un composite en fausses couleurs de l'échantillon 15-91-FE. Canal rouge 876 nm, vert 734 nm et bleu 611 nm. Les curseurs pour le rouge, le vert et le bleu peuvent être utilisés pour modifier les bandes des canaux de couleur RVB.

Machine Translated by Google

Dans le menu Fichier, deux options peuvent être définies pour contrôler si les bandes doivent être triées avant la lecture ou pour utiliser le BBL d'une image.

Le programme d'affichage peut ouvrir deux images en même temps. La première image est toujours visible dans la visionneuse. La deuxième image, appelée « image fantôme », n'est jamais visible. Cependant, lorsque l'image d'ombre est présente, les spectres sont obtenus à partir de l'image d'ombre plutôt que de l'image visible. C'est comme un SIG dans lequel la première couche est au-dessus et la deuxième couche est cachée sous la première image. Cette option est utile pour charger une image de classification ou une image de masque dans la première couche pendant que les spectres sont extraits de l'image d'ombre, ce qui permet d'évaluer si une classification ou une image de masque calculée a du sens. De plus, cette option est utile pour explorer une carte de longueur d'onde avec une image hyperspectrale à partir de laquelle la carte de longueur d'onde a été dérivée.

Le menu Étirer propose un certain nombre de méthodes standard pour étirer l'image [17].
Des méthodes telles que 1 à 99 % et l'égalisation de l'histogramme sont mises en œuvre. En plus de cela, l'image peut être inversée, améliorée à la volée ou pseudo-colorée avec un certain nombre de schémas de couleurs disponibles.

Un clic gauche n'importe où sur l'image fera apparaître une fenêtre de tracé avec un graphique représentant le spectre du pixel à cet emplacement, voir Figure 3. Un clic supplémentaire sur l'image fera apparaître davantage de graphiques dans le tracé. En supprimant la fenêtre de tracé, le processus recommence avec une fenêtre de tracé claire. De cette façon, les spectres de différents points de l'image peuvent être visualisés ensemble.

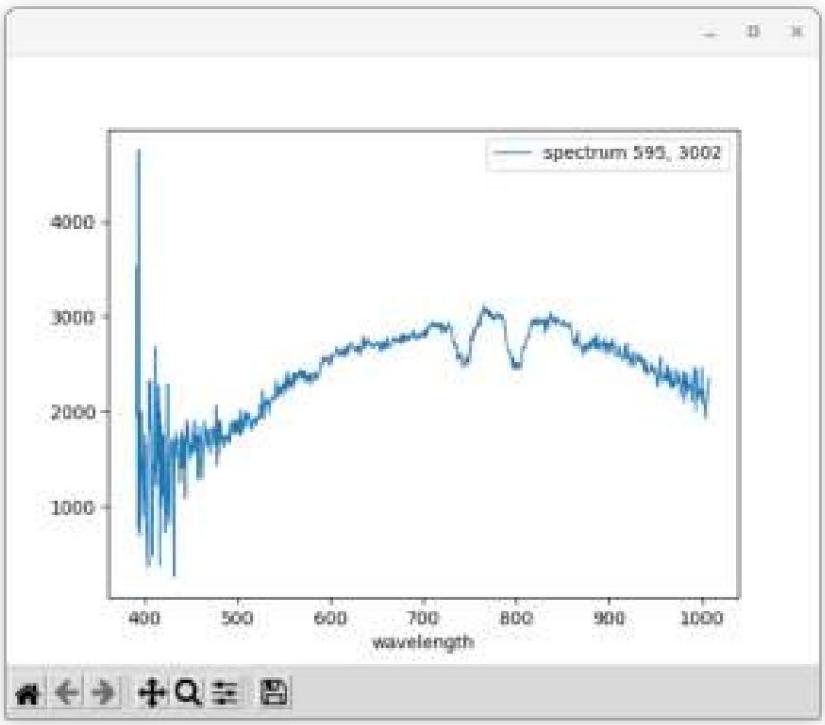


Figure 3. Fenêtre de tracé montrant le spectre de l'emplacement des pixels 595, 3002. Le spectre montre deux caractéristiques d'absorption à 740 et 800 nm. Ces caractéristiques sont dues aux éléments de terres rares (notamment le néodyme) contenus dans les cristaux de monazite. Le spectre inférieur à 450 nm est bruité en raison de la faible sensibilité dans le bleu de la caméra hyperspectrale VNIR. La barre d'outils en bas de la fenêtre est une barre d'outils de figure matplotlib standard et comporte des boutons pour naviguer dans la figure : début, arrière, avant, panoramique/zoom, zoom vers rectangle et configuration de la sous-intrigue. Le bouton Enregistrer à droite peut être utilisé pour enregistrer la figure dans différents formats tels que png ou pdf.

À l'aide de la « Fenêtre de valeurs », ces spectres peuvent également être enregistrés sous forme de fichier texte, de membres finaux ou de bibliothèques spectrales, ou importés dans un autre logiciel pour une visualisation et un traitement ultérieur. Un répertoire de spectres de membres finaux peut être utilisé comme bibliothèque spectrale dans les classificateurs HypPy, le mappeur d'angle spectral et le démixage linéaire.

Machine Translated by Google

Le programme d'affichage d'images HypPy utilise les modules Image et ImageTk de la Python Imaging Library (PIL, maintenant appelée Pillow [18]) pour afficher des images et matplotlib pour créer des graphiques 2D de spectres.

2.3.2. Visionneuse de bibliothèque spectrale

Le visualiseur de bibliothèque spectrale est un visualiseur de bibliothèques spectrales de type ENVI et ASCII. Après avoir sélectionné la bibliothèque spectrale d'entrée, les noms des spectres contenus par la bibliothèque spectrale sont affichés dans une liste. En double-cliquant sur un nom, une fenêtre de tracé apparaîtra montrant le spectre du minéral ou du matériau sélectionné. La sélection de plusieurs spectres tracera ces spectres ensemble dans la fenêtre de tracé, permettant une comparaison entre les spectres.

Le visualiseur de bibliothèque spectrale prend en charge divers formats ASCII, le format de bibliothèque spectrale CRISM (PDS), USGS SpecLib versions 6 et 7 [12], le format ASD [19] et les données en texte brut à deux colonnes. Le visualiseur de bibliothèque spectrale peut lire de manière récursive une arborescence de répertoires entière. Voir les figures 4 et 5, qui montrent l'interface du visualiseur de bibliothèque spectrale et une fenêtre de tracé montrant un exemple de spectre de monazite.

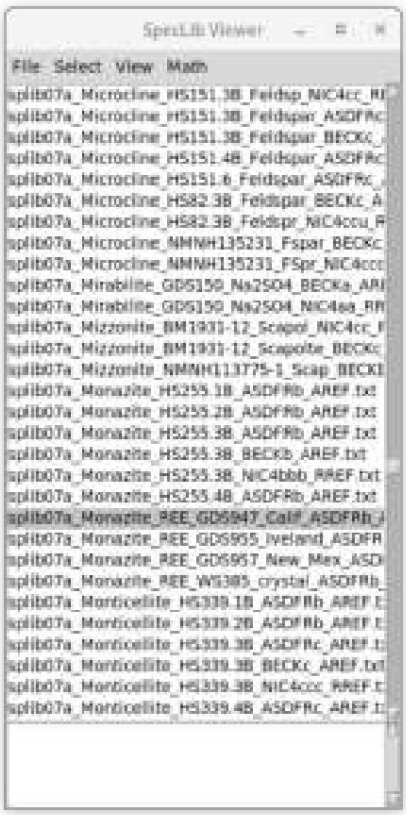


Figure 4. Interface du visualiseur de bibliothèque spectrale, dans ce cas, chargée avec la version 7 de la bibliothèque spectrale de l'USGS [12].

De plus, le visualiseur de bibliothèque spectrale prend en charge la manipulation des spectres à l'aide d'une expression « mathématique spectrale ». Un sous-ensemble de fonctions numpy peut être appliqué aux spectres. Les indices et les tranches peuvent être utilisés sur les spectres, et une multitude d'opérateurs et de fonctions mathématiques sont disponibles pour créer des opérations de traitement sur les spectres et les tester sur (une sélection de) une bibliothèque spectrale.

Les spectres avec différentes longueurs d'onde et plages de longueurs d'onde seront automatiquement rééchantillonnés à la volée. Toutes ces fonctions sont disponibles dans l'API et sont implémentées dans la classe spectre de HypPy (voir aussi la section 2.4.2 ci-dessous). Les résultats de l'opération mathématique spectrale dans la visionneuse de bibliothèque spectrale peuvent être enregistrés en définissant d'abord un répertoire de sortie.

Machine Translated by Google

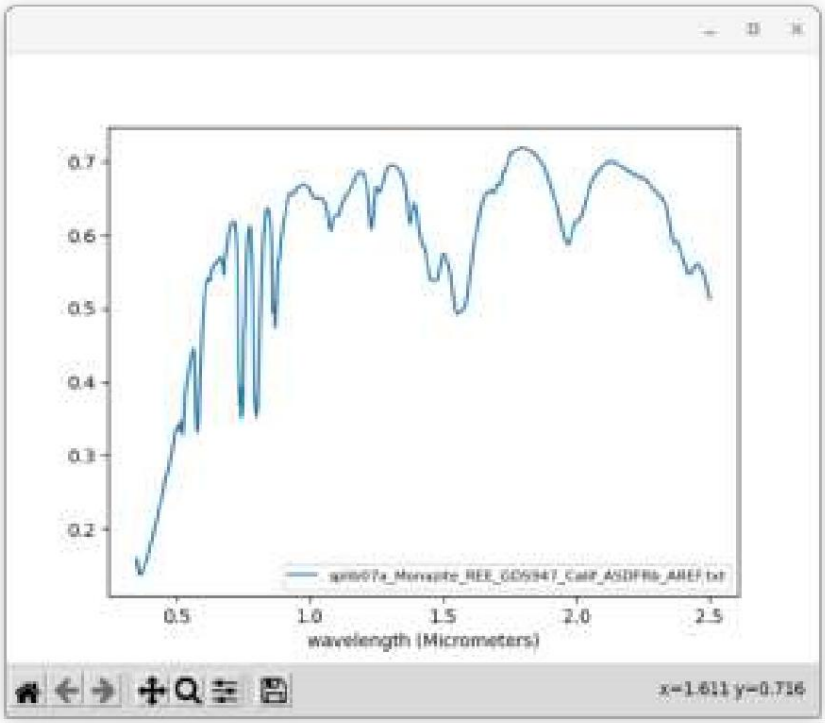


Figure 5. Fenêtre de tracé montrant le spectre Monazite_REE_GDS947_Calif_ASDFRb_AREF. Cet échantillon est connu pour contenir du Ce, La, Nd et Th. Notez les caractéristiques d'absorption du Nd à 740 nm et 800 nm.

2.4. Outils

Dans ce qui suit, un sous-ensemble des outils de HypPy sera abordé plus en détail.

2.4.1. Outil mathématique de bande

Les images multispectrales et hyperspectrales peuvent être considérées comme un ensemble de bandes spectrales, et des fonctions sont souvent créées pour traiter ces données bande par bande. L' outil mathématique de bande implémente cela.

Un exemple d'une telle opération mathématique de bande est le calcul du NDVI, dans lequel la différence entre une bande proche infrarouge et une bande rouge est normalisée en divisant par la somme de la bande proche infrarouge et de la bande rouge. Pour le NDVI, généralement pour la bande rouge, une plage de 400 nm à 700 nm est utilisée, et pour l'infrarouge, la plage est de 700 nm à 1 100 nm Dans l'exemple suivant, pour le rouge, nous choisissons une bande autour de 675 nm, et pour l'infrarouge, une bande autour de 1000 nm est choisie. Dans HypPy, le NDVI d'une image AVIRISNG peut être calculé à l'aide de l'une des expressions suivantes : (1) ou (2).

$$(i1[59] - i1[124])/(i1[59] + i1[124]), (i1(675.0) \quad (1)$$

$$- i1(1000.0))/(i1(675.0) + i1(1000.0)). \quad (2)$$

En mathématiques de bande, la séquence de noms de fichiers image est liée aux variables i1, i2, etc. Dans l'exemple du NDVI, une seule image d'entrée est nécessaire, liée à la variable i1. Les images sont numérotées à partir de 1 et les numéros de bande à partir de 0. Dans l' exemple, la bande 59 est la bande avec une longueur d'onde centrale de 676,9 nm et la bande 124 a une longueur d'onde centrale de 1 002,4 nm. Veuillez noter que, dans cet exemple, le BBL est utilisé et que les sept premières bandes, marquées comme mauvaises, ne sont pas présentes.

En mathématiques de bandes, il est également possible, et même préféré, d'utiliser des longueurs d'onde au lieu de numéros de bandes. La bande avec la longueur d'onde centrale la plus proche sera automatiquement sélectionnée. Notez que, dans le calcul des bandes, lorsque des longueurs d'onde sont utilisées plutôt que des numéros de bande, ces longueurs d'onde doivent être entourées de crochets et non de crochets. En Python,

Machine Translated by Google

ceci est implémenté à l'aide d'un « appel » sur la classe image. Les objets Python peuvent être conçus pour se comporter comme des listes ou des itérateurs, ayant un index entre crochets ou une fonction avec une liste de paramètres entre crochets. La figure 6 montre un exemple de rapport de bande.

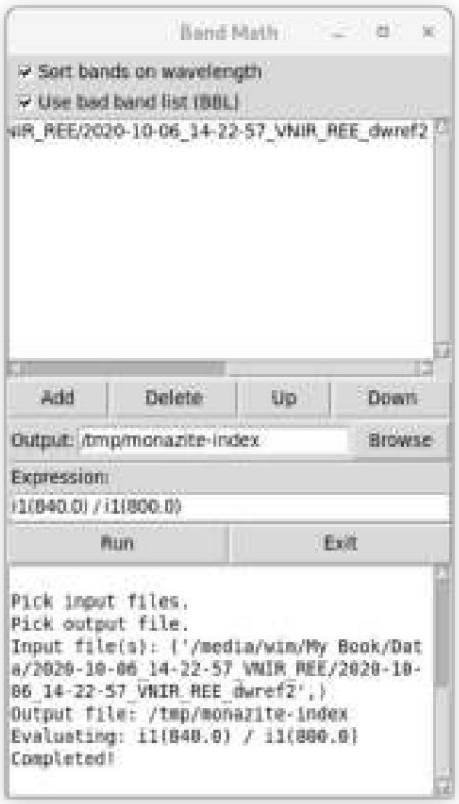


Figure 6. Interface de l'outil mathématique de bande. La formule montre un exemple de rapport de bande pour un indice de monazite simple basé sur l'épaulement et la caractéristique d'absorption du Nd à 840 nm et 800 nm, respectivement. Dans ce cas, il n'existe qu'un seul fichier d'entrée i1.

HypPy utilise la fonction Python eval() pour exécuter des expressions comme l'exemple NDVI ci-dessus. Les expressions mathématiques de bande sont essentiellement des expressions Python. HypPy configure les variables i1, i2, etc. Ce sont des objets image HypPy, qui peuvent être appelés pour renvoyer des bandes d'une certaine longueur d'onde. La liaison des variables d'image aux fichiers image est une étape essentielle dans l'utilisation des expressions mathématiques de bande. Notez que la syntaxe utilisée dans les mathématiques de bande diffère de celle utilisée dans les mathématiques spectrales. Les mathématiques de bande utilisent des objets images, tandis que les mathématiques spectrales utilisent des objets de spectre ; voir ci-dessous.

2.4.2. Classe de spectre et mathématiques spectrales

Les mathématiques spectrales sont utilisées pour appliquer des expressions mathématiques à des spectres d'images ou à des bibliothèques spectrales. Les mathématiques de bande utilisent le paradigme de bande ; les mathématiques spectrales utilisent le paradigme du spectre. En mathématiques spectrales, l'image est traitée pixel par pixel. En calcul de bande, l'image est traitée bande par bande. Les mathématiques spectrales reposent sur la fonctionnalité telle qu'elle est implémentée dans la classe Spectrum. L'interface de l'outil mathématique spectral est similaire à celle de l'outil mathématique de bande (voir Figure 6), mais l'expression doit suivre la syntaxe de la classe spectre.

La classe Spectrum a été créée pour la manipulation des spectres. L'idée est qu'un objet spectral contient des informations sur les longueurs d'onde ainsi que des valeurs, qu'il s'agisse de la réflectance, de la radiance ou d'autres données. La classe Spectrum offre un accès aux données, ainsi que des fonctions pour opérer sur les données spectrales.

La liste ci-dessous résume certaines des propriétés de l'objet conteneur de spectre :

• L'objet spectre contient des longueurs d'onde et des valeurs. • L'objet spectre peut contenir un nom et une description.

Machine Translated by Google

• Un index peut être utilisé pour sélectionner des valeurs ou créer des sous-ensembles du spectre. • Un nombre entier comme indice désigne un numéro de bande ; un flotteur désigne une longueur d'onde. • Les opérateurs mathématiques comme +, −, * et/peuvent être utilisés sur les spectres. • L'objet spectre contient une multitude d'opérateurs et de méthodes. Actuellement, quelque 150 fonctions et 20 opérateurs sont implémentés (voir le manuel de mathématiques spectrales de HypPy). • Lorsque vous travaillez avec des spectres de longueurs d'onde différentes, les spectres sont automatiquement rééchantillonné à la volée.

Comme indiqué précédemment, les indices spectraux peuvent être implémentés à l'aide des mathématiques de bande. Un indice spectral est censé avoir une valeur élevée si un certain minéral ou matériau est présent et une valeur faible s'il est absent. Un exemple d'un tel indice a été donné dans la section précédente dans laquelle nous avons présenté l'exemple du NDVI.

Cependant, si l'image contient du bruit, le NDVI résultant peut également contenir du bruit. Dans un tel cas, il peut être avantageux de travailler avec des valeurs moyennes dans des plages de bandes plutôt qu'avec des bandes uniques. En mathématiques de bandes, il n'est pas possible d'utiliser des plages de bandes car les bandes d'une plage doivent être adressées une par une. Cependant, en mathématiques spectrales, une plage spectrale peut être obtenue en utilisant une tranche du spectre.

L'expression mathématique spectrale pour NDVI peut maintenant ressembler à l'expression suivante (3) :

```
(S1[700.0 : 1100.0].mean() − S1[400.0 : 700.0].mean())/ (S1[700.0 :
1100.0].mean() + S1[400.0 : 700.0].mean()),                    (3)
```

dans laquelle S1 désigne les spectres de la première image. Les spectres des images d'entrée suivantes sont numérotés S2, S3, etc. Cependant, dans cet exemple, une seule image est utilisée. Les crochets contiennent ce qu'on appelle en Python des tranches, désignées par l'utilisation de deux points ':'. Notez que les flotteurs sont utilisés comme indices, qui désignent les longueurs d'onde. Les entiers comme indices sont interprétés comme des numéros de bande. Le résultat d'une tranche sur un spectre est un nouvel objet spectre qui contient un sous-ensemble de l'objet spectre d'origine. Dans ce cas, le sous-ensemble contiendra les valeurs spectrales environ entre 400 nm et 700 nm et entre 700 nm et 1100 nm.

Les longueurs d'onde sont approximatives car ces longueurs d'onde correspondent aux numéros de bande les plus proches, qui peuvent ne pas avoir exactement les mêmes longueurs d'onde que les longueurs d'onde centrales. HypPy ne génère pas d'erreurs ou d'avertissements pour la conversion des longueurs d'onde en bandes afin de donner la priorité à la vitesse plutôt qu'aux contrôles de qualité, car les contrôles consommeraient beaucoup de temps CPU. Cependant, les utilisateurs et les programmeurs peuvent vérifier les longueurs d'onde demandées et obtenues à l'aide des fonctions length2index() et index2wavelength().

Par la suite, la fonction de moyenne moyenne() est utilisée sur ces sous-ensembles. Autrement dit, le résultat sera le NDVI calculé à partir du spectre moyen entre 400 nm et 700 nm et du spectre entre 700 nm et 1100 nm. De cette manière, il est possible de construire des indices spectraux fonctionnant sur des plages de longueurs d'onde plutôt que sur des bandes uniques. L'avantage évident est que cela réduit le bruit en sortie. L'inconvénient est que ces expressions sont plus lentes que les opérations bande par bande. De plus, dans l'exemple ci-dessus, les tranches et les moyennes sont calculées deux fois car l'outil mathématique spectral ne prend en charge qu'une seule expression. Cependant, en utilisant intelligemment l'opérateur morse ':=', introduit dans la version 3.8 de Python [20], le double calcul peut être évité. L'expression du NDVI ressemblerait alors à l'expression (4) :

```
((ir := S1[700.0 : 1100.0].mean()) − (rouge := S1[400.0 : 700.0].mean()))/ (ir + rouge),
                                                               (4)
```

dans lequel la moyenne de la plage infrarouge est stockée dans la variable ir et la moyenne de la plage rouge est stockée dans la variable rouge, respectivement, et ces valeurs sont récupérées ultérieurement dans l' expression de calcul du dénominateur.

Veuillez noter que nous utilisons une large gamme pour le rouge et l'infrarouge dans les équations (3) et (4), en utilisant l'absorption de la chlorophylle dans la partie visible du spectre (400 nm à 700 nm) comme rouge et la réflexion par la structure cellulaire de la feuille. dans la partie proche infrarouge du spectre

Machine Translated by Google

(700 nm à 1 100 nm) comme ir. Les plages de longueurs d'onde spécifiques peuvent varier en fonction du capteur de télédétection ou de la technologie utilisée. Par exemple, Tucker [21] a utilisé la plage de 630 nm à 690 nm pour le rouge et de 750 nm à 800 nm pour l'ir, qui sont les plages des bandes spectrales Landsat MSS 5 et 7, respectivement.

HypPy utilise la fonction Python eval() pour exécuter des expressions comme l' exemple de moyenne NDVI ci-dessus. Les expressions mathématiques spectrales sont des expressions Python. Cependant, HypPy configure les variables S1, S2, etc., qui sont des objets du spectre HypPy qui peuvent être utilisés comme classe conteneur Python. En ce sens, le comportement de l'objet spectre est similaire à celui d'un objet liste Python, qui peut être indexé, découpé, bouclé, etc.

Dans HypPy, les mathématiques spectrales peuvent être utilisées dans l'outil de mathématiques spectrales ainsi que dans la visionneuse de bibliothèque spectrale.

2.4.3. Filtres spatio-spectraux

Les filtres spatio-spectraux prennent en compte à la fois les informations spectrales et les informations spatiales . HypPy propose trois types de filtres spatio-spectraux : les filtres dégradés, les filtres moyens et les filtres de regroupement. Les filtres à gradient sont implémentés en considérant un noyau spatial local de spectres comme vecteurs et en calculant la différence pondérée entre ces vecteurs. Les filtres moyens calculent la somme pondérée de ces vecteurs. Les filtres de regroupement réduisent le nombre d'échantillons en agrégeant les valeurs.

HypPy dispose d'un outil de filtrage des bords hyperspectraux. De tels filtres peuvent être utilisés pour détecter la variabilité spectrale dans une scène. Les zones avec une valeur élevée sont spectralement inhomogènes et les zones avec des valeurs faibles sont spectralement homogènes. Les résultats de ces filtres peuvent servir de première étape dans l'évaluation, la classification ou la segmentation de la qualité de l'image. Le résultat des filtres dégradés est toujours une image à niveau de gris monobande. L'outil HypPy appelé gradient hyperspectral implémente un certain nombre de filtres directionnels (haut, bas et deux diagonales) et non directionnels (énervé et Sobel). Un certain nombre de mesures de distance ont été mises en œuvre et peuvent être utilisées conjointement avec ces filtres.

La mesure de distance de l'angle spectral (SA) ne prend en compte que les informations spectrales de l'image d'entrée car l'angle spectral est insensible à l'intensité du spectre d'entrée. La mesure de distance de différence d'intensité (ID), en revanche, ne prend en compte que les informations d'intensité du spectre, et non les informations spectrales. La mesure de la distance euclidienne (ED) est sensible à la fois aux informations spectrales et d'intensité des spectres, bien que la plupart du temps, l'ED semble être dominée par l'intensité des spectres d'entrée. En plus des mesures de différence spectrale standard évoquées ci-dessus, deux autres mesures de distance, largement utilisées en analyse hyperspectrale, sont disponibles : la distance Bray-Curtis (BC) et la divergence d'information spectrale (SID) [22].

Toutes ces mesures de distance ont leurs propres propriétés concernant la similarité spectrale.

Là où les filtres à gradient prennent en compte un ensemble bidimensionnel 3 × 3 de spectres complets, les filtres moyens prennent en compte un ensemble tridimensionnel plus petit de valeurs spectrales. Les filtres moyens appliquent un noyau 3 × 3 × 3 sur le cube de données hyperspectral. Le filtre peut prendre en compte toutes les valeurs du noyau ou un ensemble limité de noyaux. Les noyaux actuels dans HypPy calculent la moyenne pondérée de 27, 19 ou 7 des valeurs du noyau. Deux fonctions de moyenne sont mises en œuvre : la moyenne et la médiane. Un filtre moyen est un filtre linéaire et, par conséquent, son effet sur l'image est prévisible. Un filtre médian réduit l' effet des valeurs aberrantes. À cet égard, un filtre médian peut être utile pour filtrer le bruit du poivre et du sel. Cependant, l'effet d'un filtre médian n'est pas aussi facile à prédire qu'un filtre moyen, ce qui peut entraîner des effets inattendus sur l'image de sortie. Dans certains cas, les filtres médians peuvent avoir un effet comparable aux filtres morphologiques dans le sens où des zones semblent rétrécir ou s'agrandir, ce qui peut prêter à confusion et ainsi compliquer l'interprétation de l' image résultante.

Le résultat d'un filtre moyen spatio-spectral est toujours un cube hyperspectral de mêmes dimensions que l'entrée. Cependant, comme pour les filtres spatiaux bidimensionnels classiques, les effets de bord doivent être pris en compte. Les première et dernière valeurs dans l'une des trois dimensions (x, y et bande) sont traitées différemment du reste du cube hyperspectral.

Machine Translated by Google

En plus de ces filtres, HypPy dispose d'un certain nombre d'outils de regroupement, spatial, spectral et spatial-spectral, pour réduire le nombre d'échantillons dans le cube d'image. Ces outils peuvent être utilisés pour diminuer le volume de données, réduire le bruit en faisant la moyenne et accélérer les étapes de traitement ultérieures. Par exemple, le regroupement spatio-spectral de xy = 3 × 3 et z = 11 réduira le volume de données de près d'un facteur 100 (3 × 3 × 11 = 99) et réduira l'écart type du bruit de presque un facteur de 100. 10 ($\sqrt{99}$).

2.4.4. Mappeur de longueur

d'onde La cartographie des propriétés physiques à partir d'images hyperspectrales est compliquée et repose souvent sur la disponibilité de bibliothèques spectrales de matériaux cibles. De nombreuses méthodes existantes dépendent des décisions subjectives de l'interprète pour la sélection des membres finaux et le post-traitement des images de règles. Si aucune information préalable n'est disponible, les informations doivent être extraites par caractérisation des caractéristiques d'absorption spectrale. Sur les images multispectrales, ces caractéristiques peuvent être trouvées à l'aide d'indices basés sur les rapports de bande et les profondeurs de bande ; voir, par exemple, [23]. Cependant, les images hyperspectrales montrent davantage de détails spectraux des caractéristiques, ce qui ouvre la possibilité d'utiliser des méthodes plus sophistiquées.

Les caractéristiques d'absorption des courbes spectrales sont importantes pour l'analyse et l'interprétation. De telles caractéristiques peuvent être paramétrées comme la position de la longueur d'onde centrale, la profondeur, la surface et l'asymétrie [24]. Habituellement, les caractéristiques les plus importantes sont la position centrale de la longueur d'onde et la profondeur de la principale caractéristique d'absorption. La position de la longueur d'onde d'un élément d'absorption peut être utilisée pour l'identification des minéraux, car de nombreux minéraux ont leurs propres longueurs d'onde d'absorption uniques. La profondeur d'une telle caractéristique est une indication de l'abondance d'un minéral.

La longueur d'onde centrale et la profondeur de la caractéristique la plus profonde et des caractéristiques consécutives les plus profondes peuvent être automatiquement déterminées. Ces paramètres peuvent être calculés pour tous les pixels (spectres) d'une image hyperspectrale et présentés sous forme d'images en niveaux de gris représentant la longueur d'onde et la profondeur de l'entité.

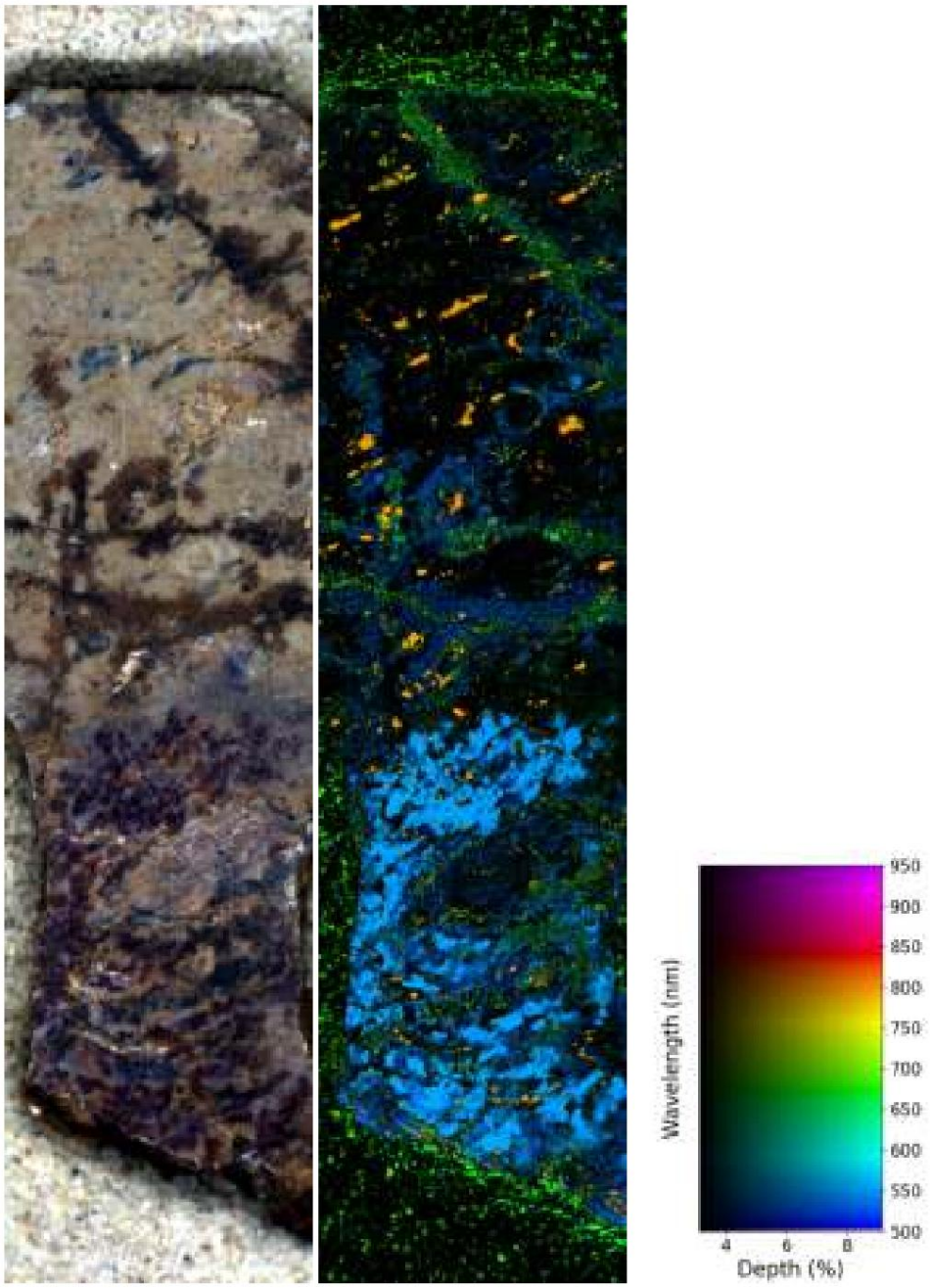
Par la suite, les images de longueur d'onde et de profondeur résultantes peuvent être fusionnées à l'aide d'une transformation HSI (teinte-saturation-intensité) [17] de telle sorte que l'image résultante montre la longueur d'onde en couleur (la teinte). L'intensité représente la profondeur du trait. La saturation est définie sur 1. L'outil de cartographie de longueur d'onde de HypPy détermine automatiquement la longueur d'onde et la profondeur de la caractéristique d'absorption la plus profonde de tous les pixels d'un cube hyperspectral et fusionne les deux en une seule image couleur.

La figure 7a montre un échantillon de roche dans un composite de couleurs naturelles. La figure 7b montre le résultat du mappeur de longueur d'onde ; La figure 7c montre la légende de la carte de longueur d'onde.
La monazite contenant des ETR est visible sous forme de grains orange. Les verts et les bleus sont des carbonates. Pour une discussion plus approfondie de cet échantillon de roche, voir [16].

Le mappeur de longueur d'onde est non supervisé, simple et reproductible ; de vastes zones peuvent être cartographiées à la fois et les détails sont préservés dans le cadre de la plage spectrale choisie. Le résultat, appelé carte de longueur d'onde, peut être compris par les utilisateurs ayant une expérience limitée en traitement d'image. Cependant, même si les couleurs peuvent suggérer la présence de certains minéraux, la carte de longueur d'onde qui en résulte n'est pas une carte minérale. La carte de longueur d'onde peut être utilisée comme étape de pré-classification afin de comprendre la variabilité spectrale de la scène. Les connaissances obtenues à partir de la carte de longueur d'onde sont utiles pour les étapes supplémentaires de production d'une carte minérale.

Le mappeur de longueur d'onde s'appuie uniquement sur deux paramètres d'entrée : la plage de longueurs d'onde sur laquelle se concentrer et l'étirement à utiliser pour la profondeur de l'entité. Le mappeur de longueur d'onde a été publié pour la première fois dans van Ruitenbeek et al. [25], et un tutoriel sur la méthode a été publié dans [24].

Machine Translated by Google



(a) Composition de couleurs (b) Carte de longueur d'onde (c) Légende

Figure 7. (a) Composite de couleur naturelle de l'échantillon de roche 15-91-FE (rouge 650 nm, vert 550 nm et bleu 450 nm). (b) La carte de longueur d'onde de la plage de 500 nm à 950 nm montre la longueur d'onde et la profondeur de la caractéristique la plus profonde. La monazite est visible en orange (montrant la caractéristique de 800 nm de Nd) ou en jaune-vert lorsque la caractéristique de 740 nm s'avère être plus profonde que la caractéristique de 800 nm. (c) Légende de la carte de longueur d'onde.

2.4.5. Arbres de décision

Un arbre de décision est un type de classificateur constitué d'une série de décisions binaires utilisées pour déterminer la classe de chaque pixel d'entrée. Les décisions sont basées sur les caractéristiques des ensembles de données. Chaque décision divise les données en deux classes ou groupes de classes potentiels. Les décisions sont des expressions appliquées à une seule image ou à un ensemble Chacune de ces images représente une propriété des pixels d'entrée.

Une telle propriété, par exemple, pourrait être la position en longueur d'onde de la caractéristique d'absorption la plus profonde du spectre du pixel d'entrée. Une expression pourrait être construite pour vérifier

Machine Translated by Google

si le pixel d'entrée peut contenir un minéral argileux avec une caractéristique d'absorption la plus profonde d' environ 2,2 µm. Des décisions ultérieures pourraient vérifier les caractéristiques de l'eau de ce pixel pour voir si le matériau a un indice de cristallinité élevé. Après un certain nombre de décisions, une classe sera attribuée au pixel.

Les arbres de décision sont utiles pour créer des classificateurs basés sur des propriétés physiques plutôt que sur des inférences statistiques. Bien entendu, cela ne peut être réalisé que si toutes les propriétés nécessaires aux arbres de décision peuvent être établies avec suffisamment de confiance.

Un package comme ENVI prend en charge la construction et la visualisation de nouveaux arbres de décision. Cependant, ENVI présente un certain nombre d'inconvénients. La manière dont les arbres de décision sont implémentés dans ENVI est plutôt rigide car une fois qu'un nœud de l'arbre est construit, sa place dans l' arbre de décision est fixée. Dans un arbre de décision ENVI finalisé, il est impossible de déplacer des éléments. Dans ENVI, si vous souhaitez modifier l'arbre de décision, vous devez le reconstruire à partir de zéro.

HypPy peut être utilisé pour exécuter des arbres de décision ENVI. La seule condition est que les expressions utilisées pour les décisions puissent être traduites en expressions en Python.

En plus du format d'arbre de décision ENVI, HypPy possède son propre format d'arbre de décision. Ce format est un format texte qui peut être modifié à l'aide d'un éditeur de texte.

Pour la visualisation des arbres de décision, HypPy utilise le package Graphviz [26]. À cette fin, l'arbre de décision ENVI ou HypPy doit d'abord être converti au format « point ». Le langage par points est un langage étendu pour décrire des graphiques. Les arbres de décision binaires utilisés par ENVI et HypPy présentent un petit sous-ensemble de tous les graphiques et types de graphiques possibles pris en charge par le package Graphviz.

2.4.6. Statistiques zonales

L'outil de statistiques zonales calcule des statistiques sur les pixels d'une image d'entrée dans les zones définies par l'image de zone. L'image d'entrée est une image hyperspectrale ou multispectrale régulière. L'image de zone peut être le résultat d'une image de classification ou d'une image de masque. Pour chaque valeur distincte dans l'image de zone, une fonction d'agrégation est exécutée sur l'image d'entrée pour tous les emplacements ayant cette valeur distincte dans l'image de zone. Les fonctions d'agrégation actuellement mises en œuvre sont les suivantes : moyenne, médiane, minimum, maximum, écart type et moyenne plus ou moins deux fois l'écart type.

Les résultats des statistiques zonales sont utiles pour déterminer les propriétés des zones dans l'image d'entrée. Par exemple, dans le cas de bruit dans l'image d'entrée, le spectre moyen d'une zone peut révéler plus de détails spectraux que le spectre d'un seul pixel, dans lequel les détails spectraux des caractéristiques peuvent être noyés par le bruit.

3. Domaines d'application

Ci-dessous, l'application de HypPy dans plusieurs projets dans les domaines de l'éducation, du conseil et la recherche sera discutée plus en détail.

3.1. Éducation

HypPy est l'un des progiciels utilisés dans nos cours de géologie spectrale à l' Université de Twente [27]. La clarté de l'interface HypPy en fait un excellent outil pour les étudiants débutant en géologie spectrale. Un exemple de recherche publiée par des étudiants utilisant HypPy est [28]. Pour un didacticiel sur l'analyse des caractéristiques d'absorption pour la recherche de minerai à l'aide de la télédétection géologique, voir [24].

HypPy peut être utilisé dans les cours courts pour apprendre les concepts de base de l'imagerie hyperspectrale et réaliser des opérations spécialisées sur le traitement et l'interprétation des données hyperspectrales. Par exemple, HypPy a été utilisé dans le cours « Télédétection et spectroscopie minérale pour le géologue d'exploration et d'exploitation minière » proposé à Istanbul [29].

En décembre 2022, HypPy a été utilisé dans un atelier spectral lors de la conférence annuelle du GRSG [30]. L'atelier comprenait un aperçu de la théorie et de la mise en œuvre pratique des méthodes spectrales pour l'identification des minéraux, accompagné d'une séance pratique en laboratoire impliquant l'application du traitement analytique analytique à un ensemble de données hyperspectrales aéroportées à haute résolution récemment obtenues qui couvre le visible et près-

Machine Translated by Google

spectres infrarouges (VNIR 400 nm à 1 µm), infrarouges à ondes courtes (SWIR 1 µm à 2,5 µm) et infrarouges à ondes longues (LWIR 8 µm à 15 µm). La session de laboratoire a utilisé le logiciel de traitement hyperspectral HypPy et a fourni un accès sans entrave aux données d'échantillon.

3.2. Consultant

Nous avons exploré la viabilité de l'utilisation d'images hyperspectrales infrarouges à ondes courtes (SWIR) en laboratoire pour un partenaire commercial afin d'identifier et de cartographier la composition minéralogique d'échantillons de carottes de forage sédimentaires spécifiques provenant du sous-sol des Pays-Bas. Les capacités de script de HypPy se sont avérées essentielles dans ce projet.

Nos résultats indiquent que l'utilisation d'images hyperspectrales infrarouges à ondes courtes (SWIR) de laboratoire avec un échantillonnage spatial de 200 µm offre une approche non destructive précieuse pour identifier et cartographier les minéraux de mica blanc, d'argile et de carbonate dans les forages sédimentaires. noyaux. En collaboration avec des partenaires de l'Université de Groningue, grâce à HypPy, nous avons réussi à construire un prototype de chaîne de traitement pour la quantification de la teneur en minéraux et à déduire un certain nombre de propriétés géophysiques des carottes de forage.

3.3. Recherche

Les premiers travaux sur HypPy ont commencé avec le projet pilote de cartographie planétaire de Mars pour l'Agence spatiale européenne (ESA). L'étude pilote a été conçue pour transférer les méthodes et flux de travail numériques modernes utilisés dans la cartographie géologique terrestre à la cartographie planétaire. Cette étude a été axée sur les ensembles de données de l'ESA provenant de la caméra stéréo haute résolution (HRSC) et de l'instrument hyperspectral du spectromètre de cartographie minéralogique visible et infrarouge (OMEGA) à bord de Mars Express [31]. La première étape du flux de travail consistait à traiter les données brutes pour mettre en évidence les caractéristiques intéressantes des produits d'information dérivés Ceux-ci comprenaient des modèles numériques d'élévation extraits des données stéréo et des cartes minérales réalisées à l'aide des données hyperspectrales. Ces outils ont été utilisés pour interpréter les données et produire une carte géologique de la région de Nili Fossae. Cette étude a ouvert la voie à l'adoption de la technique de cartographie des longueurs d'onde pour cartographier régulièrement les planètes Mars et la Terre [7].

HypPy introduit une méthodologie de traitement conçue pour transformer les données OMEGA de la radiance au niveau du capteur à la réflectance de surface. Pour relever les défis particuliers associés aux données OMEGA, nous avons conçu des approches innovantes pour déterminer un spectre de transmission par scène , effectuer une correction atmosphérique et filtrer le bruit systématique et aléatoire. Voir, par exemple, van Ruitenbeek et al. [25]. Ces filtres font désormais partie de HypPy. La méthodologie de traitement génère des images de réflectance de surface, permettant l'extraction de spectres à pixel unique pour l'identification des caractéristiques de surface. Conformément à diverses études antérieures, nos résultats suggèrent, entre autres observations, l'existence d' argiles riches en fer sur la surface martienne.

La cartographie des propriétés physiques à partir d'images hyperspectrales est compliquée et dépend souvent de l'accessibilité des bibliothèques spectrales des matériaux cibles. De nombreuses méthodes existantes dépendent des décisions subjectives de l'interprète pour la sélection des membres finaux et le post-traitement des images de règles. Si aucune information préalable n'est disponible, les informations doivent être extraites par la caractérisation des caractéristiques d'absorption spectrale. Dans van Ruitenbeek et al. [25], ils décrivent une méthode appelée mappeur de longueur d'onde (voir section 2.4.4) pour déterminer automatiquement la réflectance minimale dans une plage de longueurs d'onde. Nous concluons que le mappeur de longueur d'onde s'avère être un outil utile pour la sélection impartiale des membres terminaux, la cartographie des minéraux et l'interprétation des zones démontrant le zonage des altérations minérales à l'aide de l'imagerie hyperspectrale. L'utilité du mappeur de longueur d'onde a déjà été reconnue dans d'autres recherches. Par exemple, il a été utilisé pour cartographier les affleurements minéraux de pegmatites contenant du lithium dans une carrière pilote située près d'Uis, en Namibie [3

Dans une autre étude, la fonction d'entropie a été ajoutée à la classe spectrale pour permettre le calcul d'un indice pour les processus de tri dans les systèmes hydrothermaux. Les processus hydrothermaux affectent la composition des roches par des réactions chimiques et l'écoulement des fluides, et cela

Machine Translated by Google

L'article a montré que les modèles restants de tels processus peuvent être visualisés en utilisant l' entropie de Shannon. L'application de l'entropie est utile pour étudier les gisements minéraux hydrothermaux et peut également être utile dans les études des environnements des premiers stades de la vie.

Les fonctions déjà implémentées dans la classe spectre de HypPy pourraient être utilisées pour ajouter du bruit aux données. Ceci a été utilisé pour la simulation de capteurs hyperspectraux SWIR aéroportés pour la cartographie du gaz méthane (CH4) [33]. Les résultats de l'étude ont indiqué que trois spectromètres d'imagerie aéroportés actuellement actifs n'ont pas réussi à détecter le panache de CH4. Le document souligne l'importance de l'échantillonnage spectral et de l'étalonnage radiométrique pour une cartographie réussie du méthane.

Les spectromètres d'imagerie contemporains sont compacts, légers, économiques et applicables en laboratoire. Des images hyperspectrales précises nécessitent un bruit minimal, un étalonnage précis et une uniformité de réponse constante. Des contrôles périodiques sont conseillés en raison du vieillissement des composants optiques et électroniques. Dans une autre étude, nous introduisons une technique pour quantifier les distorsions géométriques du sourire et de la clé de voûte à partir d'images de test [34]. Les mesures sont acquises grâce à l'utilisation de modèles simples pour le sourire et la clé de voûte, ainsi que d'outils préexistants dans HypPy. De plus, des fonctions d'interpolation spline quadratique , de corrélation croisée normalisée et d'optimisation sont appliquées pour obtenir les valeurs de sourire et de clé de voûte.

Les résultats indiquent que le sourire et la clé de voûte d'un spectromètre d'imagerie de laboratoire peut être déterminé avec une précision inférieure au pixel.

4. Discussion

En plus de HypPy, plusieurs autres packages offrent des fonctionnalités de traitement d'images hyperspectrales.

Pour n'en nommer que quelques-uns, QUANTools a été implémenté à l'aide du langage de programmation ENVI/IDL pour la détection automatique de plusieurs paramètres de caractéristiques d'absorption [35]. Cependant , pour exécuter QUANTools, une version d'exécution d'ENVI/IDL est nécessaire, et le package QUANTools offre des fonctionnalités limitées, similaires à un outil spécifique dans HypPy, appelé mappeur de longueur d'onde.

PySptools est un module Python qui implémente des algorithmes spectraux et hyperspectraux [36]. Cependant, il s'agit d'un package ou d'une bibliothèque et vous devez coder avant de pouvoir l'utiliser. En fait, HypPy utilise pysptools pour son classificateur de démixage linéaire. Contrairement aux pysptools, HypPy offre une interface utilisateur pour la fonctionnalité.

Une autre boîte à outils remarquable est connue sous le nom de Python spectral (SPy) [37]. SPy est un module Python conçu pour gérer les données d'images hyperspectrales, doté de fonctions pour des tâches telles que la lecture, l'affichage, la manipulation et la classification d'images hyperspectrales. Notamment, SPy est limité à une utilisation interactive à partir de l'invite de commande Python ou via des scripts Python, contrairement à HypPy.

Les progiciels Orfeo et EnMAPBox sont tous deux des outils puissants utilisés dans les applications de télédétection et de traitement d'images [38,39]. Ils sont largement utilisés dans la communauté de la télédétection et offrent des capacités étendues de traitement et d'analyse des images satellite. Alors que ces packages se concentrent sur le traitement des images satellite, HypPy vise également le traitement des images hyperspectrales générales, telles que les images proximales enregistrées avec des caméras hyperspectrales en laboratoire.

Tout comme le langage Python lui-même [6], HypPy a été conçu dans un souci de simplicité plutôt que de performance [40]. Certaines opérations dans HypPy peuvent être relativement lentes. Cependant, la modularité des fonctions et des outils permet de construire des programmes expérimentaux ou même des prototypes de chaînes de traitement fonctionnant sur n'importe quel matériel.

Concernant la vitesse, les futures versions de Python pourraient être plus rapides. Pour citer un projet en cours, le langage Mojo est un surensemble du langage Python avec la vitesse du langage C [41].

Machine Translated by Google

5. Conclusions

Dans cet article, nous avons décrit la conception, la mise en œuvre et l'application du package Python appelé Hyperspectral Python, HypPy. L'idée de HypPy est de pouvoir traiter des images hyperspectrales à l'aide d'un logiciel gratuit et open source. HypPy a été développé en utilisant le langage Python et s'appuie fortement sur les capacités de traitement de tableaux de packages tels que NumPy et SciPy. HypPy peut être utilisé pour l'analyse précoce des données et le prototypage des chaînes de traitement. HypPy a été conçu dans un souci de simplicité et de spectrométrie d'imagerie pratic Simultanément, il intègre plusieurs concepts innovants et puissants.

HypPy est un package en développement et pour le développement, et a prouvé ses mérites. Sa simplicité et sa structure flexible de menus de premier niveau rendent HypPy attrayant pour une utilisation dans l'éducation et les cours de courte durée. De nouvelles méthodes de recherche peuvent être rapidement programmées et testées. Des exemples de tels projets de recherche sont le projet pilote de cartographie planétaire de Mars de l'ESA et le projet de recherche visant à déterminer la clé de voûte et le sourire des caméras à ligne hyperspectrale.

Dans des projets, il a été utilisé pour construire des chaînes de traitement entières, depuis les données hyperspectrales jusqu'à l'information. En bref, HypPy peut être utilisé comme plateforme pour générer et tester de nouvelles idées pour le traitement des données hyperspectrales. Le noyau de la lecture et de l'écriture d'images hyperspectrales, le mappeur de longueur d'onde, les filtres de gradient hyperspectraux, les filtres spatio-spectraux, les outils de mathématiques de bande et de mathématiques spectrales avec leur propre langage d'expression, les outils d'arbre de décision, l' utilisation de statistiques zonales et une bibliothèque d'images et spectrale dédiée. les téléspectateurs en sont de bons exemples .

Nous espérons qu'HypPy apportera également une grande valeur ajoutée à d'autres organisations et individus dans les domaines de la recherche, de l'éducation et des projets utilisant des données hyperspectrales.

Contributions des auteurs : Conceptualisation, WB et FvR ; méthodologie, BM; logiciel, WB; validation, WB, FvR, HvdW et CH ; ressources, AD; rédaction – préparation du projet original, WB ; rédaction – révision et édition, HvdW ; visualisation, WB ; supervision, HvdW et FvdM ; administration de projet , BM ; acquisition de financement, FvdM et FvR Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Nous reconnaissons le soutien financier reçu de l'Agence spatiale européenne (projet de cartographie planétaire de Mars) et de Wijnand van Eindthoven de Deep Atlas. Le financement fourni par ces organismes a joué un rôle crucial dans la réussite de notre projet de recherche.

Déclaration de disponibilité des données : HypPy peut être téléchargé depuis Github et Zenodo, voir [42] ; Nom : HypPy (Python hyperspectral) ; Licence : Licence publique générale GPU version 3 ; Contact : whbakker@utwente.nl ; Configuration matérielle requise : ordinateur portable ou de bureau Windows, Mac ou Linux ; Langage du programme : Python version 3 ; Configuration logicielle minimale requise : numpy, scipy, Pillow, matplotlib, pysptools ; Taille du programme : 12,2 Mo.

Remerciements : Nous sommes reconnaissants à l'Université de Twente d'avoir fourni les ressources et les installations essentielles qui ont grandement soutenu l'avancement fluide de notre recherche. De plus, nous remercions tous les étudiants qui ont contribué directement ou indirectement à ce travail en utilisant HypPy au cours de leurs études. Leur collaboration et leurs encouragements ont joué un rôle déterminant dans la création d'HypPy.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Abréviations

Les abréviations suivantes sont utilisées dans ce manuscrit :

API Une interface de programmation
Liste des mauvais groupes BBL
BIL Bande entrelacée par ligne
Séquentiel de bande BSQ
Interface de ligne de commande CLI
Interface utilisateur graphique GUI
SWIR Infrarouge à ondes courtes
VNIR visible et proche infrarouge

Machine Translated by Google

Références

1. NV5 Geospatial Solutions Inc. ENVI—Environnement de visualisation d'images. 2023. Disponible en ligne : https://www.nv5 geospatialsoftware.com/docs/using_envi_Home.html (consulté le 24 novembre 2023).
2. Prokakis, E. Logiciels libres et open source : liberté, transparence et efficacité à l'ère de la numérisation. J. Éthique politique Nouvelle technologie. AI 2022, 1, e31230. [Référence croisée]
3. Harris, CR ; Millman, KJ ; van der Walt, SJ; Gommers, R. ; Virtanen, P. ; Cournapeau, D. ; Wieser, E. ; Taylor, J. ; Berg, S. ; Smith, New Jersey ; et coll. Programmation de tableaux avec NumPy. Nature 2020, 585, 357-362. [Référence croisée] [Pub Med]
4. Virtanen, P. ; Gommers, R. ; Oliphant, TE; Haberland, M. ; Reddy, T. ; Cournapeau, D. ; Burovski, E. ; Peterson, P. ; Weckesser, W. ; Lumineux, J. ; et coll. SciPy 1.0 : Algorithmes fondamentaux pour le calcul scientifique en Python. Nat. Méthodes 2020, 17, 261-272. [Référence croisée] [Pub Med]
5. Hunter, JD Matplotlib : un environnement graphique 2D. Calculer. Sci. Ing. 2007, 9, 90-95. [Référence croisée] 6. van Rossum, G. ; Drake, FL Manuel de référence Python 3 ; CreateSpace : Scotts Valley, Californie, États-Unis, 2009. Disponible en ligne : https : //dl.acm.org/doi/book/10.5555/1593511 (consulté le 31 juillet 2024).
7. Tragheim, DG ; Marais, SH; Pedley, RC; Napier, BR; Bateson, L. ; Smith, AG; Marchant, AP; Gunnink, JL; Oosthoek, JHP; Muller, JP Le projet pilote de cartographie planétaire de Mars ; Agence spatiale européenne (ESA) : Paris, France, 2010.
8. Konstantinides, K. ; Rasure, JR L'environnement de développement logiciel Khoros pour le traitement de l'image et du signal. IEEETrans. Processus d'images. 1994, 3, 243-252. [Référence croisée] [Pub Med]
9. Jeune, M. ; Argiro, D. ; Kubica, S. Cantate. Calcul ACM SIGGRAPH. Graphique. 1995, 29, 22-24. [Référence croisée]
10. Ousterhout, JK ; Jones, K. ; Foster-Johnson, E. ; Boursiers, D. ; Griffin, B. ; Welton, D. Tcl et Tk Toolkit, 2e éd. ; Addision-Wesley Série informatique professionnelle ; Addison-Wesley : Upper Saddle River, NJ, États-Unis, 2009.
11. Roseman, M. Modern Tkinter pour les développeurs Python occupés : apprenez rapidement à créer de superbes interfaces utilisateur pour Windows, Mac et Linux utilisant la boîte à outils GUI standard de Python ; Presse de fin d'après-midi : Victoria, Colombie-Britannique, Canada, 2020.
12. Kokaly, RF ; Clark, inf.; Swayze, Géorgie ; Livo, KE ; Hoefen, TM; Pearson, Caroline du Nord ; Sage, RA ; Benzel, WM; Abaisse, HA ; Driscoll, RL; et coll. Bibliothèque spectrale USGS version 7 ; Rapport technique ; USGS : Reston, VA, États-Unis, 2017. [CrossRef]
13. Bibring, JP; Langevin, Y. ; Altieri, F. ; Arvidson, R. ; Beilud, G. ; Berthél, M. ; Douté, S. ; Drossart, P. ; Encrenaz, T. ; Oubliez, F. ; et coll. OMEGA : Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité ; Agence spatiale européenne, ESA (publication spéciale) : Paris, France, 2009 ; SP-1291 ; pp. 75-95. Disponible en ligne : https://sci.esa.int/s/WnjXOYW (consulté le 31 juillet 2024).
14. Vane, G. ; Vert, RO ; Chrien, TG; Enmark, HT; Hansen, EG; Porter, WM Le spectromètre aéroporté d'imagerie visible/infrarouge (AVIRIS). Télédétection Environ. 1993, 44, 127-143. [Référence croisée]
15. Jia, G. ; Huéni, A. ; Tao, D. ; Geng, R. ; Schaepman, MOI ; Zhao, H. Récupération de la réflectance spectrale en super-résolution à partir des données d'un spectromètre d'imagerie de télédétection. Opter. Express 2016, 24, 19905-19919. [Référence croisée] [Pub Med]
16. Marien, C. ; Dijkstra, AH; Wilkins, C. L'altération hydrothermale de la carbonatite dans le complexe Fen, Norvège : minéralogie, géochimie et implications pour la formation de ressources en éléments de terres rares. Minéral. Mag. 2018, 82, S115-S131. [Référence croisée]
17. Richards, JA; Jia, X. Analyse d'images numériques par télédétection ; Springer : Berlin/Heidelberg, Allemagne, 1999. [CrossRef]
18. Driscoll, M. Pillow : Traitement d'images avec Python ; Leanpub : Victoria, Colombie-Britannique, Canada, 2021. Disponible en ligne : https://leanpub.com/oreiller (consulté le 31 juillet 2024).
19. Malvern Panalytical Ltd. Format de fichier ASD — Version 8. Rapport technique, Malvern Panalytical. 2017. Disponible en ligne : https : //www.malvernpanalytical.com/en/learn/knowledge-center/user-manuals/asd-file-format-v8 (consulté le 31 juillet 2024).
20. Hettinger, R. Quoi de neuf dans Python 3.8. 2019. Disponible en ligne : https://docs.python.org/3/whatsnew/3.8.html (consulté sur 24 novembre 2023).
21. Tucker, CJ Combinaisons linéaires infrarouges rouges et photographiques pour la surveillance de la végétation. Télédétection Environ. 1979, 8, 127-150. [Référence croisée]
22. Chein-I Chang. Divergence d'informations spectrales pour l'analyse d'images hyperspectrales. Dans Actes du Symposium international de géoscience et de télédétection IEEE 1999, IGARSS'99 (Cat. No.99CH36293), Hambourg, Allemagne, 28 juin-2 juillet 1999 ; Volume 1, p. 509-511. [Référence croisée]
23. Hewson, RD; Cudahy, TJ; Mizuhiko, S. ; Ueda, K. ; Mauger, AJ Génération transparente de cartes géologiques à l'aide d'ASTER dans le Broken Province de Hill-Curnamona en Australie. Télédétection Environ. 2005, 99, 159-172. [Référence croisée]
24. Hecker, C. ; van Ruitenbeek, FJA; van der Werff, HMA ; Bakker, WH; Hewson, RD; van der Meer, FD Analyse des caractéristiques d'absorption spectrale pour la recherche de minerai : un didacticiel sur l'utilisation de la méthode en télédétection géologique. Géosci IEEE. Télédétection Mag. 2019, 7, 51-71. [Référence croisée]
25. van Ruitenbeek, FJA; Bakker, WH; van der Werff, HMA ; Zegers, TE; Oosthoek, JHP; Omer, ZA; Marais, SH; van der Meer, FD Cartographie de la position de longueur d'onde des caractéristiques d'absorption les plus profondes pour explorer la diversité minérale dans les images hyperspectrales. Planète. Science spatiale. 2014, 101, 108-117. [Référence croisée]
26. Ellson, J. ; Gansner, ER; Koutsofios, E. ; Nord, Caroline du Sud ; Woodhull, G. Graphviz et Dynagraph — Dessin graphique statique et dynamique Outils. Dans un logiciel de dessin graphique ; Lien Springer : New York, NY, États-Unis, 2004 ; 127-148. [Référence croisée]
27. CCI. Programme de maîtrise ès sciences en sciences de la géo-information et observation de la Terre — Année universitaire 2023-2024. 2023. Disponible en ligne : https://studyguide.itc.nl/m-geo (consulté le 24 novembre 2023).

Machine Translated by Google

28. Portela, B. ; Sepp, MD; van Ruitenbeek, FJA; Hecker, C. ; Dilles, JH Utilisation de l'imagerie hyperspectrale pour l'identification des intercroissances de pyrophyllite-muscovite et d'alunite dans l'environnement épithermal peu profond du district de cuivre porphyre de Yerington. Minerai Géol. Rév. 2021, 131, 104012. [CrossRef]

29. Portela, B. Cours de courte durée — Télédétection et spectroscopie minérale pour le géologue d'exploration et d'exploitation minière, du 7 au 11 novembre 2022, Istanbul. 2022. Disponible en ligne : https://www.linkedin.com/posts/brunobvportela_geologicalremotesensing-istanbul-turquie-activity-7001884993555693568-bkXG (consulté le 24 novembre 2023).
30. Groupe de télédétection géologique. Conférence annuelle et AGA du GRSG 2022 : Orbit to Outcrop. 2022. Disponible en ligne : https://www.grsg.org.uk/grsg-agm-conference-2022 (consulté le 24 novembre 2023).
31. Chicarro, A. ; Martin, PD ; Trautner, R. La mission Mars Express : un aperçu. Dans Mars Express : Une mission européenne sur la planète rouge ; ESA : Noordwijk, Pays-Bas, 2004 ; Tome SP-1240. Disponible en ligne : https://sci.esa.int/documents/33745/35957/ 1567254632829-OverviewWeb.pdf (consulté le 31 juillet 2024).

32. Booysen, R. ; Lorenz, S. ; Thiele, ST; Fuchsloch, WC; Marais, T. ; Suivant, PAM ; Gloaguen, R. Imagerie hyperspectrale précise des affleurements minéralisés : un exemple de pegmatites contenant du lithium à Uis, Namibie. Télédétection Environ. 2022, 269, 112790. [Référence croisée]
33. Scafutto, RDM ; van der Werff, H. ; Bakker, WH; van der Meer, F. ; de Souza Filho, CR Une évaluation des spectromètres d'imagerie SWIR aéroportés pour la cartographie du CH4 : implications du positionnement de la bande, de l'échantillonnage spectral et du bruit. Int. J.Appl. Obs. de la Terre. Géoinfo. 2021, 94, 102233. [Réf. croisée]
34. Bakker, W. ; van der Werff, H. ; van der Meer, F. Détermination du sourire et de la clé de voûte des caméras à ligne hyperspectrale de laboratoire. Dans les actes du 10e atelier 2019 sur l'imagerie hyperspectrale et le traitement du signal : évolution de la télédétection (WHISPERS), Amsterdam, Pays-Bas, 24-26 septembre 2019 ; p. 1 à 5. [Référence croisée]
35. Kopacková, V. ; Koucká, L. Cartographie minérale basée sur la détection automatique de multiples caractéristiques d'absorption. EARSeL EProc. 2014, 13, 95-99. [Référence croisée]
36. Therien, C. Bienvenue dans la documentation PySptools. 2022. Disponible en ligne : https://pysptools.sourceforge.io/ (consulté sur 24 novembre 2023).
37. Boggs, T. Guide de l'utilisateur de Spectral Python (SPy). 2020. Disponible en ligne : https://www.spectralpython.net/user_guide.html (accédé le 24 novembre 2023).
38. Grizonnet, M. ; Michel, J. ; Poughon, V. ; Inglada, J. ; Savinaud, M. ; Cresson, R. Orfeo ToolBox : Traitement open source des données distantes capter des images. Ouvrez Géospat. Logiciel de données. Rester. 2017, 2, 15. [Réf. croisée]
39. van der Linden, S. ; Rabe, A. ; Jakimow, B. ; Thiel, F. ; Cooper, S. ; Okujeni, A. ; Hostert, P. Intégration de la spectroscopie d'imagerie et du SIG — Analyse d'images gratuite et open source dans QGIS avec EnMAP-Box 3. Dans les actes du congrès OSA sur les capteurs optiques et la détection 2021 (AIS, FTS, HISE, SENSORS, ES), Washington, DC, États-Unis, 19-23 juillet 2021 ; Groupe d'édition Optica : Washington, DC, États-Unis, 2021 ; p. HF4E.2. [Référence croisée]

40. van Rossum, G. ; Fridman, L. Pourquoi Python 3.11 est si rapide. 2022. Entretien sur YouTube. Disponible en ligne : https://youtu.be/TLhRuZ9cJWc (consulté le 31 juillet 2024).
41. Modular Inc. Mojo : un nouveau langage de programmation. 2023. Disponible en ligne : https://www.modular.com/mojo (consulté le 24 novembre 2023).
42. Bakker, WH avec mhbakker/hyppy : HypPy3. Zenodo 2024. [CrossRef]

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Conception de systèmes d'intelligence spectrale multicanaux utilisant Algorithme de transformation de Fourier discret approximatif pour antenne Applications de perception du spectre basées sur les baies

Arjuna Madanayake [1], Keththura Lawrance 1,* , Bopage Umesha Kumarasiri 3 et [1], Sivakumar Sivasankar [1], Ainsihara Gunaratne [2], US Edussooriya Renato J. Cintra Chamira [4]

1 Département de génie électrique et informatique, Florida International University, Miami, FL 33174, États-Unis ; amadanay@fiu.edu (AM); ukuma003@fiu.edu (BUK); ssiva011@fiu.edu (SS)
2 Centre de recherche Herzberg en astronomie et en astrophysique, Conseil national de recherches Canada, Penticton, Colombie-Britannique V0H 1K0, Canada ; donchara.gunaratne@nrc-cnrc.gc.ca
3 Département d'ingénierie électronique et des télécommunications, Université de Moratuwa, Moratuwa 10400, Sri Lanka ; chamira@uom.lk
4 Département de technologie, Universidade Federal de Pernambuco, Caruaru 55015, PE, Brésil ; rjdsc@de.ufpe.br
* Correspondance : klawr041@fiu.edu

Résumé : Le spectre radioélectrique est une ressource rare et extrêmement précieuse qui nécessite une attention particulière. surveillance en temps réel et allocation dynamique des ressources. L'accès dynamique au spectre (DSA) est un nouveau paradigme de gestion du spectre radioélectrique, qui nécessite des algorithmes basés sur l'IA/ML pour un fonctionnement optimal. performances dans des conditions de canal en évolution rapide et d'éventuelles cyberattaques dans le domaine électromagnétique. Détection rapide dans plusieurs directions à l'aide de processeurs matriciels, avec suivi Algorithmes basés sur l'IA/ML pour la détection et la perception des formes d'onde mesurées à partir du L'environnement est essentiel pour fournir une aide à la décision en DSA. Dans le cadre du directionnel et du large bande la perception du spectre, la capacité de canaliser finement les entrées à large bande en utilisant une analyse de Fourier efficace est indispensable. Cependant, une transformation de Fourier rapide (FFT) à grain fin sur un grand nombre des directions nécessite des calculs intensifs et conduit à une surface de puce et une consommation d'énergie élevées. Nous abordons ce problème en exploitant la transformée de Fourier discrète approchée récemment proposée. (ADFT), qui possède sa propre factorisation clairsemée pour une implémentation en temps réel avec une faible complexité et la consommation d'énergie. L'ADFT est utilisé pour créer un formateur de faisceaux numérique RF multifaisceau à large bande et une unité d'attention basée sur le spectre temporel qui surveille 32 directions discrètes sur 32 sous-bandes en temps réel à l'aide d'un algorithme sans multiplicateur avec une faible complexité de calcul. La sortie de cette unité d'attention spectrale est appliquée comme variable de décision à un récepteur intelligent qui adapte son fréquence centrale et résolution de fréquence via des canaliseurs FFT conçus sur mesure pour le temps réel surveillance à haute résolution. Ce processus en deux étapes permet d'appliquer la FFT à gain fin uniquement à directions et bandes d'intérêt déterminées par l'espace-temps 2D de faible complexité basé sur l'ADFT unité d'attention. La FFT à grain fin fournit une signature spectrale qui peut trouver de futurs cas d'utilisation dans moteurs de réseau neuronal pour réaliser la reconnaissance de modulation, l'identification des appareils IoT et la RFI identification. Des algorithmes de formation de faisceaux et de canalisation spectrale, une architecture informatique numérique, et les premiers prototypes utilisant un récepteur multicanal entièrement numérique à 32 éléments et programmable sur le terrain La radio logicielle (SDR) haute vitesse basée sur un réseau de portes (FPGA) est présentée.

Mots clés : transformée de Fourier discrète approchée ; perception du spectre à large bande ; large bande formation de faisceaux numérique RF multifaisceaux ; algorithmes de faible complexité

check for updates

Citation : Madanayake, A. ; Laurent, K. ; Kumarasiri, BU ; Sivasankar, S. ; Gunaratne, T. ; Edussooriya, CUS ; Cintra, RJ Conception du multicanal Systèmes d'intelligence spectrale utilisant Fourier discret approximatif Algorithme de transformation pour l'antenne Perception du spectre basée sur les baies Applications. Algorithmes 2024, 17, 338. https://doi.org/10.3390/a17080338

Rédacteur académique : Frank Werner

Reçu : 3 juillet 2024
Révisé : 24 juillet 2024
Accepté : 26 juillet 2024
Publié : 1 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.

Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Dans le classique de Tolkien Le Seigneur des Anneaux, un « œil qui voit » omniprésent et inspire jusqu'à présent les habitants de la Terre du Milieu à agir en fonction de leur proximité avec les objets sous l'influence du Seigneur Sauron [1]. Dans notre monde, nous allons permettre des

Machine Translated by Google

connaissance omniprésente du spectre électromagnétique via la détection dans l'espace, le temps et la forme d'onde grâce à la perception des radiofréquences (RF) et permet ainsi une action intelligente pour le DSA basé sur l'IA. Nous avons besoin d'un œil qui voit tout RF pour le spectre. La fusion de la détection du spectre RF avec l'IA et l'apprentissage automatique (ML) permet une perception complète du spectre. L'omniprésence, qui implique partout, et la persistance, qui implique d' avoir constamment accès à des informations sur ce qui pourrait se produire dans l'environnement électromagnétique, conduiront à des capacités incroyablement puissantes pour les systèmes sans fil intelligents de demain. Le spectre est une ressource naturelle extrêmement précieuse et rare ; cependant, la croissance exponentielle de la connectivité sans fil conduit inévitablement à une pénurie de spectre. Le problème de la pénurie de spectre peut en théorie être atténué en déplaçant les réseaux de communication vers la bande des ondes millimétriques dans la gamme de 30 à 300 GHz . Néanmoins, l'industrie a récemment appris une amère leçon avec les ondes millimétriques. En effet, la technologie des ondes millimétriques fonctionne en laboratoire ; Pourtant, un certain nombre de problèmes sont présents : (i) il est difficile de le faire fonctionner dans des environnements réels, car les faisceaux RF pointus nécessaires au maintien de la connectivité sont sujets aux blocages ; (ii) la recherche de faisceaux est un problème difficile à résoudre, notamment si l'on considère les aspects de mobilité ; (iii) les amplificateurs de puissance (PA) sont inefficaces et gaspillent beaucoup d'énergie ; et (iv) les temps de cohérence du canal sont très faibles, ce qui nécessite des mises à jour plus rapides des informations sur l'état du canal (CSI), entraînant une charge de calcul élevée.

1.1. Groupe FR3 : Le spectre Boucle d'or

D'une part, les bandes de fréquences FR1/sub-6 GHz (le spectre existant) ne sont tout simplement pas disponibles pour une croissance accrue. En effet, la sous-bande de 100 MHz à 3,5 GHz a été vendue pour la somme époustouflante de 29 milliards de dollars, démontrant clairement la demande de l'industrie du sans fil. Cependant, tout l'argent du monde ne peut pas vous fournir des ressources qui n'existent pas. Étant donné qu'il est peu probable que les ondes millimétriques conduisent à des systèmes pratiques dans un avenir proche, l'industrie se tourne vers une solution intermédiaire. Ce juste milieu spectral se situe dans la bande FR3 (6 GHz à 24 GHz), affectueusement surnommée la bande moyenne supérieure par les acteurs de l'industrie. Dans nos travaux récents, nous avons montré que la bande moyenne supérieure n'est pas trop basse au point d'être aussi surpeuplée que la FR1 (bande héritée) et n'est pas trop haute pour que la construction d'une infrastructure pratiquement évolutive ne soit pas actuellement viable. La bande FR3 est « parfaite » car elle se situe dans une plage favorable aux systèmes évolutifs. De plus, la bande passante est disponible dans FR3 pour la prochaine génération de technologie sans fil, à condition que l'on puisse naviguer dans l'enchevêtrement de la politique du spectre [2].

1.2. Une économie à spectre partagé

Il existe un problème urgent lié à l'utilisation de la bande moyenne supérieure pour la connectivité sans fil. La gamme de 6 GHz à 24 GHz a déjà été attribuée à des applications critiques et ses licences appartiennent au ministère de la Défense (DoD) et à d'autres entités gouvernementales. La plupart des capteurs critiques du DoD, notamment les radars en bande X, les réseaux de communication, les bandes expérimentales scientifiques (par exemple, la radioastronomie) et les capteurs radar météorologiques existent également dans la bande FR3. Les utilisateurs primaires (PU) n'abandonneront probablement jamais leur « précieux » spectre de façon permanente ; cependant, l'avenir nécessitera probablement le DSA, où les utilisateurs secondaires (SU) tels que les réseaux sans fil commerciaux peuvent utiliser les bandes FR3 d'une manière qui ne présente que peu ou pas de menace pour les applications critiques poursuivies par les PU. Même si les DSA peuvent apporter d'énormes bénéfices et une croissance économique à la société, un tel pouvoir s'accompagne de grandes responsabilités dans la mesure où le respect des accords de partage doit être assuré avec non-répudiation et confiance zéro.

1.3. Le besoin d'une perception spectrale par l'IA en temps

réel « Le DoD a déterminé que le partage est réalisable si certaines fonctionnalités avancées d'atténuation des interférences et un cadre de coordination pour faciliter le partage du spectre sont mis en place » proclame la stratégie nationale du spectre 2023 de la Maison Blanche [3]. La détection précoce des actes de guerre figure plus haut sur la liste des priorités que la qualité de service des réseaux sans fil commerciaux. Même si le risque d'un événement cataclysmique est faible, il n'est pas nul. La seule façon de garantir que les SU n'empiètent pas sur les allocations de spectre des PU dans l'espace, le temps,

Machine Translated by Google

et les domaines de forme d'onde se font au moyen de mesures indépendantes, en supposant une confiance zéro pour garantir pleinement le respect des accords de partage de spectre. Aux erreurs réelles et aux interférences RF nuisibles involontaires s'ajoutent des cybermenaces imminentes émanant d'acteurs malveillants jusqu'ici inconnus. La possibilité de brouiller les infrastructures critiques sans fil et de défense ne doit pas être prise à la légère. Des mesures fiables et à haute sensibilité, avec une détection RF partout et à tout moment, permettent une meilleure résilience à de telles menaces futures du point de vue de la guerre du spectre.

1.4. Les capteurs de spectre avec perception omniprésente

L'autonomie RF grâce aux systèmes d'IA créés à l'aide de l'apprentissage profond (DL) et de l'apprentissage par renforcement (RL) devrait générer une connaissance approfondie – c'est-à-dire la « perception » – de l'environnement électromagnétique détecté [4]. La stratégie nationale du spectre (Maison Blanche , 11/2023) appelle à des études initiales dans les bandes de 3,1 GHz à 3,45 GHz, de 5,030 GHz à 5,091 GHz, de 7,125 GHz à 8,400 GHz et de 18,1 GHz à 18,6 GHz en vue d'un DSA entre le secteur commercial et le DoD. entités. La perception basée sur l'IA nécessite une détection du spectre RF et la détection des paramètres de forme d'onde (par exemple, les paramètres de modulation et peut-être même de codage [5]), l'agrégation de porteuses et les parties de bande passante ou des informations similaires, les directions de propagation, les niveaux de puissance et les comportements des canaux tels que les trajets multiples. , Doppler et métriques d'interférence inter-symboles (ISI) [6]. Les capteurs de spectre typiques ne fournissent probablement pas d'informations exploitables sur le spectre, à l'exception des statistiques d'utilisation [7-10], où la détection de « trous de spectre » détermine la ou les sous-bandes éventuellement disponibles pour les SU [11-17]. La fraude au spectre est un problème majeur [18]. La présence d'interférences RF nuisibles (RFI) ne peut être prouvée que par une détection et une perception par un tiers indépendant. Pour internaliser un modèle du paysage spectral, nous aurons besoin d'une bonne compréhension de l'environnement RF mesuré ; Le simple regroupement d'énergie à l'aide d'une FFT est trop limitant pour les cas d'utilisation futurs. Atténuer les événements RFI partout et à tout moment est un défi et nécessite la détection du brouillage, des armes à impulsions et de la fraude spectrale.

Nous avons besoin d'une perception en temps réel utilisant l'intelligence artificielle, car les événements se produisent partout et à tout moment, à des vitesses de plusieurs milliers d'événements par seconde, excluant ainsi toute approche impliquant l'humain dans la boucle. La perception entièrement basée sur l'IA de l'environnement RF est d'une importance cruciale et sert de source d'informations intelligente pour piloter le DSA automatisé [19].

1.5. Contribution du document

Nous proposons dans cet article un système de détection de spectre intelligent et efficace sur le plan informatique . Nous incorporons un formateur de faisceaux multifaisceau basé sur DFT pour détecter directionnellement les signaux RF. Afin de réduire la complexité de calcul du formateur de faisceaux multifaisceaux, nous utilisons la transformée de Fourier discrète approximative (ADFT) au lieu de la DFT. Le back-end de traitement numérique du formateur de faisceaux multifaisceau proposé est implémenté sur un réseau de portes programmables par l'utilisateur (FPGA). Le back-end de traitement numérique connecté à un réseau d'antennes composé de 32 antennes patch fonctionnant à 5,7 GHz confirme la détection du spectre en temps réel . En outre, la configuration expérimentale démontre l'application de la formation de faisceaux multifaisceaux basée sur l'ADFT dans la détection du spectre, offrant une approche prometteuse pour une perception du spectre directionnel en temps réel, de faible complexité

1.6. Organisation du journal

Le reste de cet article est organisé comme suit. La section 2 propose un examen complet, commençant par une exploration de l'accès dynamique au spectre, se concentrant à la fois sur les techniques de traitement du signal et les approches d'apprentissage automatique/d'apprentissage profond, suivie d'une discussion sur la radioastronomie et les interférences radiofréquences. La section 3 se penche sur la transformée de Fourier discrète approchée, offrant un aperçu du contexte théorique et de la mise en œuvre spécifique de l'ADFT. Dans la section 4, nous présentons un aperçu détaillé du système, couvrant l'architecture proposée, le frontal analogique, le noyau ADFT et l'estimation spectrale ADFT numérique, le spectromètre FFT numérique et la connectivité numérique à hau La section 5 présente les résultats expérimentaux, y compris les mesures d'étalonnage et de faisceau,

Machine Translated by Google

démontrant les performances pratiques de la méthode proposée. Enfin, la section 6 discute des implications de nos résultats et décrit les orientations potentielles pour les recherches futures.

2. Révision

Nous passons en revue les travaux liés à l'accès dynamique au spectre dans cette section. Nous considérons d'abord les techniques de traitement du signal utilisées pour la détection spectrale. Nous examinons plus en détail l' apprentissage automatique et les techniques d'apprentissage profond récemment proposées dans la première sous-section. Nous passons ensuite en revue la radioastronomie avec les radiotélescopes de nouvelle génération, en particulier la détection spectrale des RFI.

2.1. Accès dynamique au spectre : traitement du signal et approches ML/DL Le DSA a

été envisagé pour la première fois il y a vingt ans sur la base des mesures de l'utilisation réelle du spectre effectuées par le groupe de travail sur la politique du spectre de la Federal Communications Commission (FCC) [20]. Contrairement à la croyance majoritaire selon laquelle le spectre est très encombré , les mesures réelles ont indiqué que de nombreuses parties du spectre sont inactives à un moment et à un endroit donnés, conduisant à des trous spectraux. Cela résulte de la politique d'accès au spectre fixe. Sous DSA, les utilisateurs secondaires (c'est-à-dire les utilisateurs sans licence) sont autorisés à utiliser les trous spectraux des utilisateurs principaux (c'est-à-dire les utilisateurs sous licence) sans interférer avec le fonctionnement des utilisateurs principaux via une radio cognitive [21-24], qui est un système intelligent construit sur une radio définie par logiciel. Cependant, il existe une limite supérieure à la capacité de canal qui peut être atteinte avec DSA [25]. Ici, nous passons brièvement en revue les techniques de détection du spectre associées au DSA, et le lecteur peut se référer à [26-28] pour d'autres aspects du DSA, tels que la gestion du spectre et les politiques de partage du spectre.

Plusieurs techniques ont été proposées pour la détection du spectre à bande étroite et à large bande . Pour la détection à bande étroite, les principales techniques non ML sont la détection d'énergie [29-32], la détection basée sur des filtres adaptés [33-35], la détection de caractéristiques cyclostationnaires [36,37] et la détection basée sur la covariance [38-40]. Les approches de détection d'énergie et de détection basée sur la covariance ne nécessitent pas de connaissance préalable du signal de l'utilisateur principal ; cependant, le premier n'est pas fiable à de faibles valeurs de rapport signal/bruit (SNR). La détection basée sur un filtre adapté offre une détection optimale et une meilleure détection à de faibles valeurs SNR, mais cette technique nécessite une connaissance préalable du signal de l'utilisateur principal, qui n'est pas toujours disponible.
La détection des caractéristiques cyclostationnaires est généralement robuste face à l'incertitude du bruit et fonctionne bien à de faibles valeurs SNR ; cependant, il souffre d'un temps de détection important. Dans la détection à large bande, comme approche simple, un signal à large bande peut être divisé en un ensemble de signaux à bande étroite via une DFT temporelle ou un banc de filtres [41-44] et détecté séquentiellement ou en parallèle à l'aide de techniques de détection à bande étroite. Cependant, la détection séquentielle nécessite plus de temps, tandis que la détection parallèle nécessite plus de ressources. Parce qu'un signal à large bande avec spectre peut être considéré comme un signal clairsemé dans le domaine fréquentiel, plusieurs techniques basées sur la détection compressive ont été proposées pour la détection à large bande (45-48). Ces approches nécessitent des taux d'échantillonnage plus faibles et ont une consommation d'énergie plus faible en plus d'une détection rapide . Ainsi, les approches basées sur la détection compressive éliminent la plupart des inconvénients de la détection séquentielle ou parallèle à bande étroite d'un signal à large bande.

Au cours de la dernière décennie, un nombre considérable de techniques ML et DL ont été proposées pour la détection spectrale. Les algorithmes ML ont été principalement utilisés pour déterminer les modèles d'occupation des canaux d'un canal principal utilisé et pour classer le canal comme libre ou occupé. Dans [49–53], les modèles de regroupement de K-moyennes et de mélange gaussien ont d'abord été utilisés pour déterminer la présence d'utilisateurs primaires, et un classificateur tel qu'une machine à vecteurs de support et le K-voisin le plus proche est ensuite utilisé pour déterminer si un canal est libre ou occupé. Des techniques DL telles que les réseaux neuronaux convolutifs et la mémoire à long terme ont été utilisées pour la détection du spectre dans [54-57], qui ont surpassé les techniques conventionnelles. Récemment, une architecture DL basée sur un transformateur a été proposée dans [58]. Cette approche DL surpasse les approches proposées précédemment basées sur les réseaux de neurones convolutifs. Une autre approche DL utilisée pour la détection spectrale de manière coopérative est l'apprentissage par renforcement profond [59-6

Machine Translated by Google

les approches d'apprentissage améliorent la robustesse du système de détection spectrale et lui permettent de prendre des décisions plus précises dans des environnements dynamiques. Le lecteur peut se référer à [63–65] pour une revue complète des techniques de détection du spectre.

La modélisation spatio-temporelle des systèmes radio cognitifs utilisant des concepts de traitement du signal multidimensionnel a été présentée dans [66]. Ici, la détection directionnelle a été exploitée en utilisant des filtres tridimensionnels à réponse impulsionnelle infinie (IIR) comme formateurs de faisceaux. Des approches de détection directionnelle similaires utilisant des filtres multidimensionnels ont été présentées dans [67–69].
De plus, des systèmes de réseau d'antennes ont été utilisés pour la détection directionnelle dans [70–72].
Dans [73], un multifaisceau basé sur DFT a été utilisé pour la détection directionnelle en temps réel.
Par rapport au formateur multifaisceaux utilisé avec ADFT dans cette étude, la DFT exacte a été utilisée dans [73] avec 16 faisceaux simultanés à 2,4 GHz. Nous notons que la limite supérieure de la capacité du canal qui peut être atteinte avec DSA [25] n'est pas modifiée en utilisant l'ADFT au lieu du DFT.

2.2. Radioastronomie et RFI

La radioastronomie s'appuie sur un spectre radio peu contaminé pour observer l'univers [74]. La détection et l'estimation des signaux célestes faibles provenant de galaxies lointaines sont particulièrement nécessaires pour atteindre les objectifs scientifiques clés des radiotélescopes de nouvelle génération, à savoir le Square Kilometer Array (SKA) (https://www.skao.int/ consulté le 24 juillet 2024) [75], le 2030 Atacama Large Millimeter/sub-millimeter Array — Wideband Sensitivity Upgrade (ALMA-WSU) (https://science.nrao.edu/facilities/alma/ science_sustainability/wideband-sensitivity-upgrade consulté le 24 juillet 2024) [76], et le Very Large Array ngVLA de nouvelle génération) (https://ngvla.nrao.edu/ consulté le 24 juillet 2024) [77]. Bien qu'ils soient situés dans des régions isolées et peu peuplées du monde, l'intensité, la fréquence, la bande passante et les occurrences croissantes de RFI menacent l' utilité des radiotélescopes de nouvelle génération [78]. Les RFI du système mondial de navigation par satellite (GNSS) [79], des systèmes de communications sans fil terrestres et aéroportés, des méga constellations de satellites [80] et d'autres technologies empiètent sur les spectres qui ont été exclusivement utilisés par les radioastronomes, ce qui rend plus difficile la distinction. faibles signaux célestes de RFI.

L'Union internationale des télécommunications (UIT) (https://www.itu.int/en/Pages/ par défaut.aspx consulté le 24 juillet 2024) et la FCC (https://www.fcc.gov/ consulté le 24 juillet 2024) jouent un rôle important dans la protection du spectre pour la radioastronomie et d'autres utilisateurs passifs. L'UIT, à travers son secteur des radiocommunications (ITU-R), est chargée d' élaborer des réglementations et des recommandations internationales pour l'utilisation des radiofréquer
Il reconnaît l'importance de protéger le spectre pour les services passifs, tels que la radioastronomie, les services d'exploration de la Terre par satellite et les services météorologiques par satellite.
L'UIT-R identifie des bandes de fréquences spécifiques pour ces services passifs et établit des dispositions réglementaires pour garantir leur protection contre les interférences nuisibles. Ces dispositions comprennent des procédures de coordination, des limites de puissance et des exigences de séparation de fréquence pour protéger les observations et mesures sensibles effectuées par les radioastronomes et autres utilisateurs passifs [81,82].
De même, la FCC aux États-Unis reconnaît l'importance de protéger le spectre pour la radioastronomie et d'autres services passifs.
L'Office of Engineering and Technology (OET) de la FCC formule des politiques et des règles pour éviter les interférences nuisibles avec ces services. Il établit des normes techniques et des conditions d'autorisation qui tiennent compte des besoins des utilisateurs passifs, notamment des radioastronomes . En outre, la FCC travaille en étroite collaboration avec la National Science Foundation (NSF) et d'autres agences compétentes pour coordonner l'utilisation du spectre et protéger l'intégrité des observations scientifiques. Cependant, les deux organisations subissent des pressions incessantes de la part de l'industrie et des gouvernements pour une utilisation plus commerciale du spectre précédemment alloué aux utilisateurs non commerciaux.

Par conséquent, idéalement, la connaissance des caractéristiques clés de la RFI (c'est-à-dire la gamme de fréquences, le schéma de modulation, l'emplacement des sources, l'intensité, etc.) aiderait les opérateurs de radiotélescopes à planifier et à programmer leurs observations. Néanmoins, ce n'est pas pratique pour tout le monde

Machine Translated by Google

cas de RFI ; par conséquent, les enregistrements de l'intensité spectrale radio en termes d'intensité, de direction, de durée et d'heure de l'incident aideraient également les radioastronomes à signaler et à exciser les observations contaminées [83]. Dans ce cas, la perception omniprésente de l'environnement EM serait une énorme aubaine dans l'étape de post-traitement des observations radioastronomiques [84].

3. DFT approximative

Dans cette section, nous passons en revue les principes fondamentaux de l'approximation DFT en décrivant sa justification technique et son formalisme mathématique en termes de problème d'optimisation. Ensuite, nous nous concentrons sur la DFT approximative à 32 points de pointe, qui est décrite numériquement avec son algorithme rapide associé. Une analyse de complexité arithmétique pour l' ADFT en 32 points discutée est également fournie avec une comparaison avec les méthodes de référence.

3.1. Contexte théorique La

transformée de Fourier discrète est un opérateur linéaire qui relie un signal d'entrée à N points au x = x[0] x[1] . . . x[N − 1] selon signal de sortie X = X[0] X[1] . . . X[N − 1] l'expression suivante :

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot \omega_N^{-nk} \qquad k = 0, 1, \ldots, N-1, \tag{1}$$

où ωN = exp(2πj/N) est la Nième racine de l'unité, et j $\sqrt{-1}$. Calculée par définition, la DFT présente une complexité de calcul en O(N2), prohibitive pour les applications temps réel.

En pratique, la DFT est calculée au moyen de l'une des nombreuses routines numériques efficaces : des algorithmes rapides, collectivement appelés « FFT ». Les choix particulièrement populaires de FFT sont les algorithmes basés sur la base 2, tels que la FFT Cooley – Tukey, qui peuvent décomposer de manière récursive le bloc DFT à N points en deux sous-blocs DFT à N/2 points [85].

En général, les FFT sont capables de réduire la complexité du calcul DFT de O(N2) à O(N log N) [86]. Dans de nombreux cas [87–92], une telle réduction des coûts de calcul atteint la complexité multiplicative minimale théorique de la DFT [93].
En conséquence, il est impossible de réduire davantage la complexité multiplicative. Il s'agit d'un domaine de recherche assez mature, et la proposition de nouvelles FFT capables d'offrir des réductions significatives des coûts de calcul est un événement peu probable.

Toutefois, des réductions notables peuvent être obtenues dans le cadre d'un paradigme de tolérance aux pannes. Si le calcul DFT est autorisé à un niveau d'erreur donné, alors l'effort de calcul peut être réajusté pour atteindre le niveau de précision inférieur, mais acceptable. En conséquence, moins d' opérations arithmétiques sont nécessaires pour l'estimation DFT [94].

3.2. Problème d'optimisation

De manière systématique, un moyen d'obtenir de tels estimateurs DFT de faible complexité consiste à recourir à la théorie de l'approximation matricielle. Dans le formalisme matriciel, la DFT est décrite par l' opération suivante :

$$X = F_N \cdot x, \tag{2}$$

où la matrice DFT à N points, notée FN, est la matrice N × N avec des éléments donnés par fk,n = ω N , k, n = 0, 1, . . . , Dans le contexte N − 1.

des approximations des transformations discrètes, qualitativement, un N approximatif est une transformée discrète, Fˆ Matrice de transformation Fˆ, les signaux exacts et approximatifs du domaine de transformation sont « proches » dans un sens pertinent d'être définis quantitativement. Généralement, une transformation d'approximation doit satisfaire aux conditions suivantes : (i) elle peut largement préserver les propriétés significatives de la transformation exacte ; (ii) il existe une proximité mathématique entre la matrice de transformation exacte et approximative selon une fonction métrique contextuellement pertinente - généralement un chiffre de performance de

Machine Translated by Google

mérite; et (iii) il possède un coût de calcul nettement inférieur, par rapport au modèle exact calcul par FFT. Dans [95], un aperçu de la théorie est disponible.

Habituellement, afin de préserver l'interprétation physique du spectre approximatif, on cherche à ce que les matrices de transformation approchées soient proches au sens euclidien de la associer une matrice de transformation exacte. Ceci peut être accompli grâce à la minimisation de la norme de Frobenius de la différence entre les matrices exacte et candidate pour approximation, sous réserve de la restriction d'une faible complexité. Par conséquent, une faible complexité l'approximation peut être obtenue en limitant ses éléments à des ensembles numériques de triviaux multiplicandes, tels que {0, ±1} ou {0, ±1, ±2}.

En symboles, une formulation possible pour le problème de dérivation d'une approximation DFT est donné par [95]

$$\min_{\hat{f}_{k,n}} \quad \left\| F_N - S \cdot \hat{F}_N \right\|_F^2 \tag{3}$$

$$\text{sous réserve de} \quad \hat{f}_{k,n} \in \{0, \pm 1\}, \tag{4}$$

· $\hat{F}$ où $\hat{f}_{k,n}$, où $k, n = 0, 1, \ldots, N-1$, désigne les entrées de la matrice approchée

· $\|\cdot\|_F$ désigne la norme de Frobenius matricielle, et S est une matrice de normalisation utilisée pour garantir que les vecteurs de base de $S \cdot \hat{F}_N$ ont une énergie égale ou proche de un.

3.3. ADFT en 32 points

Dans ce travail, nous avons adopté l'ADFT en 32 points introduit mathématiquement dans [96] et développé plus en détail en théorie et en matériel dans [97,98]. La forme numérique explicite du L'ADFT à 32 points sélectionnée est donnée dans l'Annexe A. La figure 1 montre une comparaison des réponses en amplitude pour l'ADFT exacte à 32 points et l'ADFT à 32 points. Les réponses de phase le long des maxima, c'est-à-dire les faisceaux, restent inchangés et toute déviation de phase uniquement se produit dans les régions de la bande d'arrêt profonde. Par conséquent, l'influence de la nature de l'approximation de l'ADFT sur les phases des signaux formés en faisceau est négligeable.

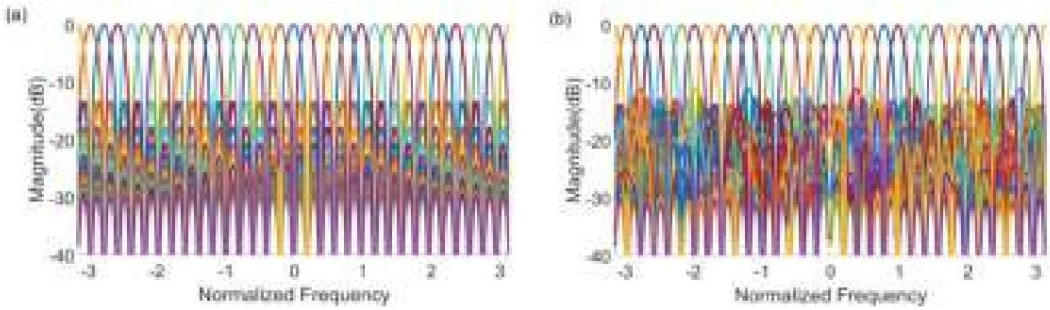


Figure 1. Réponse d'amplitude simulée de (a) une DFT idéale à 32 points et (b) une faible complexité à 32 points ADFT [97,98].

3.4. Algorithme rapide et analyse de complexité

Représenté par la matrice de faible complexité $\hat{F}_{32}$, l'ADFT sélectionné admet le factorisation suivante :

$$\hat{F}_{32} = W8 \cdot W7 \cdot W6 \cdot W5 \cdot W4 \cdot W3 \cdot W2 \cdot W1, \tag{5}$$

où Wi , je = 1, 2, . . . , 8, sont des matrices clairsemées et de faible complexité. Dans l'Annexe A, nous fournir la définition explicite des matrices Wi , i = 1, 2, . . . , 8.

A titre de comparaison, le tableau 1 répertorie le nombre d'opérations arithmétiques pour la DFT exacte à 32 points. calculé selon sa définition et la FFT Cooley – Tukey radix-2. Pour cette opération en comptant, nous supposons que (i) le signal d'entrée a une valeur purement réelle ; (ii) multiplications triviales par 0, ±1 ou ±j ne sont pas comptés ; et (iii) une multiplication entre un nombre réel et un nombre complexe le nombre est effectué comme deux multiplications réelles – et non comme une multiplication complexe [99]. En raison de (i) et (iii), les 704 multiplicandes non triviaux de $\hat{F}_{32}$ équivalent à 1408 multiplications réelles.

Machine Translated by Google

L'évaluation du nombre d'ajouts a considéré la somme du nombre de réels et
ajouts de parties imaginaires. Il a également été fourni la complexité arithmétique du problème discuté.
ADFT 32 points par sa définition et grâce à son algorithme rapide (factorisation). Le
L'analyse de complexité de la FFT Cooley – Tukey radix-2 est détaillée dans [99].

Tableau 1. Performances du DFT, du Cooley – Tukey FFT et de l'ADFT en 32 points discuté. Banal
les multiplications par ±1 ou ±j ne sont pas comptées ; seuls les signaux d'entrée à valeur réelle sont supposés.

Méthode	Multiplications réelles	De vrais ajouts
DFT exacte de 32 points (F32)	1408	1666
FFT Radix-2 Cooley – Tukey (N = 32) [99]	88	408
DFT approximatif (F^ 32) [96]	0	1282
Algorithme rapide pour F^32 ($\prod_{je=1}^{8}$ Wi) [97]	0	144

Le tableau 2 montre la complexité arithmétique pour chaque terme matriciel individuel du F^ 32 fac-
torisation. La matrice W8 ne présente pas de coûts arithmétiques car elle est constituée uniquement de
opérations de changement de signe et d'échange de données (échange de parties réelles et imaginaires).
L'échange de données est une conséquence de la multiplication par j (une rotation par π/2).

Tableau 2. Performances des facteurs matriciels de l'algorithme rapide ADFT. Multiplications triviales par ±1
ou ±j non compté ; seul le signal d'entrée à valeur réelle est supposé.

Opération	W1	W2	W3	W4	W5	W6	W7	W8
De vraies multiplications	0	0	0	0	0	0	0	0
De vrais ajouts	30	30	14	14	30	14	12	0

4. Présentation du système

Dans cette section, nous décrivons la conception et la réalisation des antennes, des récepteurs micro-ondes ,
des circuits convertisseurs de signaux mixtes et des réalisations numériques basées sur FPGA qui constituent
le système de perception du spectre spatio-temporel proposé. Les antennes et le micro-onde
les systèmes ont été conçus et réalisés en interne, tandis que le système FPGA a été acheté auprès de
la communauté des développeurs d'instruments de radiotélescope open source (CASPER) qui est en
charge des principaux instruments radio, tels que le Square Kilometer Array (SKA), Murchison
Réseau à champ large, réseau de télescopes Allen et autres.

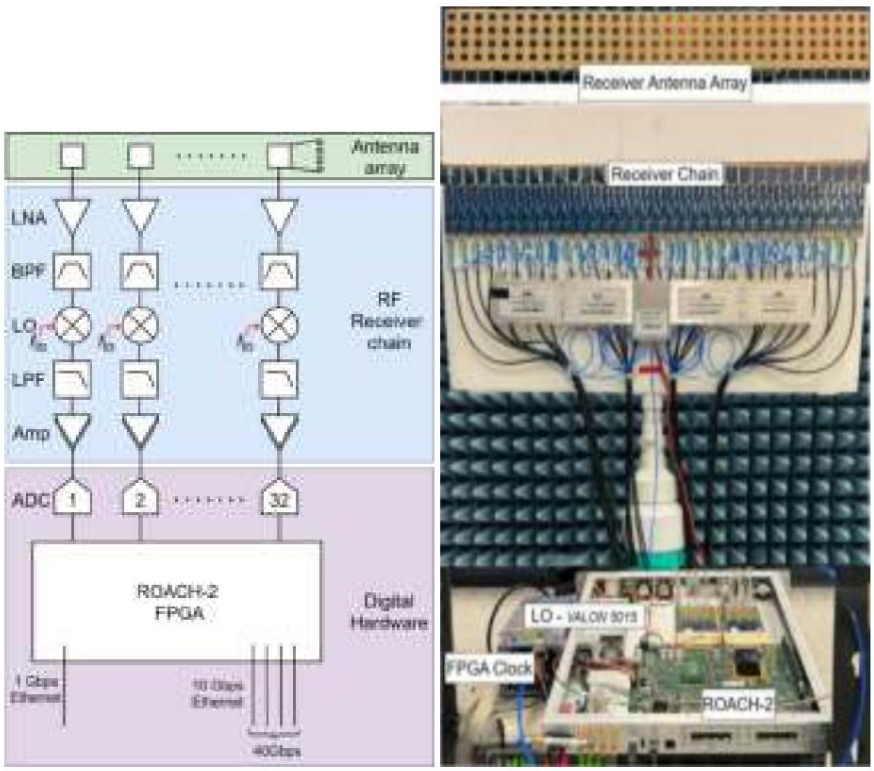
4.1. Architecture proposée

Valider et mesurer expérimentalement les faisceaux générés par l'ADFT proposé
algorithmes, un système RF de 5,7 GHz avec un back-end de traitement numérique a été méticuleusement
conçu et mis en œuvre. La figure 2a montre l'architecture globale du système basée sur
un réseau d'antennes à 32 éléments, englobant des sous-systèmes clés tels que le réseau d'antennes lui-même,
la chaîne de réception RF et l'unité matérielle numérique. La figure 2b montre la configuration expérimentale
de l'architecture proposée. En plus des noyaux numériques chargés d'exécuter le
Pour la DFT spatiale des signaux échantillonnés, des circuits supplémentaires ont été développés au sein du FPGA.
Ce circuit supplémentaire facilite la manipulation des données en temps réel et les mesures de faisceaux
sous différents angles, améliorant ainsi la flexibilité et la capacité du système.

Le matériel informatique à architecture ouverte reconfigurable (ROACH-2) basé sur le numérique
La plate-forme de traitement du signal (DSP) est illustrée à la figure 3 et joue un rôle crucial en
servant de couche intermédiaire, permettant la communication entre le matériel FPGA
(plus précisément sa mémoire) et l'ordinateur de bord. La carte ROACH-2 s'interface avec
le serveur Linux hôte principal via une connexion Ethernet 1 Gbps . Une routine Python en cours d'exécution
sur le serveur hôte est programmé pour interagir avec la plateforme ROACH-2 , effectuant
fonctions de contrôle et récupération des données de la mémoire FPGA tout en balayant de manière itérative
sous différents angles. Cette architecture système unique garantit non seulement un faisceau précis

Machine Translated by Google

mesures, mais améliore également l'efficacité globale et l'automatisation de l'expérience, en fournissant une plate-forme robuste pour valider les algorithmes avancés de formation de faisceaux.



(un) (b)

Figure 2. (a) L'architecture du système ; (b) la configuration expérimentale du récepteur réseau à 32 faisceaux fonctionne sur une bande de 5,700 GHz à 5,800 GHz, suivie du back-end numérique du système FPGA ROACH-2 (Xilinx Virtex-6 sx475T FPGA).



Figure 3. Plate-forme DSP basée sur ROACH-2 basée sur le FPGA Xilinx Virtex-6 Sx475 et une carte ADC à 32 canaux .

Nous remercions chaleureusement le Dr Dan Werthimer de l'UC Berkeley et la communauté CASPER pour leurs contributions aux outils ROACH-2 et CASPER.

Machine Translated by Google

4.2. Frontaux analogiques

Le frontal RF du formateur de faisceaux en mode réception a été construit à l'aide d'un réseau linéaire uniforme (ULA) à 32 éléments fonctionnant à 5,7 GHz, associé à 32 chaînes de récepteurs RF à conversion directe sur des cartes de circuits imprimés (PCB). Les éléments du réseau sont espacés de 0,6λ, ce qui équivaut à environ 32 mm à 5,7 GHz. Chaque élément d'antenne est conçu comme un sous-réseau vertical composé d'antennes patch 4 × 1, facilitant la formation de faisceau passive dans le plan vertical. Cette formation de faisceaux analogique fonctionne indépendamment des algorithmes de formation de faisceaux spatiaux basés sur la FFT discutés dans ce travail.

Les sorties des éléments d'antenne sont directement connectées à 32 récepteurs hétérodynes . Ces récepteurs utilisent un système d'oscillateur local (LO) centralisé, qui comprend un oscillateur à faible bruit de phase relié à un réseau diviseur de puissance à 32 sorties. L'étage initial de chaque récepteur comprend un amplificateur à faible bruit (LNA) qui offre un gain de 16 dB à 5,8 GHz avec un facteur de bruit de 2,4 dB. Après l'amplification, les signaux sont filtrés à travers un filtre passe-bande allant de 4,7 GHz à 6 GHz pour éliminer le bruit et les interférences hors bande. Les signaux filtrés sont ensuite mélangés et passés à travers des filtres passe-bas pour produire des entrées basse fréquence intermédiaire (IF) converties vers le bas.

Par la suite, ces 32 signaux Low-IF sont amplifiés d'environ 30 dB et numérisés à l'aide de deux cartes de conversion analogique-numérique (ADC) ADC16x250-8 . Chaque carte ADC prend en charge 16 canaux d'entrée asymétriques, offre une résolution de 8 bits et fonctionne à des taux d'échantillonnage allant jusqu'à 250 Méchantillons/s par canal. Cette configuration garantit la numérisation haute fidélité des signaux reçus pour un traitement ultérieur [97].

4.3. Noyaux ADFT et estimation spectrale ADFT numérique

L'ADFT en 32 points discuté a été implémenté à l'aide des outils Xilinx. Les matrices clairsemées de la factorisation présentée dans (5) ont été structurées sous forme de blocs d'entrée-sortie parallèles pour optimiser les performances [97, 98]. La vérification de ces blocs a été réalisée via MATLAB Simulink 8.5, garantissant la fiabilité de la conception avant le déploiement matériel. Par la suite, la conception ADFT vérifiée a été mappée sur le FPGA Xilinx Virtex-6 sx475T.

Pour intégrer le cœur ADFT dans la plate-forme FPGA ROACH-2 , nous avons utilisé deux bibliothèques de la communauté Collaboration for Astronomy Signal Processing and Electronics Research (CASPER) [100,101]. Cette intégration a permis au système de générer 32 faisceaux RF simultanés en parallèle. Chaque faisceau correspond à une direction discrète, calculée à l'aide de la fonction θk = arcsin(k/32), où k varie de −16 à 15. La mise en œuvre réussie de l'ADFT sur le FPGA ROACH-2 a non seulement validé l'efficacité de l'algorithme mais également a démontré l'application pratique de techniques avancées de conception de matériel numérique . Ce travail met en évidence le potentiel de l'ADFT dans diverses applications nécessitant une formation de faisceaux RF efficace.

Nous avons mis en œuvre une technique de fenêtrage spatial pour résoudre le problème des lobes secondaires élevés dans l'ADFT, où le lobe secondaire le plus défavorable a été observé comme étant 1,5 dB supérieur à celui de la DFT (exacte). Quatre configurations ont été testées : trois fonctions de fenêtrage (Butterworth, Hamming et Hanning) et une ligne de base sans fenêtrage [85]. La sélection parmi ces fonctions de fenêtrage était gérée à l'aide d'un multiplexeur équipé d'une broche de sélection de commande.

La figure 4 montre l'unité de fenêtrage spatial couplée aux matrices clairsemées ADFT. En termes mathématiques, le fenêtrage est présenté comme une matrice diagonale. L' estimation spectrale basée sur l'ADFT peut être décrite en combinant la matrice diagonale de fenêtrage (Dwindowing), le noyau ADFT de faible complexité (Fˆ 32) et une matrice diagonale de normalisation de puissance (Dnormalizer). La représentation mathématique de l' estimation approximative du spectre proposée est donnée par

$$\text{ADFTestimator} = \text{Dnormalizer} \cdot F\hat{}\,32 \cdot \text{Dwindowing}. \quad (6)$$

Machine Translated by Google

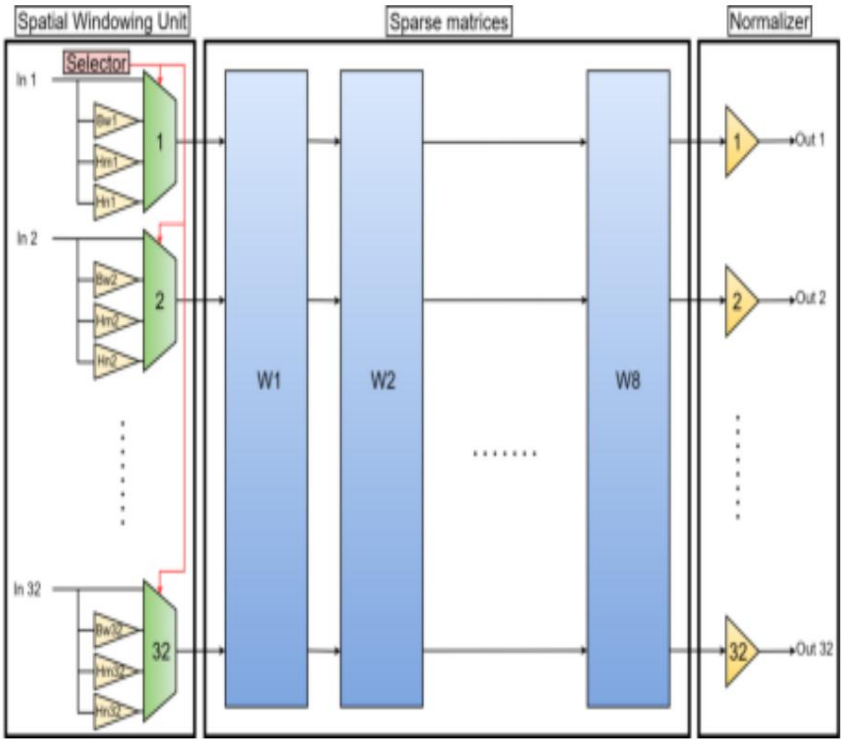


Figure 4. Architecture ADFT avec fenêtrage spatial et unités de normalisation de puissance.

La figure 5 illustre la réponse ADFT et ses modifications sous les techniques de fenêtrage. Il convient de noter que l'introduction de fonctions de fenêtrage a abouti à des lobes principaux plus larges, ce qui constitue un compromis pour supprimer les lobes secondaires. Ce compromis entre la largeur du lobe principal et la suppression des lobes secondaires est à prendre en compte dans l'application des techniques de fenêtrage spatial. Cependant, la mise en œuvre du fenêtrage spatial augmente la complexité matérielle en raison des multiplexeurs introduits par l'unité de fenêtrage. Par conséquent, l'utilisateur doit sélectionner la configuration appropriée en fonction de ses exigences et contraintes spécifiques.

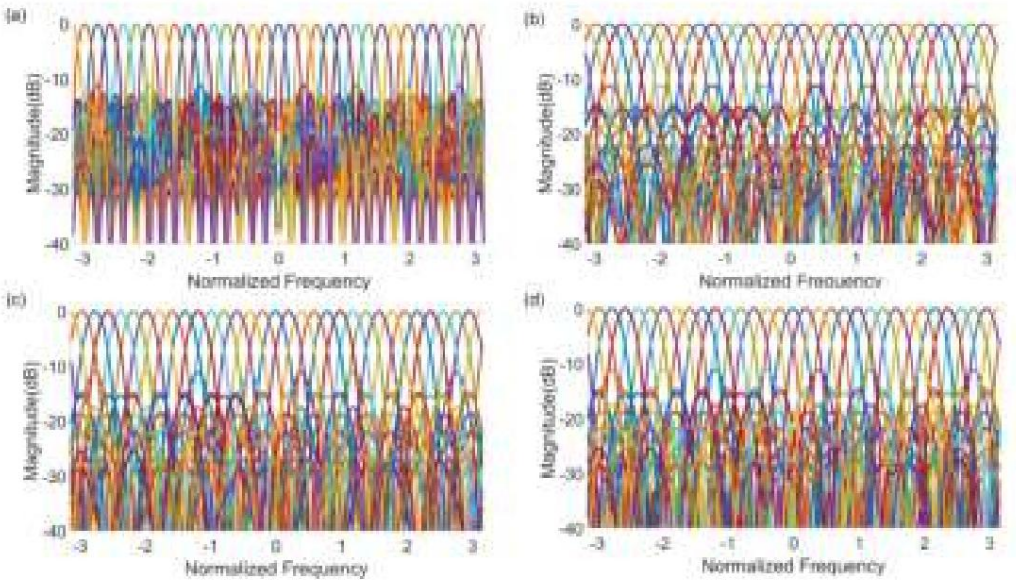


Figure 5. Réponse d'amplitude simulée de l'ADFT à 32 points et ses modifications sous trois techniques de fenêtrage différentes. Le sous-tracé (a) montre la réponse de base sans aucun fenêtrage. Les sous-parcelles (b – d) démontrent les réponses ADFT après avoir appliqué respectivement les fenêtres Butterworth, Humming et Hanning .

Machine Translated by Google

4.4. Spectromètres FFT numériques

Le système décrit implique la mise en œuvre d'une FFT spatiale qui génère 32 sorties parallèles, chacune correspondant à un faisceau RF spatial distinct. Suite à cette FFT spatiale, un spectromètre FFT temporel est appliqué à chacun de ces 32 faisceaux RF spatiaux, réalisant ainsi une canalisation dans le domaine des fréquences temporelles. Chaque FFT temporelle est calculée avec 1024 points, nécessitant l'utilisation de 32 cœurs FFT parallèles. Ce traitement parallèle est facilité par l'exploitation des blocs FFT biplex échantillonnés réels disponibles dans le FPGA Xilinx Virtex-6 SX475T.

Les blocs FFT biplexes échantillonnés réels exploitent la méthode de conjugaison hermitienne, permettant le calcul de la FFT échantillonnée réelle. Cette technique utilise un noyau FFT complexe pour transformer simultanément deux flux de données à valeur réelle. En conséquence, l'opération FFT produit toutes les fréquences, à la fois positives et négatives, les fréquences négatives étant les images miroir de leurs homologues positives [100].

Les sorties FFT temporelles pour chaque faisceau donnent 1 024 valeurs complexes correspondant à 1 024 cases de fréquence discrètes. La densité spectrale de puissance (PSD) est estimée en mettant au carré l'amplitude de chaque groupe de fréquences. Cette estimation est obtenue en faisant la moyenne de 1 024 évaluations FFT temporelles, ce qui donne lieu à une nouvelle mesure PSD tous les 1 million de cycles d'horloge. Un bloc accumulateur vectoriel de mémoire vive (BRAM) collecte les valeurs PSD et les transfère au processus logiciel pris en charge par le flux c
La figure 6 montre le flux global ADFT et FFT temporel, y compris l'intégration d'énergie et les registres BRAM. La bibliothèque Python casperfpga est utilisée pour interagir avec le matériel CASPER, en extrayant les valeurs du BRAM partagé.

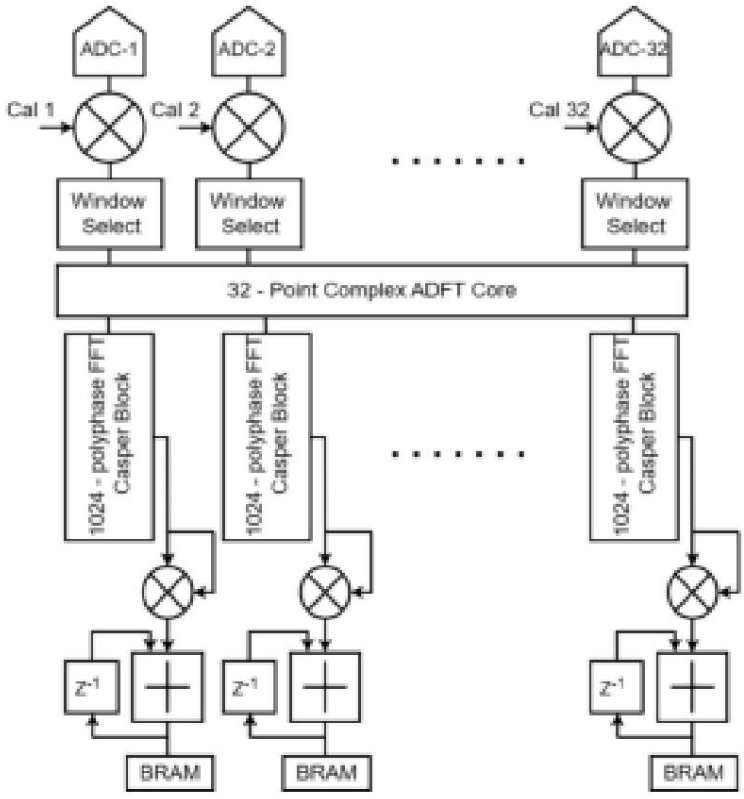


Figure 6. Étalonnage ADFT et FFT, intégration énergétique et bloc système global, y compris les registres BRAM.

Le FPGA fonctionne avec une fréquence d'horloge réglée à 200 MHz. Compte tenu de cette configuration, la résolution en fréquence du système pour chaque faisceau est déterminée comme étant ΔF = 100/1 024 MHz, soit environ 100 kHz. La bande passante Nyquist du système est donc de 100 MHz.
Cette configuration garantit une canalisation précise et une analyse de fréquence haute résolution pour chacun des 32 faisceaux RF spatiaux.

Machine Translated by Google

4.5. Connectivité numérique haut débit

Les cartes ROACH-2 sont équipées de quatre ports SFP+ sur une seule carte mezzanine 10 GbE , où l'interface Ethernet est pilotée par un oscillateur à cristal intégré de 156,25 MHz . Cette fréquence d'horloge est multipliée en interne par un facteur de 66 au sein du FPGA, ce qui donne une vitesse de fil effective de 10,3125 Gbit/s.

Le cœur CASPER 10 GbE Simulink facilite la transmission et la réception de paquets UDP sur IPv4, encapsulés dans des trames Ethernet. Chaque trame Ethernet entraîne une surcharge de 38 octets pour l'en-tête de trame, de 20 octets pour l'en-tête IPv4 et de 16 octets pour l'en-tête UDP, totalisant au moins 74 octets de surcharge par paquet. Pour s'aligner sur les limites de 64 bits, certains en-têtes peuvent également inclure un remplissage nul, ce qui a un impact supplémentaire sur le débit effectif. La taille maximale de la charge utile pour le cœur CASPER 10 GbE est de 8 192 octets, implémentés dans la BRAM, avec 512 octets supplémentaires disponibles dans la RAM distribuée pour les en-têtes spécifiques à l'application [100].

5. Résultats expérimentaux 5.1. Étalonnage

L'étalonnage du système de réseau RF est essentiel pour des performances optimales de formation de faisceau et a été réalisé en deux phases distinctes. Initialement, l'étalonnage s'est concentré sur les ADC, en utilisant des routines open source précédemment développées par les membres du groupe CASPER pour ce matériel. La deuxième phase concernait la correction numérique des disparités dans le frontal micro-ondes. Les inadéquations relatives de gain et de phase des sorties IF pour chaque chaîne ont été déterminées par rapport à une chaîne de référence en utilisant une entrée porteuse de référence de 5, 71 GHz . Les différences de gain relatives ont été compensées à l'aide de multiplicateurs numériques et les facteurs d'échelle de gain ont été calculés en mesurant les valeurs de puissance moyennes de chaque sortie ADC. Idéalement, pour la compensation de différence de phase, des filtres à retard fractionnaire sont nécessaires, car les chaînes RF pourraient avoir des retards inférieurs à la valeur de la période d'échantillonnage. En guise de solution provisoire, des filtres à retard entier ont été conçus et implémentés dans le matériel pour atténuer les différences de phase.

5.2. Mesures de faisceau

Les modèles de puissance ont été obtenus en transmettant un signal à onde continue (CW) à fRF = 5,71 GHz. La fréquence du signal LO fLO détermine le FI comme fRF - fLO. Pour les mesures de cette étude, fLO a été réglé sur 5,7 GHz, ce qui donne une FI de 10 MHz, avec les sorties converties vers le bas numérisées à fCLK = 200 MHz.

Des intégrateurs numériques à chaque sortie de bac ADFT ont été utilisés pour calculer l'énergie reçue sur une durée fixe. La formation de faisceaux spatiaux se produit à une fréquence de 200 MHz et la PSD est calculée en temps réel en mettant au carré l'amplitude de chaque groupe de fréquence. Cette estimation est dérivée de la moyenne de 1 024 calculs FFT temporels effectués à 200 MHz. La géométrie du dispositif de mesure influence de manière significative le facteur de réseau mesuré des faisceaux. Idéalement, l'émetteur et le récepteur devraient être suffisamment éloignés l'un de l'autre pour que les ondes incidentes sur le réseau de récepteurs puissent être assimilées à des ondes planes. Selon les directives standard [102], la séparation entre l'émetteur et le récepteur doit dépasser 20 m à 5,7 GHz pour garantir que l'ouverture du récepteur se trouve dans le champ lointain. Les faisceaux ont été mesurés sur un parking ouvert avec une séparation émetteur-récepteur d'environ 20 m.

La figure 7 illustre les modèles de facteurs de réseau mesurés pour les faisceaux 0 à 16. En raison de l'entrée de signaux IF à valeur réelle dans l'ADFT, les modèles de faisceaux résultants présentent une symétrie, produisant des images miroir pour les bacs restants. De plus, la présence de deux pics de miroir est observée dans les groupes 0 à 16. Les faisceaux mesurés se conforment étroitement aux modèles attendus, maintenant une séparation de plus de 10 dB par rapport à l'autre faisceau.

Chaque bac de sortie ADFT correspond à un faisceau RF en champ lointain dirigé selon un angle discret donné par $\theta k = \arcsin(k/N)$, où $k = -16, -15, \ldots, 14, 15$ et $N = 32$ (nombre d'éléments du tableau). Cette configuration se traduit par une séparation des faisceaux de $180^\circ/32 = 5{,}625^\circ$, offrant une résolution angulaire de $\pm 2{,}8125^\circ$.

Machine Translated by Google

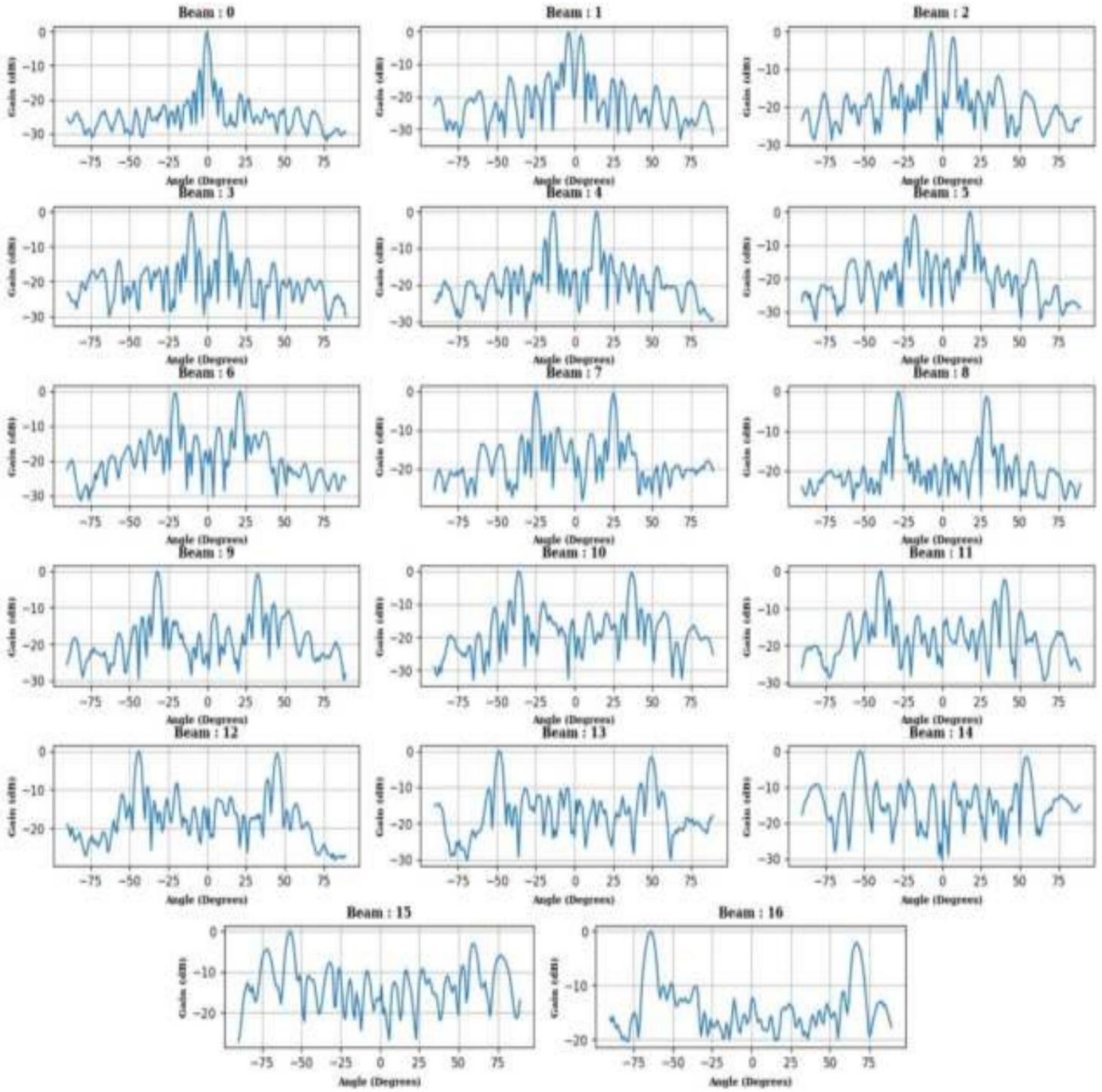


Figure 7. Seize faisceaux mesurés à partir du réseau 5,7 GHz (l'axe vertical est en décibels et l' axe horizontal est l'angle azimutal [−90◦ , 90◦]). Chaque sous-figure contient les modèles de facteurs de réseau mesurés du faisceau en utilisant les entrées réelles du noyau ADFT. La composante imaginaire de chaque entrée ADFT est ici mise à zéro, ce qui donne lieu à des lobes principaux symétriques comme prévu.

La figure 8 représente la distribution PSD temporelle mesurée des faisceaux respectifs à 35◦ plusieurs fréquences lorsque la source du signal est positionnée à un angle latéral de 15◦ , , et 50◦ . La configuration expérimentale correspondante avec les emplacements relatifs des sources et les fréquences RF émises à des fins de test est illustrée à la figure 9.

Machine Translated by Google

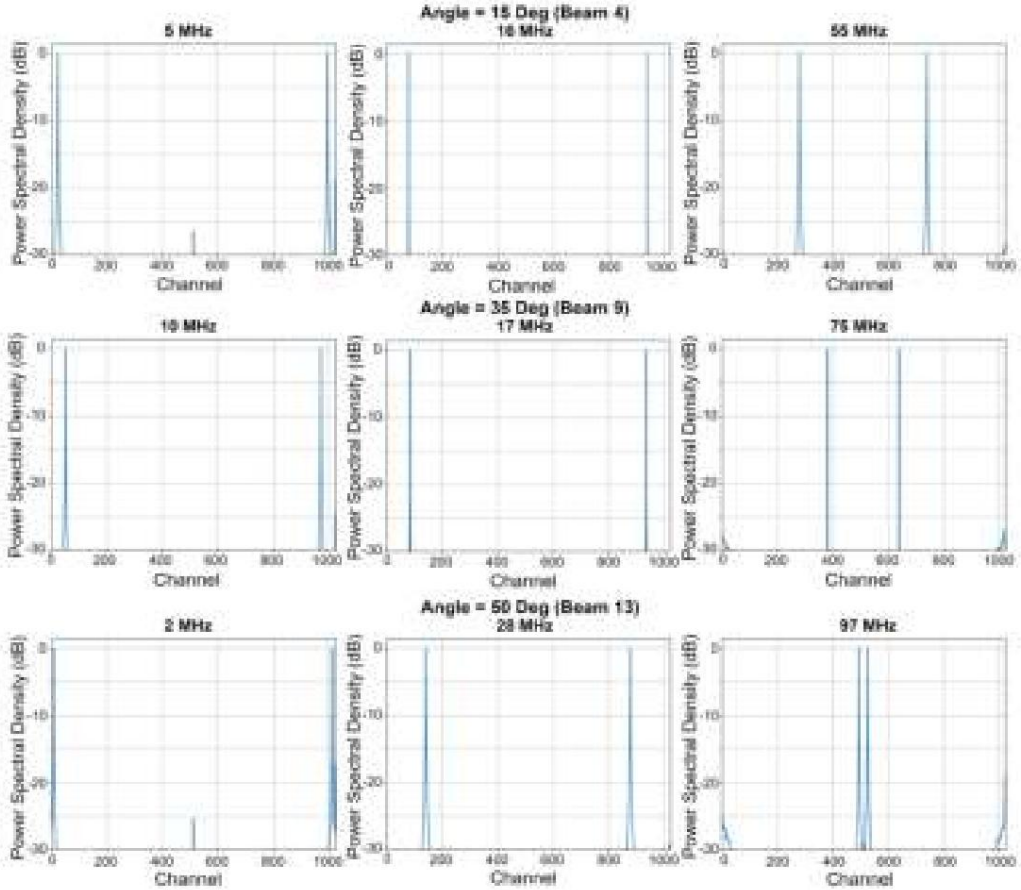


Figure 8. PSD temporelle sur les 1 024 intervalles de fréquences discrètes pour un faisceau RF particulier à plusieurs fréquences lorsque la source est placée à 15◦, 35◦, et des angles de récepteur latéraux de 50◦.

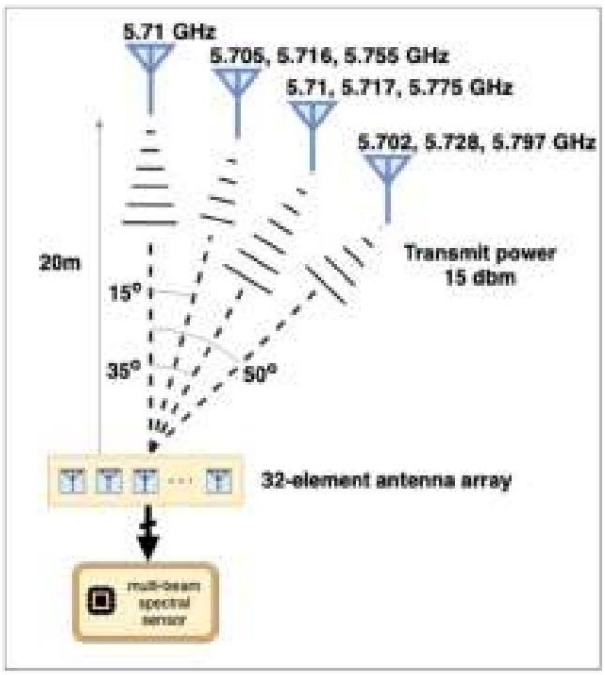


Figure 9. Illustration de la configuration globale montrant les antennes d'émission et le réseau d'antennes de réception et le capteur spectral multifaisceau. Le récepteur réseau d'antennes à 32 éléments capture les signaux de directions différentes, les émetteurs étant situés à 20 m du récepteur.

Machine Translated by Google

6. Discussion

Cet article présente une architecture de formation de faisceaux numérique de faible complexité conçue pour générer 32 faisceaux RF simultanés en utilisant une approximation DFT en 32 points, éliminant ainsi efficacement les opérations de multiplication. Nous avons validé la méthode proposée sur un modèle pleinement opérationnel Réseau de récepteurs numériques 1D à 32 éléments fonctionnant à 5,7 GHz. Le réseau 1-D utilise 32 parallèles CAN pour échantillonner les sorties d'antenne et utilise un ADFT implémenté sur un FPGA Xilinx pour calculer 32 faisceaux RF en temps réel. De plus, cet article discute de l'amélioration de notre système de détection de spectre existant à 32 canaux de 5,7 GHz grâce à la mise en œuvre de 32 FFT temporelles parallèles de 1024 bin sur une plateforme matérielle ROACH-2 . Cet expérimental la vérification avec des mesures en temps réel sur une bande de fréquences et d'angles d'incidence démontre la capacité de l'architecture à détecter des informations provenant de plusieurs directions avec une résolution en fréquence de 100 kHz et une résolution directionnelle d'environ 6◦ .

Les auteurs prévoient d'étendre l'ADFT à un plus grand nombre d'éléments et de créer formateurs de faisceaux ADFT multidimensionnels, pour l'orientation du faisceau en élévation et en azimut avions. Par exemple, en utilisant un réseau 32 × 8, nous pourrions étendre l'ensemble des faisceaux en un rectangle. grille de taille 32 × 8 où les FFT de 32 points et de 8 points sont remplacées par les FFT correspondantes. Algorithmes ADFT basés sur la factorisation matricielle clairsemée [103–106].

Contributions de l'auteur : Conceptualisation et conception de systèmes par AM Théorie et conception d'ADFT et algorithme rapide par RJC Revue du traitement du signal multidimensionnel par CUSE Revue de la radio Problèmes d'interférence en astronomie par TG Digital, implémentation du noyau ADFT par KL Experimental configuration, conception numérique des outils CASPER, calibrage et expérimentation en temps réel par SS et BUK Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Ce travail a été soutenu en partie par NSF (dans le cadre des subventions FuSE 2329012, SWIFT 2229471, ICORPS 2243346, SpectrumX), NTIA, CNPq (Brésil) dans le cadre de la subvention 315047/23-2 et Senate Research Comité, Université de Mortuwa dans le cadre de la subvention SRC/LT/2020/08.

Déclaration de disponibilité des données : Toutes les contributions originales de cette étude sont documentées dans le article. Pour plus de détails, veuillez contacter les auteurs correspondants.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Abréviations

Les abréviations suivantes sont utilisées dans ce manuscrit :

CDA	Convertisseur analogique-numérique ;
ADFT	Transformée de Fourier discrète approximative ;
IA	Intelligence artificielle;
ALMA	Atacama grand réseau millimétrique ;
BRAM	Bloquer la mémoire vive ;
CASPER	Collaboration pour le traitement du signal astronomique et la recherche en électronique ;
CSI	Informations sur l'état du canal ;
CW	Onde continue ;
TFD	Transformée de Fourier discrète ;
DL	Apprentissage profond ;
DSA	Accès dynamique au spectre ;
DSP	Traitement du signal numérique ;
FCC	Commission fédérale de la communication ;
FFT	Transformée de Fourier rapide ;
FPGA	Réseau de portes programmables sur site ;
GNSS	Système mondial de navigation par satellite ;
SI	Fréquence intermédiaire ;

Machine Translated by Google

IdO	Internet des objets ;
ISI	Interférence entre symboles ;
UIT	Union internationale des télécommunications;

LNA Amplificateur à faible bruit ;
LO Oscillateur local ;

ML	Apprentissage automatique ;

ngVLA Très grande matrice de nouvelle génération ;
Fondation nationale de la science NSF ;

Bureau d'ingénierie et de technologie de l'OET ;

PCB	Cartes de circuits imprimés ;
PSD	Densité spectrale de puissance ;
Unité centrale	Utilisateur principal ;
RF	Radiofréquence;
RFI	Interférences radiofréquences ;
R.L.	Apprentissage par renforcement ;

ROACH Matériel informatique à architecture ouverte reconfigurable ;
Radio définie par logiciel SDR ;

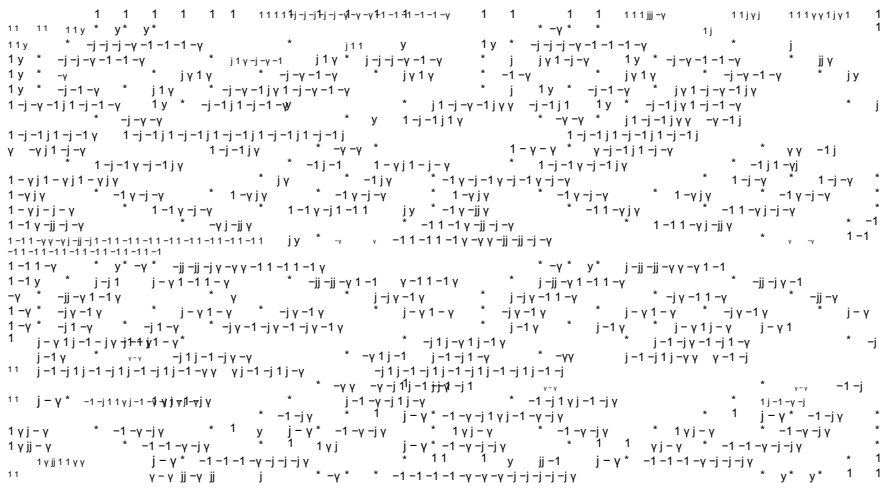
SFP	Petit facteur de forme enfichable ;
SKA	Tableau d'un kilomètre carré ;
RSB	Rapport signal/bruit ;
SU	Utilisateur secondaire ;

Réseau linéaire uniforme ULA ;
Mise à niveau de la sensibilité WSU Wideband.

Annexe A

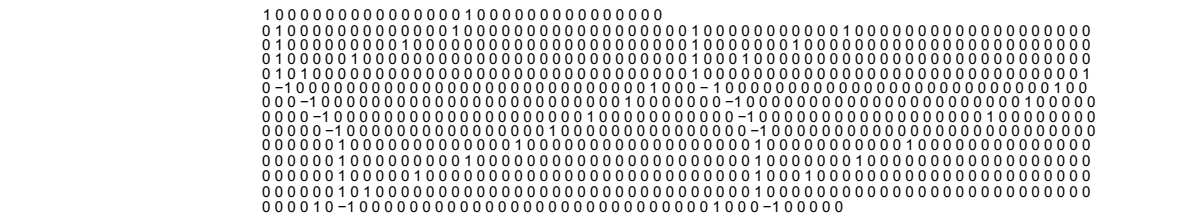
Annexe A.1. Algorithme rapide ADFT en 32 points

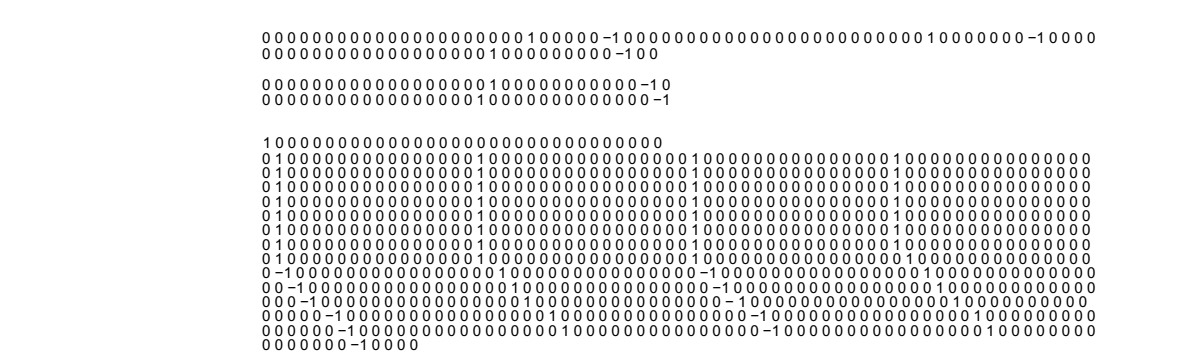
L'algorithme rapide pour la DFT approximative discutée en 32 points F^ 32 consiste en une factorisation dans des matrices clairsemées de faible complexité, comme le montre (5). La matrice approximative et son les termes de factorisation sont explicitement donnés ci-dessous :

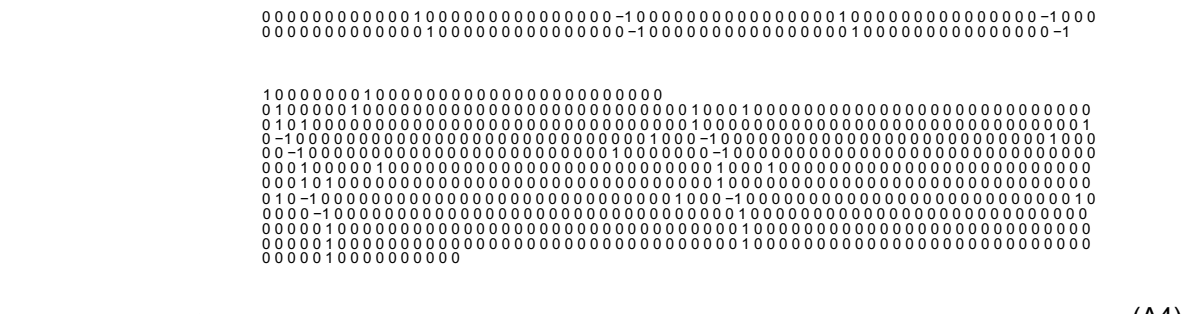
F¨32 =  , (A1)

Machine Translated by Google

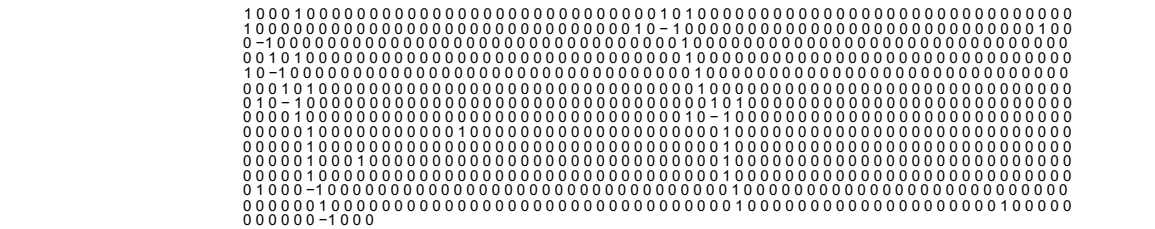
où γ = 1 + j, et * désigne une conjugaison complexe :

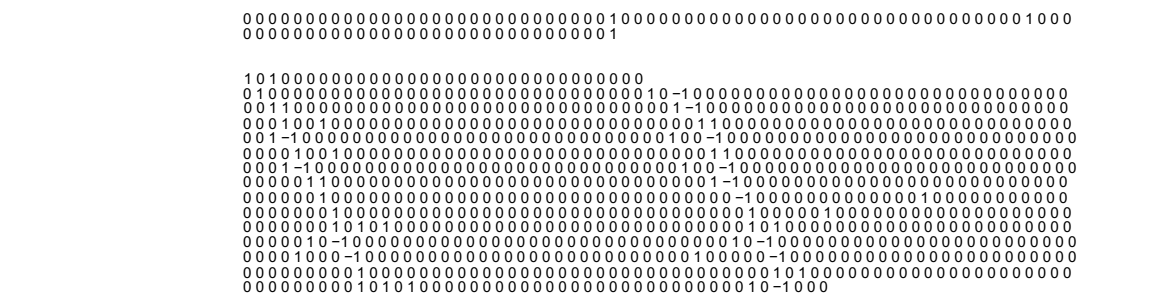
W1 =  , (A2)

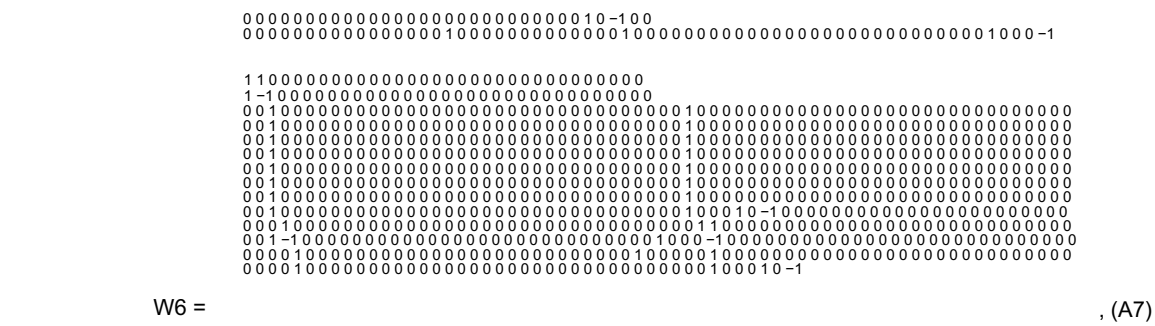
W2 =  , (A3)

W3 =  , (A4)

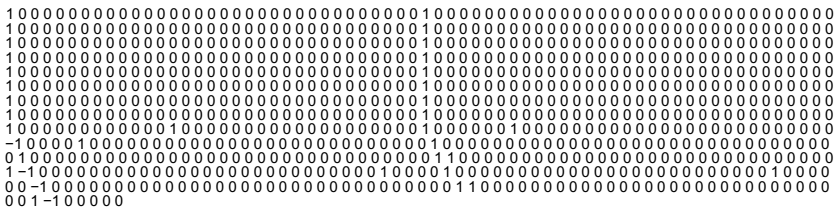
Machine Translated by Google

W4 =  , (A5)

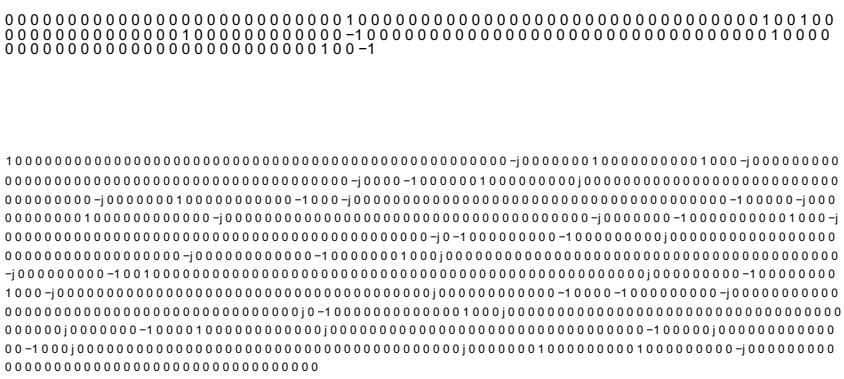
W5 =  , (A6)

W6 =  , (A7)

Machine Translated by Google

W7 =  , (A8)

et

W8 =  . (A9)

Références

1. Tolkien, J. Le Hobbit et le Seigneur des Anneaux ; Houghton Mifflin Harcourt : Boston, MA, États-Unis, 1954.
2. Kang, S. ; Mezzavilla, M. ; Rangan, S. ; Madanayake, A. ; Venkatakrishnan, SB; Hellbourg, G. ; Ghosh, M. ; Rahmani, H. ; Dhananjay, A. Réseaux cellulaires sans fil dans la bande moyenne supérieure. IEEE Ouvert J. Commun. Soc. 2024, 5, 2058-2075. [Référence croisée]
3. La Maison Blanche. Stratégie nationale du spectre ; La Maison Blanche : Washington, DC, États-Unis, 2023.
4. Restuccia, F. ; Melodia, T. Deep Learning au niveau physique : défis système et applications pour la 5G et au-delà. IEEE Commun. Mag. 2020, 58, 58-64. [Référence croisée]
5. Triaridis, K. ; Doumanidis, C. ; Chatzidiamantis, ND; Karagiannidis, GK MM-Net : Une approche multimodale vers la classification automatique des modulations. Communique IEEE. Lett. 2023, 28, 328-331. [Référence croisée]
6. Baldesi, L. ; Restuccia, F. ; Melodia, T. ChARM : Partage de spectre NextG grâce au contrôle dynamique O-RAN en temps réel basé sur les données. Dans Actes de la Conférence internationale de l'IEEE sur les communications informatiques (INFOCOM), Londres, Royaume-Uni, 2-5 mai 2022 ; IEEE : Piscataway, New Jersey, États-Unis, 2022 ; pp. 240-249.
7. Zhang, L. ; Xiao, M. ; Wu, G. ; Alam, M. ; Liang, YC ; Li, S. Une enquête sur les techniques avancées de partage de spectre dans les réseaux 5G. Fil IEEE. Commun. 2017, 24, 44-51. [Référence croisée]
8. Hu, F. ; Chen, B. ; Zhu, K. Partage complet du spectre dans les réseaux radio cognitifs vers la 5G : une enquête. Accès IEEE 2018, 6, 15754-15776. [Référence croisée]
9. Shokri-Ghadikolaei, H. ; Boccardi, F. ; Fischione, C. ; Fodor, G. ; Zorzi, M. Partage de spectre dans les réseaux cellulaires mmWave via l'association cellulaire, la coordination et la formation de faisceaux. IEEE J. Sel. Zones Commun. 2016, 34, 2902-2917. [Référence croisée]
10. Lv, L. ; Chen, J. ; Ni, Q. ; Ding, Z. ; Jiang, H. Accès multiple cognitif non orthogonal avec relais coopératif : une nouvelle frontière sans fil pour le partage du spectre 5G. Communique IEEE. Mag. 2018, 56, 188-195. [Référence croisée]
11. Agarwal, S. ; De, S. eDSA : Protocoles d'accès au spectre dynamique économes en énergie pour les réseaux radio cognitifs. IEEETrans. Foule. Calculer. 2016, 15, 3057-3071. [Référence croisée]

Machine Translated by Google

12. Chiwewe, TM; Hancke, GP Accès coopératif au spectre dynamique à convergence rapide pour les réseaux radio cognitifs. IEEE Trans. Ind. Informer. 2017, 14, 3386-3394. [Référence croisée]
13. Chang, HH ; Chanson, H. ; Yi, Y. ; Zhang, J. ; Lui, H. ; Liu, L. Accès distributif au spectre dynamique grâce à l'apprentissage par renforcement profond : une approche basée sur le calcul de réservoir. IEEE Internet Things J. 2018, 6, 1938-1948. [Référence croisée]
14. Nguyen, QG ; Nguyen, BT; Dong, TQ; Ngo, DT; Nguyen, TA Deep Q-Learning avec détection multibande pour un accès dynamique au spectre. Dans les actes du Symposium international de l'IEEE sur les réseaux d'accès dynamique au spectre (DySPAN), Séoul, République de Corée, 22-25 octobre 2018 ; p. 1 à 5. [Référence croisée]
15. Naparstek, O. ; Cohen, K. Apprentissage par renforcement multi-utilisateurs approfondi pour l'accès au spectre dynamique distribué. IEEETrans. Fil. Commun. 2019, 18, 310-323. [Référence croisée]
16. Wang, S. ; Liu, H. ; Gomes, PH; Krishnamahari, B. Apprentissage par renforcement approfondi pour un accès multicanal dynamique dans le sans fil Réseaux. IEEETrans. Cogn. Commun. Réseau. 2018, 4, 257-265. [Référence croisée]
17. Yu, Y. ; Wang, T. ; Liew, SC Apprentissage par renforcement profond Accès multiple pour les réseaux sans fil hétérogènes. IEEE J. Sel. Zones Commun. 2019, 37, 1277-1290. [Référence croisée]
18. Reston, VA Ligado Networks poursuit le gouvernement américain pour prise illégale et sans compensation du spectre exclusif sous licence FCC de Ligado. 2023. Disponible en ligne : https://ligado.com/press/ligado-networks-sues-us-government-for-unlawful-and-uncompensated-taking-of-ligados-exclusive-fcc-licensed-spectrum/ (consulté le 28 mai 2024).
19. Barden, S. ; Dewdney, P. ; Friesen, R. ; Murowinski, R. ; Savanandam, S. Rapport d'examen stratégique de la R&D d'ATD — Portefeuille de traitement du signal ; Centre de recherche Herzberg en astronomie et en astrophysique-Conseil national de recherches Canada : Victoria, Colombie-Britannique, Canada, 2022.
20. Groupe de travail sur la politique du spectre de la FCC. Rapport du Groupe de travail sur l'efficacité du spectre, novembre 2002 ; FCC : Washington, DC, États-Unis, 2024.
21. Mitola, J. ; Maguire, G. Radio cognitive : Rendre les radios logicielles plus personnelles. IEEE Pers. Commun. 1999, 6, 13-18. [Référence croisée]
22. Haykin, S. Radio cognitive : communications sans fil assistées par le cerveau. IEEE J. Sel. Zones Commun. 2005, 23, 201-220. [Référence croisée]
23. Liang, YC; Chen, KC ; Li, GY; Mahonen, P. Réseaux radio cognitifs et communications : un aperçu. IEEETrans. Véh. Technologie. 2011, 60, 3386-3407. [Référence croisée]
24. Wang, B. ; Liu, KR Progrès dans les réseaux radio cognitifs : une enquête. IEEE J. Sel. Haut. Processus de signal. 2011, 5, 5-23. [Référence croisée]
25. Talebi, SP Panne du service primaire et performances du service secondaire dans les réseaux radio cognitifs. Fil. Commun. Foule. Calculer. 2015, 15, 1982-1990. [Référence croisée]
26. Zhao, Q. ; Sadler, BM Une enquête sur l'accès dynamique au spectre. Processus de signal IEEE. Mag. 2007, 24, 79-89. [Référence croisée]
27. Ji, Z. ; Liu, KR Partage dynamique du spectre : un aperçu de la théorie des jeux. Communique IEEE. Mag. 2007, 45, 88-94. [Référence croisée]
28. Bhattarai, S. ; Parc, JMJ ; Gao, B. ; Bian, K. ; Lehr, W. Un aperçu du partage dynamique du spectre : initiatives en cours, défis, et une feuille de route pour la recherche future. IEEETrans. Cogn. Commun. Réseau. 2016, 2, 110-128. [Référence croisée]
29. Zeng, Y. ; Liang, YC ; Zhang, R. Détection d'énergie combinée aveuglément pour la détection du spectre dans la radio cognitive. Signal IEEE Processus. Lett. 2008, 15, 649-652. [Référence croisée]
30. Ling, X. ; Wu, B. ; Wen, H. ; Ho, PH ; Bao, Z. ; Pan, L. Contrôle de seuil adaptatif pour la détection du spectre basée sur la détection d'énergie dans Radios cognitives. Fil IEEE. Commun. Lett. 2012, 1, 448-451. [Référence croisée]
31. Ranjan, A. ; Anurag ; Singh, B. Conception et analyse de la détection spectrale en radio cognitive basée sur la détection d'énergie. Dans Actes de la Conférence internationale sur le traitement du signal et de l'information, Nanded, Inde, 6-8 octobre 2016 ; p. 1 à 5.
32. Arjoune, Y. ; Mrabet, ZE; Ghazi, IL ; Tamtaoui, A. Détection spectrale : Technique améliorée de détection d'énergie basée sur la mesure du bruit. Dans les actes de l'atelier et de la conférence annuels de l'IEEE sur l'informatique et la communication, Las Vegas, NV, États-Unis, 8-10 janvier 2018 ; pp. 828-834.
33. Jiang, C. ; Li, Y. ; Bai, W. ; Yang, Y. ; Hu, J. Détection spectrale robuste basée sur un filtre adapté statistique dans un environnement d'incertitude de bruit . Dans Actes de la Conférence internationale sur les technologies de la communication, Chengdu, Chine, 9-11 novembre 2012 ; pages 1209 à 1213.
34. Zhang, X. ; Chaise.; Gao, F. Détection de spectre basée sur un filtre adapté et détection du niveau de puissance pour le réseau radio cognitif. Dans les actes de la conférence mondiale de l'IEEE sur le traitement du signal et de l'information, Atlanta, GA, États-Unis, 3-5 décembre 2014 ; pages 1267 à 1270.
35. Lv, Q. ; Gao, F. Détection de spectre basée sur un filtre adapté et reconnaissance du niveau de puissance avec plusieurs antennes. Dans Actes du sommet IEEE Chine et de la conférence internationale sur le traitement du signal et de l'information, Chengdu, Chine, 12-15 juillet 2015 ; pp. 305-309.
36. Ilyas, I. ; Paul, S. ; Rahman, A. ; Kundu, RK Évaluation comparative de la détection du spectre cognitif basée sur la détection cyclostationnaire. Dans les actes de la conférence annuelle de l'IEEE sur l'informatique omniprésente, l'électronique et les communications mobiles, New York, NY, États-Unis, 20-22 octobre 2016 ; p. 1 à 7.
37. Cohen, D. ; Eldar, YC Détection cyclostationnaire compressée pour Cognitive Radio. Dans les actes de la conférence internationale IEEE 2017 sur l'acoustique, la parole et le traitement du signal (ICASSP), Nouvelle-Orléans, LA, États-Unis, 5-9 mars 2017 ; pages 3509 à 3513.
38. Zeng, Y. ; Liang, YC Algorithmes de détection du spectre pour la radio cognitive basés sur des covariances statistiques. IEEETrans. Véh. Technologie. 2009, 58, 1804-1815. [Référence croisée]
39. Zeng, Y. ; Liang, YC Algorithmes de détection de spectre basés sur les valeurs propres pour la radio cognitive. IEEETrans. Commun. 2009, 57, 1784-1793. [Référence croisée]
40. Zayen, B. ; Hayar, A. ; Kansanen, K. Détection du spectre aveugle pour la radio cognitive basée sur l'estimation de la dimension spatiale du signal. Dans les actes de la conférence internationale IEEE 2009 sur les communications, Dresde, Allemagne, 14-18 juin 2009 ; p. 1 à 5.

Machine Translated by Google

41. Farhang-Boroujeny, B. Détection du spectre par banque de filtres pour les radios cognitives. IEEETrans. Processus de signal. 2008, 56, 1801-1811. [Référence croisée]
42. Quan, Z. ; Cui, S. ; Il a dit : AH ; Mauvaise détection conjointe multibande optimale HV pour la détection du spectre dans les réseaux radio cognitifs. IEEETrans. Processus de signal. 2009, 57, 1128-1140. [Référence croisée]
43. Raghu, moi; Chaudrée, SS ; Elias, E. Détection efficace du spectre pour la radio cognitive à l'aide de banques de filtres modulés en cosinus. Dans Actes de la conférence IEEE Région 10, Singapour, 22-25 novembre 2016 ; pages 2086 à 2089.
44. Kumar, A. ; Saha, S. ; Bhattacharya, R. Détection de bord améliorée basée sur la transformation d'ondelettes pour la détection de spectre à large bande dans Cognitive Radio. Dans Actes de la réunion scientifique radio USNC-URSI, Fajardo, Porto Rico, 26 juin-1er juillet 2016 ; p. 21-22.
45. Polo, YL; Wang, Y. ; Pandharipande, A. ; Leus, G. Détection de spectre compressif à large bande. Dans les actes de la conférence internationale de l'IEEE sur l'acoustique, la parole et le traitement du signal, Taipei, Taiwan, 19-24 avril 2009 ; pages 2337 à 2340.
46. Arjoune, Y. ; Kaabouch, N. Détection du spectre à large bande : une approche bayésienne de détection compressive. Capteurs 2018, 18, 1839. [Référence croisée] [Pub Med]
47. Li, Z. ; Chang, B. ; Wang, S. ; Liu, A. ; Zeng, F. ; Luo, G. Détection dynamique de spectre à large bande par compression basée sur la reconstruction de l'énergie des canaux dans l'Internet cognitif des objets. IEEETrans. Ind. Informer. 2018, 14, 2598-2607. [Référence croisée]
48. Yang, L. ; Croc, J. ; Duan, H. ; Li, H. Estimation rapide du spectre de puissance compressé : vers une solution pratique pour le large bande Détection du spectre. IEEETrans. Fil. Commun. 2020, 19, 520-532. [Référence croisée]
49. Ding, G. ; Wu, Q. ; Yao, YD; Wang, J. ; Chen, Y. Apprentissage basé sur le noyau pour le traitement statistique du signal dans les réseaux radio cognitifs : fondements théoriques, exemples d'applications et orientations futures. Processus de signal IEEE. Mag. 2013, 30, 126-136. [Référence croisée]
50. Thilina, KM; Choi, KW; Saquib, N. ; Hossain, E. Techniques d'apprentissage automatique pour la détection coopérative du spectre dans les domaines cognitifs Réseaux radio. IEEE J. Sel. Zones Commun. 2013, 31, 2209-2221. [Référence croisée]
51. Mikaeil, AM; Guo, B. ; Wang, Z. Apprentissage automatique et approche de fusion de données pour la détection coopérative du spectre. Dans Actes de la Conférence internationale sur l'informatique distribuée cybernétique et la découverte des connaissances, Shanghai, Chine, 13-15 octobre 2014 ; pp. 429-434.
52. Ghazizadeh, E. ; Nikpour, B. ; Moghadam, DA; Nezamabadi-pour, H. Une méthode de pondération basée sur PSO pour améliorer les techniques d'apprentissage automatique pour la détection de spectre coopérative dans les réseaux CR. Dans Actes de la conférence sur l'intelligence en essaim et le calcul évolutionnaire, Bam, Iran, 9-11 mars 2016 ; p. 113-118.
53. Khalfi, B. ; Zaid, A. ; Hamdaoui, B. Quand l'apprentissage automatique rencontre l'échantillonnage compressif pour la détection de spectre à large bande. Dans les actes de la Conférence internationale sur les communications sans fil et l'informatique mobile, Valence, Espagne, 26-30 juin 2017 ; pp. 1120-1125.
54. Lee, W. ; Kim, M. ; Cho, DH Deep Cooperative Sensing : détection de spectre coopérative basée sur des réseaux de neurones convolutifs. IEEETrans. Véh. Technologie. 2019, 68, 3005-3009. [Référence croisée]
55. Liu, C. ; Wang, J. ; Liu, X. ; Liang, YC Deep CM-CNN pour la détection du spectre dans la radio cognitive. IEEE J. Sel. Zones Commun. 2019, 37, 2306-2321. [Référence croisée]
56. Soni, B. ; Patel, Danemark ; López-Benítez, M. Schéma de détection du spectre basé sur la mémoire à long terme et à court terme pour l'utilisation de la radio cognitive Statistiques d'activité primaire. Accès IEEE 2020, 8, 97437-97451. [Référence croisée]
57. Peng, Q. ; Gilman, A. ; Vasconcelos, N. ; Cosman, PC ; Milstein, LB Détection profonde robuste grâce à l'apprentissage par transfert dans la radio cognitive. Fil IEEE. Commun. Lett. 2020, 9, 38-41. [Référence croisée]
58. Zhang, W. ; Wang, Y. ; Chen, X. ; Cai, Z. ; Tian, Z. Spectrum Transformer : un détecteur de spectre à large bande basé sur l'attention. IEEE Trans. Fil. Commun. 2024, 1-1. [Référence croisée]
59. Sarikhani, R. ; Keynia, F. La détection coopérative du spectre rencontre l'apprentissage automatique : approche d'apprentissage par renforcement profond. IEEE Commun. Lett. 2020, 24, 1459-1462. [Référence croisée]
60. Cai, P. ; Zhang, Y. ; Pan, C. Apprentissage par renforcement profond basé sur des graphiques de coordination pour la détection coopérative du spectre sous Décoloration corrélée. Fil IEEE. Commun. Lett. 2020, 9, 1778-1781. [Référence croisée]
61. Gao, A. ; Du, C. ; Ng, SX; Liang, W. Une détection coopérative du spectre avec une approche d'apprentissage par renforcement multi-agents dans les réseaux radio cognitifs. Communique IEEE. Lett. 2021, 25, 2604-2608. [Référence croisée]
62. Ngo, QT ; Jayawickrama, BA; Hé.; Dutkiewicz, E. Détection de spectre assistée par RIS multi-agents basée sur DRL dans le domaine cognitif Réseaux satellite-terrestre. Fil IEEE. Commun. Lett. 2023, 12, 2213-2217. [Référence croisée]
63. Ali, A. ; Hamouda, W. Avancées sur la détection du spectre pour les réseaux radio cognitifs : théorie et applications. Communique IEEE. Survivre. Tuteur. 2017, 19, 1277-1304. [Référence croisée]
64. Arjoune, Y. ; Kaabouch, N. Une enquête complète sur la détection du spectre dans les réseaux radio cognitifs : progrès récents, nouveaux Défis et orientations futures de la recherche. Capteurs 2019, 19, 126. [CrossRef]
65. Zhang, Y. ; Luo, Z. Une revue de la recherche sur la détection du spectre basée sur l'apprentissage profond. Électronique 2023, 12, 4514. [CrossRef]
66. Wijenayake, C. ; Madanayake, A. ; Kota, J. ; Bruton, L. Espaces blancs spectraux spatio-temporels dans la radio cognitive : théorie, algorithmes, et Circuits. IEEE J. Emerg. Sel. Haut. Système de circuits. 2013, 3, 640-653. [Référence croisée]
67. Madanayake, A. ; Wijenayake, C. ; Tran, N. ; Cooklev, T. ; Hum, S. ; Bruton, LT Détection de spectre directionnel à l'aide de filtres discrets spatio-temporels multi-D accordables. Dans Actes du Symposium international de l'IEEE sur un monde de réseaux sans fil, mobiles et multimédias, San Francisco, Californie, États-Unis, 25-28 juin 2012 ; p. 1 à 6.

Machine Translated by Google

68. Wijenayake, C. ; Madanayake, A. ; Bruton, LT; Devabhaktuni, V. Estimation DOA et localisation de sources dans les réseaux CR à l'aide de filtres de faisceau IIR 2D orientables. Dans Actes du Symposium international de l'IEEE sur les circuits et les systèmes, Pékin, Chine, 19-23 mai 2013 ; pp. 65-68.

69. Gunaratne, savoirs traditionnels ; Bruton, LT Filtres trapézoïdaux FIR 2D adaptatifs à coefficient complexe pour la formation de faisceaux à large bande dans les domaines cognitifs systèmes radio. Système de circuits. Processus de signal. 2011, 30, 587-608. [Référence croisée]
70. Wilcox, D. ; Tsakalaki, E. ; Kortun, A. ; Ratnarajah, T. ; Papadias, CB; Sellathurai, M. Sur l'utilisation de la radio cognitive dans le domaine spatial Réseaux d'antennes parasites à une seule radio. IEEE J. Sel. Zones Commun. 2013, 31, 571-580. [Référence croisée]
71. Qian, R. ; Sellathurai, M. ; Ratnarajah, T. Détection de spectre directionnel pour la radio cognitive utilisant des réseaux ESPAR avec une seule chaîne RF. Dans Actes de la Conférence européenne 2014 sur les réseaux et les communications (EuCNC), Bologne, Italie, 23-26 juin 2014 ; p. 1 à 5.

72. Yazdani, H. ; Vosoughi, A. Sur les systèmes radio cognitifs avec antennes directionnelles et détection de spectre imparfaite. Dans les actes de la conférence internationale IEEE 2017 sur l'acoustique, la parole et le traitement du signal (ICASSP), Nouvelle-Orléans, LA, États-Unis, 5-9 mars 2017 ; pp. 3589-3593.

73. Pulipati, S. ; Ariyarathna, V. ; Edussooriya, CUS ; Wijenayake, C. ; Wang, X. ; Madanayake, A. Détection directionnelle multifaisceau basée sur FPGA en temps réel de sources RF ISM 2,4 GHz. Dans Actes de la Moratuwa Engineering Research Conference 2019 (MERCon), Moratuwa, Sri Lanka, 3-5 juillet 2019 ; pp. 129-134.

74. Richard Thompson, A. ; Gergely, TE; Vanden Bout, PA Interférence et radioastronomie. Phys. Aujourd'hui 1991, 44, 41-49. [Référence croisée]
75. Observatoire des objectifs scientifiques - Réseau de kilomètres carrés. Disponible en ligne : https://www.skao.int/en/explore/science-goals (accédé le 27 mai 2024).
76. Charpentier, J. ; Brogan, C. ; Iono, D. ; Mroczkowski, T. La mise à niveau de la sensibilité à large bande ALMA2030. arXiv 2022, arXiv :2211.00195.
77. La science avec un très grand réseau de nouvelle génération. Disponible en ligne : https://ngvla.nrao.edu/page/science (consulté le 27 mai 2024).
78. Sihlangu, I. ; Oozeer, N. ; Bassett, BA Nature et évolution des interférences radio fréquence UHF et bande L au radiotélescope MeerKAT. Dans Actes de l'atelier RFI 2022, Centre européen pour les prévisions météorologiques à moyen terme (ECMRWF), en ligne, 14-18 février 2022.

79. Galt, J. Contamination par les satellites. Nature 1990, 345, 483-483. [Référence croisée]
80. De Pree, CG; Anderson, CR ; Zheleva, M. 2023. L'astronomie est menacée par les interférences radio des satellites : voici ce qui peut être fait à ce sujet. Disponible en ligne : https://www.weforum.org/agenda/2023/03/astronomy-radio-interference-satellites-technology/ (consulté le 27 mai 2024).

81. Bandes de fréquences privilégiées pour les mesures radioastronomiques inférieures à 1 THz. Recommandation UIT-R RA.314-11, Union internationale des télécommunications — Secteur des radiocommunications. 2023. Disponible en ligne : https://www.itu.int/dms_pubrec/itu-r/rec/ ra/R-REC-RA.314-11-202312-I!!PDF-E.pdf (consulté le 27 mai 2024).

82. Critères de protection utilisés pour les mesures radioastronomiques. Recommandation UIT-R RA.769-2, Union internationale des télécommunications — Secteur des radiocommunications. 2003. Disponible en ligne : https://www.itu.int/dms_pubrec/itu-r/rec/ra/R-REC-RA. 769-2-200305-I!!PDF-E.pdf (consulté le 27 mai 2024).
83. Fridman, Pennsylvanie ; Baan, WA Méthodes d'atténuation des RFI en radioastronomie. Astrone. Astrophysique. 2001, 378, 327-344. [Référence croisée]
84. Zhang, L. ; Jin, R. ; Zhang, Q. ; Wang, R. ; Zhang, H. ; Wen, Z. Méthode de fusion de détection, de localisation et de suppression des RFI en combinant des radiomètres à synthèse d'ouverture unidimensionnels et bidimensionnels. Remote Sens.2024 , 16, 667. [CrossRef]
85. Proakis, JG; Manolakis, DG Traitement du signal numérique ; Pearson Prentice Hall : Upper Saddle River, NJ, États-Unis, 2007.
86. Oppenheim, AV; Schäfer, RW; Buck, J. Traitement du signal à temps discret ; Éditions internationales Prentice Hall ; Prentice Hall : Upper Saddle River, New Jersey, États-Unis, 1999.
87. Baghaie, R. ; Dimitrov, V. Algorithme DHT basé sur le codage d'entiers algébriques. Électron. Lett. 1999, 35, 1303-1305. [Référence croisée]
88. Britanak, V. ; Yip, P. ; Rao, KR Transformations discrètes en cosinus et sinus ; Presse académique : Cambridge, MA, États-Unis, 2007.
89. Feig, E. ; Winograd, S. Sur la complexité multiplicative des transformations en cosinus discrètes. IEEETrans. Inf. Théorie 1992, 38, 1387-1391. [Référence croisée]
90. Malvar, HS; Hallapuro, A. ; Karczewicz, M. ; Kerofsky, L. Transformation et quantification de faible complexité en H.264/AVC. IEEETrans . Système de circuits. Technologie vidéo. 2003, 13, 598-603. [Référence croisée]
91. Winograd, S. Complexité arithmétique des calculs ; Série de conférences régionales CBMS-NSF en mathématiques appliquées ; SIAM: Philadelphie, Pennsylvanie, États-Unis, 1980 ; Tome 33.
92. Winograd, S. Sur le calcul de la transformée de Fourier discrète. Mathématiques. Calculer. 1978, 32, 175-199. [Référence croisée]
93. Heideman, MT Complexité multiplicative, convolution et DFT ; Springer : Heidelberg, Allemagne, 1988.
94. Oppenheim, AV; Verghese, Signaux, systèmes et inférence GC ; Série de traitement du signal Prentice-Hall ; Pearson : Selle supérieure River, New Jersey, États-Unis, 2016.
95. Cintra, RJ Une méthode d'approximation entière pour les transformations sinusoïdales discrètes. Système de circuits. Processus de signal. 2011, 30, 1481-1501. [Référence croisée]
96. Suárez Villagrán, DM Approximations de transformation de Fourier discrète avec applications en détection et estimation. Master Thèse, Universidade Federal de Pernambuco, Recife, Brésil, 2015.

Machine Translated by Google

97. Madanayake, A. ; Ariyarathna, V. ; Madishetty, S. ; Pulipati, S. ; Cintra, RJ; Coelho, D. ; Oliveira, R. ; Bayer, FM; Belostotski, L. ; Mandal, S. ; et coll. Vers un réseau numérique à 1 024 faisceaux à faible SWaP : un sous-système à 32 faisceaux à 5,8 GHz. IEEETrans. Antennes Propag. 2020, 68, 900-912. [Référence croisée]

98. Madanayake, A. ; Cintra, RJ; Akram, N. ; Ariyarathna, V. ; Mandal, S. ; Coutinho, Virginie ; Bayer, FM; Coelho, D. ; Rappaport, TS DFT approximatifs Fast Radix-32 pour la formation de faisceaux RF numérique à 1024 faisceaux. Accès IEEE 2020, 8, 96613-96627. [Référence croisée]

99. Blahut, RE Algorithmes rapides pour le traitement du signal ; Cambridge University Press : Cambridge, Royaume-Uni, 2010. [CrossRef]

100. La Collaboration pour le traitement du signal astronomique et la recherche électronique. 2024. https://casper.berkeley.edu/ (consulté sur 23 mai 2024).

101. Le flux d'outils CASPER. GitHub. 2024. https://github.com/casper-astro/mlib_devel/ (consulté le 24 mai 2024).

102. Stutzman, WL; Thiele, GA Théorie et conception des antennes ; Wiley : Hoboken, New Jersey, États-Unis, 2012.

103. Kulasekera, S. ; Madanayake, A. ; Suarez, D. ; Cintra, RJ; Bayer, ouvertures du récepteur FM multifaisceau utilisant une DFT approximative à 8 points sans multiplicateur. Dans Actes de la conférence IEEE Radar 2015 (RadarCon), Arlington, VA, États-Unis, 10-15 mai 2015 ; pages 1244 à 1249. [Référence croisée]

104. Zhao, H. ; Madanayake, A. ; Cintra, RJ; Mandal, S. Formeur multifaisceau DFT approximatif à 8 points en mode courant analogique avec 4,7 Capacité du canal Gbit/s. Accès IEEE 2023, 11, 53716-53735. [Référence croisée]

105. Kulasekera, S. ; Madanayake, A. ; Wijenayake, C. ; Bayer, FM; Suarez, D. ; Cintra, RJ Formeurs de faisceaux numériques à ouverture RF multifaisceaux 8 × 8 utilisant des approximations FFT 2D sans multiplicateur. Dans Actes de la Moratuwa Engineering Research Conference 2015 (MERCon), Moratuwa, Sri Lanka, 7 et 8 avril 2015 ; pp. 260-264. [Référence croisée]

106. Ariyarathna, V. ; Madanayake, A. ; Tang, X. ; Coelho, D. ; Cintra, RJ; Belostotski, L. ; Mandal, S. ; Rappaport, algorithmes, architectures et circuits CMOS à FFT approximative analogique TS à 8/16 faisceaux pour les émetteurs-récepteurs MIMO à formation de faisceaux 5G. IEEE J. Émerger. Sel. Haut. Système de circuits. 2018, 8, 466-479. [Référence croisée]

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Projection complète de sous-couverture pour les automates de couverture par étapes †

Antonio Al Serhali* et Joachim Niehren*

Centre Inria, Université de Lille, 59000 Lille, France *
Correspondance : antonio.al-serhali@inria.fr (AAS) ; joachim.niehren@inria.fr (JN)
† Cet article est une version étendue de notre article publié au International Symposium on Fundamentals of Computation Theory, Trèves, Allemagne, du 18 au 21 septembre 2023.

Résumé : Nous démontrons comment évaluer les automates de couverture pas à pas (SHA) avec projection de sous-couverture tout en projetant complètement les sous-couvertures non pertinentes. Puisque cela nécessite de transmettre des informations d'état fini de haut en bas, nous introduisons la notion d'automates de couverture pas à pas descendants. Nous les utilisons pour définir des évaluateurs en mémoire et en streaming avec une projection complète de sous-couverture pour les SHA. Nous adaptons ensuite les évaluateurs afin qu'ils puissent décider de l'adhésion le plus tôt possible. Nous appliquons nos algorithmes au problème de répondre aux requêtes XPath régulières sur les flux XML . Nos expériences montrent que la projection complète de sous-couvertures de SHA peut en effet accélérer les premières réponses aux requêtes sur les flux XML, de sorte qu'elle devienne compétitive avec les meilleurs outils de streaming existants pour les requêtes XPath

Mots-clés : automates ; algorithme de projection ; algorithme de diffusion en continu ; XML

1. Introduction

Les haies sont des séquences de lettres de l'alphabet et des arbres h où h est à nouveau une haie. Les haies sont abstraites des détails syntaxiques des documents XML ou JSON tout en étant capables de les représenter. La linéarisation d'une haie est un mot imbriqué pouvant être écrit dans un fichier. Dans cet article, nous étudions le problème de la correspondance de modèles de haie, c'est-à-dire si une haie donnée h appartient à un langage de haie régulier donné L. Les langages de haie réguliers peuvent être définis notamment dans XPath ou JSONPath.

La projection est nécessaire à l'efficacité de nombreux algorithmes sur des mots, des arbres, des haies ou des mots imbriqués. Intuitivement, un algorithme projette s'il ne visite qu'un fragment de la structure d'entrée : dans le meilleur des cas, la partie pertinente pour le problème considéré.
La pertinence de la projection pour le traitement XML a déjà été remarquée par [1–4]. L'évaluateur en mémoire de Saxon , par exemple, projette des documents XML d'entrée relatifs à un programme XSLT , qui contient une collection de requêtes XPath auxquelles il faut répondre simultanément [5]. L' outil QuiXPath [4] évalue les requêtes XPath en mode streaming avec sous-arbre et projection descendante. La projection lors de l'évaluation des requêtes JSONPath sur des documents JSON en mode streaming est appelée fast-forwarding [6].

La projection de l'évaluation en mémoire suppose que le graphique complet de la couverture d'entrée soit construit au préalable. Néanmoins, la projection peut encore faire gagner du temps si l'on doit faire correspondre plusieurs modèles sur la même couverture d'entrée ou si le graphique a de toute façon été construit pour des raisons différentes. En mode streaming, la situation est similaire : l'ensemble de la couverture d'entrée sur le flux doit être analysé, mais les évaluateurs doivent inspecter uniquement les nœuds qui ne sont pas projetés. Étant donné que l'analyse pure est de deux ou trois ordres de grandeur plus rapide que l'évaluation de modèles, le gain de temps de projection peut être considérable.

Le point de départ de cet article est la notion de non-pertinence des positions de couverture h par rapport à un langage de couverture L que nous introduisons. Ce sont des positions de h où, pour toutes les continuations possibles, l'insertion d'une sous-couverture n'affecte pas l'appartenance à L. Nous proposons un algorithme pour la mise en correspondance de modèles de haie avec une projection complète de sous-couverture. Notre algorithme décide de la correspondance des modèles de couverture tout en projetant complètement les sous-couvertures situées à des positions de sous-couverture non pertinentes. Autrement dit, il décid

check for updates

Citation : Al Serhali, A. ; Niehren, J. Projection complète de sous-couverture pour les automates de couverture par étapes. Algorithmes 2024, 17, 339. https://doi.org/10.3390/a17080339

Rédacteur académique : Henning Fernau et Klaus Jansen

Reçu : 29 janvier 2024
Révisé : 23 juillet 2024
Accepté : 23 juillet 2024
Publié : 2 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

Machine Translated by Google

si une haie h correspond à un modèle L sans visiter une sous-couverture de h située à une position non pertinente par rapport à L.

Les langages de haie réguliers peuvent être représentés par des expressions régulières imbriquées, qui peuvent être dérivées de requêtes XPath régulières [7] ou par une sorte d'automates de haie, qui peuvent être compilés à partir d'expressions régulières imbriquées. Nous utilisons des automates de couverture pas à pas (SHA) [8], une variante récente des automates de couverture standard, qui, eux, remontent aux années 1960 [9,10]. Les SHA mélangent le traitement ascendant des automates d'arbres standard avec le traitement de gauche à droite des automates de mots finis (DFA). Ils ne prennent pas en charge le traitement descendant et n'ont pas de pile explicite, contrairement aux NWA d'automates de mots imbriqués [11-15]. Les SHA ont une bonne notion du déterminisme ascendant et de gauche à droite, évitant la notion problématique de déterminisme pour les automates de couverture standards et la notion trop coûteuse de déterminisme pour les NWA. De plus, des compilateurs allant des expressions régulières imbriquées aux SHA sont disponibles [8], et des algorithmes de détermination [16] produisent de petits SHA pour toutes les requêtes XPath régulières dans les benchmarks pratiques existants [7,17].

La motivation du présent travail est d'ajouter la projection de sous-couverture à un algorithme de streaming récent pour les SHA déterministes qui peuvent évaluer les requêtes XPath régulières le plus tôt possible . La plupart des approches alternatives de streaming évitent les réponses les plus précoces aux requêtes en considérant sans délai les sous-langages des requêtes régulières de sorte que la sélection du nœud dépend uniquement du passé du nœud mais pas du futur [18, 19] ou en admettant une sélection tardive [20].
En particulier, il a été récemment démontré que la première requête répondant aux requêtes monadiques régulières définies par les SHA déterministes [21] avait une complexité dans le pire des cas inférieure à celle des NWA déterministes [22]. Ainsi, la première réponse aux requêtes régulières est devenue réalisable dans la pratique pour la première fois. D'un autre côté, elle est encore plus lente expérimentalement que les meilleures approches existantes, non les plus anciennes, pour les requêtes monadiques régulières générales sur les flux XML (voir [23] pour une comparaison) puisque cette dernière prend en
La manière dont les projections pourraient être faites pour les SHA n'a pas été étudiée avant la version conférence du présent article [24].

Prenons l'exemple du filtre XPath [self::list/child::item] qui est satisfait par un document XML si sa racine est un élément de liste qui a un élément enfant. Avec l'encodage du document XML au fur et à mesure que les haies sont choisies dans le présent article, les haies satisfaisantes ont la forme liste · h1 · item · h2 · h3 · h4, et pour certaines haies, h1, h2, h3, h4. Lors de l'évaluation de ce filtre sur une haie, il suffit d'inspecter le premier sous-arbre pour avoir la liste d'étiquettes racine , puis tous ses enfants jusqu'à ce qu'un élément avec un élément d'étiquette racine soit trouvé. Les sous-haies de ces enfants, c'est-à-dire la haie h3 et les sous-haies des arbres de niveau supérieur en h1 et h3, peuvent être projetées, ainsi que la haie h4 : elles n'ont pas besoin d'être inspectées pour évaluer ce filtre. Cependant, il faut mémoriser si le niveau du nœud actuel est 0
Ce niveau d'information peut être naturellement mis à jour de manière descendante. Toutefois, les évaluateurs des SHA opèrent exclusivement de bas en haut et de gauche à droite. Par conséquent, les évaluateurs projetés pour les SHA doivent être basés sur des machines plus générales. Nous proposons des automates de couverture descendante par étapes (SHA↓ s), une variante des SHA qui prennent en charge le traitement descendant. Ce sont essentiellement les automates forestiers pushdown de Neumann et Seidl [25], sauf qu'ils sont appliqués à des haies non étiquetées au lieu de forêts étiquetées. De plus, les NWA sont connues pour fonctionner fondamentalement de la même manière sur les mots imbriqués [26] tout en permettant des refoulements visibles légèrement plus généraux. Nous distinguons ensuite les états de projection de sous-couverture pour les SHA↓ et montrons comment les utiliser pour évaluer les SHA avec projection de sous-couverture à la fois en mémoire et en mode streaming.

Comme première contribution, nous présentons le compilateur safe-no-change des SHA aux SHA↓ qui peut distinguer les états de projection de sous-couverture appropriés. L'idée est que le SHA↓ sûr et sans changement peut distinguer les contextes dans lesquels les états du SHA ne changeront pas en toute sécurité.
Par exemple, le filtre XPath [self::list/child::item] peut être défini par un SHA déterministe fonctionnant de bas en haut que notre compilateur mappe au SHA↓ fonctionnant de haut en bas. Les informations contextuelles rendues explicites par le traitement descendant concernent les niveaux des États.
Cela nous permet de distinguer les états de projection à partir du niveau 2, et dans lesquels les sous-couvertures peuvent être ignorées. Nous prouvons la solidité de notre compilateur en nous basant sur un invariant non trivial qui

Machine Translated by Google

nous établissons (notons que la preuve a nécessité une adaptation du compilateur original de FCT [24]). Nous présentons également un contre-exemple montrant que la projection sans changement n'est pas complète pour la projection de sous-couverture. Cela montre qu'une sous-couverture peut ne pas être pertinente même si son état est encore en train de changer.

Comme deuxième contribution, nous proposons l'algorithme de projection de congruence. Il compile à nouveau les SHA en SHA↓ mais s'appuie sur les relations de congruence des états des automates. Nous prouvons ensuite que la projection de congruence produit non seulement un algorithme solide, mais que cet algorithme est également complet pour la projection de sous-couvertures, c'est-à-dire que toutes les sous-couvertures fortement non pertinentes (voir Définition 9) sont évaluées dans un état de boucle. La projection de congruence commence au niveau supérieur des haies avec la congruence Myhill – Nerode bien connue de la minimisation des automates.

Pour les langues avec des mots L, cette congruence identifie les préfixes v qui ont le même (L) langage résiduel $v^{-1} = \{w \mid v \cdot w \in L\}$. Un préfixe v est alors un suffixe non pertinent si son (L) langage résiduel v sur $^{-1}$ est soit universel, soit vide. Dans le cas des haies, qui s'étendent les mots en ajoutant une imbrication hiérarchique via un constructeur d'arbre, la congruence doit être adaptée lors de la descente dans les sous-arbres.

Nous montrons que les deux compilateurs peuvent augmenter la taille des automates de manière exponentielle puisqu'ils doivent être capables de convertir des automates déterministes ascendants sur des arbres monadiques en automates déterministes descendants. Cela revient à inverser des automates déterministes sur des mots, ce qui est bien connu pour provoquer une explosion exponentielle dans le pire des cas (par exemple, pour la famille de langues $(a+b)^*a(a+b)^n$ où $n \in \mathbb{N}$, la gauche déterministe minimale l'automate déterministe minimal de droite à gauche a 2n+1 états tandis que l'automate déterministe minimal de droite à gauche a n + 1 états).

L'explosion exponentielle peut être évitée lorsqu'on s'intéresse à l'évaluation des haies par automate avec projection de sous-couverture, en ne construisant pas les SHA↓ complets de manière statique. Au lieu de cela, seule la partie requise des SHA↓ peut être construite à la volée lors de l'évaluation d'une couverture.

Notre troisième contribution est le raffinement des deux compilateurs précédents afin qu'ils puissent également distinguer les états sûrs pour une sélection au plus tôt. Pour cela, nous combinons ces compilateurs avec une variante du premier testeur d'appartenance de [21] qui fonctionne en mémoire en compilant des SHA↓ au lieu de NWA. De plus, l'échec de l'adhésion est également détecté dès la première position. De cette façon, nous obtenons le premier testeur d'appartenance en mémoire pour les SHA déterministes .

Notre quatrième contribution est de montrer comment exécuter les évaluateurs en mémoire précédents en mode streaming tout en lisant la linéarisation de la haie dans un mot imbriqué en entrée. Ainsi, nous pouvons améliorer le tout premier testeur d'adhésion en streaming derrière l'outil AStream [21] en y ajoutant une projection de sous-couverture.

Notre cinquième et dernière contribution est une implémentation et une évaluation expérimentale du premier algorithme de réponse aux requêtes en streaming pour les dSHA en introduisant la projection de sous-couverture dans l'outil AStream [21]. Nous avons mis en œuvre à la fois une projection sans changement et une projection de congruence. Pour cela, nous avons élevé notre premier testeur d'appartenance avec projection de sous-couverture vers un premier algorithme de réponse aux requêtes monadiques. Pour l'expérimentation, nous avons commencé avec des SHA déterministes construits avec le compilateur de [8] pour les requêtes XPath régulières directes du benchmark XPathMark [27] et les requêtes XPath du
Il s'avère que la projection de congruence se projette beaucoup plus fortement que la projection sans changement pour au moins la moitié des requêtes de référence. Cela réduit considérablement le temps d'exécution de toutes les requêtes XPath classiques , car une grande partie des haies d'entrée peut être projetée . Dans notre benchmark, le pourcentage projeté varie de 75,7 % à 100 % du flux d'entrée. Pour les requêtes XPath qui contiennent uniquement des axes enfants, le premier algorithme de réponse aux requêtes d'AStream avec projection de congruence devient compétitif en efficacité avec le meilleur outil de streaming existant appelé QuiXPath [23], qui n'est pas toujours le plus précoce pour certaines requêtes. Notre implémentation actuelle de la première projection de congruence dans AStream est 1,3 à 2,9 fois plus lente que QuiXPath sur les requêtes de référence. L'amélioration est moindre pour les requêtes XPath avec l'axe descendant, où la projection de sous-couverture est moindre.

Machine Translated by Google

possible. Au lieu de cela, une sorte de projection descendante serait nécessaire. Pourtant, même dans le pire des cas de notre benchmark, la première projection de congruence avec AStream n'est actuellement que d'un facteur 13,8 plus lente que QuiXPath.

Contour. Nous commençons par des préliminaires sur les haies et les mots imbriqués dans la section 2. Dans la section 3, nous rappelons le problème de la correspondance des modèles de haies et définissons formellement les sous-couvertures non pertinentes. Dans la section 4, nous rappelons la définition des SHA, introduisons les SHA↓ et discutons de leurs évaluateurs en mémoire. Dans la section 5, nous introduisons la notion d'états de projection de sous-couverture, définissons quand un SHA↓ est terminé pour une projection de sous-couverture et présentons un évaluateur en mémoire pour les SHA↓ avec projection de sous-couverture. Dans la section 6, nous introduisons la projection sans changement, déclarons sa solidité et montrons son caractère incomplet. Dans la section 7, nous introduisons la projection de congruence et prouvons qu'elle est solide et complète pour la projection de sous-couverture. Les premiers évaluateurs en mémoire pour les SHA↓ avec projection de sous-couverture sont présentés dans la section 8. Les variantes de streaming de nos évaluateurs en mémoire sont systématiquement dérivées dans la section 9. La section 10 explique comment faire passer nos algorithmes des tests d'appartenance aux requêtes monadiques. La section 11 discute de nos expériences pratiques. La section 12 discute d'autres travaux connexes sur les notions d'automates liées aux SHA et aux SHA↓ . L'Annexe A présente la reformulation du premier algorithme de réponse aux requêtes de [21] basé sur les SHA↓ et avec des restrictions de schéma. L'annexe B fournit une preuve sonore pour la projection sans danger et sans changement .

Commentaires sur les publications. Cette version originale de l'article de la revue a été publiée lors de la conférence FCT [24]. Par rapport à la publication précédente, les contributions suivantes sont nouveaux :

- La définition des préfixes non pertinents des mots imbriqués et la définition des exhaustivité pour la projection de sous-couverture (Section 3).
- L'ajout de restrictions de schéma tout au long de l'article et la notion de schéma exhaustivité (section 4.3).
- La notion de complétude pour la projection en sous-couverture (Section 5.2).
- La preuve de solidité de l'algorithme de projection sans changement dans la section 6.2, déplacée à l'Annexe B.
- L'algorithme de projection de congruence et la preuve de sa solidité et de son exhaustivité pour la projection de sous-couverture (Section 7).
- Une méthode systématique pour ajouter une projection de sous-couverture à un premier SHA↓ (Section 8).
- Une méthode systématique pour réduire l'exactitude des évaluateurs d'automates en streaming aux évaluateurs en mémoire correspondants (Section 9).
- Une discussion sur la façon de traiter les requêtes monadiques tout en exploitant la disponibilité de restrictions de schéma (Section 10).
- Une implémentation de projection de congruence avec des résultats expérimentaux pour l'évaluation XPath sur des flux XML (Section 11).
- Une discussion plus longue sur les travaux connexes (Section 12).
- Dans l'Annexe A, nous montrons également comment obtenir les premiers dSHA↓ à partir des dSHA en adaptant le compilateur de [21] mappant les dSHA aux premiers dNWA.

2. Préliminaires

Soient A et B des ensembles. Si A est un sous-ensemble de B, alors le complément de A dans B est noté $\overline{A} = B \setminus A$ tout en gardant l'ensemble par rapport auquel le complément est pris implicite. Le domaine de la relation binaire $r \subseteq A \times B$ est $dom(r) = \{a \in A \mid \exists b \in B.\ (a, b) \in r\}$. Une fonction partielle $f : A \to B$ est une relation binaire $f \subseteq A \times B$ fonctionnelle. Une fonction totale $f : A \to B$ est une fonction partielle $f : A \to B$ avec $dom(f) = A$.

Mots. Soit $\mathbb{N}$ l'ensemble des nombres naturels, dont 0. Soit l'alphabet Σ un ensemble. L' ensemble des mots sur Σ est Σ^* Un mot est $(a_1, \ldots, a_n) = \Sigma^n$ où $n \in \mathbb{N}$ s'écrit comme $a_1 \ldots a_n$. On note le mot vide de longueur 0 par $\varepsilon \in \Sigma^0$, et par $v_1 \cdot v_2 \in \Sigma^*$ le concaténation de deux mots $v_1, v_2 \in \Sigma^*$.

Machine Translated by Google

Si v = u · v ' · w est un mot, alors nous appelons u un préfixe de v et v' un facteur de v et w un suffixe de v. Étant donné tout sous-ensemble *, nous désignons l'ensemble des préfixes de mots dans L par prefs(L) et L Σ l'ensemble des suffixes de mots dans L par suffs(L).

Pour tout sous-alphabet Σ ' Σ, la projection d'un mot v Σ) * à Σ ' est le mot projΣ′(v) dans 'obtenu à partir de v en supprimant toutes les lettres de Σ \ Σ '.

(Σ Haies. Les haies sont des séquences de lettres et d'arbres h avec une haie h. Plus formellement, une haie h HΣ a la syntaxe abstraite suivante :

h, h' HΣ ::= ε | un | h | h · h ' où a Σ

Nous supposons ε · h = h · ε = h et (h · h1) · h2 = h · (h1 · h2). Par conséquent, nous considérons n'importe dans Σ* comme haie dans HΣ, c'est-à-dire Σ* quel mot aab = a · a · b HΣ. Toute couverture peut être convertie ou stockée sous forme de graphique. Pour la signature Σ = {liste, item, A, . . . , Z}, le graphique d'un exemple de couverture dans HΣ est présenté dans la figure 1. Il code la séquence de documents XML dans la figure 2 :

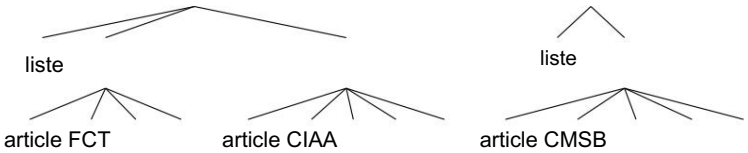


Figure 1. Le graphique de la haie list · item · F · C · T · item · C · I · A · A · list · item · C · M · S · B est une séquence d'arbres.

```
<list>
   <item>FCT</item>
   <item>CIAA</item> </
list>
<liste>
   <item>CMSB</item> </
list>
```

Figure 2. La séquence de documents XML codés par la haie de la figure 1.

Mots imbriqués. Un mot imbriqué sur Σ est un mot sur l'alphabet Σˆ = Σ { , } dans lequel toutes les parenthèses sont bien imbriquées. Intuitivement, cela signifie que pour toute parenthèse ouvrante, il existe une parenthèse fermante correspondante et vice versa.

L'ensemble des mots imbriqués sur Σ peut être défini comme les sous-ensembles de mots sur Σˆ qui sont une linéarisation de la haie, où la fonction de linéarisation nw : HΣ → (Σ { , }) est définie comme suit :

nw(ε) = ε, nw(h) = · nw(h) · , nw(a) = a, nw(h · h ') = nw(h) · nw(h ').

Ainsi, l'ensemble de tous les mots imbriqués sur Σ est nw(HΣ). Soit hdg l'inverse de la fonction injective nw restreinte à son image, donc le mappage des mots imbriqués aux haies avec hdg(nw(h)) = h pour tout h HΣ.

Préfixes de mots imbriqués. Les préfixes, suffixes et facteurs de mots imbriqués ne peuvent pas être eux-mêmes des mots imbriqués. Par exemple, la haie h = a · b · c a la linéarisation nw(h) = a bc .
Son préfixe a b n'est pas bien imbriqué car il a une parenthèse ouvrante pendante. Son suffixe c n'est pas non plus bien imbriqué puisqu'il comporte une parenthèse fermante pendante.

Tout algorithme qui traverse une haie h de haut en bas et de gauche à droite inspecte tous les préfixes de nw(h). Tout préfixe v de nw(h) distingue non seulement la position de la haie h mais précise également la partie de la haie qui se situe avant cette position dans la linéarisation nw(h). Un exemple est illustré graphiquement dans la figure 3. Cela est particulièrement vrai pour les algorithmes de streaming qui reçoivent le mot imbriqué nw(h) en entrée sur un flux.

Machine Translated by Google

et ne peut être inspecté qu'une seule fois de gauche à droite. Mais cela vaut également pour les algorithmes en mémoire qui reçoivent la couverture d'entrée h sous la forme d'une structure hiérarchique dont le graphe est stocké en mémoire et parcourt ensuite le graphe de haut en bas et de gauche à droite. Tout nœud sera alors visité deux fois : une fois lors de la lecture d'une parenthèse ouvrante et en descendant dans une sous-haie, et une autre fois en remontant vers la parenthèse fermante après avoir traité la sous-haie.

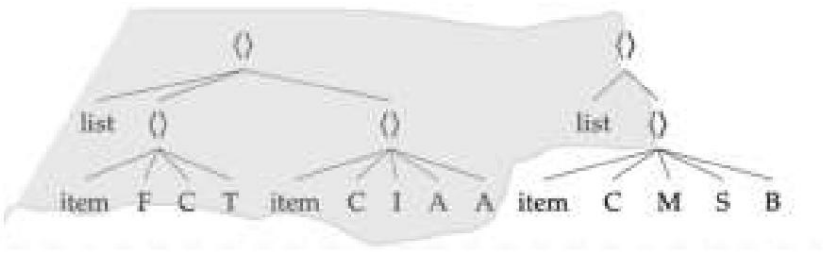


Figure 3. La partie de la haie distinguée par le préfixe de mot imbriqué list · item · F · C · T · item · C · I · A · A · list · .

3. Problème

Nous définissons formellement le problème du pattern matching pour les couvertures et la notion de non-pertinence des sous-couvertures, ce qui nous permet de caractériser les parties de la couverture d'entrée qui n'ont pas besoin d'être visitées lors du pattern matching.

3.1. Correspondance de modèles

de couverture Nous nous intéressons aux algorithmes qui font correspondre un modèle de couverture à une autre couverture. Nous commençons par des algorithmes qui testent si une telle correspondance existe. Ceci est parfois appelé filtrage ou, alternativement, réponse aux requêtes booléennes. Dans la section 10, nous considérons le problème plus général du retour de l'ensemble des correspondances. Ce problème est parfois appelé réponse monadique aux requêtes. Nos expériences de la section 11 appliquent la réponse aux requêtes monadiques pour les requêtes XPath classiques.

Nous partons de modèles de haie qui sont des expressions régulières imbriquées avec des lettres dans la signature Σ. Pour ceux-ci, nous fixons un ensemble V de variables de récursion et supposons qu'il est disjoint de Σ.

e, e′ nRegExpΣ ::= ε | un | z | e · e′ | e + e′ | e | e | μz.e où a Σ, z V

Chaque expression régulière imbriquée e nRegExpΣ est satisfaite par un sous-ensemble de couvertures dans e HΣ V. Nous rappelons la définition de cet ensemble de manière informelle. L'expression ε n'est satisfaite que par le mot vide, l'expression a Σ uniquement par la couverture a uniquement et l'expression z V uniquement par la couverture à une lettre z. Une expression e · e′ avec l'opérateur de concaténation et , respectivement. e,e avec l'opérateur étoile de Kleene et e nRegExpΣ est satisfait par des nRegExpΣ est satisfait par les concaténations de couvertures satisfaisant e et e′ répétitions Une expression nRegExpΣ, est satisfaisant e. L'étoile Kleene permet une récursion horizontale sur les haies. L' expression μ μz.e, où z V et satisfaite par toutes les couvertures sans variable de récursion x dans lesquelles z est remplacé à plusieurs reprises par une couverture satisfaisant e jusqu'à ce qu'elle disparaisse. Nous adoptons la restriction habituelle [28] selon laquelle dans toute expression régulière imbriquée μz.e, toutes les occurrences de z dans e sont gardées par un constructeur d'arbre . . De cette façon, l' opérateur μ fournit une récursion verticale appropriée. Notez également que, pour définir les sous-langages de HΣ, nous n'avons besoin que de telles expressions dans nRegExpΣ dans lesquelles toutes les occurrences des variables de récursion z sont liées dans la portée du classeur μz.

Il est bien connu que les expressions régulières imbriquées dans nRegExpΣ peuvent définir tous les sous-langages réguliers des haies dans HΣ, c'est-à-dire qu'elles ont la même expressivité que les automates de haie avec l'alphabet Σ. Des résultats analogues pour les expressions régulières d'arbres et les automates d'arbres sont du folklore [29].

Machine Translated by Google

Exemple 1 (Expressions régulières imbriquées de filtres XPath réguliers). Par exemple, nous pouvons définir des sous-ensembles d'encodages de couverture de documents XML avec deux types d'éléments dans Σ = {list, item} par des expressions dans nRegExpΣ . Pour cet article, nous utiliserons un encodage du document XML , qui produit des haies qui satisfont l'expression régulière suivante doc nRegExpΣ , où z V :

=def arbre où arbre doc =déf μz. (liste + élément) · z

Dans notre application, les encodages sont un peu plus riches afin de distinguer différents types de nœuds XML tels que les éléments, les attributs, les commentaires et les textes. Un filtre XPath tel que [self::list/child::item] peut alors être exprimé par l'expression régulière imbriquée :

filter =def liste · doc · élément · doc · doc · doc

Des compilateurs généraux allant des requêtes de filtre XPath descendantes régulières aux expressions régulières imbriquées ont été proposés dans [8]. Une implémentation a été développée dans le contexte de l' outil AStream [21].

Nous utiliserons les schémas S HΣ pour restreindre les entrées de notre problème de correspondance de modèles. Comme nous nous intéressons principalement aux schémas réguliers, nous ne supposons pas de régularité globale puisque la projection sans danger ne repose pas sur elle. Le problème de correspondance de modèles pour les expressions régulières imbriquées par rapport à un schéma S HΣ reçoit les entrées suivantes :

- une expression régulière imbriquée e nRegExpΣ et une haie h
- S.

Il renvoie ensuite la valeur de vérité du jugement h e. La couverture d'entrée h S peut être donnée soit par un graphe résidant en mémoire, soit par un flux contenant la linéarisation nw(h) de la couverture d'entrée, qui ne peut être lue qu'une seule fois de gauche à droite.

3.2. Sous-couvertures non

pertinentes Nous définissons le concept d'occurrences non pertinentes de sous-couvertures par rapport à un modèle de couverture donné. Ce que cela signifie dépend du type d'algorithme que nous utiliserons pour la correspondance de modèles. Nous utiliserons des algorithmes qui opèrent sur la couverture des entrées de haut en bas, de gauche à droite et de bas en haut. Cela vaut en particulier pour les algorithmes de streaming.

Intuitivement, lorsque l'algorithme de correspondance de modèles atteint un nœud de haut en bas ou de gauche à droite dont la sous-couverture ultérieure n'est pas pertinente, il peut alors sauter par-dessus sans inspecter sa structure. Ce que signifie sauter devrait être clair si la haie est donnée par un graphique stocké en mémoire. Notez que le graphique complet doit être construit au préalable, même si certaines parties peuvent s'avérer non pertinentes. Néanmoins, on peut gagner beaucoup de temps en sautant par-dessus des parties non pertinentes si le graphique a déjà été construit pour d'autres raisons ou si de nombreux problèmes de correspondance de modèles doivent être résolus sur la même haie.

Dans le cas des flux, la sous-couverture non pertinente doit encore être analysée, mais elle ne le sera pas autrement. Le plus souvent, l'analyse possible est effectuée par des automates, ce qui peut prendre deux fois plus de temps que nécessaire pour l'analyse. Par conséquent, ne pas avoir à effectuer d'analyse peut considérablement accélérer l'algorithme de streaming.

Définition 1. Soient S HΣ un schéma et L S un langage satisfaisant ce schéma. Nous définissons diff prefs(NΣ) : $\overset{L}{S}$ comme relation la moins symétrique sur les préfixes des mots imbriqués u, u ′

$$\text{différence}_{\overset{L}{S}}(u, u') \quad w \quad \text{suffs}(N\Sigma).\ u \cdot w \quad \text{nw}(L) \quad u' \cdot w \quad \text{nw}(S \setminus L).$$

Un préfixe de mot imbriqué u est appelé sous-couverture pertinente pour L avec le schéma S s'il existe des mots $v, v' \quad N\Sigma$ tel que diff $\overset{L}{S}(u \cdot v, u \cdot v')$ imbriqués). Sinon, le préfixe u est appelé sous-couverture, sans importance pour de schéma S.

Donc, $\text{différence}_{\overset{L}{S}}(u, u')$ déclare que toi et toi ′ ont des continuations qui se comportent différemment par rapport à L et S. De plus, un préfixe u n'est pas pertinent si l'appartenance au langage ne dépend pas de l'insertion de haie en u sous la condition que l'appartenance au schéma

Machine Translated by Google

garanti. Le cas sans restrictions de schéma est englobé par la définition ci-dessus en choisissant S = HΣ. Dans ce cas, le complément diff est la congruence Myhill-Nerode HΣ de L, qui est bien connue dans la théorie des langages formels grâce à la minimisation des DFA (voir n'importe quel manuel standard sur la théorie des langages formels ou Wikipédia (théorème de Myhill-Nerode : https : //en.wikipedia.org/wiki/Myhill-Nerode_theorem, consulté le 23 juillet 2024)).

Cette congruence sert de base à l'apprentissage de style Gold des langues régulières à partir d'exemples positifs et négatifs (voir [30] ou Wikipédia (apprentissage de style Gold dans la limite https://en.wikipedia.org/wiki/Language_identification_in_the_limit, consulté le 23 juillet 2024)).

Des restrictions de schéma sont nécessaires pour notre application aux modèles XPath standards.

Exemple 2 (filtres XPath réguliers). Considérons le schéma S = doc pour les couvertures représentant les documents XML avec la signature Σ = {list, item}, et le langage de couverture standard L = filtre, où

filtre = liste · arbre · élément · arbre · arbre · arbre

est l'expression régulière imbriquée de l'exemple 1 pour le filtre XPath [self::list/child::item]. Rappelons que cette expression régulière imbriquée peut être appliquée aux haies représentant uniquement des documents XML . Le préfixe de mot imbriqué u = list · item n'est pas pertinent pour le langage L avec le schéma S. Notez que sa continuation list · item avec le suffixe w = appartient à L. Par conséquent, pour tout h HΣ, si la suite list · item · h appartient à S = doc, alors elle appartient aussi à L. Néanmoins, pour la haie h1 = , la suite list · item · h1 n'existe pas appartient à L puisqu'il ne satisfait pas au schéma S = doc.

Le préfixe item n'est pas non plus pertinent pour le langage L, même indépendamment du schéma. En effet, L ne contient aucune continuation de ce préfixe vers une haie. Le préfixe list n'est pas sans importance pour la langue L par rapport à S. Cela peut être vu avec le suffixe w = , puisque list n'appartient pas à L tandis que la continuation list · h avec h = item appartient à L et les deux continuations satisfont au schéma S = doc.

Cependant, les restrictions de schéma peuvent également avoir des conséquences surprenantes qui rendent le problème plus fastidieux.

Exemple 3 (Conséquences surprenantes des restrictions de schéma). Considérons la signature Σ = {a}, le modèle L = { a } et le schéma S = L { · a}. Le préfixe est en effet une sous-couverture sans rapport avec L et S. Pour voir cela, nous considérons tous les suffixes de fermeture possibles un par un. Le plus petit préfixe de fermeture est . Notez qu'une haie h L si et seulement si h = a. Ainsi, l'appartenance à L semble dépendre de la sous-couverture h. Cependant, notons également que h S si et seulement si h = a. Par conséquent, en supposant que h appartient au schéma S, il doit également appartenir à L. Ainsi, le modèle doit correspondre en supposant l'appartenance au schéma. Le suffixe de fermeture suivant est · a. Notez qu'une haie h · a S si et seulement si h = ε. Cependant, h · a L, donc le modèle ne correspondra à aucune sous-couverture h avec ce préfixe en supposant que la couverture d'entrée satisfait le schéma. La situation est similaire pour les suffixes plus grands, donc en aucun cas l'appartenance à la langue ne dépend de la sous-couverture au préfixe en supposant que la couverture complète satisfait le schéma S.

3.3. Propriétés de base

Nous montrons d'abord que les préfixes non pertinents pour les sous-couvertures restent non pertinents lorsqu'ils sont étendus par le mot imbriqué d'une haie. Cette propriété devrait être attendue pour toute notion raisonnable de non-pertinence de la sous-couverture.

Lemme 1. Pour tout préfixe de mot imbriqué u, langage L, S HΣ et mot imbriqué v NΣ, si le préfixe u n'est pas pertinent pour L avec le schéma S, alors le préfixe u · v l'est aussi.

Preuve. Soit u une sous-couverture non pertinente pour L avec le schéma S et v NΣ un mot imbriqué. Nous NΣ et imbriqués arbitrairement v ', v " un suffixe de mot imbriqué w suffs(NΣ) tels que corrigent des mots

Machine Translated by Google

tu · v · v ′ · w nw(S) et u · v · v · w nw(S). Puisque v · v, la sous- » ′ et v · v » sont aussi des mots imbriqués, couverture de u donne ce qui suit :

$$tu \cdot v \cdot v' \cdot w \in nw(L) \quad u \cdot v \cdot v'' \cdot w \in nw(L).$$

Par conséquent, le préfixe de mot imbriqué u · v n'est pas pertinent pour L avec le schéma S. □

Définition 2. On appelle une relation binaire D sur les préfixes de mots imbriqués une relation de différence sur les préfixes si pour tous les préfixes de mots imbriqués u, u′ ∈ prefs(NΣ) et mots imbriqués v ∈ NΣ :

$$(tu \cdot v, tu' \cdot v) \in D \iff (u, u') \in D.$$

Lemme 2. $\overline{\text{diff}}^{L}_{S}$ est une relation de différence.

Preuve. Il faut montrer pour tous les préfixes u, u′ ∈ prefs(NΣ) et mots imbriqués v ∈ NΣ qui sont des préfixes de mots imbriqués et v soit un mot imbriqué tel que $(u \cdot v, u' \cdot v) \in \overline{\text{diff}}^{L}_{S}$. Alors, il existe un suffixe imbriqué w tel que

$$u \cdot v \cdot w \in nw(L) \quad u' \cdot v \cdot w \in nw(S \setminus L).$$

Le suffixe w˜ = v · w satisfait alors u · w˜ ∈ nw(L) ∧ u′ · w˜ ∈ nw(S \ L). Par conséquent, $(u, u') \in \overline{\text{diff}}^{L}_{S}$ requis. □

Dans le cas des mots sans schéma, le complément $\overline{\text{diff}}^{L}_{\Sigma}$ est une relation d'équivalence connue sous le nom de congruence Myhill – Nerode. Cependant, dans le cas des schémas, le complément $\overline{\text{diff}}^{L}_{S}$ peut ne pas être transitif ; il se peut donc même qu'il ne s'agisse pas d'une relation d'équivalence. Cela pourrait surprendre à première vue.

Exemple 4. Soit Σ = {a, b}, L = {ε, b, aa} et S = L ∪ {aab}. Alors, (ε, a) ∈ $\overline{\text{diff}}^{L}_{S}$ car il n'y a pas de continuation qui étende à la fois ε et a dans le schéma. Pour la même raison, on a (a, aa) ∈ $\overline{\text{diff}}^{L}_{S}$. Cependant, la différence (ε, aa) ∈ $\overline{\text{diff}}^{L}_{S}$ puisque ε · b ∈ L et aa · b ∈ S \ L. Ainsi, (ε, aa) ∈ $\overline{\text{diff}}^{L}_{S}$, donc $\overline{\text{diff}}^{L}_{S}$ n'est pas transitif.

Il convient de noter pour tous les langages L que $\overline{\text{diff}}^{L}_{H\Sigma}$ est réflexif et transitif, il est donc une relation d'équivalence et donc une congruence. Pour les schémas généraux S, $\overline{\text{diff}}^{L}_{S}$ peut ne pas être transitif, comme le montre l'exemple 4, et peut donc ne pas être congruent.

Cela indique que les schémas peuvent rendre plus difficile la détermination de la non-pertinence des sous-couvertures. En effet, la manière naturelle dont on pourrait espérer éliminer les schémas basés sur des compléments ne fonctionne pas, comme le confirme le lemme suivant (notamment le contre- exemple de la preuve 1. ⇒ 2.) :

Lemme 3. Pour tout préfixe de mot imbriqué u ∈ suffs(NΣ) et L, S ∈ HΣ, considérons les deux propriétés suivantes :

1. u n'est pas pertinent pour L de schéma S.
2. u n'est pas pertinent pour $\overline{L}$ S de schéma HΣ.

Ensuite, la propriété 2. implique la propriété 1., mais pas toujours l'inverse.

Preuve. 1. ⇒ 2. : Ici, un contre-exemple est présenté. Soit L = {ε} et S = {ε, a}. Alors, le préfixe a n'est pas pertinent pour L de schéma S. Cependant, le préfixe a est pertinent pour $\overline{L}$ S = HΣ \ {a} de schéma HΣ, puisque pour les mots imbriqués v = ε et v = a nous avons a · v ∈ $\overline{L}$ S et a · v ′ ∉ $\overline{L}$ S.

Machine Translated by Google

2. : Ensuite, nous supposons la propriété 2. Considérons des mots arbitrairement imbriqués v, v · ' NΣ et 1.
un suffixe de mot imbriqué w suffit(NΣ) tel que u · v · w, u · v pour montrer que u · v · w ' w S. Il suffit
S et u · v ' · w S implique u · v · w L u · v ' · w L.
Cette implication est triviale si u · v · w S ou u · vu · v · w S et u · v ' · w S. Sinon, nous avons · w S.
' La propriété 2., comme supposé, implique alors u · v · w L
' tu · v · w L. Ainsi, la propriété 1. est vraie.
□

4. Automates de couverture

Nous rappelons la notion d'automates de couverture pas à pas (SHA), introduisons leur extension vers le bas (SHA↓), discutons de l'exhaustivité du schéma pour les deux types d'automates, définissons les états de projection de sous-couverture pour les SHA↓ et montrons comment utilisez-les pour évaluer les SHA↓ avec une projection de sous-couverture. Nous nous limitons à l'évaluation en mémoire dans cette section, mais discuterons de l'évaluation en continu dans la section 9. La relation avec plusieurs notions étroitement liées d' automates sur les haies, les forêts ou les mots imbriqués est discutée plus en détail dans la section 12.

4.1. Automates de couverture par étapes

Les SHA sont une notion récente d'automates pour haies [8] qui mélangent de manière naturelle des automates d'arbres ascendants et des automates de mots de gauche à droite. Ils étendent les automates arborescents pas à pas [31] en ce sens qu'ils opèrent sur des haies plutôt que sur des arbres non classés, c'est-à-dire sur des séquences de lettres et des arbres contenant des haies. Ils diffèrent des automates de mots imbriqués (NWA) [11, 15] en ce sens qu'ils traitent les haies directement plutôt que leurs linéarisations en mots imbriqués.

Définition 3. Un automate de couverture pas à pas (SHA) est un tuple A = (Σ, Q, Δ, I, F) où Σ et Q sont des ensembles finis, I, F Q et Δ = ((a Δ)a Σ, Δ, @Δ), où a Q × Q, Δ Q et @Δ (Q × Q) × Q. Un SHA est déterministe ou de manière équivalente un dSHA si I et Δ contiennent au plus un élément, et toutes les relations (a Δ)a Σ et @Δ sont des fonctions partielles.

L'ensemble des états Q englobe un sous-ensemble I de l'état initial, un sous-ensemble F des états finaux, un sous-ensemble Δ d'états initiaux d'arbres. Les règles de transition dans Δ ont trois formes : si (q, q alors nous avons ' et) a Δ, une règle de lettre que nous écrivons sous la forme q a−→ q) @Δ, alors nous avons par dans Δ. Enfin, si q Δ Q, alors nous q@p → q ' avons une règle d'application initiale de l'arbre que nous écrivons sous la forme
règle que nous désignons par −→ q dans Δ.

Dans la figure 4, nous illustrons le graphique d'un dSHA pour le filtre XPath [self::list/child::item].
Le graphique de SHA est obtenu et dessiné comme d'habitude pour un automate d'état sur des mots. Les états sont les nœuds du graphe ; dans l'exemple, Q = {0, 1, 2, 3, 4, 5}. Les règles de transition définissent les arêtes. Le bord d'une règle de lettre q a−→ q dans Δ est annoté par la lettre a Σ. Dans l'exemple, Σ = {liste, élément}. Toute règle applicable q@p → q annotée par l'état p Q. L'état initial 0 I est ' p dans Δ est dessiné comme une arête q −→ q ' indiqué par une flèche grise entrante, et l'état initial de l'arbre 0 Δ par une flèche grise entrante annotée par le symbole . Les états finaux , tels que 4 F, sont indiqués par un double cercle.

Machine Translated by Google

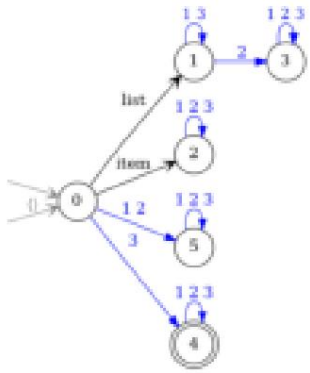


Figure 4. Un dSHA pour le filtre XPath [self::list/child::item] qui est complet pour le schéma doc.

4.1.1. Sémantique

Nous définissons ensuite la sémantique des SHA, c'est-à-dire les transitions sur les haies qu'ils définissent et le langage de haies qu'ils acceptent. Pour toute couverture h HΣ nous définissons les étapes de transition Q, a Σ, et h, h

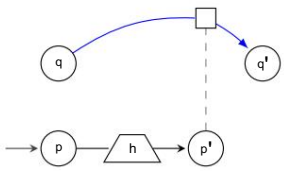
h
−→ p par rapport à Δ tel que pour tout q, qq ', p, p ' ' HΣ

$$\frac{q\ a{-\!\!\rightarrow}\ '\ \text{en } \Delta}{qq\ a{-\!\!\rightarrow}\ q'^{\ \text{par rapport à } \Delta}} \qquad \frac{h\ q -\!\!\rightarrow q'\ \text{par rapport à } \Delta\ q'\ h -\!\!\rightarrow q''^{\ \text{par rapport à } \Delta}}{{}^{h\,h\,-}_{-\!\!\rightarrow qq}\ ''^{\ \text{par rapport à } \Delta}}$$

$$\frac{q \quad Q\ q}{\varepsilon{-\!\!\rightarrow}\ q\ \text{par rapport à } \Delta} \qquad \frac{h -\!\!\rightarrow p\ \text{dans } \Delta\ p \quad '}{-\!\!\rightarrow p \qquad q@p\ '\ \rightarrow q\ '\ \text{en } \Delta}$$

$$-\!\!\rightarrow qq\ -\ '^{\ \text{par rapport à } \Delta}$$

Les transitions peuvent être utilisées pour évaluer les couvertures de manière non déterministe, de bas en haut et de gauche à droite, à l'aide de SHA. Les trois premières règles indiquent comment évaluer des séquences d'arbres et de lettres, comme un automate à états finis habituel, tout en supposant que les arbres ont déjà été évalués en états. La dernière règle indique comment évaluer un arbre h d'un état q à un état q
Cela peut être visualisé comme suit :



Pour tout état initial d'arbre p Δ, il faut évaluer la sous-couverture h de manière non déterministe pour obtenir, on revient ensuite à un état q
'. Pour chaque p ' un état p de manière non déterministe. ' tel que q@p ' → q ' en Δ

Une couverture est acceptée par A si elle permet une transition d'un état initial à un état final.
État. Le langage L(A) est l'ensemble de toutes les couvertures acceptées :

$$L(\text{UNE}) = \{h \quad H\Sigma \mid q -\!\!\rightarrow q \overset{h}{\ } ' \text{ par rapport à } \Delta,\ q \quad I,\ q' \quad F\ \}.$$

4.1.2. Fonctionne

Les exécutions peuvent représenter toute l'histoire d'un choix unique de l'évaluateur non déterministe sur une haie. Par exemple, l'unique exécution réussie du dSHA dans la figure 4 sur la haie h =' h = liste item ' est la haie 0 · 0 · liste · 1 · 0 · élément · 2 · 3 · 4 dont le calcul
avec sous-haie h
est illustré sur la figure 5. Au niveau supérieur, l'exécution justifie la transition −→ 4 par rapport à Δ de
h
0 l'état initial 0 à un état final 4. Lorsqu'elle est démarrée à l'état 0, la sous-couverture doit être évaluée
h ' en premier, donc le l'exécution au niveau supérieur doit être suspendue jusqu'à ce que

Machine Translated by Google

est fait. Le subrun sur h ' justifie finalement la transition 0 −→ 3 par rapport à Δ. A ce moment, l'exécution du niveau supérieur peut reprendre et continuer dans l'état 4 puisque 0@3 → 4 en Δ.

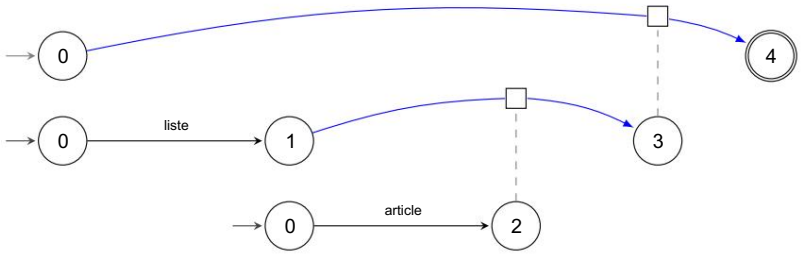


Figure 5. L'unique exécution réussie du dSHA dans la figure 4 sur la haie list · item est la haie 0 · 0 · liste · 1 · 0 · item · 2 · 3 · 4 calculée, comme illustré ci-dessus.

Lors du dessin, nous illustrons tout mécanisme de suspension/reprise avec une boîte, par exemple la boîte sur le bord 0 4. La boîte représente une valeur future calculée par le niveau inférieur, provoquant la suspension du niveau supérieur jusqu'à son arrivée. Finalement, la case est remplie par l'état 3, comme illustré par 3 afin que le calcul puisse reprendre au niveau supérieur.

Nous définissons ensuite formellement les séries de SHA sur les couvertures. Qu'une haie avec des lettres et des états R HΣ Q soit un Δ-run ou un simple run sur une haie h HΣ, noté R runΔ(h), est défini par les règles suivantes :

$$\frac{q\ a{-\!\!\rightarrow}\ ' \text{ en } \Delta}{q \cdot une \cdot q'\ q \quad run\Delta(a)} \qquad \frac{q \cdot R \cdot q\ ' \quad course\Delta(h)\ q\ ' \cdot R\ ' \cdot q\ » \quad course\Delta(h\ ')}{q \cdot R \cdot q\ ' \cdot R\ ' \cdot q\ » \quad course\Delta(h \cdot h')}$$

$$\frac{\text{vrai}}{q \quad run\Delta(\varepsilon)\ q \cdot \quad R \quad \cdot q} \qquad \frac{-\!\!\rightarrow p \text{ dans } \Delta\ p \cdot R \cdot p \quad course\Delta(h)\ q@p \quad ' \rightarrow q\ ' \text{ en } \Delta}{}$$

Notez que si R runΔ(h), ' courseΔ(h)

alors h peut être obtenu en supprimant tous les états de R, c'est-à-dire projΣ (R) = h. Toute exécution R runΔ(h) peut être identifiée grâce au mappage des préfixes du mot imbriqué nw(h) aux états dans Q. Notez que nw(h) = projΣ (nw(R)). Pour tout préfixe v de nw(h), il existe un préfixe unique r du mot imbriqué nw(R) qui se termine par un état q Q et satisfait projΣ (r) = v. Nous appelons alors q l'état de R à préfixe v. L'exécution 0 · 0 · liste · 1 · 0 · élément · 2 · 3 · 4 sur la haie list item de la figure 5, par exemple, attribue l'état 0 aux préfixes de mots imbriqués ε et et état 1 au préfixe de mot imbriqué · liste, etc.

Une Δ-exécution est appelée une exécution de A = (Σ, Q, Δ, I, F) si elle commence par un état dans I. Une exécution de A réussit s'il se termine par un état dans F.

Lemme 4. h L(A) ssi il existe une exécution réussie R runΔ(h) de A.

Preuve. Simple avec induction sur h. □

4.1.3. Détermination

Les SHA peuvent toujours être déterminés à l'aide de la construction de sous-ensembles connue des détermination d'automates à états finis sur des mots ou d'automates arborescents.

Étant donné un SHA A = (Σ, Q, Δ, I, F), l'ensemble d'états de sa détermination est Qdet = 2 Q. Les règles de transition det Δ états de la détermination sont présentés dans la figure 6.

Machine Translated by Google

$$\frac{Q \quad Q \text{ une } \quad \Sigma\, Q \qquad ' = \{q \quad ' \quad Q \mid q\, a{-\!\!\rightarrow} q \quad ' \quad \Delta, q \quad Q\}}{Q\, a{-\!\!\rightarrow} Q' \ \text{det} \quad \Delta}$$

$$\frac{\Delta \quad =}{\Delta \quad \Delta}\ \text{dét} \qquad \frac{Q1, Q2 \quad QQ \qquad ' = \{q \quad ' \quad Q \mid q1@q2 \rightarrow q \quad ' \quad \Delta, q1 \quad Q1, q2 \quad Q2\}}{Q1@Q2 \rightarrow Q' \quad \Delta}\ \text{dét}$$

Figure 6. La détermination Δ^{det} des règles de transition Δ d'un SHA.

4.1.4. Test d'adhésion en mémoire

Tester l'appartenance h L(A) pour une couverture h, dont le graphique est stocké en mémoire, peut être effectué en calculant l'exécution unique de la détermination de A de manière ascendante et de gauche à droite, et en testant si c'est réussi. Alternativement, on peut calculer l'ensemble unique P Q tel que I $-\!\!\rightarrow^{h}$ P par rapport à la déterminisation de A et tester si P ∩ F = . Les calculs sont essentiellement les mêmes, mais l'historique complet du calcul est perdu lors du renvoi uniquement de l'ensemble P d'états atteints et non de l'exécution.

Pour déterminer les SHA, nous pouvons nous appuyer sur une variante de la construction habituelle de sous-ensembles qui est bien connue des automates à états finis sur mots ou des automates arborescents standards (voir par exemple [8,16]). Lorsqu'il s'agit de tests d'appartenance, la détermination à la volée est suffisante, ce qui crée uniquement la partie de l'automate déterministe utilisée par l'exécution sur h. Cette partie est de taille linéaire O(|h|) tandis que la détermination complète de A peut être de taille exponentielle O(2 | A|). De cette manière, l'appartenance h L(A) peut être testée en temps O(|h||A|).

4.1.5. Accessibilité de la couverture

Pour tout ensemble P Q nous définissons l'ensemble des états accessibles depuis P via une couverture comme suit :

$$\text{acc}\Delta\,(P) = \{q \quad ' \quad h \quad Q \mid \quad q \quad P.\ h \quad H\Sigma.\ q -\!\!\rightarrow q \quad '\ \text{par rapport à } \Delta\}.$$

Nous notons que acc∆(P) peut être calculé à partir de ∆ et P en temps linéaire dans la taille de ∆. Un état est accessible depuis P si et seulement s'il est accessible pour P dans le graphe de ∆. Soit en calculant le point le plus petit fixe du programme Datalog sol suivant en fonction de P et ∆ qui est de taille linéaire dans A :

$$\frac{q}{\text{Pacc (p).}} \qquad \frac{q\, a{-\!\!\rightarrow} \ ' \ \text{en } \Delta}{\text{acc(q } 'q\text{) :– acc(q).}} \qquad \frac{q@p \rightarrow q \quad ' \ \text{en } \Delta}{\text{acc(q } ') \text{ :– acc(q), acc(p).}}$$

Clairement, L(A) = si et seulement si F ∩ acc∆(I) = . Par conséquent, la vacuité peut être décidée en temps linéaire pour les SHA. Ceci contraste avec la notion plus générale de SHA↓ que nous présenter ensuite.

4.2. Automates de couverture par étapes vers le bas

L'évaluation des SHA s'effectue de bas en haut et de gauche à droite, mais ne peut transmettre aucune information de haut en bas. Nous proposons ensuite une extension des SHA aux SHA↓ , ajoutant la possibilité de transmettre des informations d'état fini de haut en bas.

Les SHA↓ sont fondamentalement égaux aux automates forestiers pushdown de Neumann et Seidl [25] mais déplacés des forêts (étiquetées) vers les haies (non étiquetées). Voir la section 12 pour une discussion plus détaillée sur la relation des SHA↓ avec d'autres notions d'automates et de NWA en particulier. lar.

Définition 4. Un automate de couverture pas à pas descendant (SHA↓) est un tuple A = (Σ, Q, ∆, I, F), où Σ et Q sont des ensembles finis, I, F Q et ∆ = ((a ∆)a Σ, ∆, @∆). De plus, a ∆ Q × Q, et @∆ (Q × Q) × Q. Un $^{\Delta}$ Q × Q, SHA↓ est déterministe ou de manière équivalente un dSHA↓ si I contient au plus un élément, et toutes les relations, ∆, a ∆ et @∆ sont des fonctions partielles.

Machine Translated by Google

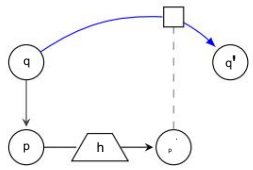
La seule différence par rapport aux SHA réside dans la forme des règles d'ouverture des arbres. Si (q, q ′) ∈ Δ Q, alors nous avons une règle initiale d'arbre que nous notons comme suit : q −→ q où commence ′ en Δ. Alors, indiquez ′ l'évaluation d'une sous-couverture, dépend de l'état q du q parent.

4.2.1. Sémantique

La définition des étapes de transition pour les SHA↓ diffère de celle des SHA uniquement par l'utilisation des règles d'ouverture dans la règle d'inférence suivante :

$$\frac{h \longrightarrow p \text{ dans } \Delta \quad p \longrightarrow pq \quad \text{par rapport à } \Delta \quad q@p \longrightarrow q' \text{ en } \Delta}{q \longrightarrow qq' \text{ par rapport à } \Delta}$$

Cela signifie que l'évaluation de la sous-couverture h démarre dans un état p tel que q −→ p dans Δ.



Ainsi, l'état de redémarrage p choisi peut désormais dépendre de l'état q ci-dessus. C'est ainsi que les informations sur l'état fini sont transmises de haut en bas par les SHA↓ . Les SHA, en revanche, fonctionnent uniquement de bas en haut et de gauche à droite.

4.2.2. Fonctionne

La notion de runs peut être facilement adaptée des SHA aux SHA↓ . Dans l'état q, il suffit de relancer le calcul en sous-couvertures sur l'état p, tel que

$$\frac{p \longrightarrow p \text{ dans } \Delta \quad p \cdot R \cdot pq \quad \text{course}_\Delta(h)\ q@p \longrightarrow q' \text{ en } \Delta}{q \cdot R \cdot q' \quad \text{course}_\Delta(h)}$$

Un exemple d'exécution du dSHA↓ dans la figure 7 est présenté dans la figure 8. Ici, les informations sur le niveau des nœuds sont transmises. Il est représenté par le nombre de nombres premiers. Par exemple, lors de l'ouverture du premier sous-arbre étiqueté avec item, nous utilisons la règle de transition $1' \xrightarrow{»}_0$ déclarant que l'on passe de l'état 1 au premier niveau à l'état 0 au deuxième niveau.

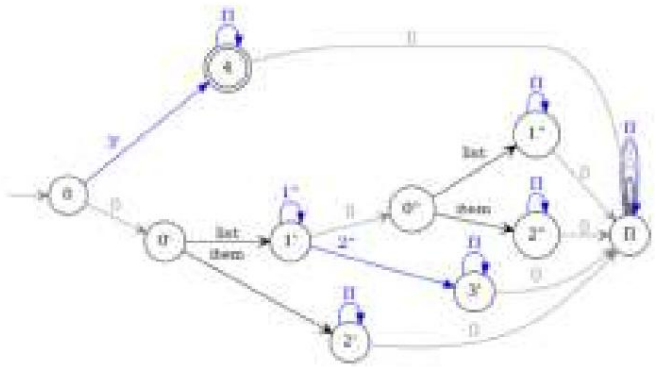


Figure 7. Le SHA↓ déterministe avec état de projection de sous-couverture Π obtenu par projection sûre sans changement .

Machine Translated by Google

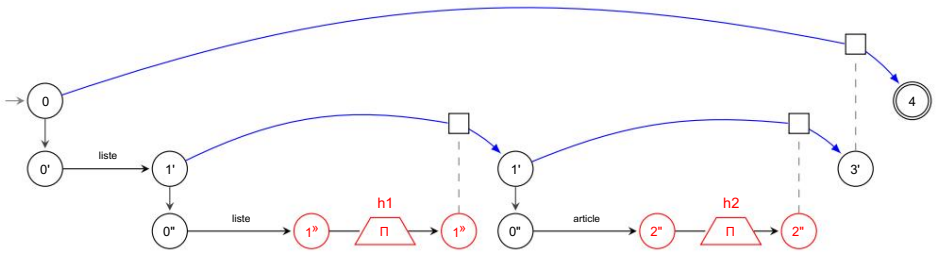


Figure 8. Une exécution réussie du SHA↓ dans la figure 7 sur la haie list · list · h1 · item · h2 .

On peut voir que tous les nœuds des sous-arbres h1 et h2 sont évalués à l' état de projection Π. Ainsi, on peut sauter par-dessus ces sous-arbres sans les lire. Ceci se justifie par le fait qu'ils sont au niveau 2, l'information que transmettait l'évaluateur de ce SHA↓ vers le bas.

4.2.3. Test d'adhésion en mémoire

Tester l'appartenance h L(A) pour un SHA↓ s en mémoire peut à nouveau être effectué en calculant l'exécution unique de la détermination de A. Comme auparavant, le testeur d'appartenance peut être basé sur une détermination à la volée. Cependant, la procédure de détermination des SHA↓ est plus compliquée que celle des SHA puisque les SHA↓ peuvent transmettre des informations de haut en bas et pas seulement de bas en haut et de gauche à droite. La détermination ne repose pas seulement sur la construction pure de sous-ensembles, mais utilise également des paires d'états dans les sous-ensembles, essentiellement de la même manière que pour les automates à mots imbriqués (NWA) [14, 15]. Cela est nécessaire pour gérer la pile d'états vue ci-dessus. Par conséquent, chaque étape de détermination peut coûter du temps O(|A|) comme indiqué par exemple dans [23]. La limite de temps supérieure pour les tests d'appartenance pour SHA↓ s est donc dans le temps O(|h||A| ^{5}) et non plus combinés linéairement comme c'est le cas pour les SHA.

4.2.4. Relation avec les SHA

Tout SHA A = (Σ, Q, Δ, I, F) peut être mappé à un SHA↓ A tout $^{\text{vers le bas}}$ = (Σ, Q, Δ $^{\text{vers le bas}}$, SI) en préservant les exécutions et le déterminisme. Le seul changement $^{\text{vers le bas}}$ comparé à Δ est de Δ décrit par la règle suivante :

$$\frac{-\!\!\rightarrow p \text{ dans } \Delta\ q \qquad Q}{\Delta\ q \qquad -\!\!\rightarrow p \text{ dans}^{\text{vers le bas}}}$$

Indépendamment de l'état actuel q Q, le SHA↓ Une sous-couverture $^{\text{down}}$ peut démarrer l'évaluation du d'un sous-arbre dans tout état d'arbre ouvert p Δ. Par exemple, le dSHA↓ A dSHA A de la figure 4 de $^{\text{vers le bas pour}}$ l'introduction est fourni dans la figure 9. Inversement, on peut dSHA↓ A en un SHA équivalent en introduisant le non-déterminisme pour éliminer tout traitement descendant. Voir la section 12.2 pour la construction.

Machine Translated by Google

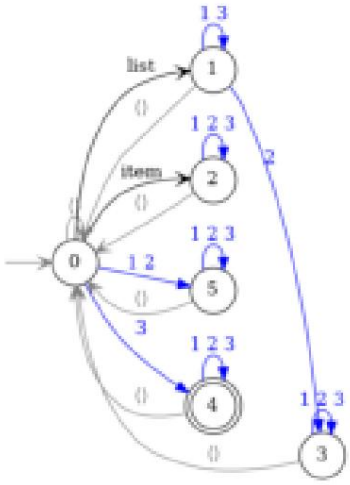


Figure 9. Le SHA↓ A vers le bas pour le dSHA A pour le filtre [self::list/child::item] dans la figure 4.

4.2.5. Accessibilité des haies

L'accessibilité des haies peut être définie pour les SHA↓ de la même manière que pour les SHA. Cependant, l'ensemble accΔ(P) ne peut plus être calculé en temps linéaire dans la taille de Δ, contrairement aux SHA. Cela peut toujours être fait en temps cube, de la même manière que pour les NWA [22]. Le problème est que l'accessibilité des couvertures pour les SHA↓ ne coïncide plus avec la notion d'accessibilité du graphe de l'automate.

Par exemple, dans le SHA↓ de la figure 7, nous avons accΔ({0}) = {4}. Notez que 0′ fait dans Δ, ce qui fait cet ensemble, même s'il existe une règle de transition 0 −→ 0 accessible depuis le nœud 0′ n'appartient pas à 0 dans le graphe de l'automate. En effet, 0′ n'est accessible sur aucune haie à partir de 0 ; c'est-à-dire qu'il n'y a pas de couverture h telle que 0 par rapport à Δ. Notez que 0 · · 0 est un facteur du mot imbriqué d'une série de Δ. Cela montre que 0′ est néanmoins accessible à partir de 0 sur un préfixe de mot imbriqué. Pour tout SHA↓ A, l'accessibilité des haies peut être calculée par le programme Datalog au sol suivant, de taille cubique en fonction du nombre d'états de A :

$$\frac{p@q' \quad \rightarrow q \text{ dans } \Delta \; p \dashrightarrow p' \text{ en } \Delta}{\text{acc}(p, q) :\!- \text{acc}(p', q').} \qquad \frac{p}{Q \text{ acc}(p, p).} \qquad \frac{p \; a \dashrightarrow q \text{ dans}}{\Delta \text{ acc}(p, q).}$$

$$\frac{p, q, r \quad Q}{\text{acc}(p, q) :\!- \text{acc}(p,r), \text{acc}(r, q).}$$

En raison du temps cube d'accessibilité des haies pour les SHA↓ , nous utiliserons uniquement l'accessibilité des haies. pour les SHA où c'est en temps linéaire.

4.3. Complétude du schéma

Les schémas sont des entrées du problème d'appartenance que nous considérons tandis que la complétude est une propriété fondamentale des automates. Pour combiner les deux aspects de manière appropriée, nous proposons la notion de complétude du schéma, que nous définissons uniformément pour les SHA et les

Nous appelons un ensemble de Δ règles de transition SHA↓ complet si toutes ses relations (a Δ)a Σ, Δ et @Δ sont totales. Pour les SHA, nous exigeons que Δ soit un ensemble non vide (au lieu d'une relation totale). Un SHA ou SHA↓ A est dit complet si Δ est complet et I = . Nous pouvons compléter n'importe quel SHA ou SHA↓ en ajoutant un nouvel état récepteur ainsi que des règles de transition vers l'état récepteur pour tous les côtés de gauche, pour lesquels aucune autre règle de transition n'existe. L'état récepteur est rendu initial si et seulement s'il n'y a pas d'état initial auparavant. Ce n'est cependant pas définitif. On note la complétion de A par compl(A).

Définition 5. Une exécution partielle de A sur une haie h est un préfixe r d'une exécution de compl(A) sur h, tel que

- r se termine par un état de Q, et

Machine Translated by Google

- r ne contient pas l'état puits de compl(A).

Une exécution partielle r de A sur h est dite bloquante s'il n'existe pas d'exécution partielle r ′ de A sur h, tel qui r soit un préfixe strict de r′ .

Par exemple, considérons la haie h = list · item . Le dSHA de la figure 4 a l' exécution unique R = 0 · · 0 · liste · 1 · · 0 · élément · 2 · 3 · · 4 sur h illustré sur la figure 5.
Les exécutions partielles de h sont donc tous des préfixes de nw(R) qui se terminent par un état, c'est-à-dire 0, 0 · · 0, 0 · · 0 · liste · 1, etc. Aucune de ces exécutions partielles n'est bloquante.

Définition 6. Un schéma pour un automate A est un langage de couverture S HΣ, tel que L(A) S.
Nous appelons l'automate A schéma-complet pour le schéma S si aucune haie h S n'a une exécution partielle bloquante de A.

Les schémas S sont généralement utilisés pour restreindre le type de haies qu'un problème peut accepter en entrée. Le problème de correspondance de modèles dSHA pour un schéma S prend deux entrées : une haie h S et un automate A, plutôt qu'une expression régulière imbriquée e. L'automate A sélectionne ensuite les couvertures d'entrée h S qui correspondent au modèle, c'est-à-dire telles que h L(A). Pour cette raison, nous supposons que S est un schéma pour A, c'est-à-dire L(A) S. Nous supposerons souvent que A est un schéma complet pour S, de sorte qu'aucune exécution partielle de A sur une couverture d'entrée h S peut un jour bloquer. Cela peut toujours être réalisé ensuite sur la base du prochain Lemme 5.

Exemple 5. Le dSHA de la figure 4 pour le filtre XPath [self::list/child::item] a la signature Σ = {item, list}. Il s'agit d'un schéma complet pour le document de schéma. Pour y parvenir, nous avons ajouté l'état 5 à cet automate, qui n'est utilisé par aucune exécution réussie. Pourtant, ce SHA n'est pas complet. Pour le compléter, nous devrions exécuter 5 dans des états puits en y ajoutant de nombreuses transitions, telles que 0@4 → 5 et 0@5 → 5, et également ajouter pour tous les autres états q {1, 2, 3, 4, 5} les règles de transition
$\overset{\text{liste}}{\longrightarrow} 5, q \overset{\text{article}}{\longrightarrow} 5, q@4 \to 5$ et $q@5 \to 5$. q

Comme nous le verrons, l'achèvement de l'automate rendrait l'algorithme de sécurité sans changement incomplet pour la projection de sous-couverture pour l'exemple 5. Cependant, sans l'exhaustivité du schéma, la projection sans changement peut s'avérer douteuse : elle pourrait négliger le rejet des couvertures à l'intérieur du schéma. C'est pourquoi il est bien préférable de supposer uniquement l'exhaustivité du schéma plutôt que l'exhaustivité totale pour les automates d'entrée de la projection sûre et sans changement. Pour la projection de congruence, l'intégralité du schéma sera supposée implicitement.

Nous notons que la complétude du schéma est bien définie même pour les schémas non réguliers. Pour une projection sûre et sans changement, nous n'avons en effet pas besoin de restreindre les schémas pour qu'ils soient réguliers, tandis que pour la projection de congruence, seuls les schémas réguliers sont considérés.

De plus, si A est schéma-complet pour le schéma universel S = HΣ et ne ont des états inaccessibles, alors A est complet.

L'exhaustivité du schéma des automates déterministes pour les schémas réguliers peut toujours être assurée comme nous le montrerons ensuite. Pour cela nous définissons que A est compatible avec le schéma S si pour toute couverture h S il existe une exécution de A dans run∆(h). L'exhaustivité du schéma implique la compatibilité, comme le montre le lemme suivant. L'inverse n'est vrai qu'en supposant le déterminisme.

Lemme 5. Soit A un SHA déterministe de schéma S. Alors, A est compatible avec S si A est schéma-complet pour S.

Preuve. Supposons que A soit un schéma complet pour S Wlog., nous pouvons supposer S = . Notez que si I = , alors ε serait une exécution partielle bloquante pour toute couverture dans S \ {ε}. Puisque S = , il s'ensuit qu'il existe q0 I. Donc, q0 est une exécution partielle pour toute couverture h S \ {ε}. Puisqu'il n'existe pas d'exécutions partielles bloquantes par complétude du schéma, cette exécution peut être étendue étape par étape à une exécution sur n'importe quel h S. Par conséquent, pour toute couverture h S, il existe des exécutions par A, montrant la compatibilité.

Machine Translated by Google

A l'inverse, soit A compatible avec S et déterministe. Nous devons montrer que A est un schéma complet pour S. Soit v une exécution partielle de A sur h S. Par compatibilité, il existe une exécution R runΔ(h) qui commence dans un état initial de A. Par déterminisme, v doit être un préfixe de nw(R). Ainsi, v n'est pas bloquant. □

Proposition 1. Tout SHA A avec un schéma régulier S peut être rendu schéma-complet pour S.

Preuve. D'après le lemme 5, il suffit de rendre A compatible avec S. Soit B un dSHA avec S = L(B). L'automate B est compatible avec S. Puisque B est déterministe, le lemme 5 implique que B est un schéma complet pour S. Puisque S est un schéma pour A avec L(A) L(B). Nous calculons ensuite la complétion de compl(A). Le produit C = B × compl(A) avec les états finaux est schéma-complet pour le schéma S. De plus, puisque L(A) S = L(B), CB F = F
suit que L(C) A × F il
= L(B) ∩ L(A) = L(A). □

Enfin, notons que le schéma S peut être reconnu par le même produit C = B × par l'ensemble des états finaux du compl(A) sauf pour remplacer l'ensemble des états finaux F C schéma × (QA {sink}). On a S = L(C[F C/F S = F B F S]), où C[F C/F S] est le dSHA
S obtenu à partir de C en remplaçant l'ensemble des états finaux F C avec F .

5. Projection de sous-couverture

Nous introduisons ensuite le concept d'états de projection de sous-couverture pour les SHA↓ afin de distinguer les préfixes de haies qui ne sont pas pertinents pour les sous-couvertures, et d'évaluer les SHA↓ avec la projection de sous-couverture.

5.1. États de projection des sous-couvertures

Intuitivement, un état de projection de sous-couverture ne peut boucler sur lui-même que jusqu'à ce qu'une haie se ferme (puis qu'elle soit appliquée). Lors de la descente dans un sous-arbre, l'état de projection du sous-couverture peut toujours être modifié en un autre état de boucle. Cependant, en quittant le sous-arbre, il faut revenir à l'état de projection initial du sous-arbre. De toute évidence, tout puits est un état de projection de sous-couverture. Les cas encore plus intéressants, cependant, sont les états de projection de sous-couverture qui ne sont pas des puits.

Nous commençons avec les SHA↓ sans aucune restriction mais imposerons l'intégralité du schéma et le déterminisme plus tard.

Définition 7. Nous appelons un état q Q un état de projection de sous-couverture de Δ s'il existe q ' Q appelé témoin de q, tel que l'ensemble des règles de transition de Δ contenant la position ' ou q à l'extrême gauche q est inclus dans l'ensemble suivant :

$$\{q \xrightarrow{} q',\ q@q' \to q,\ q' \xrightarrow{} q'@q' \to q'\}$$
$$q \{q\xrightarrow{a} ',\ q \text{ une} \xrightarrow{} q \mid \text{une } \Sigma\}$$

L'ensemble de tous les états de projection de sous-couverture de Δ est noté $Q_{\Delta_{shp}}$.

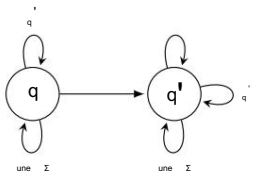
Pour tout SHA↓ complet , l'ensemble de règles de transition ci-dessus doit être égal à celui défini à la toutes les règles de transition de Δ avec q ou q ' position la plus à gauche. Mais, si le SHA↓ n'est que l'une de schéma complet pour un schéma S, alors toutes ces transitions ne doivent pas être présentes. Notez également qu'un état de projection de sous-couverture q peut être égal à son témoin q d'un '. Ainsi, tout témoin q' certain état de projection de sous-couverture qui est lui-même un état de projection de sous-couverture '.

Dans l'exemple dSHA↓ A = (Σ, Q,Δ, I, F) sur la figure 7, nous avons QΔ = {Π, 4, 2′ ,avec le témoin q 3′ , 1′
shp Le témoin de tous ces états de projection de sous-couverture est Π. Notez que toutes les transitions possibles ne sont pas présentes pour les états dans QΔ \ {Π}, étant donné que cet automate n'est pas complet (mais toujours shp schema-complete pour le schéma doc).

Lemme 6. Si q $Q_{\Delta_{shp}}$ est un état de projection de sous-couverture et $q \xrightarrow{h} p$ par rapport à Δ, alors q = p.

Machine Translated by Google

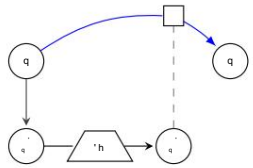
Preuve. Supposons que $q \xrightarrow{h} p$ par rapport à Δ et que q soit un état de projection de sous-couverture de Δ avec témoin sur q'. Donc les seules transitions possibles avec q ou q' le membre de gauche sont les suivants :



que $h = t1 \cdot \cdots \cdot tn$ est une séquence d'arbres ou de lettres ti $H\Sigma \Sigma$, où $1 \le i \le n$. Alors, il Supposons existe des exécutions Ri $run\Delta(ti)$ et une exécution R $run\Delta(h)$ qui a la forme $h = q0 \cdot R1 \cdot q1 \cdot \cdot Rn \cdot qn$, où $q0 = q$ et $qn = p$. On prouve pour tout $0 \le i \le n$ que $qi = q$ par récurrence sur i. Pour $i = 0$, c'est trivial. Pour $i > 0$, l'hypothèse d'induction montre que $qi-1 = q$.

Cas ti = a Σ. Alors, $qi-1 \xrightarrow{a} qi$ dans Δ. Puisque $qi-1 = q$ est un état de projection de sous-couverture de Δ, il s'ensuit que $qi = q$ également.

Cas ti = h $'$ pour certains h' $H\Sigma$. Puisque $qi-1 = q$ est un état de projection de sous-couverture de Δ avec le sous-témoin q', ensemble de R reconnaissant ti doit être justifié par le diagramme suivant :



Donc, le sous-ensemble de R de h' ne peut contenir que le témoin q' et, de plus, $qi = q$.

□

5.2. Complétude pour la projection de sous-couverture

Avant de définir quand un SHA↓ est terminé pour une projection de sous-couverture, nous montrons que les états de projection de sous-couverture dans les séries de SHA déterministes ne peuvent se produire qu'à des préfixes non pertinents.

Proposition 2. Soit $A = (\Sigma, Q, \Delta, I, F)$ un dSHA↓ de schéma S, R $run\Delta(h)$ est l'exécution de compl(A) sur une haie h S, et q est un état de projection de sous-couverture pour Δ. Alors, pour tout préfixe $r \cdot q$ de nw(R), le préfixe $proj\Sigma$ (r) de nw(h) est une sous-couverture non pertinente pour L(A) pour le schéma S.

Preuve. Soit R $run\Delta(h)$ l'unique exécution de compl(A) sur une haie h S. Supposons que $r \cdot q \cdot s = nw(R)$ pour certains états de projection de sous-couverture q de Δ et soit $v = proj\Sigma$ (r) et $w = proj\Sigma$ (s). Puisque $v \cdot w = nw(h)$ et h S, il s'ensuit que $v \cdot w$ nw(S). Nous fixons) $\cdot$ w nw(S) arbitrairement. Soit ˜h $H\Sigma$, tel qu'une couverture h nw(˜h) = v · nw(h $'$ $H\Sigma$, tel que v · nw(h $'$) · w. Pour la non-pertinence de v en tant que sous-couverture, nous devons montrer que

h L(A) ˜h L(A).

« » Nous supposons que h L(A) et devons montrer que ˜h L(A). Pour l'exécution partielle $r \cdot q$ de $\cdot p$ $run\Delta(h')$. A sur ˜h, il doit exister R et p, tels que q · R état de projection de $'$ Puisque q est une sous-couverture) $\cdot$ q $\cdot$ s est une Δ, il s'ensuit par le Lemme 6 que q = p. Par conséquent, r · q · nw(R exécution de A sur ˜h. Puisque R $'$ a réussi pour A, s doit se terminer par F ; ainsi, l'exécution ci-dessus sur ˜h est également réussie pour A. Par conséquent, ˜h L(A), comme nous l'avons montré. « » Nous supposons que

˜h L(A) et devons montrer que h L(A) est l'unique exécution de compl (A) sur ˜h. déterminisme, R˜ doit avoir le préfixe r · q Donc, il existe.

Machine Translated by Google

R. ' ' ' ' ' , p ,s, tel que p · R · q runΔ(h ′) et r · q · R · p · s = R˜. Le lemme 6 montre que p = q. Par conséquent, r · q · s est l'exécution unique de compl(A) sur h et cette exécution est acceptante. Donc h L(A).

□

Nous notons que la proposition 2 ne serait pas valable sans supposer le déterminisme. Pour voir cela, nous pouvons ajouter un récepteur à n'importe quel dSHA comme deuxième état initial. On peut alors toujours accéder à ce puits, qui est un état de projection de sous-couverture. Néanmoins, aucun préfixe allant à l'évier ne peut être sans importance.

La proposition 2 montre que les états de projection de sous-couverture permettent de distinguer les préfixes qui ne sont pas pertinents pour les SHA↓ déterministes . Une question intéressante est de savoir si tous les préfixes non pertinents pour les sous-couvertures peuvent être trouvés de cette manière. Une question étroitement liée est de savoir si, pour tous les modèles réguliers, il existe des dSHA qui projettent toutes les sous-couvertures non pertinentes.

Exemple 6. Reconsidérez l'exemple 3 avec Σ = {a}, un modèle régulier L = { a } et un schéma régulier S = L { · a}. Le préfixe n'est pas pertinent pour L et S, puisque toutes ses continuations se terminant par le schéma sont rejetées : il n'y a qu'une seule de ces continuations, qui est · a. Cependant, un dSHA pour L et S avec une projection de sous-couverture maximale est donné dans la figure 17 mais n'entrera pas dans un état de projection de sous-couverture à ce préfixe. La raison en est qu'il entrera dans le même état {3, 4}D0 pour tout préfixe dans NΣ , en ignorant le mot imbriqué de la sous-couverture non pertinente. Par conséquent, cet état doit accepter a . Mais, lors de la lecture d'un a dans cet état, le dSHA passe à l'état de rejet {5}D0, montrant que la couverture du schéma · a S est rejetée.

L'exemple montre qu'on finit par raisonner avec des ensembles de préfixes. On lève d'abord le notion de sous-couverture non pertinente pour les ensembles de préfixes.

Définition 8. Soient S HΣ un schéma et L S un langage satisfaisant ce schéma. Un ensemble de préfixes de mots imbriqués U prefs(NΣ) est appelé sous-couverture pertinente pour L, de schéma S s'il existe des mots imbriqués v NΣ préfixes u · v "). Sinon, l'ensemble U, tel que diff U est L S (tu · v · "tu " appelé sous-couverture non pertinent pour L par rapport à S.

Afin d'expliquer le problème de l'exemple 6, nous définissons pour chaque préfixe de mot imbriqué w prefs(NΣ) une classe classL (w) de préfixes similaires S en remplaçant le mot imbriqué de toute sous-couverture non pertinente par NΣ de gauche à droite. Pour tous les préfixes w prefs(NΣ), les mots imbriqués v NΣ, les lettres a Σ, les sous-ensembles de préfixes U prefs(NΣ) et l Σˆ lettres ou parenthèses, ce qui suit s'applique :

$$\text{classe}_{L_S}(w) = \text{classe_pr}_{\{\varepsilon\}}(w)$$

$$\text{classe_prU}(\varepsilon) = \begin{cases} U \cdot N\Sigma & \text{si U n'est pas pertinent pour L par rapport à S} \\ U & \text{sinon} \end{cases}$$

$$\text{class_prU}(w \cdot \cdot v) = \text{class_insU}'(, v) \text{ où } U' = \text{class_prU}(w)\ \text{class_prU}(a \cdot v) =$$

$$\text{class_insU}'(a, v) \text{ où } U' = U \cdot a\ \text{class_prU}(\text{ v}\ \text{class_insU}(l, v) =$$

$$\cdot v) = \text{class_insU}'(, v) \text{ où } U' = \text{class_prU}(\cdot v$$

$$U \cdot N\Sigma \quad) \text{ si U n'est pas pertinent pour L par rapport à S}$$

$$\text{class_prU} \cdot l\ (v) \text{ sinon}$$

Définition 9. Un préfixe u prefs(NΣ) est fortement sans importance pour la sous-couverture pour L par rapport à S si l'ensemble classL (u) est sans importance pour la sous-couverture pour L par rapport à S.

Exemple 7. Dans notre exemple 6, où L = { a }, S = L { · a} et Σ = {a}, nous avons classL ( ) = · NΣ, donc le préfixe est fortement hors de propos. L'ensemble classL (a) = S NΣ est pertinent pour la sous-couverture, donc le préfixe a n'est pas fortement sans importance pour la sous-couverture. L'ensemble classL (· a) = NΣ · a · NΣ n'est pas pertinent pour la sous-couverture, donc le préfixe · a est fortement non pertinent pour la sous-couverture.

Machine Translated by Google

Définition 10. A dSHA↓ A = (Σ, Q, Δ, I, F) est complet pour la projection de sous-couverture pour le schéma S si A est un schéma complet pour S et pour toutes les haies h S et tous les préfixes v de nw(h) qui sont fortement sans rapport avec la sous-couverture pour Δ, l'exécution unique de A sur h passe à un état de projection de sous-couverture de Δ en v.

Exemple 8. Le dSHA de la figure 17 est complet pour la projection de sous-couverture pour notre exemple 6 en cours où L = { a } et S = L { · a}. Elle a été obtenue par projection de congruence. Sur tous les préfixes de la classe non pertinente de sous-couverture classL ( ), il passe à l'état de projection de sous-couverture {0}D1. Sur tous les préfixes de la classe classL (a), qui n'est pas sans importance pour la sous-couverture, il passe à l'état {3, 4}D0, qui n'est pas un état de projection de sous-couverture. Sur tous les préfixes de la classe non pertinente de sous-couverture classL (· a), il passe à l'état de projection de sous-couverture {5}D0.

Exemple 9. Le dSHA↓ de la figure 7 est complet pour la projection de sous-couverture avec le schéma doc. Il peut être obtenu par projection sûre et sans changement. En général, cependant, les dSHA↓ obtenus par la projection sans changement peuvent ne pas être complets pour la projection de sous-couverture, comme nous le verrons à la section 6.3.

5.3. Évaluateur en mémoire avec projection de sous-couverture Nous montrons ensuite comment affiner les transitions pour les SHA↓ par projection de sous-couverture. Nous définissons la relation de transition avec la projection de sous-couverture $\xrightarrow{h}_{shp}$ $Q \times Q$ par rapport à Δ, tel que, pour toutes les haies h, h H Σ et les lettres a Σ,

$$\frac{q \quad Q\Delta \quad shp \quad h \quad H\Sigma}{\xrightarrow{h}_{shp} q \text{ par rapport à } \Delta} \qquad \frac{Q\Delta \ni shp\ q \xrightarrow{a} q\ q' \text{ en } \Delta}{a \xrightarrow{}_{shp} q\ q' \text{ par rapport à } \Delta} \qquad \frac{q \quad Q\Delta_{shp}}{q \xrightarrow{\varepsilon}_{shp} q \text{ par rapport à } \Delta}$$

$$\frac{q \quad Q\Delta \xrightarrow{h}_{shp} q\ shp\ q \quad \text{par rapport à } \Delta\ q' \xrightarrow{h'}_{shp} q\ »^{\text{par rapport à } \Delta}}{h \cdot h' \xrightarrow{}_{shp} qq\ »^{\text{par rapport à } \Delta}}$$

$$\frac{q \quad \xrightarrow{}_{shp} p \text{ dans } \Delta\ p\ shp\ q \quad \xrightarrow{h}_{shp} \quad q@p \quad \xrightarrow{} q' \text{ en } \Delta}{h \quad - \quad \text{par rapport à } \Delta}$$

$\xrightarrow{}_{shp} qq$ La première règle dit que les transitions de projection de sous-couverture restent dans les états de projection de sous-couverture jusqu'à ce que la fin de la sous-couverture actuelle soit atteinte. Ceci est correct d'après le lemme 6 à condition qu'il n'y ait pas d'exécutions bloquantes. Les autres règles stipulent que l'évaluateur se comporte sans projection de sous-couverture dans d'autres États.

Proposition 3. Soit A = (Σ, Q, Δ, I, F) un SHA↓ dont le schéma est complet pour S. Alors, pour toutes les couvertures h S et les états q, p Q, ce qui suit s'applique :

$$\xrightarrow{h} p \text{ par rapport à } \Delta \text{ ssi } q \xrightarrow{h}_{shp} p \text{ par rapport à } \Delta\ q$$

Preuve. Nous distinguons deux cas comme suit :

Cas q $Q\Delta$ Alors pour tout p Q, $q \xrightarrow{h} p$ par rapport à Δ ssi $q \xrightarrow{h}_{shp} p$ par rapport à Δ par définition du shp. projeter les transitions.

Cas q $Q\Delta$ Alors, $q \xrightarrow{h}_{shp} q$ par rapport à Δ. Puisque h S et A sont schéma-complets pour S, il y a shp. h existe une série de A sur h ; ainsi, l'état est q, tel que $q \xrightarrow{} q$ que q par rapport à Δ. Le lemme 6 montre $q' = q$, donc $q \xrightarrow{h} q$ par rapport à Δ.

□

Pour évaluer un SHA↓ non déterministe A = (Σ, Q, Δ, I, F) avec projection de sous-couverture sur une couverture h dont le graphique est stocké en mémoire, nous pouvons calculer le sous-ensemble unique P Q, tel que $I \xrightarrow{h}_{shp} P$ par rapport à la relation de transition Δ $\overset{\text{dét}}{}$ de détermination

Machine Translated by Google

de A et tester si P ∩ F = . Lorsqu'il atteint un état de projection de sous-couverture P , l'évaluateur peut directement sauter alors que à la fin de la sous-couverture actuelle en appliquant la règle de « séjour » de la détermination de A :

$$\frac{P. \quad Q\Delta \overset{\text{dét}}{\underset{\text{chp}}{}} \quad h \quad H\Sigma}{P. \quad h \quad - \rightarrow \text{shp } P \quad \text{par rapport à } \Delta^{\text{dét}}}$$

Ainsi, il évitera de visiter une sous-couverture dans un état de projection de sous-couverture de la détermination de A. Le temps d'exécution dépendra donc linéairement de la taille de la partie non projetée de h sous l'hypothèse que la détermination de A est disponible.

Comme auparavant, seule la partie nécessaire de la détermination de A pour évaluer h doit être produite à la volée. Globalement, cela coûte un temps polynomial en termes de taille de la partie nécessaire de la détermination (mais pas nécessairement un temps linéaire puisque la détermination des SHA↓ est plus coûteuse que celle des SHA). Notez également que chaque fois que nous découvrons une nouvelle détermination de la d'état P de A, nous devons tester s'il s'agit d'un état de projection de sous-couverture. Pour cela, on détermine et doit calculer l'état P ", tel que les règles P autorisent les $-\rightarrow^{\text{»}}_{P}$ par rapport à $\Delta^{\text{dét}}$ teste si seule la transition avec le témoin P états de projection de sous-couverture pour P » sont disponibles par rapport à $\Delta^{\text{dét}}$.

6. Projection sûre et sans changement

Nous voulons résoudre le problème habituel de correspondance de motifs pour les haies avec projection de sous-couverture. Nous supposons que le modèle régulier est donné par un dSHA alors que le schéma peut être arbitraire, sauf que le dSHA doit être complet (voir Section 4.3). Nous présentons ensuite l'algorithme de projection sans changement, qui compile le dSHA en dSHA↓ tout en introduisant des états de projection de sous-couverture.

L'idée de la projection sans changement est de pousser les informations de haut en bas afin de distinguer les états de boucle qui ne seront plus modifiés en toute sécurité. Ainsi, des états de projection de sous-couverture sont produits, comme nous l'illustrerons par des exemples. La preuve de solidité d' une projection sûre sans changement n'est pas triviale et instructive pour la preuve de solidité d'une projection de congruence dans la section 7.3. On ne peut pas s'attendre à ce que la projection de sous-couverture soit complète, puisque la projection sans risque ne prend en compte que les boucles simples. Les boucles plus générales ne peuvent pas être traitées et la manière dont cela pourrait être fait n'est pas claire. L' algorithme de projection de congruence de la section 7.2 apportera éventuellement une réponse à cette question.

Nous gardons le compilateur sécurisé sans changement indépendant du schéma et, par conséquent, nous n'avons pas à assumer sa régularité. Mais nous devons supposer que le schéma est complet du dSHA d'entrée afin de montrer que le dSHA↓ de sortie reconnaît le même langage dans le schéma. Le problème régulier de correspondance de modèles peut ensuite être résolu en mémoire et avec une projection de sous-couverture en évaluant le dSHA↓ sur la couverture d'entrée en mémoire et avec une projection de sous-couverture, comme discuté dans la section 5.3.

6.1. Algorithme Nous

décrivons maintenant l'algorithme de projection sans changement. Soit A = (Σ, Q, Δ, I, F) un SHA avec le schéma S. L'algorithme prend le dSHA A en entrée et le compile en un dSHA↓ A snc. Nous dirons pourquoi A est valable pour la projection de sous-couverture à condition que A soit un schéma complet pour le schéma S. Nous n'avons même pas besoin de supposer que le schéma S est régulier. La preuve fait huit pages et est donc déléguée en annexe.

Pour tout ensemble P Q, nous définissons l'ensemble des états qui mènent en toute sécurité à P :

$$\text{sûr}\Delta(P) = \{q \quad Q \mid \text{acc}\Delta(\{q\}) \quad P\}.$$

Ainsi, l'état q est sûr pour P si toute couverture lue à partir de q sur laquelle le run avec Δ ne bloque pas doit atteindre un certain état dans P. Notons que safeΔ (P) peut être calculé en temps linéaire dans le

Machine Translated by Google

taille de ∆ en utilisant l'accessibilité de couverture inverse à partir de P. Nous définissons l'ensemble des états qui ne changeront plus :

aucun changement∆ = {q | q sûr∆ ({q})}.

Donc, q aucun changement∆ si et seulement si acc∆({q}) {q}. Cela signifie que toutes les règles de transition commençant par q doivent boucler dans q. Dans l'exemple d'automate de la figure 4, les états d'auto-bouclage sont ceux sans changement∆ = {2, 3, 4, 5}.

Lemme 7. Soit A = (Σ, Q, ∆, I, F) un SHA dont le schéma est complet pour le schéma S. Pour toute couverture h S et tout état q I ∩ no-change∆ , il tient alors que q $\xrightarrow{h}$ q par rapport à ∆.

Preuve. Le schéma-complétude de A pour S appliqué à h S et q I donne l'existence acc∆(q). = q. Ce d'un état q Q, tel que q $\xrightarrow{h}$ q ′ par rapport à ∆. Soit q ′ être un tel état. Notez que q acc∆({q}) ′ Puisque q aucun changement∆ , nous avons q safe∆ ({q}), de sorte que q ′ implique q à ∆. prouve q $\xrightarrow{h}$ q par rapport □

Pour tout état q Q et sous-ensemble d'états Q Q, nous définissons ce qui suit :

s-down∆ (q, Q) = sûr∆ ({p Q | q@∆p Q}) s-no-change∆ (q) = s-down∆ (q, {q})

Un état appartient à s-down∆ (q, Q) si tous les états p acc∆(q) satisfont q@∆p Q. Ainsi, p est un état dans un niveau bas qui montera en toute sécurité vers un état dans Q. Un état p appartient à s-no-change∆ (q) s'il ne change pas q en toute sécurité, c'est-à-dire si {q}@∆acc∆({p}) {q}.

Nous compilons ensuite le SHA A en un SHA↓ A = (Σ, Qsnc, Δ^{snc}, I^{snc}, Fsnc). Pour cela, soit Π un nouveau symbole et considérons l'ensemble d'état comme suit :

$$Qsnc = \{\Pi\} \quad (Q \times 2^{Q}).$$

En pratique, nous limitons l'espace d'état aux états accessibles ou nettoyons les dSHA en ne conservant que les états utilisés lors d'une exécution réussie. Mais, dans le pire des cas, la construction peut effectivement être exponentielle. L'exemple 15 peut être adapté pour le montrer.

Une paire (q, P) signifie que l'état q ne changera plus en toute sécurité dans la sous-couverture actuelle si q P aucun changement∆ . Dans ce cas, la sous-haie peut être projetée. Les ensembles de initial et final les états sont définis comme suit :

$$I^{snc} = I \times \{\} \quad F^{snc} = F \times \{\}.$$

Comment générer les règles de transition de A^{snc} à partir de celles de A est décrit dans la figure 10. Sur les états assignés au niveau supérieur (q, P), l'ensemble P est vide de sorte que seuls les états en non-changement∆ sont sûrs pour aucun changement. C'est pourquoi la définition de I^{snc} et F à partir utilisez P = . En partant d'un état (q, P) pour lequel q peut ne pas changer en toute sécurité, c'est-à-dire q P aucun alors le changement∆ , la règle suivante est appliquée :

$$\frac{q \quad P \quad \text{aucun changement}\Delta}{(q, P) \longrightarrow \Pi \text{ dans } \Delta^{snc}}.$$

L'évaluation au niveau inférieur passe à l'état supplémentaire Π, où elle boucle ensuite jusqu'à revenir à q au niveau supérieur. En descendant d'un état (q, P), tel que q P no-change∆ , alors la règle suivante est appliquée :

$$\frac{q' \text{ dans } \Delta \quad q \quad P \quad \text{aucun changement}\Delta}{\Delta\ (q, P) \longrightarrow (q, \text{s-pas de changement}\Delta\ (q)) \text{ dans } ^{snc}}$$

Machine Translated by Google

Les états dans l'ensemble s-no-changeΔ (q) au niveau inférieur ne feront pas en toute sécurité changer q au niveau supérieur pour aucune sous-couverture à venir (nous pourrions détecter des sous-couvertures plus non pertinentes en permettant à q de changer mais seulement de manière unique). Cela peut être obtenu avec la définition alternative : s-no-change′Δ (q) = r Qs-downΔ (q, {r}). Le calcul de cet ensemble nécessiterait un temps quadratique O(n |A|).), tandis que le calcul de s-no-changeΔ (q) peut être effectué en temps linéaire O(| A|)).

$$\frac{q \xrightarrow{a} q' \text{ dans } \Delta \quad q \quad P \quad \text{aucun changement}_\Delta}{(q, P) \xrightarrow{a} (q', P) \text{ en } \Delta}\ \text{snc} \qquad \frac{a \quad \Sigma \quad q \quad P \quad \text{aucun changement}_\Delta \ (q,}{P) \xrightarrow{a} (q, P) \text{ dans } \Delta}\ \text{snc}$$

$$\frac{q' \text{ dans } \Delta \quad q \quad P \quad \text{aucun changement}_\Delta}{\Delta\ (q, P) \xrightarrow{} (q, q' \text{,s-pas de changement}_\Delta\ (q)) \text{ dans}}\ \text{snc} \qquad \frac{q \quad P \quad \text{aucun changement}_\Delta}{(q, P) \xrightarrow{} \Pi \text{ dans } \Delta}\ \text{snc}$$

$$\frac{q@p \to q\ (q, q' \text{ dans } \Delta \quad q \quad P \quad \text{pas de changement}_\Delta}{P)@(p,\text{s-pas de changement}_\Delta\ (q)) \to (q, P) \text{ en } \Delta}\ \text{snc} \qquad \frac{q \quad P \quad \text{aucun changement}_\Delta}{(q, P)@\Pi \to (q, P) \text{ dans } \Delta}\ \text{snc}$$

$$\frac{\text{une} \quad \Sigma}{\Pi \xrightarrow{a} \Pi \text{ dans } \Delta}\ \text{snc} \qquad \frac{\text{vrai}}{\Pi@\Pi \to \Pi \text{ dans } \Delta}\ \text{snc} \qquad \frac{\text{vrai}}{\Pi \xrightarrow{} \Pi \text{ dans } \Delta}\ \text{snc}$$

Figure 10. Les règles de transition du SHA↓ A snc déduits de ceux du SHA A.

Lorsqu'elle est appliquée au SHA de la figure 4 pour [self::list/child::item], la construction donne le SHA↓ de la figure 11, qui est en effet égal au SHA↓ de la figure 7 jusqu'au changement de nom de l'état. Lorsqu'il est exécuté sur la haie list · list · h1 · item · h2 , comme le montre la figure 8, il n'est pas nécessaire de visiter les sous-haies h1 ou h2 puisque toutes seront atteintes à partir de l'état de projection Π .

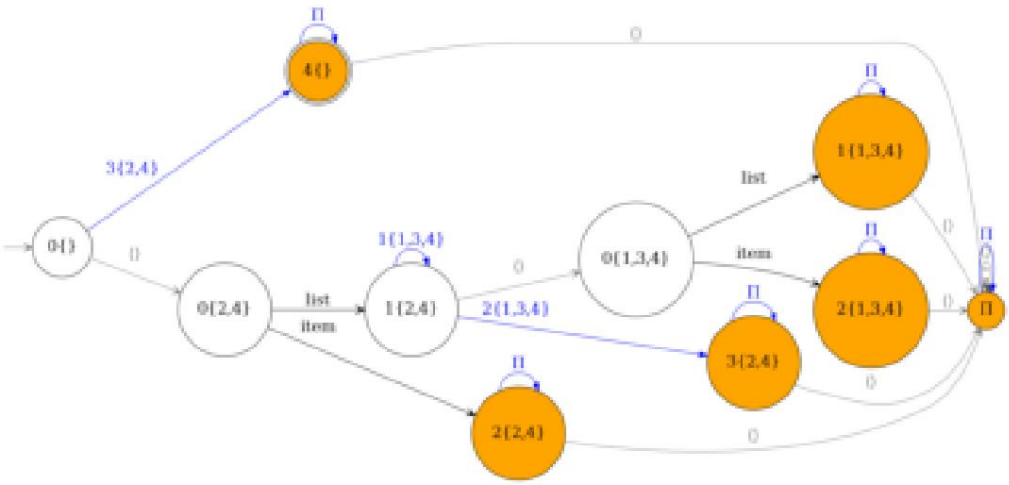


Figure 11. La projection sans changement dSHA↓ A snc construit à partir du dSHA A pour la requête [self::list/child::item] est illustrée à la figure 4. Les états de projection de sous-couverture sont colorés en orange. Les états inutiles et les transitions sortant du document de schéma sont omis. L'état Π a une transition étiquetée par la lettre générique « - », qui signifie soit une liste de lettres , soit un élément. Ce dSHA↓ est égal au dSHA↓ de la figure 7 jusqu'à l'état renommé 0 = (0, {}), 0′ = (0, {2, 4}), 0″ = 1' (0, {1, 3 , 4}), = (1, {2, 4}), 1″ = (1, {1, 3, 4}), 2′ = (2, {2, 4}), 2″ = (2, {1, 3, 4}), 3′ = (3, {2, 4}) et 4 = (4, {}).

Rappelez-vous que aucun changementΔ = {2, 3, 4, 5}.

6.2. Solidité

Nous énonçons et prouvons ensuite un résultat de solidité pour une projection sûre et sans changement.

Machine Translated by Google

Théorème 1 (Finalité de la projection sûre et sans changement.). Si un SHA A est complet pour un schéma S, alors la projection sans changement pour A préserve le langage dans ce schéma : L(A snc) ∩ S = L(A).

L'évaluateur projetant en mémoire de A snc sera plus efficace que l' évaluateur non projetant de A. Notez que la taille de A de A. Par conséquent, pour évaluer un snc peut être exponentiellement plus grand que ça dSHA A avec une projection de sous-couverture sur une couverture h donnée , nous préférerons peut-être seulement la partie nécessaire de A à la volée. snc Cette partie a une taille de O(|h|) et peut être calculée en temps O(|A| |h|) afin d'éviter le temps de prétraitement exponentiel.

6.3. Caractère

incomplet La projection sans changement peut être incomplète pour la projection de sous-couverture, de sorte que tous les préfixes qui sont fortement non pertinents pour la sous-couverture ne sont pas mappés aux états de projection de sous-couverture. Ceci est illustré par le contre-exemple dSHA A pour le filtre XPath [child::item] dans la figure 12. Sa projection sûre et snc est donné sur la figure 13. Notez que le préfixe u = sans changement A item. item a la classe S classL(A) (u) = item. item.NΣ, qui n'est pas pertinent pour la sous-couverture, donc u est fortement non pertinent pour le filtre XPath [child::item]. Néanmoins, cela conduit à l'état (2, {1, 3, 5, 6}), qui n'est pas un état de projection de sous-couverture puisque 2 {1, 3, 5, 6}. Le problème est que cet état peut toujours être modifié en (4, {1, 3, 5, 6}). Cet état est en quelque sorte équivalent par rapport au filtre mais n'est pas égal à (2, {1, 3, 5, 6}).

Un autre problème d'incomplétude doit être mentionné : la projection sûre sans changement est sensible à l'achèvement des automates comme indiqué précédemment dans l'exemple 5. En effet, un état peut appartenir à aucun changementΔ avant l'achèvement mais pas plus après l'ajout d'un puits. Néanmoins, de tels états ne changent jamais sur aucun arbre satisfaisant le schéma. Ce problème peut être appliqué, par exemple, à l'exemple dSHA dans la figure 4 pour la requête XPath [self::list/child::item]. Par conséquent , il était important de supposer uniquement l'exhaustivité du schéma pour une projection sûre et sans changement et de ne pas imposer l'exhaustivité totale.

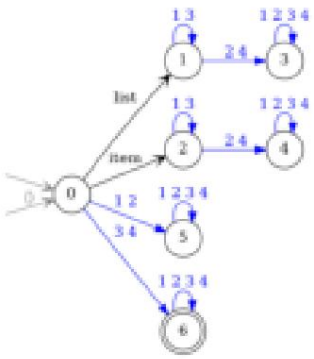


Figure 12. Un dSHA complet pour le filtre XPath [child::item]. Il s'agit d'un contre-exemple de l'exhaustivité de la projection sans changement, voir Figure 13.

Machine Translated by Google

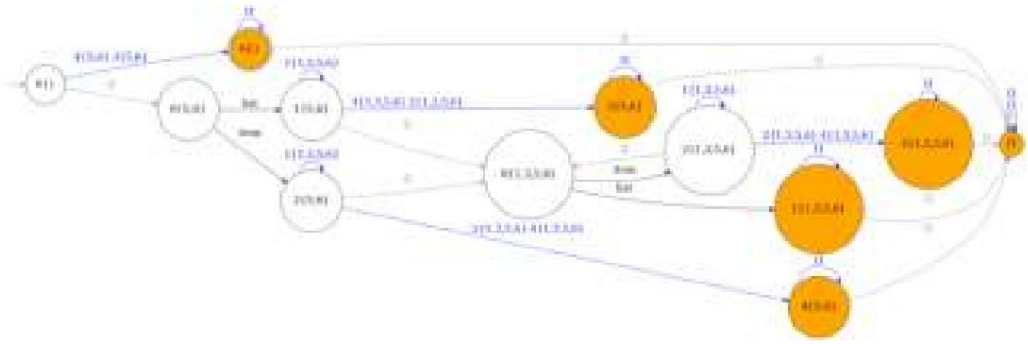


Figure 13. La projection sûre et sans changement Une ^{snc} du dSHA A dans la figure 12. Il est incomplet pour projection de sous-couverture avec un schéma doc à l'état (2, {1, 3, 5, 6}) : cet état n'est pas un état de projection de sous-couverture, même si les préfixes list · item et item · item qui y mène ne sont pas pertinents pour la sous-couverture.

7. Projection de congruence

Nous présentons l'algorithme de projection de congruence pour détecter les sous-couvertures non pertinentes pour des modèles de couverture réguliers avec des restrictions de schéma régulières. Nous prouvons que la projection de congruence est complète pour la projection de sous-couverture, résolvant ainsi le caractère incomplet de la projection sans changement. Pour cela, la projection de congruence ne peut plus ignorer le schéma, nous devons donc supposer que le schéma d'entrée est également régulier.

7.1. Idées

L'idée de départ de la projection de congruence est de résoudre le contre-exemple pour l'exhaustivité de la projection sans changement de la figure 13. Ici, un état est modifié lors de la lecture d'une sous-couverture non pertinente. Mais le changement d'état passe à un état en quelque sorte équivalent, ce n'est donc pas vraiment pertinent. L'idée est donc de détecter quand un état reste toujours équivalent plutôt que inchangé lors de la lecture d'une sous-couverture arbitraire.
Ceci est finalement réalisé par la projection de congruence de la figure 14 que nous allons construire.

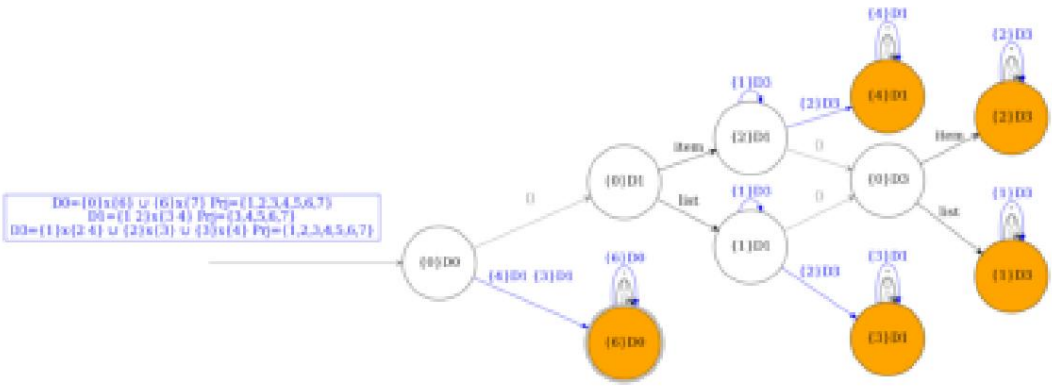


Figure 14. La projection de congruence dSHA↓ A cgr(doc) pour le contre-exemple de l'exhaustivité de la projection sans changement, c'est-à-dire le dSHA A pour le filtre [child::item] dans la figure 12 avec les états finaux du schéma FS = {0, 5, 6}. Notez que l'état {1}D3 qui est atteint sur le préfixe list · item et l' état {2}D3 qui est atteint sur le préfixe item · item ne sont pas pertinents pour la sous-couverture, et donc, les états de projection de sous-couverture.

La question évidente est la suivante : quelle relation d'équivalence choisissons-nous ? Supposons que nous voulions tester si une haie satisfaisant un schéma régulier S correspond à un modèle régulier L. Dans le cas restreint de mots sans restrictions de schéma, l'idée pourrait être d'utiliser la différence de congruence de Myhill – Nerode $\overline{L}^{\Sigma}$ mais mappé aux États. Cependant, dans le cas général, la situation devient plus complexe, étant donné que dans le cas général diff $\overline{L}^{S}$ peut ne pas être un

Machine Translated by Google

relation d'équivalence. Il se peut donc qu'il ne s'agisse pas d'une congruence, comme déjà illustré dans l'exemple 4. De plus, le traitement de l'imbrication des haies nécessitera de mettre à jour la relation considérée lors de la descente d'une haie.

7.1.1. Relations de différence entre les États

L'algorithme de projection de congruence maintiendra une relation de différence sur les états de le dSHA qui au départ sera induit par la relation de différence sur les haies diff et mis à jour à chaque descente d'une haie. L S

Définition 11. Soit (Σ, Q, ∆, _, _) un dSHA. Une relation de différence pour ∆ est une relation symétrique sur les états D Q × Q, telle que pour tout h HΣ,

$$(q \xrightarrow{h} q' \text{ par rapport à } \Delta \quad p \xrightarrow{h} p' \text{ par rapport à } \Delta \quad (q', p')) \quad D) \quad (q, p) \quad D.$$

L'ensemble de toutes les relations de différence pour ∆ est noté D∆.

On appelle un sous-ensemble Q Q compatible avec une relation de différence D D∆ si Q 2 ∩ D = .
Cela signifie qu'il n'y a pas deux états de Q qui peuvent être différents par rapport à D.

Définition 12. Soit (Σ, Q, ∆, _, _) un SHA et D une relation de différence pour ∆. Un sous-ensemble d'états Q Q est une sous-couverture non pertinente pour D par rapport à ∆ si acc∆(Q) est compatible avec D. L'ensemble de tous les sous-ensembles d'états non pertinents de sous-couverture pour D par rapport à D. ∆ est donc la suivante :

$$dPrj_\Delta(D) = \{Q \quad Q \mid acc_\Delta(Q) \quad ^2 \cap D = \quad \}.$$

Un état q Q est une sous-couverture non pertinente pour D si le singleton {q} est une sous-couverture non pertinente pour D. L'ensemble de tous les états de sous-couverture non pertinents de Q est noté Prj∆ (D).

Nous considérons la non-pertinence de sous-couverture pour des sous-ensembles d'états puisque l'algorithme de projection de congruence doit éliminer le non-déterminisme qu'il introduit par une sorte de déterminisation SHA . La détermination est nécessaire afin de reconnaître correctement les préfixes non pertinents pour la sous-couverture : tous les états uniques d'un sous-ensemble peuvent être non pertinents pour la sous-couverture alors que l'ensemble du sous-ensemble ne l'est pas.

7.1.2. Relations de moindre différence

Pour toute relation binaire R Q × Q, soit ldr∆ (R) la relation de moindre différence sur les états qui contient R.

Lemme 8. (p1, p2) ldr∆ (R) (q1, q2) R. h HΣ. p1 $\xrightarrow{h}$ q1 par rapport à ∆ p2 $\xrightarrow{h}$ q2 par rapport à ∆.

Preuve. L'ensemble {(p1, p2) Q2 | (q1, q2) R. h HΣ. p1 $\xrightarrow{h}$ q2 par rapport à ∆ q1 $\xrightarrow{h}$ q1 par rapport à ∆ p2 est de différence qui contient R, et contient donc la moindre relation de différence de ce type ldr∆ (R). Inversement, chaque paire de l'ensemble ci-dessus doit être contenue dans toute relation différentielle contenant R, et donc dans ldr∆ (R).

□

Lemme 9. Pour tout R Q × Q, la relation de différence ldr∆ (R) est la valeur du prédicat D au point le moins fixe du programme d'enregistrement de données au sol généré par les règles d'inférence suivantes :

$$\frac{p, q \quad Q}{D(p, q) :- R(p, q).} \qquad \frac{p1 \xrightarrow{a} p2 \text{ par rapport à } \Delta \; q1 \xrightarrow{a} q2 \text{ par rapport à } \Delta}{D(p1, q1) :- D(p2, q2).}$$

$$\frac{p1@p \to p2 \text{ par rapport à } \Delta \; q1@q \to q2 \text{ par rapport à } \Delta}{D(p1, q1) :- D(p2, q2).} \qquad \frac{p, q \quad Q}{D(q, p) :- D(p, q).}$$

Machine Translated by Google

Preuve. La première règle d'inférence garantit que R D. Les trois règles d'inférence suivantes caractérisent les relations de différence D DΔ. Les deuxième et troisième règles stipulent que les différences de D se propagent à rebours sur les haies h. Cela se fait de manière récursive en traitant les haies de lettres h = a par la deuxième règle et les haies d'arbres h = h exprime la symétrie des relations ' par la troisième règle. Les quatrièmes règles de différence D. Ainsi, le point le moins fixe du programme d'enregistrement de données généré par les règles ci-dessus contient R et est le la moindre relation qui satisfait tous les axiomes des relations de différence, c'est donc ldrΔ (R). □

Proposition 4. Pour tout R Q × Q, la relation de moindre différence ldrΔ (R) peut être calculée en temps $O(|A|^2)$

Preuve. La taille du programme d'enregistrement de données au sol calculant ldrΔ (R) à partir du lemme 9 est au plus quadratique en taille de A, donc son point le plus petit fixe peut être calculé dans le temps $O(|A|^2)$. □

7.2. Algorithme

Nous définissons maintenant l'algorithme de projection de congruence par un compilateur à partir de dSHA et un ensemble d'états finaux de schéma en SHA↓ s, qui, lorsqu'il est exécuté sur une haie, peut détecter tous les préfixes non pertinents de sous-couverture.

7.2.1. Entrées et sorties En

tant qu'entrées, l'algorithme de projection de congruence reçoit un dSHA A = (Σ, Q, Δ, I, F) pour le modèle régulier et un ensemble FS avec F FS Q. Le dSHA définit le langage régulier L = L(A) tandis que le schéma régulier S est reconnu par le dSHA A[F/FS] = (Σ, Q, Δ, I, FS).
Ainsi, le même automate de base définit le langage régulier et le schéma régulier en choisissant différents ensembles d'états finaux.

Exemple 10. Dans l'exemple dSHA de la figure 4 pour le filtre XPath [self::list/child::item] avec schéma doc, nous avons F = {4} et FS = {4, 5}. Dans l'automate de l'exemple de compteur [child::item] pour la projection sans changement de la figure 12, nous avons F = {6} et FS = {5, 6}.

Nous notons que, si L et S étaient donnés par des SHA indépendants, alors nous pouvons obtenir un dSHA commun A et un ensemble FS comme ci-dessus en calculant le produit des dSHA pour L et S et en le complétant. Comme remarqué dans [16], il peut être plus efficace de déterminer le SHA pour S dans une première étape, de construire le produit du SHA pour L avec le dSHA pour S dans une seconde, et de déterminer ce produit dans la troisième étape.

La projection de congruence de A par rapport à. FS maintiendra dans sa configuration actuelle un sous-ensemble d'états Q Q et une relation différence sur les états dans D DΔ. Ainsi, la projection de congruence dSHA↓ peut toujours détecter si le préfixe actuel n'est pas pertinent pour L par rapport à L. schéma S en testant si l'ensemble actuel d'états Q est une sous-couverture
sans importance pour la relation de différence actuelle D.

7.2.2. Relation de différence initiale

La relation de différence initiale sur les états D init ^{A,FS} DΔ est induit par la différence relation sur les préfixes diff $^{L}_{S}$ comme suit:

$$D^{A,FS}_{\text{init}} = \{(q, q') \mid (v, v') \in \text{diff}^{L}_{S} . \; q_0 \xrightarrow{hdg(v)} q \text{ par rapport à } \Delta \text{ et } q_0 \xrightarrow{hdg(v')} q' \text{ par rapport à } \Delta\}.$$

Le lemme suivant indique comment D peut être défini à ^{A,FS} partir de A et FS sans référence à init
les langages L = L(A) et S = L(A[F/FS]).

Lemme 10. $D^{A,FS}_{\text{init}} = \text{ldr}_\Delta (F \times (FS \setminus F)) \cap \text{acc}_\Delta(I)^2$.

Machine Translated by Google

Preuve. Pour une direction soit (p ′A,FS , p ″) D init . Alors, il existe des mots imbriqués (v ′, v ″) diff L S et un état initial q0 I, tels que q0 −−−→ p sont des mots imbriqués, ′ par rapport à ∆ et q0 , hdg(v ″) −−−→ p » par rapport à ∆. Depuis hdg(v) ′ v , v ″ l'accessibilité des haies (pv ″) diff S nécessite l'existence ′ p ″) acc∆(I) 2 suit. De plus, (v. ′, d'une haie h L HΣ, telle que hdg(v hdg (v ″) · h S \ L. Il existe donc des états q ′) · h L et ′ F et q , p » FS \ F, tel que p ′ h − → q ′ et p » h −→ q ″. En utilisant le lemme 8, cela implique que (p , p ″) ′″) ldr∆ (F × (FS \ F)).

(p . En utilisant le lemme 8, , p ″) ldr∆ (F × (FS \ F)) (nous l'avons) qu'il existe deux états tels que FS \ F et , p ″) acc∆(I) 2 propriété (p ′ F, q » h HΣ, tel que p ′ h − → q ′ par rapport à ∆ et p » h − → q » par rapport à ∆. De (p ′ , ça suit ′ qu'il y a des mots imbriqués v hdg(v ″) −−− ·. v ″ NΣ et un état initial q0 I, tel que q0 hdg(v − −−→) ′ p par rapport à ∆ et q′ 0 −→ p » par rapport à ∆. Par conséquent, (v ′, v ″) diff L S , de sorte que (q ′, q ′ ′) D A,FS init . □

Suite au lemme 10 et à la proposition 4, la relation de différence initiale D) de A et FS. A,FS initialisation peut être calculé en temps O(|A| 2

Proposition 5 (Justice de la relation différentielle initiale). Soit A = (Σ, Q, ∆, I, F) un dSHA complet , F FS Q, L = L(A) et S = L(A[F/FS]). Pour toute couverture h HΣ et état −→ q par rapport à ∆ pour certains q0 I, le mot imbriqué nw(h) est une sous-couverture non pertinente h couverture non pertinente pour q Q, de telle sorte que q0 L et S si et seulement si q est une sous-pour DA ,FS initialisation par rapport. ∆.

Preuve. Nous montrons que nw(h) est une sous-couverture pertinente pour L et S si et seulement si q est une sous-couverture pertinent pour D par rapport à D. ∆. A,FS initialisation

« » Soit nw(h) une sous-couverture pertinente pour L et S. Alors, il existe des couvertures h. h » HΣ et un mot imbriqué w NΣ, tel que

$$nw(h) \cdot nw(h') \cdot w \quad L \quad nw(h) \cdot nw(h'') \cdot w \quad S \setminus L.$$

Puisque A est déterministe et q0 −→ q par rapport à ∆ pour l'unique q0 I, il s'ensuit qu'il existe des états p Q, q F et q ′, p » ′ » FS \ F, tel que

$$q \xrightarrow{h'} p \xrightarrow{hdg(w)} q \text{ par rapport à } \Delta \quad q \xrightarrow{h''} p \xrightarrow{hdg(w)} q \text{ par rapport à } \Delta.$$

De q ′ F et q » FS \ F, il s'ensuit que (q ′, q ′ ′) D A,FS A,FS initialisation . Puisque D est une différence rapport, p′ hdg(w) −−−→ q ′ par rapport à ∆ et p » hdg(w) −−−→ q » par rapport à ∆, cela implique que (p ′, p ″) D A,FS initialisation , aussi. Par conséquent, (p.′A,FS , p ″) acc∆({q}) 2 ∩ D init , c'est-à-dire q, est pertinent pour D A,FS initialisation par rapport à. ∆.

« » Soit q une sous-couverture pertinente pour D par rapport à D. ∆. Alors, il existe un couple (p A,FS initialisation ∩ D ′, p ″) acc∆({q}) 2 init . Donc, il y a des haies h ′ et h ″, tel que

$$h' \xrightarrow{} pq \text{ par rapport à } \Delta \quad q \xrightarrow{h''} p \text{ » par rapport à } \Delta.$$

Depuis D init A,FS = ldr∆ (F × (FS \ F)) d'après le lemme 10, il existe un mot imbriqué w et a) appartient à F × (FS \ F) et paire (q ′, q ″) Q2 de sorte que soit (q ′, q ″) ou (q », q ′

$$p \xrightarrow{hdg(w)} q \text{ par rapport à } \Delta \quad p \text{ » } \xrightarrow{hdg(w)} q \text{ » par rapport à } \Delta.$$

Par conséquent, nw(h) · nw(h ′) · w dans L et nw(h) · nw(h ″) · w S \ L ou vice versa. Ce montre que nw(h) est pertinent pour L et S.

□

Machine Translated by Google

7.2.3. Mise à jour de la relation de différence

Pour toute relation de différence D D∆ et sous-ensemble d'états Q Q, nous définissons une relation binaire down∆ Q(D) Q × Q, telle que pour tous les états p1, p2 Q :

$$(p1, p2) \quad \text{down}\Delta\ Q(D) \text{ ssi } q1, q2 \quad Q.\ (q1 @ \overset{\Delta}{} p1, q2 @ \overset{\Delta}{} p2) \quad D .$$

Pour tout sous-ensemble d'états Q Q et la relation de différence D D∆, soit DQ D∆ la plus petite relation de différence qui contient down∆ Q(D) :

$$DQ = \text{ldr}\Delta\ (\text{down}\Delta\ Q(D)).$$

Lemme 11. La relation de différence DQ peut être calculée en temps O(|A| ^{2}) de Q, ∆ et D.

Preuve. La relation binaire down∆ Q(D) peut être calculée dans le temps O(|∆| La relation ^{2}) de Q, D et ∆.
de moindre différence ldr∆ (down∆ Q(D)) peut être calculée par un enregistrement de données au sol dans le temps)
O(|UNE| ^{2} à partir de down∆ Q(D) en utilisant Proposition 4. □

Exemple 11. Reconsidérez le dSHA pour le filtre XPath [self::list/child::item] dans la figure 4 avec l'ensemble des états finaux du schéma FS = {4, 5}. Puisque F = {4}, nous avons FS \ F = {5}. La relation de différence initiale A,FS qui est la fermeture symétrique sous-couverture sont Prj∆ (D0) = {1, 2, 3, 4, 5} de initialisation((0} × {1, 4, 5}) est D0 = D {(1, 4),(4, 5)}. Les états non pertinents de la puisque seul l'état 0 peut accéder à deux états dans la relation de différence D0, l'état final dans F = {4} et un état non final . c'est le schéma final dans FS \ F = {5} de certaines haies. La relation de différence down∆ {0} (D la fermeture symétrique de {1, 2} × {3}, qui est égale à la relation de différence D1 = D0{0} = ldr∆ (down∆ {0} (D0)) . Par conséquent, les états de projection sont Initialisation A,FS) est Prj∆ (D1) = {2, 3, 4, 5} puisqu'à partir des états 0 et 1 on peut toujours atteindre à la fois 1 et 3 tandis que (1, 3) D1 = . down∆ {1} (D1) est la fermeture symétrique de {(1, 2), (2, 3)} Par conséquent, les états de projection sont Prj∆ (D2) = {1, 2, 3, 4, 5}. .

A partir de l'état 1, on ne peut accéder qu'aux états 1 et 3, qui ne sont pas différents pour D2.

Les états de projection pour la relation de différence initiale contiennent tous les états qui sont sûrs pour sélection ou coffre-fort pour rejet par rapport au schéma :

Lemme 12. Tous les états dans safe∆ (F) et safe∆ (FS \ F) sont des sous-couvertures non pertinentes pour Dinit A,FS .

Nous omettons la preuve puisque le résultat n'est pas utilisé par la suite.

Étant donné un dSHA A = (Σ, Q, ∆, I, F) et un ensemble F FS Q, nous construisons maintenant la projection de congruence A cgr(S) comme le dSHA↓ suivant :

$$A\ \text{cgr}(S) = (\Sigma, Q\text{cgr}\ , \Delta\text{cgr}, \text{Je cgr}(S), \text{Fcgr}\ (S)),$$

où l'ensemble des états, des états initiaux et des états finaux est le suivant :

$$Q\text{cgr} \quad 2\ Q \times D\Delta$$

$$\text{Je cgr}(S) = \{(I, D)\} \overset{A,FS}{\text{initialisation}}$$

$$\text{Fcgr}\ (S) = (Q, \text{ré}) \mid \overset{A,FS}{\text{initialisation}} \quad Q \quad \text{dPrj}\Delta\ (D \overset{A,FS}{\text{initialisation}}) \quad Q \cap F =$$

$$Q \quad \text{dPrj}\Delta\ (D \overset{A,FS}{\text{initialisation}}) \quad \text{acc}\Delta(Q) \cap F =$$

Les règles de transition dans ∆ cgr sont fournies par les règles d'inférence de la figure 15. Pour illustrer la construction et pourquoi la détermination est nécessaire, nous reconsidérons l'exemple 3, où les restrictions de schéma ont des conséquences qui à première vue peuvent être contre-intuitives.

Machine Translated by Google

$$\frac{a \quad \Sigma\, Q \quad dPrj_{\Delta}(D)}{(Q, D)\, a{-\!\!\rightarrow}\, (Q, D) \text{ dans } \Delta \text{ cgr}} \qquad \frac{Q \quad dPrj_{\Delta}(D)}{(Q, D) {-\!\!\rightarrow} (Q, D) \text{ dans } \Delta\text{cgr}}$$

$$\frac{Q@P \rightarrow Q' \text{ dans } \Delta \quad ^{dét} \quad Q \quad dPrj_{\Delta}(D)}{(Q, D)@(Q, D) \rightarrow (Q, D) \text{ dans } \Delta\text{cgr}}$$

$$\frac{Q\, a{-\!\!\rightarrow}\, Q' \text{ dans } \Delta^{dét} \quad Q \quad dPrj_{\Delta}(D)}{(Q, D)\, a{-\!\!\rightarrow}\, (Q', D) \text{ dans } \Delta \text{ cgr}} \qquad \frac{Q \quad dPrj_{\Delta}(D)}{(Q, D) \quad {-\!\!\rightarrow} (\ \Delta, DQ) \text{ dans } \Delta \text{ cgr}}$$

$$\frac{Q@acc_{\Delta}(P) \rightarrow Q' \text{ dans } \Delta \quad ^{dét} \quad Q \quad dPrj_{\Delta}(D)\ P \quad dPrj_{\Delta}(DQ)}{(Q, D)@(P, DQ) \rightarrow (Q', D) \text{ dans } \Delta \text{ cgr}}$$

$$\frac{Q@P \rightarrow Q' \text{ dans } \Delta \quad ^{dét} \quad Q \quad dPrj_{\Delta}(D)\ P \quad dPrj_{\Delta}(DQ)}{(Q, D)@(P, DQ) \rightarrow (Q', D) \text{ dans } \Delta \text{ cgr}}$$

Figure 15. Les règles de transitions Δ cgr de la projection de congruence A cgr(S)·

Exemple 12 (Une détermination lors de la projection de congruence est nécessaire.). Reconsidérez l'exemple 3 avec la signature Σ = {a}, le modèle L = { a } et le schéma S = L { · a}. Ce langage peut être défini par le dSHA A de la figure 16. Le schéma S peut être défini par le même automate mais avec les états finaux du schéma FS = {3, 5} et F = {3}. Le dSHA A cgr(S) produit par projection de congruence est illustré à la figure 17. La couverture unique a du langage L = L(A) est acceptée dans l'état ({3, 4}, D0), où D0= { (3, 5), (5, 3)}. L'unique haie · a dans S \ L est rejetée : après lecture de l'arbre , l'automate passe à l'état ({3, 4}, D0) où il se dirige vers un puits saillant ({5}, D0) lorsque en lisant la lettre suivante a. On note que toute haie dans Σ est acceptée dans l'état ({3, 4}, D0) par A cgr(S) même si seule la haie unique a appartient à L. Ceci est solide étant donné que les haies dans (ε + a · a · a) n'appartiennent de toute façon pas au schéma S. On remarque que ({3, 4}, D0) ne peut pas être un état de projection puisque le run on hedge · a doit continuer jusqu'au puits ({5}, D0) ;par conséquent, l'état ({3, 4}, D0) est pertinent pour la sous-couverture.

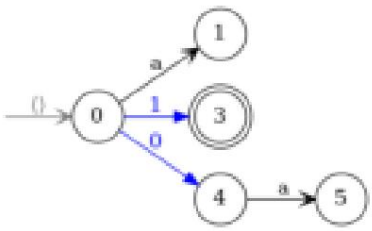


Figure 16. Un dSHA A pour L(A) = { a } et les états finaux du schéma FS = {3, 5} pour le schéma S = L(A) { · a}.

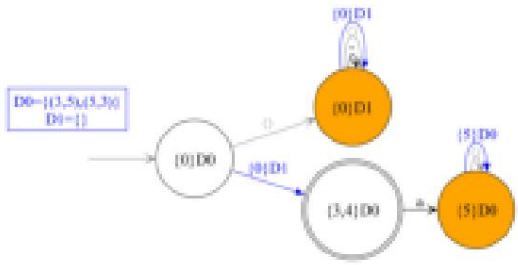


Figure 17. La projection de congruence A cgr(S) pour le dSHA A et les états finaux du schéma FS dans la figure 16.

Machine Translated by Google

Notez que les deux états individuels 3 et 4 ne sont pas pertinents pour D0 tandis que le sous-ensemble les deux états {3, 4} sont une sous-couverture pertinente pour D0 puisque (3, 5) $\text{acc}_\Delta(\{3, 4\})$ la construction déterministe est en effet nécessaire pour décider de la non-pertinence de la sous-couverture dans des cas comme dans cet exemple. Notez également que l'état 0 n'est pas pertinent pour la sous-couverture pour D1= . C'est revanche, l'acceptation des couvertures de la très bien puisque le par A cgr(S) ne dépend en effet pas de h. En forme h · h dépend de h' [illegible] le sous-ensemble d'états {3, 4} ne peut pas être une sous-couverture non pertinente pour D0.

Enfin, discutons de la manière dont les règles de transition de dSHA↓ A cgr(S) sont déduites. La règle de transition la plus intéressante est ({0}, D0)@({0}, D1) → ({3, 4}, D0) dans Δ cgr. Il est produit par la règle d'inférence suivante, où Q = P = {0}, $\text{acc}_\Delta(P) = \{0, 1, 3, 4, 5\}$, Q = {3, 4}, D = D0 et DQ = D1 :

$$\frac{Q@\text{acc}_\Delta(P) \to Q' \text{ point dans } \Delta \quad Q \quad \text{dPrj}_\Delta(D)\ P \quad \text{dPrj}_\Delta(DQ)}{(Q, D) @ (P, DQ) \to (Q', D) \text{ en } \Delta\text{cgr}}$$

Cette règle stipule que si P est une sous-couverture non pertinente pour DQ, alors tous les états accessibles depuis P doivent être essayés puisque tous pourraient être atteints par une sous-couverture différente qui a été projetée. En conséquence, on ne peut pas savoir, en raison de la projection de sous-couverture, dans quel état de Q on doit aller. Afin de rester déterministe, nous entrons donc dans tous les états possibles, c'est-à-dire que dans le sous-ensemble Q se . Dans l'exemple, ces trouvent tous les états qui peuvent être atteints lors de la lecture d'une haie dans le motif { h | h HΣ}, étant donné que les sous-couvertures h Σ ne sont pas inspectées avec la projection de sous-couverture.

Exemple 13 (filtre XPath [self::list/child::item] avec schéma doc.). La projection de congruence de l'exemple dSHA de la figure 4 est donnée dans la figure 18. Ce dSHA est similaire à la projection sans danger de la figure 7, sauf que l'état inutile 2 est supprimé et que les quatre états de projection bouclent désormais sur eux-mêmes. plutôt que d'aller dans un état de boucle partagée Π. Il convient également de noter que seuls des ensembles d'états singleton y sont utilisés, aucune détermination n'est donc nécessaire. Comme nous le verrons, cela est typique des expériences sur des requêtes XPath classiques où des sous-ensembles plus grands se produisent rarement.

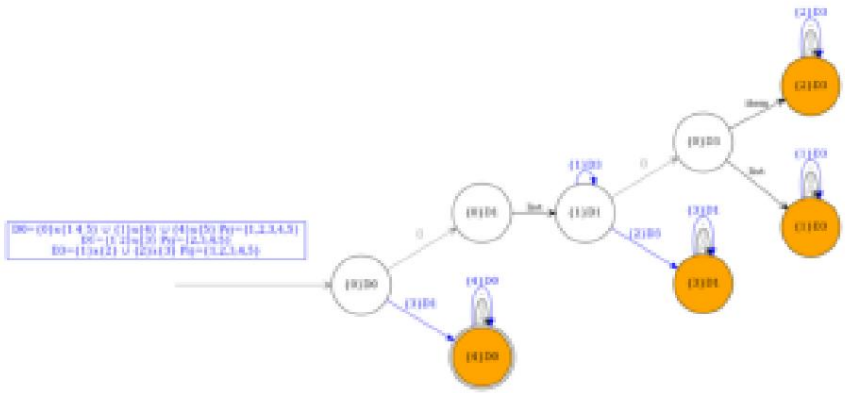


Figure 18. La projection de congruence A cgr(doc) du dSHA A dans la figure 4 pour le filtre XPath [self::list/child::item] avec schéma doc. Les états non pertinents de la sous-couverture sont colorés en orange. Le trait de soulignement représente n'importe quelle étiquette, qu'il s'agisse d'une liste ou d'un élément.

Exemple 14 (Contre-exemple pour l'exhaustivité de la projection sans changement.). Reconsidérez le contre-exemple pour l'exhaustivité de la projection sans changement, c'est-à-dire le dSHA de la figure 12 pour la requête XPath [child::item] avec les états finaux du schéma FS = {0, 5, 6}. Sa projection de congruence est représentée sur la figure 14. Nous notons que le préfixe item. item conduit à l'état {2}D3, qui est un état de projection de sous-couverture puisque 2 est une sous-couverture non pertinente pour D3. En particulier, notez que $\text{acc}_\Delta(\{2\}) = \{2, 4\}$ donc il n'y a pas deux états accessibles à partir de 2 qui soient différents dans D3. Cela signifie que l'état 2 peut toujours être changé en 4 mais qu'il ne devient alors pas différent par rapport à D3. Cela résout le problème d'incomplétude avec la projection sûre sans changement sur cet exemple.

Machine Translated by Google

La première propriété des états (Q, D) attribuée par la projection de congruence est que Q est compatible avec D. Cela signifie qu'il n'y a pas deux états dans Q qui soient différents par rapport à D. Intuitivement, chaque état de Q est aussi bon les uns que les autres, sauf pour sortir du schéma. Ainsi, si (Q, D) est attribué à un préfixe et qu'un suffixe mène d'un état de Q à F, alors il ne peut pas conduire à FS \ F à partir d'un autre état de Q.

Lemme 13. Si l'exécution partielle de A cgr(S) attribue un état (Q, D) à un préfixe, alors Q est compatible avec D.

Nous omettons la preuve puisque ce résultat instructif ne sera pas utilisé directement par la suite. Néanmoins, la compatibilité jouera un rôle important dans la preuve de solidité.

7.3. Solidité

Nous adaptons ensuite le résultat de solidité et la preuve de la projection sans changement à la projection de congruence.

Théorème 2 (Finalité de la projection de congruence.). Pour tout dSHA A = (Σ, Q, Δ, I, F) et ensemble F FS Q, la projection de congruence de A par rapport à FS préserve le langage de A dans le schéma S = L(A[F/FS]), c'est à dire,

$$L(A \text{ cgr}(S)) \cap S = L(A)$$

.

On note que S est un schéma pour A puisque F FS. De plus, A est un schéma complet pour S puisque A est déterministe et S = L(A[F/FS]).

Preuve. La preuve de l'inclusion L(A) L(A cgr(S)) ∩ S sera basée sur la formule suivante deux affirmations :

Réclamation 2.1a. Pour tout h HΣ, D DΔ et Q dPrjΔ (D), (Q, D) −→ (Q, D) par rapport à Δ cgr .

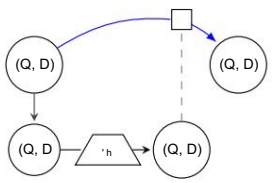
Preuve. Nous prouvons la revendication 2.1a par récurrence sur la structure de h. Supposons que Q dPrjΔ (D).

Cas h = h′ . L'hypothèse d'induction appliquée à h′ donne (Q, D) −→ (Q, D) par rapport à Δ cgr .

Nous pouvons ainsi appliquer les deux règles d'inférence suivantes :

$$\frac{Q \quad \text{dPrj}\Delta(D)}{(Q, D) \longrightarrow (Q, D) \text{ dans } \Delta\text{cgr}} \qquad \frac{Q \quad \text{dPrj}\Delta(D)}{(Q, D))@(Q, D) \rightarrow (Q, D) \text{ dans } \Delta\text{cgr}}$$

afin de fermer le diagramme suivant par rapport à Δ cgr :



Cela prouve (Q, D) −→ (Q, D) par rapport à Δ cgr, comme l'exige la revendication.

Cas h = a. Puisque q dPrjΔ (D), nous pouvons appliquer la règle d'inférence :

$$\frac{a \quad \Sigma\, Q \quad \text{dPrj}\Delta(D)}{(Q, D)\, a \longrightarrow (Q, D) \text{ dans } \Delta \text{ cgr}}$$

Cela prouve le cas de cette affirmation.

Machine Translated by Google

Cas h = ε. On a trivialement (Q, D) ε−→ (Q, D) par rapport à Δ cgr·
Cas h = h ′ · h″. Par l'hypothèse d'induction appliquée à h ′ et h ", on obtient ce qui suit :

(Q, D) $\xrightarrow{h'}$ (Q, D) et (Q, D) Δ cgr $\xrightarrow{h''}$ (Q, D) par rapport à Δ cgr. Donc (Q, D) $\xrightarrow{h'\cdot h''}$ (Q, D) par rapport .

Ceci termine la preuve de la revendication 2.1a. L'affirmation suivante est la clé de la preuve de solidité. □

Pour toute relation de différence D, on définit une relation binaire depa (D) 2 Q × 2 Q, telle que que, pour deux sous-ensembles d'états Q ′, Q″ Q,

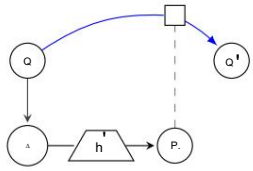
(Q ′, Q ″) depa (D) ⇔ Q ′ dPrjΔ (D) Q ′ Q″ (1
Q ′ dPrjΔ (D) (Q ′ accΔ(Q ″) Q ″ une) dPrjΔ (D)) (2 une)

De plus, notons que QQ ′) ″ dPrjΔ (D) Q ′ accΔ(Q ″) implique Q ′ dPrjΔ (D), et donc, (Q ′, depa (D).

Réclamation 2.2a. Soit h HΣ une haie, Q, Q ′ Q sous-ensembles d'états, et D DΔ une différence relation. Si Q $\xrightarrow{h}$ ′ par rapport à Δ det, alors il existe Q″ Q, tel que (Q, D) −→ (Q $\xrightarrow{h}$ ″, D) par rapport à Δ cgr et (Q ′, QQ ″) depa (D).

Preuve. Si Q dPrjΔ (D), alors (Q, D) $\xrightarrow{h}$ (Q, D) par rapport à Δ cgr d'après la revendication″ = Q. Nous aient alors Q ′ et accΔ(Q ″) et Q ″ 2.1a. Soit Q dPrjΔ (D), et donc, Q dPrjΔ (D) pour que (1a) (2a) de (QQ ″) depa (D). Il suffit donc de prouver cette affirmation sous l'hypothèse que Q dPrjΔ (D). La preuve se fait par récurrence sur la structure de h.

Cas h = h ′ h . L'hypothèse Q −→ Q ′ det montre qu'il existe un sous-ensemble d'états par rapport à Δ P Q fermant le schéma suivant :



En particulier, nous avons Q@P → Q ′ en Δ dét. Soit D′ = DQ. Puisque Q dPrjΔ (D), on qui peut déduire ce qui suit :

$$\frac{Q \quad dPrj_{\Delta}(D)}{(Q, D) \quad \xrightarrow{} (\Delta, D') \text{ dans } \Delta \text{ cgr}}$$

Il faut distinguer deux cas selon que Δ appartient à dPrjΔ (D′) ou non.

Sous-cas Δ dPrjΔ (D′). L'hypothèse de récurrence appliquée à h ′ donne l'existence de P ′ Q, telle que (P, P ′) depa (D′) et

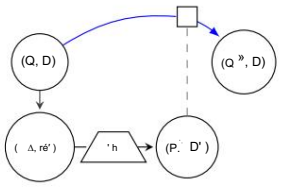
(Δ , D ′) $\xrightarrow{h'}$ (P ′ , D ′) par rapport à Δ cgr .

Sous-sous-cas P dPrjΔ (D′). Puisque (P, P ′) depa (D′), il s'ensuit que P P dans Δ ′ . Par ′ dPrjΔ (D′). Soit Q ″ Q soit tel que Q@P ′ → Q ″ det. Nous conséquent, P peut alors appliquer la règle d'inférence suivante :

$$\frac{Q@P' \to Q'' \text{ point dans } \Delta \quad Q \quad dPrj_{\Delta}(D) \; P' \quad dPrj_{\Delta}(D')}{(Q, D)@(P', D') \to (Q'', D) \text{ en } \Delta cgr}$$

Machine Translated by Google

On peut donc clôturer le diagramme comme suit :

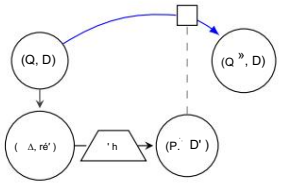


Cela montre que (Q, D) h −→ (Q », D) par rapport à Δ cgr. Puisque ', Q@P → Q ', et Q@P → Q » P P dans Δ det, il s'ensuit que Q Q ", et donc, (Q , Q ") depa (D). Cela montre la réclamation dans ce cas.

Sous-sous-cas P dPrjΔ (D′). Depuis (P, P ') depa (D′), il s'ensuit que P dPrjΔ accΔ(P ′) et règle (D′). Par conséquent, nous pouvons appliquer la déduction suivante P pour certains Q » HΣ :

$$\frac{Q@acc_\Delta(P') \to Q'' \text{ point dans } \Delta \quad Q'' \quad dPrj_\Delta(D) \; P' \quad dPrj_\Delta(D')}{(Q, D)@(P' D') \to (Q'', D) \text{ en } \Delta cgr}$$

On peut donc clôturer le diagramme comme suit :

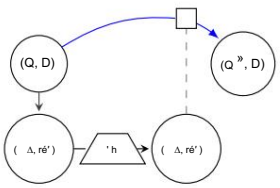


Cela montre que (Q, D) h −→ (Q et », D) par rapport à Δ cgr. Puisque P, accΔ(P ′), il → QQ ") depa (D), Q@accΔ(P ′) → Q from Q@P » dans Δdét que Q s'ensuit Q '. Ainsi, (Q , donc l'affirmation est également valable dans ce cas.

Sous-cas Δ PrjΔ (D′). La revendication 2.1a montre que (Δ, D′) −→ (h' Δ, D′) par rapport à Δ cgr. Soit » tel que Q@accΔ(Δ) → Q″ dans Δ det. Nous pouvons appliquer les déductions suivantes Règle Q ence :

$$\frac{Q@acc_\Delta(\Delta) \to Q'' \text{ point dans } \Delta \quad Q'' \quad dPrj_\Delta(D) \quad \Delta \quad Prj_\Delta(D')}{(Q, D)@(\Delta, D') \to (Q'', D) \text{ en } \Delta cgr}$$

On peut ainsi clôturer le diagramme ci-dessous comme suit par rapport à Δ cgr :



Cela montre que (Q, D) h −→ (Q », D) par rapport à Δ cgr. Puisque P accΔ(Δ), il → Q (QQ ") depa et Q@accΔ(Δ) → Q de Q@P » en Δ s'ensuit que Q Q ' ', et donc, (D). ,

Cas h = a. Puisque Q dPrjΔ (D), nous pouvons appliquer la règle d'inférence :

$$\frac{Q \; a \!\!\!\

Machine Translated by Google

Cas h = h1 · h2. Puisque Q h2 Q1 $\xrightarrow{h}$ Q' h1 par rapport à Δ det, il existe Q1 Q, tel que Q −→ Q1 et par rapport à Δ det. −→ Q tel ' Par l'hypothèse de récurrence appliquée à h1 il existe Q ' 1 Q , que (Q, D) −→ (Q h1 ' 1 , D) par rapport à Δ cgr et (Q1, Q ' 1) dépa (D).

Sous-cas Q1 dPrjΔ (D). Depuis (T1, Q ' 1) depa (D), il s'ensuit que Q1 accΔ(Q ′ 1) et Q ' 1 dPrjΔ (D). De plus, la revendication 2.1a montre que (Q accΔ(Q1) ' 1 , D) $\xrightarrow{h2}$ (Q ' 1 , D), et donc, (Q, D) −→ (Q $\xrightarrow{h}$ accΔ(Q ' 1 , D) par rapport à Δ cgr. Depuis Q ' et Q1 accΔ(Q ′ 1), il s'ensuit que Q ' ′ 1). Par conséquent, (Q Question 1 , ') depa (D) et (Q, D) $\xrightarrow{h}$ (Q ' 1 , D).

Sous-cas Q1 dPrjΔ (D). Dans ce cas, nous pouvons appliquer l'hypothèse d'induction à Q1 $\xrightarrow{h2}$ Q ' par rapport à Δ det, montrant qu'il existe Q » Q, tel que (Q ' 1 , D) $\xrightarrow{h2}$ (Q », D) par rapport à Δ cgr et ', Q ") depa (D). D'où (Q, D) −→ (Q $\xrightarrow{h}$ », D) par rapport à Δ cgr et (Q ', (QQ ") depa (D).

Ceci termine la preuve de la revendication 2.2a. □

Preuve d'inclusion L(A) L(A cgr(S)) ∩S. Puisque S est un schéma pour A, on a L(A) S, il suffit donc de montrer L(A) L(A cgr(S)). Soit h L(A). Alors, il existe q0 I et −→ q par rapport à Δ. Soit Q = {q}. Puisque A est déterministe, il s'ensuit que I q F, tel que q0 −→ Q par rapport à Δ det. En par rapport à Δ cgr. De plus, (I, D Initialisation) $\xrightarrow{h}$ (Q', D initialisation) Icgr (S) . Puisque Q contient la revendication 2.2a, cela implique l'existence d'un sous-ensemble d'états Q Q, tel que A,FS A,FS) ' (Q, Q ' A,FS) depa (D) et (I, D) initialisation

Pour prouver h L(A cgr(S)), il suffit donc de montrer que (Q distingue deux ', D $\overset{A,FS}{init}$) Fcgr (S) . Nous cas :

Cas Q ' dPrjΔ (D $\overset{A,}{\text{initialisation}}$ FS). Condition (1a) de (Q, Q ') depa (D), montre que Q Q) F ' de sorte que q Q '. Puisque q F, il s'ensuit que Q ' ∩ F = . Ainsi, (Q ', D $\overset{A,FS}{init}$ cgr(S) , de sorte que h L(A cgr(S)).

Cas Q ' dPrjΔ (D $\overset{A,}{\text{initialisation}}$ FS). Condition (2a) de (Q, Q A,FS D init ') depa (D) donne Q accΔ(Q ′). Ainsi, q accΔ(Q ′) ∩ F tel que (Q ',) Fcgr (S) , et donc, h L(A cgr(S)).

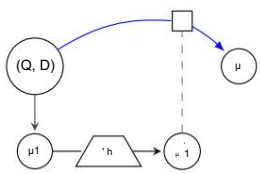
Ceci termine la preuve de la première inclusion.

Nous voulons ensuite montrer l'inclusion inverse L(A cgr(S)) ∩ S L(A). Ce sera finalement découlent des deux affirmations suivantes.

Réclamation 2.1b. Pour toute couverture h HΣ, relation de différence D DΔ, état de projection Q dPrjΔ (D) et état μ Qcgr : si (Q, D) $\xrightarrow{h}$ μ par rapport à Δ cgr, alors μ = (Q, D) .

Preuve. Par récurrence sur la structure de h HΣ, supposons que (Q, D) $\xrightarrow{h}$ μ par rapport à Δ cgr :

Cas h = h ' . Il doit exister des états μ1, μ ' 1 Qcgr fermant le schéma suivant :



Puisque Q dPrjΔ (D), la règle suivante doit avoir été appliquée pour déduire (Q, D) μ1 par −→ rapport à Δ cgr :

$$\frac{Q \quad dPrj\Delta(D)}{(Q, D) \longrightarrow (Q, D) \text{ dans } \Delta cgr}$$

Machine Translated by Google

Donc μ1 = (Q, D). L'hypothèse d'induction s'applique également à (Q, D) par rapport à Δ cgr −→ μ 1 = (Q, D) . Ainsi, μ doit avoir été déduit en appliquant la règle : montre que μ

$$\frac{Q \quad dPrj_{\Delta}(D)}{(Q, D)@(Q, D) \rightarrow (Q, D) \text{ dans } \Delta cgr}$$

Par conséquent, μ = (Q, D), comme requis.

Cas h = a. La règle suivante doit être appliquée :

$$\frac{a \quad \Sigma \, Q \quad dPrj_{\Delta}(D)}{(Q, D) \xrightarrow{a} (Q, D) \text{ dans } \Delta \text{ cgr}}$$

Par conséquent, μ = (Q, D).

Cas h = ε. Évident.

h = h1 · h2. Il doit exister des μ1, tels que (Q, D) −→ μ par rapport à Δ cgr. En utilisant −→ μ1 dans l'hypothèse d'induction appliquée à h1, avons μ1 = (Q, D). On peut ainsi appliquer l' hypothèse d'induction à h2 pour obtenir μ2 = (Q, D).

Ceci termine la preuve de la revendication 2.1b. □

Nous avons ensuite besoin d'un inverse de la revendication 2.2a. Pour relier les exécutions A cgr aux exécutions de A, nous définissons pour toute relation de différence D une autre relation binaire depb (D) 2 Q × 2 Q, telle que pour deux sous-ensembles d'états Q Q,

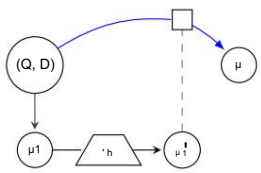
$$(Q', Q'') \quad depb(D) \quad \begin{array}{l} (Q' \times Q'') \cap D = \\ Q' \quad dPrj_{\Delta}(D) \quad Q \quad Q'' \\ Q' \quad Q \quad acc_{\Delta}(Q') \end{array} \quad \begin{array}{l} (0^{b}) \\ (1 \quad dPrj_{\Delta}(D) \quad {}^{b}) \\ (2 \quad {}^{b}) \end{array}$$

Réclamation 2.2b. Soit Q Q un sous-ensemble d'états compatible avec une relation différentielle D DΔ. Pour toute couverture h HΣ et état μ Qcgr avec (Q, D) −→ μ par rapport à Δ cgr, il existe une paire de sous-ensembles d'états (Q , Q ") depb (D), tel que μ = (Q , D) et Q −→ Q qui est vrai. ce .

Preuve. Si Q dPrjΔ (D), alors la revendication 2.1b montre que μ = (Q, D). Soit Q par = Q et Q être le sous-ensemble unique d'états, tel que Q −→ Q rapport à Δ det . , Q ") depb (D). Puisque Q −→ Q ", nous avons Q accΔ(Q ′), donc la condition (2b) est vérifiée. La condition (1b) est trivialement vérifiée puisque Q dPrjΔ (D), Pour la condition (0b), notons que Q dPrjΔ (D) implique accΔ(Q) 2 accΔ(Q) 2 × Q ") ∩ D = puisque ∩ D = . De plus, Q requis. de sorte que (Q

On peut donc supposer que Q dPrjΔ (D). La preuve se fait par récurrence sur la structure de h HΣ. On distingue toutes les formes possibles de couvertures h HΣ :

Cas h = h . Par définition de (Q, P) −→ μ par rapport à Δ cgr, il doit exister μ1, μ 1 Qcgr, tel que le schéma suivant peut être fermé :



Machine Translated by Google

Puisque Q dPrjΔ (D), la règle d'inférence suivante a été appliquée pour déduire (Q, D) −→ μ1 par rapport à Δ cgr, où D′ = DQ :

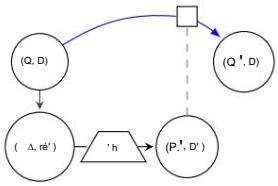
$$\frac{Q \quad dPrj_{\Delta}(D)}{(Q, D) \quad \longrightarrow (\ \Delta, D') \text{ dans } \Delta \text{ cgr}}$$

Par conséquent, μ1 = (Δ, D′). Si Δ dPrjΔ (D), alors l'hypothèse d'induction appliquée à (Δ, D′) $\xrightarrow{h} \mu_1'$ par rapport à Δ cgr montre qu'il existe (P', P'') depb (D′), tel que μ_1' (P μ ′, D′) et Δ = $\xrightarrow{h} P''$ par rapport à Δ dét. Sinon, on peut conclure la même chose que nous argumenté au début. dPrjΔ

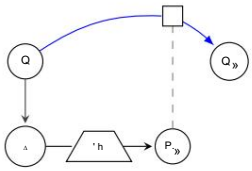
Sous-cas P ′ (D′). La règle de transition (Q, D)@μ1 → μ doit être déduite comme suit pour Q ′ Q :

$$\frac{Q@acc_{\Delta}(P') \to Q' \quad \text{point dans } \Delta \quad Q \quad dPrj_{\Delta}(D) \quad P' \quad dPrj_{\Delta}(D)}{(Q, D)@(P', D') \to (Q', D) \text{ en } \Delta cgr}$$

Cela montre que μ = (Q ′, D). On a donc le schéma suivant :



Soit Q » être le sous-ensemble unique d'états, tel que Q@P» → Q » par rapport à Δ dét. Nous pouvons alors fermez le diagramme suivant :



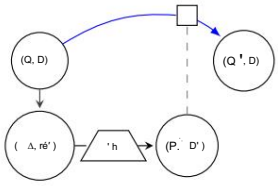
(PD = , ′, P ″) depb (D′), il s'ensuit (P From ′ P ″) ∩ D′ = , et donc, (QQ ″) depb′ × Q ″) ∩ × c'est à dire, la condition (0b) de (Q tion ′, (D). Puisque P dPrjΔ (D′), condi- accΔ(P ′). De (2b) de (PP ′) depb (D′) donne P et Q @accΔ(P», plus, Q@P dans Δ det, ce qui donne Q » → Q » ′) → Q (2b) de (QQ ″) ′ tel que les conditions (1b) et » QQ ″) depb (D). En résumé, nous depb (D) suit donc (Q h). ′,

montrer que μ = (Q ′, D), Q −−→ Q » par rapport à Δ det, et (Q′, Q ″) depb (D), selon les besoins par la

Sous-cas P revendication. dPrjΔ (D′). La règle de transition (Q, D)@μ1 → μ doit donc être déduite comme suit pour Q ′ Q :

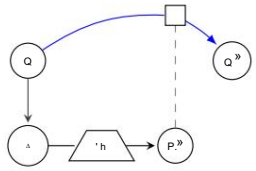
$$\frac{Q@P' \to Q' \quad \text{point dans } \Delta \quad Q \quad dPrj_{\Delta}(D) \quad P' \quad dPrj_{\Delta}(D)}{(Q, D)@(P', D') \to (Q', D) \text{ en } \Delta cgr}$$

Cela montre que μ = (Q ′, D), et on peut fermer le schéma suivant :



Machine Translated by Google

Condition (0b) de (P ′, P ″) depb (D′) donne (P × P ′) ∩ D′ = . Laissez Q par » être le sous-ensemble unique d'états, tel que Q@P » → Q » rapport. Δ dét. Puisque D′ downΔ Q(D) et (P ′ × P ″) ∩ D′ = , il découle de Q@P → Q ∩ Q ″) ∩ D′ = . ′ et Q@P Q » → Q » par rapport. Δ détermine ′ Autrement dit, la condition (0b) de (QP ″) depb (D′) est P ′, ″) depb (D). Depuis P ′ que (Q dPrjΔ (D′), condition (1b) de (P » P ′. Depuis Q@P» → Q » et Q@P ′ → Q ′, il s'ensuit donc que Q » ′. D'où les conditions (1b) et (2b) de (Q ′, Q ″) depb (D) sont valides, de sorte que (Q ′,QQ ′) depb (D) tient. En outre,



.

Cela montre que Q μ ′ h ′ −→» Q par rapport à Δ. Les deux autres exigences de la réclamation, = (Q ′, D) et (Q Cas′, Q ″) depb (D), ont été montrés plus tôt.

h = a. Puisque Q dPrjΔ (D), la règle d'inférence suivante doit être utilisée :

$$\frac{Q \xrightarrow{a} Q' \text{ dans } \Delta \quad Q \quad dPrj\Delta\ (D)}{(Q, D) \xrightarrow{a} (Q', D) \text{ en } \Delta cgr}$$

Donc μ = (Q ′, D), Q a−→ Q ′ par rapport à Δ dét. De plus, puisque Q est compatible avec D, D est une relation différentielle, Q ′ est également compatible avec D. Par conséquent, (Q ′ × Q ′et) ∩ D = , montrant la condition (0b) de (Q) depb (D). Trivialement, les conditions (1b′) et (2b′) de (Q ′, Q ′) depb (D) est également valable. Par conséquent, (Q ′, Q ′) accΔ(Q ′) donc) (Q Cas h = ε. Évident. depb (D).

Cas h = h1 · h2. Puisque (Q, D) −→ μ par rapport à. Δ cgr, il existe des μ1 Qcgr, tels que h1 h2 (Q, D) −→ μ1 et μ1 −→ μ par rapport à Δ cgr.
Nous pouvons appliquer l'hypothèse d'induction à h1.

Il existe donc des sous-ensembles d'états Q1, ′1 h1 Q, tel que μ1 = (Q1, D), Q −→ Q ′à 1 par rapport Q Δ et (Q1, Q ′1) depb (D). Nous distinguons deux cas :

h2 Sous-cas Q1 dPrjΔ (D). Puisque μ1 = (Q1, D) −→ μ par rapport à Δ cgr, la revendication 2.1b montre −→ que μ = (Q1, D) donc que (Q1, D) h2 (Q1, D) par rapport à. Δcgr . Condition (2b) de (Q1, Q ′1) depb (D) et Q1 dPrjΔ (D) impliquent que Q ′1 accΔ(Q1). Soit Q2 le sous-ensemble unique d'états, tel que Q de sorte ′1 h2 −→ Q2 par rapport à. Δ dét. Alors, Q2 accΔ(Q ′1) que Q2 accΔ(Q1), et donc la condition (2b) de (Q1, Q2) depb (D) est vérifiée.
La condition (1b) de (Q1, Q2) depb (D) est triviale puisque Q1 dPrjΔ (D). Puisque Q1 dPrjΔ (D) et Q1 × Q2 accΔ(Q) 2 il s'ensuit que Q1 × Q2 ∩ D = donc la condition (0b) de (Q1, Q2) depb (D) est également vérifiée. Donc Q (Q1, Q2) depb (D). h Q2 par rapport à Δ det et −→

Sous-cas Q1 dPrjΔ (D). Puisque Q1 h1 −→ Q ′1 et Q1 est compatible avec la différence est relation D, il s'ensuit que Q ′1 également compatible avec D. On peut ainsi appliquer l' hypothèse d'induction h2 à (Q1, D) −→ μ, montrant l'existence de sous-ensembles d'états h2) depb (D), tels que (Q2, Q ′2 = (Q2, D) et Q1 −→ Q μ ′2 par rapport à Δ. Nous avons donc Q− h → Q ′2 par rapport. Δ et (Q2, Q ′2) depb (D), selon les besoins.

Ceci termine la preuve de la revendication 2.2b. □

Preuve d'inclusion L(A cgr(S)) ∩ S L(A). Soit h L(A cgr(S)) ∩ S. Alors, il existe un dernier sous-ensemble d'états Q F cgr(S) , A,FS tel que (I, D) initialisation h −→ (Q, D) par rapport à Δ cgr. Puisque je suis initialisation A,FS

Machine Translated by Google

A,FS un singleton ou vide, il est compatible avec D init , La revendication 2.2b montre qu'il existe a) depb (D sous-ensemble d'états Q Q, tel que je $\xrightarrow{h}$ Q par rapport. Δdet et (Q, Q $\underset{\text{initialisation}}{A,}$FS). Condition (0b) de (Q, Q) depb (D $\underset{\text{initialisation}}{A,FS}$) montre que (Q, Q FS) ∩D $\underset{\text{initialisation}}{A,}$ = . Puisque h S, il s'ensuit que Q ∩FS = . Le déterminisme de A montre que Q = {q '}. FS, tel c'est un singleton. Il existe donc un état q que Q

Cas Q dPrjΔ (D $\underset{\text{initialisation}}{A,FS}$). La condition (1b) montre que Q Q pour que q Q. Puisque Q F cgr , on ait Q ∩ F = . Puisque Q est compatible avec D Q, et Q ∩ F =L(A)il init , q suit que q F, et donc h FS \ F. Puisque q FS, cela implique q

Cas Q dPrjΔ (D accΔ(Q). $\underset{\text{initialisation}}{A,FS}$). La condition (2b) montre que Q′ accΔ(Q) donc que q Puisque Q F cgr, nous avons accΔ(Q) ∩ F = . Soit q accΔ(Q) ∩ F arbitraire. Puisque Q dPrjΔ (D $\underset{\text{initialisation}}{A,FS}$) et (q , q) accΔ(Q) 2 , il s'ensuit que (q, q FS) $\underset{\text{init}}{A,D}$. De plus, FS donc q F de sorte que qh FS \ F. En combinaison avec q FS, cela implique q L(A).

Ceci termine la preuve de l'inclusion inverse, et donc de L(A) = L(A cgr(S)) ∩ S. □

7.4. Complétude

Nous montrons ensuite l'exhaustivité de la projection de congruence pour la projection de sous-couverture selon la définition 10. Soit A = (Σ, Q, Δ, I, F) un dSHA complet et F FS Q. L'automate A définit le modèle régulier L = L(A) avec les états finaux du schéma dans FS le schéma régulier S = L(A[F/FS]). Nous montrons d'abord que tous les états de relations différentielles non pertinents pour la sous-couverture sont des états de projection de sous-couverture A cgr(S) selon la définition 7.

Lemme 14. Si Q dPrjΔ (D), alors (Q, D) est un état de projection de sous-couverture de Δ cgr avec témoin (Q, D).

Preuve. Nous supposons que q PrjΔ (D) et devons montrer que (Q, D) est un état de projection de sous-couverture de Δ cgr. Nous devons montrer que toutes les règles de transition commençant par (Q, D) sont autorisées pour un état de projection de sous-couverture (Q, D) avec (Q, D). Ils sont générés par les règles suivantes, qui sont toutes en boucle :

$$\frac{a \quad \Sigma \; Q \quad dPrj\Delta\,(D)}{(Q, D) \xrightarrow{a} (Q, D) \text{ dans } \Delta \text{ cgr}} \qquad \frac{Q \quad dPrj\Delta\,(D)}{(Q, D)@(Q, D) \rightarrow (Q, D) \text{ dans } \Delta\text{cgr}} \qquad \frac{Q \quad dPrj\Delta\,(D)}{(Q, D) \longrightarrow (Q, D) \text{ dans } \Delta\text{cgr}}$$

□

exécution partielle de A cgr(S) alors sur un préfixe u attribue un état (P, D) avec P dPrjΔ (D), Donc, si une (P, D) est un état de projection de sous-couverture d'après le lemme 14, et donc , sous-couverture non pertinente par la proposition 2.

Lemme 15. Soit A = UNE[I/ Δ], L = L(A par), et S = L(UNE [F/FS]). Si (Δ, D) $\xrightarrow{hdg(u)}$ (Q, D) rapport à Acgr(S) pour un mot imbriqué u NΣ et soit • Q dPrjΔ (D) et q Q, soit • Q dPrjΔ (D) et q accΔ(Q),

alors il existe un mot imbriqué u′ classL ′(u) et q0 Δ, tel que q0 $\xrightarrow{hdg(u)}$ q par rapport à Δ.

Preuve. Par récurrence sur la longueur de u, supposons que (Δ, D) $\xrightarrow{hdg(u)}$ (Q, D) par rapport à A cgr(S) · Dans le cas de base, nous avons u = ε. Par conséquent, Q = Δ.

Cas Q dPrjΔ (D) et q Q. Alors, Δ = {q}. La course unique de A sur u = ε commence et se termine dans l'état initial de l'arbre q Δ.

Machine Translated by Google

Cas Q dPrjΔ (D) et q accΔ(Q). Puisque Q dPrjΔ (D), le lemme 14 montre que (Q, D) par ′(ε) = NΣ. De plus, est un état de projection de sous-couverture, de sorte que ε est une sous-couverture non pertinente ' et S ' pour la L Proposition 2. Par conséquent, puisque q accΔ(Q), il existe un S' la classe L couvre h et q0 I, tels que q0 −h→ q par rapport à Δ. Soit ' = nw(h). La course de A sur toi ' se termine dans q et u ' u classL'(u). S

Pour l'étape d'induction, on distingue les formes possibles du mot imbriqué u. Il existe donc u˜, v NΣ et a Σ, tels que l'un ou l'autre des cas suivants est vrai :) cgr(S)

Cas u = u˜ · · v · . Soit (Q˜ , D) l'état dans lequel l'exécution de (A ' à tes extrémités.

Sous-cas Q˜ dPrjΔ (D). Soit D′ = DQ˜ . Alors, il existe P Q, tel que (Δ, D′) v−→ (P, D′) par rapport à A cgr(S)) cgr(S) sur v passe à l'état (P, D′). Ainsi, la course de (A '

Sous-sous-cas P dPrjΔ (D′). Par construction de A cgr(S) , on a alors Q = Q˜ @ΔaccΔ(P). Puisque q Q, il existe q˜ Q˜ et p accΔ(P), tels que ′(v), tel · q˜@p → q dans Δ. Par hypothèse de récurrence, il existe v que ' classeL' l'exécution de A ' ' sur v se termine à la p. Par conséquent, l'exécution de A ' = u˜ · · v ' se termine par q, et de plus, u ' sur u classL (u). S

Sous-sous-cas P dPrjΔ (D′). Par construction de A cgr(S) , on a alors Q = Q˜ @ΔP. Puisque q Q, il existe q˜ Q˜ et p P, tels que q˜@p → q dans Δ. Par hypothèse de récurrence, il existe v ′(v), tel que l'exécution de ' classeL' · se termine par q, et A de ' ' sur v se termine à la p. Par conséquent, l'exécution de A ' = u˜ · · v ' plus, u ' sur u classL (u). S

Sous-cas Q˜ dPrjΔ (D). Alors, Q = Q˜ et q accΔ(Q˜). En utilisant l'induction hdg(u˜ (u˜), tel que Δ −−−→ q. ' hypothèse appliquée à u˜, il existe u˜ ' classeL classL (u).) Puisque u˜ n'est pas pertinent pour L et S', nous t'avons ' S S

Cas u = u˜ · a. Ceci est plus facile à réaliser que les cas précédents et est donc omis. □

Lemme 16. Une exécution partielle de A cgr(S) assigne un état (Q, D) à un préfixe u prefs(NΣ). Si Q dPrjΔ (D) et q Q, alors il existe un préfixe u classL (u), tel qu'une exécution partielle de A attribue l'état q à u′ . S

Preuve. Par induction sur la longueur de u. Dans le cas de base, nous avons u = ε. L'exécution partielle de init . A cgr(S) sur u attribue l'état (Q, D), où Q = I et D = D A,FS Supposons que Q dPrjΔ (D) et q Q. Alors, I = {q}. L'exécution partielle unique de A sur u = ε se termine dans l'état initial = A[I/ Δ], L), et S [F/ possibles de u. Ainsi, u˜ prefs(NΣ), v NΣ ' FS]. Nous q I. Pour l'étape d'induction, soit A (UNE' = L((UNE' = L(UNE' distinguer les formes et a Σ existent, tels que l'une des affirmations suivantes est vraie :

Cas u = u˜ · · v. Soit (Q˜ , De D˜) soit l'état dans lequel l'exécution partielle de A cgr(S) à tes extrémités. Q dPrjΔ (D), il s'ensuit que Q˜ dPrjΔ (D˜). De plus, (Δ, D) v− → (Q, D) par rapport à A cgr(S)) cgr(S). Ainsi, l'exécution partielle de (A sur ' v passe à l'état (Q, D). En utilisant l'hypothèse d'induction, il existe v ' classL '(v), tel que l'exécution partielle de A = u˜ · · ' sur ' v se termine par q. D'où l'exécution partielle de A sur u S ' v se termine par q, et de plus, ' tu classL (u). S

Cas u = u˜ · · v · . Soit (Q˜ , D˜) soit l'état dans lequel l'exécution partielle de A cgr(S) à tes extrémités. De Q dPrjΔ (D), il s'ensuit que Q˜ dPrjΔ (D˜). De plus, (Δ, D) v−→ (P, D) par rapport à A cgr(S) pour certains P Q.

Sous-cas P dPrjΔ (D). Alors, Q = Q˜ @ΔaccΔ(P). Donc, il existe q˜ Q et p accΔ(P), tels que q˜@p → q dans Δ. D'après le lemme 15, il existe v ′(v), ' classeL' S

Machine Translated by Google

tel que Δ il hdg(v') ⟶ p par rapport à Δ. En utilisant l'hypothèse d'induction appliquée à u˜, existe un état initial q0 I et un préfixe u˜ course de A ' classL (u˜), telle que le partiel · · v · sur u˜ ' continue en déclarant q˜. D'où l'exécution partielle de A sur u ' = tu˜' ' passe à q et toi ' classL(u).

Sous-cas P dPrjΔ (D). Alors, Q = Q˜ @ΔP. Donc, il existe q˜ Q et p P, tels classL ′(v), tels que que q˜@p → q dans Δ. D'après le lemme 15, il existe v' S hdg(v') Δ ⟶ p par rapport à Δ. En utilisant l'hypothèse d'induction appliquée à u˜, il existe un état initial q0 I et un préfixe u˜ sur u˜ ' classL (u˜), telle que l'exécution partielle de A · · ' continue en déclarant q˜. D'où l'exécution partielle de A sur u ' = tu˜' v · ' continue à q et toi ' classL (u).

Cas u = u˜ · a. Ceci est plus facile à réaliser que le cas précédent et est donc omis. □

Le lemme suivant énonce l'invariant clé de la projection de congruence, ce qui prouve finalement son exhaustivité pour la projection de sous-couverture. Pour cela, on définit pour tout F FS F la relation binaire ≈F FS comme la clôture symétrique de F × (FS \ F), c'est-à-dire pour tout (q', q ") Q2 :

$$q, \approx_F \text{ESq} \text{»} \quad \begin{matrix} (q' & F & q & FS \setminus F \text{»} & FS \setminus F) \\ (q' & q & & \text{»} & F \end{matrix})$$

Lemme 17 (Invariant clé). Soit A = (Σ, Q, Δ, I, F) un dSHA, F FS Q, L = L(A) et S = L(A[F/FS]). Si A cgr(S) a une exécution partielle sur le préfixe u prefs(NΣ) pour énoncer (P, D) 2 Q × DΔ, alors pour tout (p ', p ") P 2, (r ',r ") D, et (v ', v ") N 2 tel que Σ ,

$$p' \xrightarrow{hdg(v')} r \text{ par rapport à } \Delta \quad p \text{»} \xrightarrow{hdg(v'')} r \text{ par rapport à } \Delta$$

il existe (tu ', u ") classels (u), w suffs(NΣ), q′ ≈F FS q ", et q0 I tel que

$$hdg(u \cdot w) \; q0 \longrightarrow q \quad ' \text{ rapport à } \Delta \quad q0 \xrightarrow{hdg(u \cdot w)} q \text{ par rapport à } \Delta$$

Preuve. Induction sur le nombre de parenthèses ouvrantes pendantes de u.

Dans le cas de base, u n'a pas de parenthèse ouvrante pendante donc u NΣ est un init A,FS . Ainsi, A cgr(S) a cas, D = D A,FS à (P, D), où P Q. Soit (pune exécution partielle sur le mot imbriqué u NΣ mot imbriqué. Dans ce initialisation ', p ") P 2, (r ',r ") D A,FS initialisation , et (v ', v ") N 2 tel que Σ ,

$$p' \xrightarrow{hdg(v')} r \text{ par rapport à } \Delta \quad p \text{»} \xrightarrow[\text{See More}]{hdg(v'')} r \text{ par rapport à } \Delta$$

Il s'ensuit alors que P dPrjΔ (D il A,FS initialisation). On peut donc appliquer le lemme 16. Il montre que existe des mots imbriqués (u ', u ") (classL (s))2 et q0 I, tel que

$$hdg(u') \; q0 \longrightarrow p \quad ' \text{ par rapport à } \Delta \quad q0 \xrightarrow{hdg(u'')} p \text{ »} \text{ par rapport à } \Delta$$

Depuis D A,FS initialisation IdrΔ (F × (FS \ F)) d'après le lemme 10, il existe un mot imbriqué w NΣ et ′ ≈F états q FS q ", tel que les

$$' r \xrightarrow{hdg(w)} q \quad ' \text{ par rapport à } \Delta \quad r \text{»} \xrightarrow{hdg(w)} q \text{ »} \text{ par rapport à } \Delta$$

Nous avons alors ce qui suit :

$$hdg(u' \cdot v' \cdot w) \longrightarrow q \; q0 \quad ' \text{ par rapport à } \Delta \quad q0 \xrightarrow{hdg(u \text{»} \cdots w)} q \text{ »} \text{ par rapport à } \Delta$$

Machine Translated by Google

Pour l'étape d'induction, soit u = u1 · · u2 pour un préfixe u1 prefs(NΣ) et un mot imbriqué u2 NΣ de sorte que u1 ait une parenthèse ouverte pendante de moins que u. Soit A cgr(S) une exécution partielle sur le mot imbriqué u NΣ vers (P, D). Soit (P1, D1) l'état que cette exécution attribue au préfixe u1. Alors, D = (D1)P1 . Soit (p , p ") P 2 , (r ,r ") D, et (v , v ") N 2 tel que Σ ,

$$p \xrightarrow{hdg(v} r) \text{ par rapport à } \Delta \quad p \gg \xrightarrow{hdg(v")} r \text{ par rapport à } \Delta$$

Puisque D = (D1)P1 , il existe (p et p 1 , p .) (P1) 2 et les états (r 1 ,r 1) D1, tel que p 1@q → r 1 @q → r1 . En particulier, P1 dPrjΔ (D1) pour que l'on puisse appliquer le lemme 16. Il montre qu'il existe (u 2 , 2) classeL S (u2), tel que Δ −→ p hdg(v hdg(v et Δ −→ p ". Laisser v1 = · toi2 · v . et v par rapport à Δ. En utilisant l'hypothèse d'induction appliquée à u1, sur laquelle A cgr(S) a un parcours partiel vers l'état (P1, D1), Ensuite p 1) , pli 1)

tel que (p 1 , p 1) (P1) 2 , p . 1 hdg(v 1) 1 , p 1 , hdg(v 1 ,) et (r 1 ,r 1) D, il existe (tu 1 , 1) classeL S (u1), w1 suffs(NΣ), q ≈F FS q ", et q0 I, tel que

$$hdg(u\ 1\ w1)\ q0 \xrightarrow{} q \quad \text{rapport à } \Delta \quad q0 \xrightarrow{hdg(u\ 1\ w1)} q \text{ par rapport à } \Delta$$

Laissez -vous = tu 1 · · toi vous 2 , toi 1 See More · · toi2, et w = · w1. Ce qui précède donne alors ce qui suit :

$$hdg(u \cdot w)\ q0 \xrightarrow{} q \quad \text{rapport à } \Delta \quad q0 \xrightarrow{hdg(u \cdot w)} q \text{ par rapport à } \Delta$$

De plus, (vous , u ") (classL (u))2 , donc cela devait être montré. □

Proposition 6. Soit A = (Σ, Q, Δ, I, F) a dSHA, F FS Q, L = L(A), et S = L(A[F/FS]).

Si A cgr(S) a une exécution partielle sur le préfixe u prefs(NΣ) vers un état (P, D) 2 Q × DΔ de sorte que P dPrjΔ (D), alors la classe L (u) est une sous-couverture pertinent pour L par rapport à S.

Preuve. Supposons que A cgr(S) ait une course partielle sur le préfixe u prefs(NΣ) vers un état (P, D) × DΔ de ensemble tel que P v ") dPrjΔ (D). Puisque P dPrjΔ (D), il existe acc∆(P) 2 2 Q ∩ D = Il existe donc (p , p ") P 2 , (r ,r ") D, et (v , tel que

$$p \xrightarrow{hdg(v'} r) \text{ par rapport à } \Delta \quad p \gg \xrightarrow{hdg(v")} r \text{ See More par rapport à } \Delta$$

D'après le lemme 17, il existe (u tel que u ") classeL S ≈F (u), w suffs(NΣ), q FS q ", et q0 I,

$$hdg(u - v \cdot w) \xrightarrow{} q\ q0 \quad \text{par rapport à } \Delta \quad q0 \xrightarrow{hdg(u \cdot w)} q \text{ par rapport à } \Delta$$

Par conséquent, la classe L (u) est pertinent pour L par rapport à S. □

Théorème 3 (Exhaustivité de la projection de congruence.). Pour tout dSHA complet (Σ, Q, Δ, I, F) et un ensemble F FS Q, la projection de congruence A cgr(S) est solide et complète pour la projection de sous-couverture pour le modèle régulier L(A) par rapport au schéma régulier S = L(UNE[FS/F]).

Preuve. La solidité de la projection de congruence est montrée dans le théorème 2. Pour prouver l'exhaustivité, soit u prefs(NΣ) un préfixe de mot imbriqué qui est fortement sous-couverture non pertinent pour L par rapport à S. Par définition de forte non-pertinence, la classe classL (u) est non pertinent pour L et S. Soit (P, D) l'état unique attribué par l'exécution partielle de A cgr(S) pour préfixer u. Puisque (u) n'est pas pertinent pour L et S, la proposition 6 montre que P dPrjΔ (D). En utilisant le lemme 14, classL S (P, D) est alors un état de projection de sous-couverture de A cgr(S)

. □

Machine Translated by Google

7.5. Complexité en mémoire

Nous discutons ensuite de la complexité des tests d'appartenance avec une projection de sous-couverture complète basée sur l'évaluation en mémoire de la couverture d'entrée par la projection de congruence de l'automate d'entrée.

Lemme 18. Le nombre d'états |Qcgr| est dans O(2 n 2+n), où n est le nombre d'états de A.

Preuve. Avec la construction déterministe, les états de A cgr(S) sont des paires dans 2Q × DΔ. Ainsi, le nombre maximal d'états de la projection de congruence est |Qcgr(S) | = 2 |Q| |DΔ| ≤ $2^n\ 2^{2n}\ +n = 2^{n^2}$. □

Soit diff-rel(Qcgr(S)) l'ensemble des relations de différence utilisées dans les états de A cgr(S) .

Corollaire 1. Soit la signature Σ fixée. Pour tout dSHA A complet de schéma S HΣ, l' appartenance à toute haie h S peut être testée en mémoire en temps O(|1|) par nœud non projeté de h après un temps de prétraitement de O(|A cgr(S) | + n 3d), où d est la cardinalité de diff -rel(Qcgr(S)) et n est le nombre d'états de A.

Depuis |Qcgr(S) | est dans O(2 n 2+n) d'après le lemme 18, le temps de prétraitement O(|A cgr(S) | + n 3d|) est également dans $O(2^{n\ 2+2n})$.

Preuve. Puisque A est complet, Thereom 3 montre que A cgr(S) est complet pour la projection de sous-couverture pour le schéma S. L'évaluateur en mémoire de A cgr(S) peut être exécuté sur n'importe quelle couverture d'entrée h S au temps O(|1|) par nœud non projeté de h après un temps de prétraitement de O(|A cgr(S) | + n 3d), où d est la cardinalité de diff-rel(A cgr(S)). Puisque la signature Σ est fixe et A cgr(S) déterministe , |A cgr(S) | est dans $O(|Qcgr(S)|^2)$. La partie dominante |A cgr(S) | est en $O(|Qcgr|^2)$. □

Lorsqu'elle est appliquée à une réponse à une requête XPath standard comme dans nos expériences de la section 11, nous avons n ≤ 80 et d ≤ 25 donc la partie cubique O(n 3d) n'est pas limitative. La taille des SHA↓ est inférieure à 2 600 en comptant à la fois les États et les règles. La limite supérieure O(|A cgr(S) | + n 3d) suggère donc que le temps de prétraitement peut être réalisable en pratique, même s'il est exponentiel dans le pire des cas.

Nous devons être prudents ici puisque les tailles des SHA↓ rapportées dans la section 11 ne font pas référence à la projection de congruence pure A cgr(S) mais à la projection de congruence la plus ancienne A que cgr(S) e(S) nous introduirons uniquement dans la section 8. La non-création de règles de transition vers des états qui peuvent être rejetés en toute sécurité peut conduire à une réduction radicale de la taille. Alternativement, une réduction importante de la taille de A cgr(S) pourrait être obtenue en supprimant tous les états et règles de transition inutile Mais cela ne réduirait pas le temps de précalcul, bien au contraire : cela augmenterait le temps de prétraitement puisque le nettoyage des automates a posteriori nécessite également du temps, ce qui pour les SHA↓ peut être non linéaire.

Au lieu de précalculer A cgr(S) à partir des entrées A et FS de manière statique, nous pouvons calculer uniquement la partie nécessaire de A cgr(S) pour évaluer la couverture d'entrée h à la volée. Nous notons que le nombre de règles de transition de la partie nécessaire est limité par O(|h|) mais peut être bien plus petit en raison de la projection. Le temps de prétraitement descend alors à O(|A|). Puisque l' automate complet A cgr(S) peut être exponentiel en n dans le pire des cas, cela peut réduire le temps de traitement global.

Du côté positif, l'évaluation à la volée réduit le temps de calcul global à O(|h| + n 3d(h)), où d(h) est le nombre de relations de différence dans la partie de A cgr(S) nécessaire pour l'évaluation de h, donc d(h) ≤ min(|h|, d). Par contre, les étapes d(h) ne seront plus en temps constant mais cubique en $O(n^3)$. Il convient également de noter que la durée globale de l'évaluation à la volée de la projection sans changement est en O(|A| + |h| + n s(h)), où s(h) est le nombre de sous-ensembles d'états sûrs de A dans les états de Qcgr, c'est-à-dire s(h) ≤ |h|. En effet, chaque sous-ensemble d'états sûrs peut être calculé en temps linéaire en O(n). Comme indiqué ci-dessus, le temps global de la projection de congruence obtenue par évaluation à la volée est en $O(|A| + |h| + n^3 d(h))$. En effet, le calcul d'une relation différentielle se fait en $O(n^3)$

Machine Translated by Google

en utilisant des programmes Datalog au sol de taille cubique. Le facteur cubique supplémentaire en n d(ĥ) au lieu d'un facteur linéaire en n s(h) est le prix à payer pour réaliser une projection complète de sous-couverture avec évaluation à la volée. Notez que ce prix est largement acceptable dans les cas où n ≤ 80, d(h) ≤ 15 et s(h) ≤ 15.

Exemple 15 (Explosion exponentielle.). Afin de voir comment le pire des cas exponentiel peut se produire, nous considérons une famille de langages réguliers, pour lesquels le DFA minimal de gauche à droite est exponentiellement plus grand que le DFA minimal de droite à gauche . Les exemples de langages classiques avec cette propriété sont Ln = Σ · a, Σ où n N et Σ = {a, b}. Intuitivement, un mot dans Σ appartient à Ln si et seulement sa n + 1-ème lettre à partir de la fin est égale à « a ». Le DFA minimal de gauche à droite pour Ln a 2 n+1 nombreux états puisqu'il doit mémoriser une fenêtre de n + 1 lettres. En revanche, son DFA minimal de droite à gauche ne comporte que n + 1 états ; dans ce sens, il suffit de mémoriser la distance à l'extrémité modulo n + 1.

Considérons ensuite le schéma S défini par µz. (a + b) · z *. Il contient toutes les haies ayant dans tous ses sous-arbres une étiquette de Σ = {a, b} comme première lettre, et aucune autre occurrence de lettres de Σ ailleurs. Nous considérons ensuite la famille des langages de couverture Hn HΣ de schéma S, tels que la séquence d'étiquettes d'un chemin racine-feuille de hn appartient à Ln. Notez que Hn peut être reconnu de manière ascendante par le dSHA An avec des états O (n + 1) , qui simule le DFA déterministe minimal de Ln sur tous les chemins de la couverture d'entrée. Pour un évaluateur avec projection en sous-couverture, la situation est différente. Lors d'un déplacement de haut en bas, il doit mémoriser la séquence s, et il y a 2 n+1
d'étiquettes des n + 1 derniers ancêtres, éventuellement remplies de b de telles séquences.
Si, pour une feuille, sa séquence commence par un « a », alors les sous-haies suivantes avec les feuilles suivantes peuvent être projetées. En conséquence, il ne peut y avoir de dSHA↓ reconnaissant Hn qui
projette toutes les sous-couvertures non pertinentes avec moins de 2 n+1 états. En particulier, la taille de A doit cgr(S) n être exponentielle par rapport à la taille de An.

8. Première adhésion avec projection de sous-couverture Nous

améliorons ensuite nos compilateurs pour introduire des états de projection de sous-couverture, de sorte qu'ils puissent également détecter certaines appartenances ou non-adhésions de manière plus précoce. Pour cela, nous montrons comment combiner le premier testeur d'appartenance pour les dSHA de [21] avec n'importe quel algorithme de projection de sous-arbre afin qu'il s'applique aux deux, à la projection sûre sans changement et à la projection de congruence. En combinant les deux aspects orthogonalement, nous obtenons un premier dSHA avec projection subhedge, qui peut être évalué soit en mémoire, soit en mode streaming.

8.1. Appartenance la plus

ancienne Nous commençons par une caractérisation sémantique de l'appartenance la plus ancienne, c'est-à-dire des préfixes de mots imbriqués de haies dont les suffixes ne sont pas pertinents pour l'appartenance, en supposant un traitement descendant et de gauche à droite. Notez que la même caractérisation jusqu'à différents choix de présentation a été donnée plus tôt dans le contexte du traitement de flux [21,22].

Définition 13. Soient S HΣ un schéma de couverture et L S un langage de couverture satisfaisant ce schéma. Un préfixe de mot imbriqué v est certain d'être membre de L avec le schéma S pour tous les suffixes w de mots imbriqués tels que v · w nw(S), il considère que v · w nw(L). On l'appelle certain pour la non-appartenance à L avec le schéma S pour tous les suffixes w de mots imbriqués tels que v · w nw(S), il considère que v · w nw(L).

Nous nous intéressons maintenant aux SHA↓ capables de détecter certaines appartenances et non-appartenances. appartenance syntaxiquement au préfixe ou au nœud le plus ancien possible de l'exécution.

Définition 14. Un dSHA↓ A = (Σ, Q, Δ, I, F) de schéma S HΣ est appelé le plus tôt pour le schéma S s'il existe un état sel Q, tel que pour toutes les haies h S, l'unique l'exécution R de compl(A) sur h satisfait ce qui suit :

Machine Translated by Google

- il va à l'état sel pour exactement tous les préfixes de nw(h) dont l'appartenance à L(A) par rapport à S est certaine ; il va
- au récepteur pour tous les préfixes de nw(h) qui sont certains de non-appartenance à L(A) avec S.

Pour dSHA A, schéma complet pour le schéma S = L(A[F/FS]), nous pouvons construire un dSHA↓ Ae(S) qui est le plus ancien pour S, comme le montre l'Annexe A. L'idée est d'adapter le le premier testeur d'appartenance pour les SHA déterministes de [21] afin qu'il prenne en compte les restrictions de schéma. De plus, nous devons compiler les dSHA d'entrée en dSHA↓ plutôt qu'en dNWA, comme utilisé ici. Ce changement est simple une fois que la relation entre les deux modèles d'automates devient claire.

Étant donné que la construction de Ae(S) ne fait pas partie du cœur du présent article, nous illustrons son résultat uniquement dans notre exemple courant : pour le dSHA A de schéma S = doc, dans la figure 4, nous obtenons le SHA↓ Ae (S) sur la figure 19. Notez que, sur le préfixe list item, il va à state to sel, montrant qu'il est certain d'être membre. Sur le préfixe item, il bloque indiquant qu'il est certain de non-adhésion. Cependant, sur le préfixe list list, il ne passe pas dans un état de projection de sous-couverture même si ce préfixe n'est pas pertinent pour la sous-couverture.

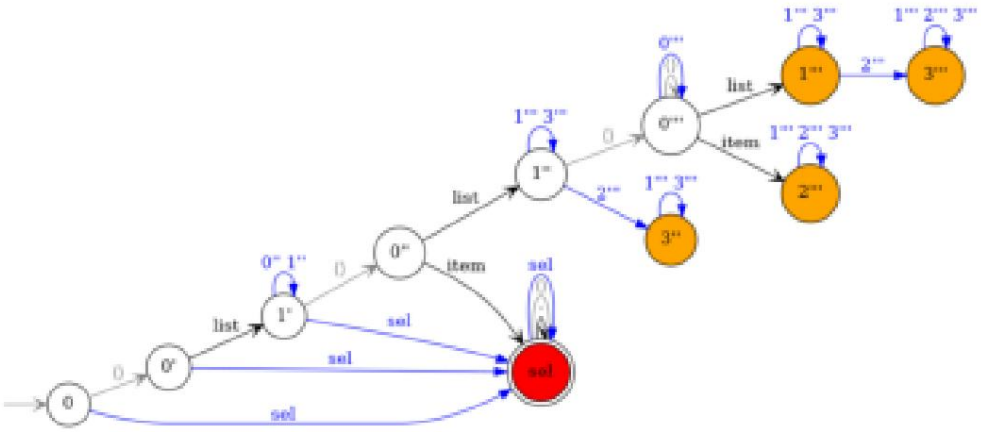


Figure 19. Le premier dSHA↓ Ae(S) pour le dSHA A de la figure 4 avec le schéma S = doc pour le filtre XPath [self::list/child::item].

8.2. Ajout d'une projection de sous-couverture

Nous améliorons désormais toute première adhésion aux dSHA avec une projection de sous-couverture. L'idée est de combiner la projection de sous-couverture avec l'adhésion la plus précoce. Supposons donc que l'on nous donne un schéma S et deux automates, A π et Ae , de schéma S qui reconnaissent le même langage jusqu'au schéma, donc L(A π) ∩ L(S) = L(Ae) ∩ L (S). De plus, supposons que Ae ait un état de sélection sel, indiquant une certaine appartenance au préfixe le plus ancien possible.

Afin d'obtenir l'adhésion la plus précoce avec la projection de sous-couverture, nous combinons les deux SHA↓ en un seul SHA↓ F e), exécutant essentiellement les deux automates en parallèle mais sous contrôle partagé. L'ensemble d'état de A $_{UN}^{\pi}$ $_{e}^{\pi}$ $_{e}^{\pi}$ $_{e

Machine Translated by Google

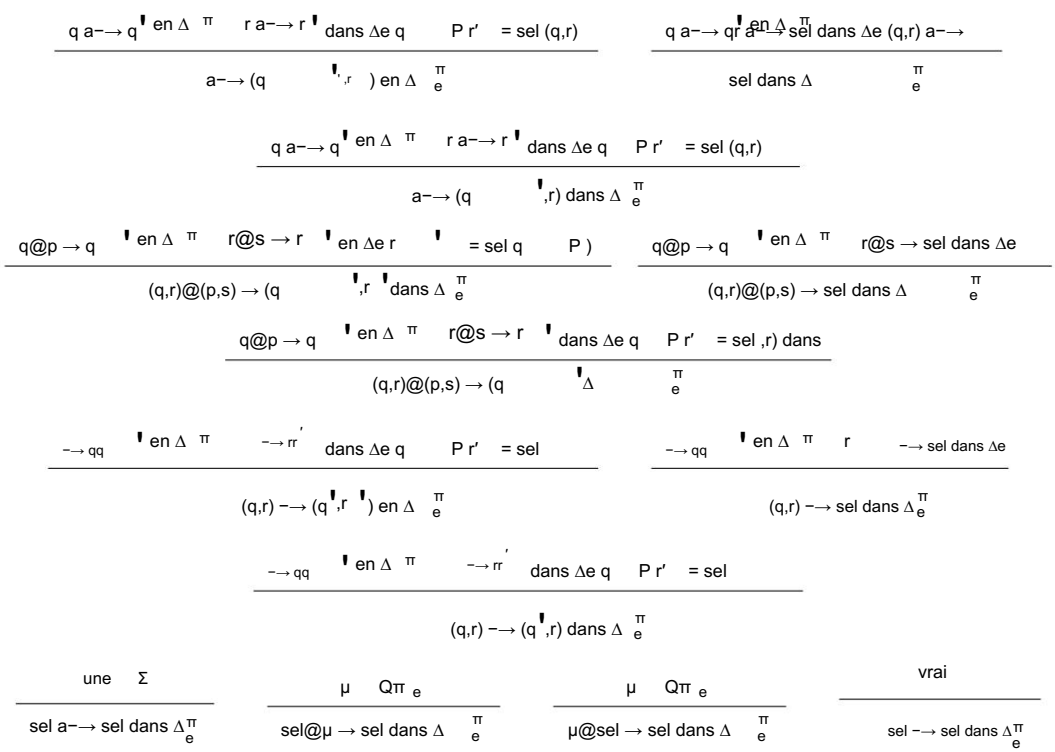


Figure 20. Les règles de transition Δ_e^{π} déduits de ceux du SHA↓ s A π avec les états de projection P et Ae avec état de sélection sel.

À titre d'illustration, nous reconsidérons le dSHA de la figure 4. Le premier dSHA↓ avec est projection de sous-couverture A^{snc} donné dans la figure 21. Ici, nous combinons le dSHA sans changement projection e(S) A^{snc} de la figure 11 avec le premier testeur d'appartenance Ae(S) de la figure 19. e(S) Leur combinaison A^{snc} représente le modèle [self::list/child::item] avec le schéma doc de manière concise. Après le préfixe list item, il détecte certaines appartenances : pour le préfixe list list, il détecte que la sous-couverture n'est pas pertinente, et après le préfixe item, il détecte certaines non-appartenances.

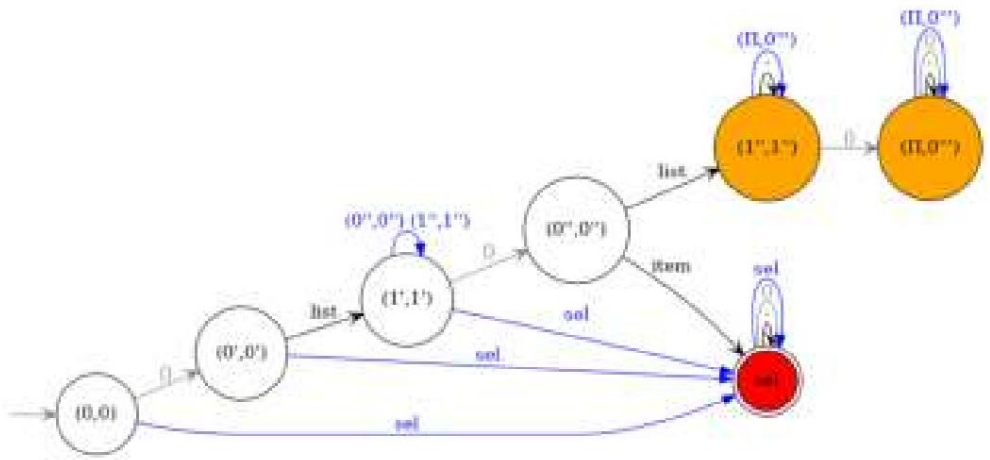


Figure 21. Le premier dSHA↓ A Figure $_{e(S)}^{snc}$ avec projection de sous-couverture sûre et sans changement pour le dSHA A en 4 avec schéma S = doc pour le filtre XPath [self::list/child::item].

Machine Translated by Google

8.3. Solidité et exhaustivité

Nous montrons ensuite que la solidité et l'exhaustivité d'un algorithme de projection sont préservées lorsqu'il est combiné avec un premier testeur d'appartenance en supposant le déterminisme.

le schéma, ie L(A π) ∩ L(S) π et Ae sont dSHA↓ s de même schéma S et de même langage jusqu'à la Proposition 7. Soit A = L(Ae) ∩ L(S). La combinaison dSHA↓ Un langage jusqu'au schéma S aussi. De plus, si Ae est le π_e alors a le même plus ancien pour S, alors A entre dans un état de projection de sous-couverture chaque fois π_e est le plus ancien pour S et va que Aπ le fait.

sous-couverture. Si q est un état de projection de sous-couverture de A π et r un état alors (q,r) est une preuve de l'état de projection de A π_e. L'évaluateur de l'automate A π_e exécute A de Ae , π et Ae en parallèle sur haie d'entrée, tout en sautant les sous-couvertures commençant dans les états de projection de sous-couverture de A π. Celles-ci incluent les sous-couvertures commençant dans un état de projection lors de π_e sur h. Candidature l'évaluation d'une proposition 2 à A π, ces sous-couvertures ne sont pas pertinentes pour L(A π) par rapport à S, π puisque A est supposé être déterministe. Puisque L(A π) ∩ S = L(Ae) ∩ S, sauter une telle sous-couverture n'affecte pas l'acceptation par Ae des couvertures à l'intérieur du schéma S. Par conséquent, l'évaluateur de A π_e est également le plus ancien par rapport à S. □

La proposition ci-dessus montre que si A est π est complet pour la projection de sous-couverture, alors A π_e également complet pour la projection de sous-couverture tout en étant également le plus ancien.

Théorème 4. Pour tout dSHA A complet avec schéma S, le dSHA A cgr(S)e(S) est le plus ancien et complet pour la projection de sous-couverture pour le schéma S.

Preuve. La projection de congruence A cgr(S) est valable d'après la proposition 2 et complète pour la projection de sous-couverture par rapport à S d'après le théorème 3. L'automate Ae(S) discuté dans la section 8.1 est le plus ancien pour S. Ainsi, leur combinaison A cgr(S)e(S) donne le premier automate avec une projection de sous-couverture complète par rapport à S en utilisant la proposition 7. □

8.4. Complexité en mémoire Nous

discutons ensuite de la complexité des premiers tests d'appartenance avec une projection complète de sous-couverture par un évaluateur en mémoire pour A cgr(S)e(S) .

Lemme 19. Le nombre d'états dans Qcgr(S)e(S) est en O(2 n 2+2n+log(n)), où n = |Q|.

Preuve. D'après le lemme 18, le nombre d'états de A cgr(S) est dans O(2 n 2+n), où n = |Q|.
Le nombre d'états de Ae(S) est en O(n2 n), et donc en O(2 n+log(n)). Le nombre d'états de Qcgr(S)e(S) est donc en O(2 n 2+n 2 n+log(n)), qui est en O(2 n 2+2n+log(n)) □

Corollaire 2. Soit la signature Σ fixée. Pour tout dSHA A complet de schéma S HΣ, la première appartenance à une couverture d'entrée h S peut être testée en mémoire dans le temps O(|1|) par nœud non projeté de h après un temps de prétraitement de O(|A cgr(S)e(S) | + n 3d + n s), où d est le nombre de relations de différence et s est le nombre de sous-ensembles d'états sûrs dans Qcgr(S)e(S) . Le temps de prétraitement est également 2n 2+4n+2log(n)), où n est le nombre d'états de A. dans O(2

Preuve. Puisque A est complet, Thereom 4 montre que A cgr(S)e(S) est la première, solide et complète pour la projection de sous-couverture pour le schéma S. L'évaluateur en mémoire de A cgr(S)e(S) peut être exécuté sur n'importe quelle couverture d'entrée h S dans le temps O(|1|) par nœud non projeté de h après un temps de prétraitement de O(|A cgr(S)e(S) | + n 3d + n s. Σ est fixé à la taille |A cgr(S)e(S)) . Le lemme 19 montre que délimité par O(|Qcgr(S)e(S) | $2n^2$ |Qcgr(S)e(S) | est dans O(2 n 2+2n+log(n)), donc |A cgr(S)e(S) | est en O(2 2+4n+2log(n)). □

Machine Translated by Google

Comme auparavant, au lieu de précalculer A cgr(S)e(S) à partir des entrées dSHA A et FS, nous pouvons calculer uniquement la partie nécessaire à l'évaluation de la couverture d'entrée h à la volée. Si la partie nécessaire de A cgr(S)e(S) pour évaluer h est petite, le temps et l'espace globaux diminuent considérablement.

9. Algorithmes de streaming Nous

montrons que les dSHA↓ peuvent être évalués sur le mot imbriqué d'une haie de manière continue , de sorte qu'ils produisent les mêmes résultats et comportement de projection qu'en les évaluant sur des haies en mémoire de manière descendante et manière de gauche à droite.

Tous les aspects liés à la projection des sous-couvertures restent inchangés. La réponse à la première requête en mode streaming signifie que les suffixes de mots imbriqués sont ignorés s'ils ne sont pas pertinents, et pas seulement les mots imbriqués comme avec la projection de sous-couverture. Ce qui change entre les deux modes d'évaluation, c'est la consommation de mémoire. Pour l'évaluation en mémoire, l'intégralité du graphique de la haie d'entrée doit être stockée, tandis qu'en mode streaming, seuls une pile et un état sont conservés à tout événement. Pour tester l'appartenance à un langage régulier, l'espace d'état est fini, tandis que l'espace de pile devient infini, puisque la profondeur de la pile dépend de la profondeur de la couverture d'entrée.

La propriété d'avoir des évaluateurs de streaming et en mémoire équivalents a déjà été remarquée dans [26] pour les automates forestiers pushdown de Neumann et Seidl [25] et pour les automates de mots imbriqués [11,15]. Nous montrons ici comment transposer ce résultat en dSHA↓ s. Cela nous permet de montrer l'exactitude de nos algorithmes d'évaluation basés sur les propriétés des automates construits. Il permet également d'obtenir les propriétés d'exactitude des algorithmes de streaming en raisonnant sur l'évaluateur en mémoire des SHA↓ .

9.1. Évaluateurs de streaming pour les automates de couverture

descendante Nous définissons l'évaluateur de streaming d'un SHA↓ A = (Σ, Q, Δ, I, F) par une machine visiblement pushdown . L'ensemble des configurations de cette machine est K = Q × Q , contenant des paires d' états et des piles d'états. La machine visiblement pushdown fournit pour tout mot v Σˆ une relation de transition en continu v−→ Σˆ σ $\xrightarrow{\text{str}}$ K × K, tel que pour tous les états q, q' Q et piles et les mots v, v ' Q :

$$\frac{\text{vrai}}{(q,\sigma)\ \xrightarrow{\varepsilon}_{\text{str}}\ (q,\sigma)\text{ par rapport à }\Delta} \qquad \frac{(q,\sigma)\ v\!-\!\xrightarrow{\text{str}}\ (q',\sigma) \qquad (q',\sigma)\ \xrightarrow{'v}_{\text{str}}\ (q'',\sigma)\text{ par rapport à }\Delta}{(q,\sigma)\ \xrightarrow[v\cdot v']{}_{\text{str}}\ (q'',\sigma)\text{ par rapport à }\Delta}$$

$$\frac{q\ a\!-\!\to q'\text{ en }\Delta}{q\ a\!-\!\xrightarrow{\text{str}}\ q'\text{ par rapport à }\Delta} \qquad \frac{q\ \!-\!\to q'\text{ en }\Delta}{(q,\sigma)\ \xrightarrow[-]{\text{str}}\ (q',\sigma\cdot q)\text{ par rapport à }\Delta} \qquad \frac{q@p\to q'\text{ en }\Delta}{(p,\sigma\cdot q)\ \xrightarrow[-]{\text{str}}\ (q',\sigma)\text{ par rapport à }\Delta}$$

Nous notons que la même machine visiblement pushdown peut être obtenue en compilant le SHA↓ dans une NWA, dont l'évaluateur de streaming est également défini par une machine visiblement pushdown (voir la section 12.3 sur les travaux connexes).

La notion d'exécutions partielles de l'évaluateur en mémoire est étroitement liée aux exécutions en continu. Pour voir cela, notons d'abord que chaque exécution partielle v par rapport à Δ définit naturellement une pile comme suit. Puisque v est un préfixe de mot imbriqué, il existe des mots imbriqués uniques v0, . . . , vn nw(HΣ Q), tel que

$$v = v0 \cdot \quad \cdot v1 \cdot \quad \cdot \quad \ldots \cdot \quad \cdot vn.$$

Ces mots imbriqués doivent également être des séquences partielles de Δ , il existe donc des états p0, . . . pn, q0, . . . , qn Q, tel que vi passe de pi à qi pour tout 0 ≤ i ≤ n. On définit la pile σ de v par · qn−1 et disons puisque · · · que v va de p0 à qn et a une pile σ. Notez que la pile est σ = q0 · vide si v est un mot imbriqué, dans ce cas n = 0 donc v = v0 et σ = ε. Par exemple, l'exécution partielle v0 = 0 · · 0 · liste · 1 · · 0 · élément · 2 a la pile σ0 = 0 · 1 tandis que l' exécution partielle v1 = v0 · · 3 · · 4 a une pile σ1 = ε.

Machine Translated by Google

Lemme 20. Considérons un SHA A = (Σ, Q, Δ, I, F), indique q, q un préfixe ' Q, et une pile σ Q . Laissez V être d'un mot imbriqué dans nw(HΣ). Alors, il existe une exécution partielle r par rapport à Δ de q à q pile σ telle que ' avec projΣ (r) = v ssi (q,ε) v−→ str ' (q , σ) par rapport à Δ.

Preuve. Ceci est réalisé par récurrence sur la longueur du préfixe v. □

Ce lemme implique que toute transition en mémoire de A sur une haie h de q à q ' peut être identifié avec une transition en continu de A sur nw(h) de (q,ε) à (q ' ,ε).

h Proposition 8. q −→ q ' par rapport à Δ (q,ε) nw(h) −−−→ str ' (q ,ε) par rapport à Δ.

h Preuve. Si q par rapport à Δ, il existe une séquence R de h avec Δ qui commence par q et se termine par ' q. Considérons l'exécution partielle v = nw(R) sur h. Puisque v est un mot imbriqué, sa pile est vide. D'après le lemme 20, nous avons (q,ε) projetΣ (v) −−−−−→ str ' (q ,ε) par rapport à Δ et projΣ (v) = nw(h). Pour l'inverse implication, supposons que (q,ε) existe nw(h) −−−→ str ' (q ,ε) par rapport à Δ. Soit u = nw(h). D'après le lemme 20, là avec une course partielle v de Δ de q à q ' pile ε. Alors, il existe un parcours R de Δ sur h tel h que v = nw(R). Donc q −→ q ' par rapport à Δ. □

Pour tout dSHA↓ A déterministe, couverture h et état q, la transition streaming sur nw(h) par rapport à Δ commençant par (q,ε) peut être calculée en temps O(1) par lettre après un précalcul en temps O (|UNE|). Ainsi, le temps de calcul global est en O(|A| + |h|). La mémoire de streaming nécessaire pour stocker une configuration est de taille O(profondeur(h) + |A|) puisque la taille de la pile visible est limitée par la profondeur de la haie d'entrée.

9.2. Ajout d'une projection de sous-couverture

Nous étendons ensuite la relation de transition en continu avec la projection de sous-couverture, par analogie avec ce que nous avons fait pour la relation de transition en mémoire. On définit une relation de transition avec projection de sous-couverture h str ' K × K par rapport à Δ, tel que pour tous les mots v, v −→ shp ' Σˆ , lettres a Σ, états p, q, q ' , q » Q, et les piles σ, σ ' Q :

$$\frac{q \quad Q\Delta_{shp} \qquad v \quad nw(H\Sigma)}{(q,\sigma)\ v{-}{\to}_{shp}^{str}\ (q,\sigma)\ \text{par rapport à } \Delta}$$

$$\frac{q \quad Q\Delta_{shp} \qquad q\ a{-}{\to}\ q' \ \text{en } \Delta}{(q,\sigma)\ a{-}{\to}_{shp}^{str}\ (q\ ,\sigma)\ \text{par rapport à } \Delta} \qquad \frac{q \quad Q\Delta_{shp}}{(q,\sigma)\ \varepsilon_{\Delta}\ (q,\sigma)\ \text{par rapport à } {-}{\to}_{shp}^{str}}$$

$$\frac{q \quad Q\Delta_{shp} \qquad (q,\sigma)\ v{-}{\to}_{shp}^{str}\ (q\ shp\ ,\sigma')\ \text{par rapport à } \Delta \quad (q',\sigma')\ {-}{\to}_{shp}^{v\ str}\ (q'',\sigma'')\ \text{par rapport à } \Delta}{(q,\sigma)\ {\xrightarrow{v\cdot v'}}_{shp}^{str}\ (q'',\sigma'')\ \text{par rapport à } \Delta}$$

$$\frac{q \quad Q\Delta_{shp} \qquad {-}{\to}\ q\ q' \ \text{en } \Delta}{(q,\sigma)\ {-}{\to}_{shp}^{str}\ (q\ ,\sigma\cdot q')\ \text{par rapport à } \Delta} \qquad \frac{p \quad Q\Delta_{shp} \qquad q@p \to q' \ \text{en } \Delta}{(p,\sigma\cdot q)\ {-}{\to}_{shp}^{str}\ (q\ ,\sigma)\ \text{par rapport à } \Delta}$$

La relation de transition de projection reste dans une configuration avec un état de projection de sous-couverture jusqu'à ce que la fin de la sous-couverture actuelle soit atteinte. Ceci est correct puisqu'un état de projection de sous-couverture ne peut de toute façon pas être modifié.

Machine Translated by Google

Proposition 9. Pour tout SHA↓ A = (Σ, Q, Δ, I, F), états q, q n'a pas' Q, mot imbriqué v sur lequel Δ d'exécution partielle bloquante à partir de q :

$$(q,\varepsilon)\ v{-\!\!\rightarrow}^{str}\ (q',\varepsilon) \text{ par rapport à } \Delta \quad (q,\varepsilon)\ v{-\!\!\rightarrow}^{str}_{shp}\ (q',\varepsilon) \text{ par rapport à } \Delta.$$

Preuve. Soit h HΣ la haie telle que v = nw(h). La preuve se fait par récurrence sur la taille de v. Pour l'implication de gauche à droite on suppose que $(q,\varepsilon)\ v{-\!\!\rightarrow}^{str}\ (q',\varepsilon)$ par rapport à Δ. Puisque v = nw(h), la proposition 8 prouve $q\ h{-\!\!\rightarrow}^{str}\ q'$ par rapport à Δ.

Cas q QΔ D'après le lemme 6, il découle de $q\ h{-\!\!\rightarrow}^{str}_{shp}\ q'$ par rapport à Δ que q = q'. Nous pouvons alors postuler la règle suivante puisque v est un mot imbriqué :

$$\frac{q \quad Q\Delta \text{ chp} \qquad v = nw(H\Sigma)}{(q,\varepsilon)\ v{-\!\!\rightarrow}^{str}\ (q,\varepsilon) \text{ par rapport à } \Delta \text{ shp}}$$

Donc $(q,\varepsilon)\ v{-\!\!\rightarrow}^{str}_{shp}\ (q,\varepsilon)$ par rapport à Δ.

Cas q QΔ shp. On distingue toutes les formes possibles de haie h.

Sous-cas h = h ' · h ». Soit v) ' et v=La règle suivante » = nw(h ") donc v = v ' v See More = nw(h) . doit avoir été appliquée :

$$\frac{(q,\varepsilon)\ {-\!\!\rightarrow}^{v'}\ (p,\sigma) \text{ par rapport à } \Delta \quad (p,\sigma)\ {-\!\!\rightarrow}^{v''}\ (q',\varepsilon) \text{ par rapport à } \Delta}{(q,\varepsilon)\ {-\!\!\rightarrow}^{v' \cdot v''}\ (q',\varepsilon) \text{ par rapport à } \Delta}$$

Depuis v ' est un mot imbriqué, il s'ensuit que σ = ε. Par l'hypothèse d'induction esis appliqué aux mots imbriqués v ' et v ', on obtient $(q,\varepsilon)\ {-\!\!\rightarrow}^{v'}_{shp}$ (p,ε) par rapport à Δ et $(p,\varepsilon)\ {-\!\!\rightarrow}^{v''}_{shp}\ (q',\varepsilon)$ wrt Δ. Par conséquent, nous pouvons appliquer la règle suivante :

$$\frac{q \quad Q\Delta_{shp} \quad (q,\varepsilon)\ {-\!\!\rightarrow}^{v'}_{shp}\ (p,\varepsilon) \text{ par rapport à } \Delta \quad (p,\varepsilon)\ {-\!\!\rightarrow}^{v''}_{shp}\ (q',\varepsilon) \text{ par rapport à } \Delta}{(q,\varepsilon)\ {-\!\!\rightarrow}^{v' \cdot v''}_{shp}\ (q',\varepsilon) \text{ par rapport à } \Delta}$$

Cela montre $(q,\varepsilon)\ v{-\!\!\rightarrow}^{str}_{shp}\ (q',\varepsilon)$ selon les besoins.

Sous-cas h = h 'a ou h = ε. Simple.

Pour l'implication de droite à gauche on suppose que $(q,\varepsilon)\ v{-\!\!\rightarrow}^{str}_{shp}\ (q',\varepsilon)$ par rapport à Δ.

Cas q QΔ shp. Alors, la règle suivante doit avoir été appliquée avec q = q ' :

$$\frac{q \quad Q\Delta_{shp} \qquad v = nw(H\Sigma)}{(q,\varepsilon)\ v{-\!\!\rightarrow}^{str}\ (q,\varepsilon) \text{ par rapport à } \Delta \text{ shp}}$$

Puisque nous supposons que v n'a pas d'exécutions partielles bloquantes avec Δ, il existe une exécution partielle sur v avec Δ qui commence par q. Puisque v est un mot imbriqué, tout partiel

Machine Translated by Google

exécuter sur v est le mot imbriqué d'une exécution, il existe donc une exécution sur la haie h qui commence par q. Cela montre l'existence d'un état p tel que q −→ p par rapport à ∆. Puisque q QΔ h

Le lemme 6 montre q = p, et donc q −→ q par rapport à ∆. La proposition 8 prouve alors shp, h

que (q,ε) $\xrightarrow{nw(h)}$ (q,ε) par rapport à ∆, qui est égal à (q,ε) v−→ str (q ',ε) par rapport à ∆, selon les besoins. str

Cas q QΔ On distingue toutes les formes possibles de la haie h. shp.

Sous-cas h = h ' · h »règle ' = nw(h 'Soit v La » = nw(h ") donc v = v) et v . ' = nw(h).

suivante doit avoir été appliquée :

$$\frac{q \quad Q\Delta \text{ shp} \qquad (q,\varepsilon) \xrightarrow{v'}_{str} (p,\sigma) \text{ par rapport à } \Delta \quad (p,\sigma) \longrightarrow_{shp} \xrightarrow{v''}_{str} (q',\varepsilon) \text{ par rapport à } \Delta}{(q,\varepsilon) \xrightarrow{v'v''}_{str} \longrightarrow_{shp} (q',\varepsilon) \text{ par rapport à } \Delta}$$

Depuis v ' est un mot imbriqué, il s'ensuit que σ = ε aussi. Par l'hypothèse d'induction

esis appliqué aux mots imbriqués v ' et v ', on obtient (q,ε) $\xrightarrow{v'}_{str}$ (p,ε) par rapport à ∆ et

(p,ε) $\xrightarrow{v''}_{str}$ (q ',ε) par rapport à ∆. On peut donc appliquer la règle suivante :

$$\frac{(q,\varepsilon) \xrightarrow{v'}_{str} (p,\varepsilon) \text{ par rapport à } \Delta \quad (p,\varepsilon) \xrightarrow{v''}_{str} (q',\varepsilon) \text{ par rapport à } \Delta}{(q,\varepsilon) \xrightarrow{v'v''}_{str} (q',\varepsilon) \text{ par rapport à } \Delta}$$

Cela montre (q,ε) v−→ str (q ',ε), selon les

Sous-cas h = h ' ou h = a ou h = ε. Direct. besoins.

Ceci termine les preuves des deux directions. □

Un évaluateur de streaming avec projection de sous-couverture pour SHA↓ s A déterministe sur les couvertures h peut ainsi être obtenu en calculant la relation de transition de streaming avec projection de sous-couverture pour A de nw(h) en commençant par la configuration initiale. Cela coûte au plus du temps O(1) par lettre de nw(h), c'est-à-dire un temps constant par événement du flux.

9.3. Complexité du streaming des premières adhésions avec projection de sous-couverture

À l'aide d'évaluateurs de streaming pour les SHA↓ , nous pouvons tester l'adhésion la plus ancienne à la projection de sous-couverture en mode streaming en exécutant le SHA↓ A cgr(S)e(S) .

Corollaire 3. Pour tout dSHA complet (Σ, Q, ∆, I, F), F FS Q, langage L = L(A) et schéma S = L(A[F/FS]), première appartenance à la projection complète de sous-couverture pour L par rapport à S peut être testée en mode streaming pour toute couverture h S dans le temps O(|1|) par lettre non projetée de nw(h) après un temps de prétraitement de O(|A cgr(S)e(S) | + n 3d + n s), où d est le nombre de (S). L'espace requis est dans des relations de différence et s est le nombre de sous-ensembles cgr sûrs dans les états de Q e(S)

O(profondeur(h) + |A cgr(S)e(S) |).

Preuve. D'après le théorème 4, le dSHA↓ A cgr(S)e(S) est le plus ancien, solide et complet pour la projection de sous-couverture. Par les propositions 8 et 9, nous pouvons ainsi obtenir un testeur d'appartenance streaming avec projection de sous-couverture pour une haie h S en exécutant l'évaluateur streaming avec projection de sous-couverture de A cgr(S)e(S) sur le mot imbriqué nw(h). Puisque ce SHA↓ est déterministe, l'évaluation en streaming peut être effectuée dans le temps O(|1|) par lettre de nw(h) après un temps de prétraitement de O(|A cgr(S)e(S) | + n 3d + ns).

□

Quant à la projection sans changement, le temps de prétraitement exponentiel peut à nouveau être évité en créant la partie du SHA↓ A cgr(S)e(S) nécessaire à l'évaluation de la haie à la volée. La taille de la pièce nécessaire est en O(| h|). Par conséquent, les besoins en espace peuvent également

Machine Translated by Google

être réduit à O(|h|), ce qui peut être trop pour les grands flux nw(h). La partie nécessaire de A cgr(S)e(S) , cependant, peut être beaucoup plus petite que |h|. De plus, une pile de taille profondeur(h) doit être maintenue mais cette dépendance à l'égard de h peut être acceptable même pour les grands flux nw(h).

10. Requêtes régulières monadiques et XPath

Nous considérons ensuite le problème de savoir comment répondre aux requêtes monadiques régulières sur les haies de manière précoce et avec une projection de sous-couverture. Chaque requête monadique définit un ensemble de positions pour chaque couverture d'entrée. Les requêtes XPath self::list[child::item], par exemple, peuvent être appliquées à la haie list. list . item . list. item , où elle sélectionne la position du premier arbre le plus haut étiqueté par liste car il contient un arbre enfant étiqueté par élément. Nous identifions les positions des couvertures par des entiers. Par exemple, l'élément de liste sélectionné dans la haie ci-dessus est identifié par l'entier 2. Par conséquent, l'ensemble de réponses de la requête ci-dessus sur la haie ci-dessus est {2}. Notez que les requêtes monadiques ne renvoient aucune autre information sur les positions sélectionnées, contrairement à la sémantique des sous-arbres de XPath dans la norme officielle du W3C.

Répondre à cette requête est plus compliqué que de vérifier si le filtre XPath [self::list/child::item] correspond à la racine du premier sous-arbre de la haie, puisque l' ensemble de réponses doit également être construit. Concernant la consommation de mémoire, les réponses candidates vivantes doivent être mises en mémoire tampon pour la réponse aux requêtes monadiques. Par conséquent, la consommation de mémoire des évaluateurs de flux de requêtes monadiques peut croître linéairement en taille du flux en fonction du nombre de candidats vivants.

10.1. Requêtes régulières monadiques

Les requêtes monadiques sur les haies dans HΣ peuvent être identifiées avec des langages annotés où x Σ dans HΣ {x} , sur les dans une variable de sélection arbitrairement fixée. Une requête monadique sur les couvertures couvertures dans HΣ peut alors être identifiée avec le langage de toutes les couvertures annotées dans HΣ {x} , dans lequel une seule position a été annotée x et cette position est sélectionnée par requête.

Par conséquent, les requêtes monadiques régulières sur les haies dans HΣ peuvent être définies par des expressions régulières imbriquées dans nRegExpΣ {x} ou par un dSHA avec la même signature. La requête XPath régulière self::list[child::item], par exemple, peut être définie en suivant l'expression régulière imbriquée :

liste · x · · · élément · · · · ,

où est comme un caractère générique acceptant toute couverture dans HΣ sans x, c'est-à-dire pour une variable de récursion z V :

= μz.(+a Σa + z) *.

Notez que l'expression régulière imbriquée pour le filtre XPath [self::list/child::item] dans nRegExpΣ peut être obtenue à partir de celle de la requête XPath self::list[child::item] en supprimant le x. De plus, pour toute requête monadique sur les haies, nous devons restreindre les haies annotées à être reconnues par le langage de la requête de telle sorte qu'exactement une position dans chaque haie soit annotée par x. Cela peut être fait par intersection avec l' expression régulière imbriquée suivante onex :

μz. (TxT + T. z .T)

Le document de schéma précédent pour la correspondance de modèles doit maintenant être modifié en S = docx ∩ onex, où docx nRegExpΣ {x} est comme doc mais sur la signature Σ {x}.

Pour la requête XPath standard self::list[child::item], nous obtenons le dSHA dans la figure 22. Les règles de transition avec la lettre spéciale x indiquent où les positions peuvent être liées à la variable de sélection : x ne peut être lié que dans l'état 2, c'est-à-dire dans les sous-arbres commençant par un élément de liste . Le même automate peut être utilisé pour définir le schéma S = docx ∩ onex en utilisant les états finaux du schéma FS = {14, 20} au lieu des états finaux F = {20}. En effet, nous avons construit l'automate en croisant l'automate obtenu à partir de l' expression XPath avec les automates calculés à partir des expressions régulières docx et onex pou

Machine Translated by Google

le schéma. Notez que 20 est un état final pour la requête (et le schéma), tandis que 14 est un état récepteur pour la requête (mais pas pour le schéma).

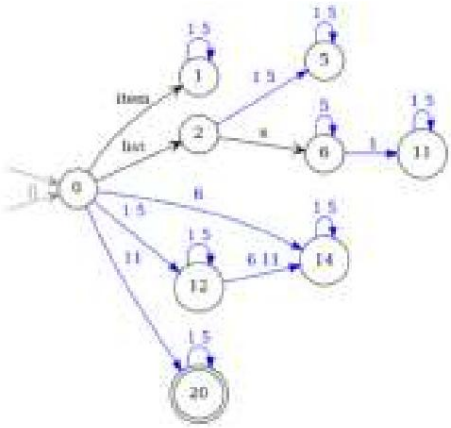


Figure 22. Le dSHA pour la requête XPath self::list[child::item]. La position de sélection est indiquée par x.

10.2. Première réponse à la requête avec projection de sous-

couverture Un algorithme naïf d'évaluation de requête pour un automate de couverture avec la signature Σ {x} reçoit une couverture d'entrée h HΣ, devine toutes les positions possibles de l'endroit où insérer x dans h et teste pour chacune d'elles si la La couverture annotée obtenue est acceptée par A. De manière moins naïve, on insère x uniquement aux positions où l'état actuel a une transition x sortante . Tout cela peut être amélioré si l'automate de couverture considéré prend en charge les états de sélection, de rejet ou de projection. Dans tous les cas, l'algorithme doit mettre en mémoire tampon tous les candidats de liaison qui ne sont pas encore dans un état de sélection ou de rejet. Dans le pire des cas, il doit tamponner les liaisons sur toutes les positions. Par conséquent, l'espace d'états de l'évaluateur de requêtes monadiques n'est plus fini.

Les états de sélection et de rejet peuvent être identifiés en adaptant le premier algorithme de réponse aux requêtes de [21]. Ceci peut être fait de telle sorte que l'on obtienne un SHA↓ Ae(S) de signature Σ {x} selon le schéma. Il convient de noter que les automates utilisés dans [21] sont des dNWA au lieu de SHA↓ , et que les schémas y sont ignorés. Les adaptations nécessaires sont donc discutées dans l'Annexe A. Comment obtenir un algorithme de réponse aux requêtes efficace en mode streaming à partir d' Ae(S) est le sujet principal de [21]. Il s'agit de l'algorithme de base que nous étendrons avec la projection de sous-couverture pour nos expériences.

Afin d'ajouter une projection de congruence au premier algorithme de réponse aux requêtes, nous devons exécuter l'automate A cgr (S). La première projection de congruence pour la requête XPath e(S) self::list[child::item] est illustrée dans la figure 23. Elle lie x dans l'état 2D1 après avoir lu un élément de liste de niveau supérieur , puis vérifie l'existence d'un élément d'élément enfant , passant dans l'état de sélection 1D3 une fois qu'il est trouvé. Tous les autres enfants de l'élément de liste de niveau supérieur sont projetés dans l'état 2D3. Pour ajouter la première requête répondant à la projection sans changement, au lieu de A cgr, nous devons exécuter l'automate π A e(S) . Nous ne donnons pas les détails, mais affirmons e(S) pour la présente requête que la puissance de projection des deux automates est la même.

Machine Translated by Google

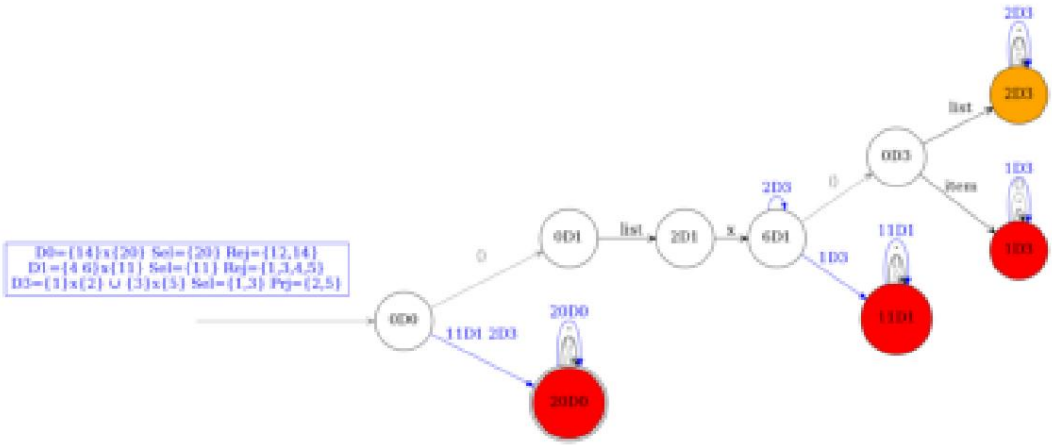


Figure 23. La première projection de congruence A cgr(doc ∩ onex)e(doc∩onex) pour le dSHA A de la figure 22 avec FS = {14, 20}. Il définit la requête du XPath self::list[child::item].

Heureusement, les théorèmes de solidité et d'exhaustivité pour les premiers tests d'appartenance avec projection de sous-couverture n'affectent que la construction de ces SHA↓ , mais pas la manière dont ils doivent être évalués. Cependant, il existe un problème supplémentaire : les préfixes de mots imbriqués ne peuvent pas être considérés comme non pertinents pour la sous-couverture suivante, si la variable de sélection x peut y être liée, même si l'acceptation ne dépend pas de l'endroit où elle sera liée. En effet, la position de la liaison doit être renvoyée par tout algorithme de réponse aux requêtes.

Définition 15. Nous appelons un préfixe v contraignant non pertinent pour L et S s'il n'existe aucune haie h contenant x et suffixe w tel que v · w nw(S) et v · nw(h) · w nw(L).

Notre évaluateur de projection de sous-couverture peut projeter uniquement sur les sous-couvertures des préfixes non pertinents qui sont également contraignants et non pertinents. Étant donné un dSHA A et un ensemble d'états finaux de schéma FS, on peut décider si un préfixe n'est pas pertinent pour la sous-couverture si l'état actuel q accepte les haies contenant x mais peut accéder à un autre état qui n'est pas sûr pour le rejet et autorise les liaisons x. . Dans le Datalog sol, nous pouvons distinguer les états non pertinents de liaison q en imposant les règles suivantes pour tous q, q

bind_irrelev(q) :- state(q), pas bind_relev(q). bind_relev(q) :- pas bind_x(q), acc(q,q'), bind_x(q'), pas rej(q').

Notez dans notre exemple de la requête XPath self::list[child::item] que l'état 2D3 du dSHA↓ A cgr(doc ∩ onex)e(doc∩onex) dans la figure 23 est une liaison non pertinente, puisque x doit être lié sur l'élément de liste de niveau supérieur dans l'état 2D1, il ne peut donc être lié à aucun élément de liste ci-dessous.

11. Expériences avec l'évaluation Streaming XPath Nous

comparons expérimentalement trois évaluateurs streaming pour dSHA, pour la première réponse aux requêtes monadiques, sans projection de sous-couverture, avec projection sûre sans changement et avec projection de congruence.

11.1. Benchmark dSHA pour les requêtes XPath

Nous partons de la définition dSHA des requêtes monadiques pour les requêtes XPath régulières A1-A8 du benchmark XPathMark [7] qui sont données dans le tableau 1. Ces requêtes XPath montrent la plupart des fonctionnalités du fragment régulier de XPath. Dans le tableau 2, nous avons ajouté 14 autres requêtes de chemin XPath que nous avons également trouvées utiles pour les tests.

Machine Translated by Google

Tableau 1. Requêtes XPath régulières A1-A8 de la collection XPathMark.

ID de requête	Expression XPath
A1	/site/closed_auctions/closed_auction/annotation/description/ texte/mot clé
A2	//enchères_fermées//mot-clé
A3	/site/closed_auctions/closed_auction//mot-clé
A4	/site/closed_auctions/closed_auction[annotation/description/ texte/mot clé]/date
A5	/site/closed_auctions/closed_auction [descendant ::mot-clé]/date
A6	/site/personnes/personne[profil/genre et profil/âge]/nom
A7	/site/personnes/personne[téléphone ou page d'accueil]/nom
A8	/site/personnes/personne[adresse et (téléphone ou page d'accueil) et (carte de crédit ou profil)]/nom

Tableau 2. Requêtes XPath régulières supplémentaires pour les documents XPathMark.

A0	/site
A1_0a	/site/*
A1_0b	/site/@*
A1_0c	/site//@*
A1_1a	//enchérisseur/personref[commence par (@person,'person0')]
A1_1d	//enchérisseur/personref[@person='person0']
A1_2	//personne
A1_3	/site/regions/afrique//@*
A1_4	/site/regions/afrique/*
A1_5	/site/régions/*
A1_6	//enchères_fermées/annotation//mot-clé
A2_1	//enchères_fermées[descendant::mot-clé]
A4_0	/site/closed_auctions/closed_auction[annotation]/date
A4_1	/site[open_auctions]/closed_auctions

Des SHA déterministes pour A1-A8 ont été fournis plus tôt dans [21]. Pour l'autre XPath requêtes, nous les avons compilées en dSHA via une expression régulière imbriquée.

Afin de produire les dSHA d'entrée pour nos évaluateurs, nous croisons ces automates avec un dSHA pour le schéma doc' $_x$ ∩ onex où doc' $_x$ est le schéma que nous utilisons pour la couverture encodages de documents XML du monde réel avec annotations x . Ainsi, nous pourrions identifier le schéma final indique FS. Nous avons minimisé et complété le résultat. Le tableau 3 indique la taille des dSHA d'entrée pour nos évaluateurs obtenus par la procédure ci-dessus. Pour chaque dSHA A, la taille est donnée par deux nombres size(n), le premier pour la taille globale et le second pour la nombre d'États n.

Pour le dSHA A d'entrée de chaque requête, nous avons calculé statiquement l'ensemble des dSHA↓ s Un cgr(S)e(S) tout en utilisant les parties nécessaires de l'algorithme de détermination de [16]. Le la taille des dSHA↓ s obtenues et le nombre d de relations de différence sont rapportés dans le tableau 3 aussi. La plus grande taille est 2091(504) pour le dSHA d'entrée pour A8. Le plus grand nombre de différence les relations d = 24 sont également obtenues pour cette requête. Ainsi, en effet, la taille de ces automates est bien

Machine Translated by Google

plus petit que ce à quoi on pourrait s'attendre d'une construction très exponentielle dans le pire des cas cas. Cependant, la seule requête pour laquelle nous n'avons pas pu obtenir son correspondant déterministe

la première projection de congruence A cgr(S)e(S) est A5.

Le temps nécessaire pour calculer la première projection de congruence a pris entre 2,3 et 26 s.

Tableau 3. Taille globale et nombre n d'états du dSHA d'entrée A pour la requête avec schéma final états FS, la taille globale et le nombre d'états du SHA↓ A cgr(S)e(S) obtenus par congruence la plus ancienne projection, et le nombre d ses relations de différence.

Requête dSHA A Taille(n) Taille(#État) Rel. d ID ID	dSHA↓ #Diff-	Requête dSHA A	dSHA↓ #Diff-Size(n) Taille(#State)	Rel. d
A1 482(68)	1296(324)	16 A1_0c 230(43)	238(62)	6
A2 224(42)	316(82)	7A1_1a 305(54)	384(101)	8
A3 320(53)	662(156)	10A1_1d 305(54)	382(101)	8
A4 629(74)	1651(404)	18 A1_2 194(39)	152(42)	5
A5 438(63)	1226(269)	13A1_3 318(53)	672(159)	10
A6 675(76)	2090(500)	22 A1_4 312(52)	479(132)	10
A7 394(59)	728(184)	12 A1_5 266(47)	293(84)	8
A8 648(79)	2091(504)	24 A1_6 265(47)	588(142)	9
A0 203(40)	158(44)	5A2_1 232(42)	295(78)	7
A1_0a 224(42)	145(44)	6A4_0 392(59)	719(184)	12
A1_0b 203(39)	68(23)	5A4_1 292(49)	287(78)	8

Nous notons que nous n'avons pas encore calculé statiquement les dSHA↓ complets pour la projection sans changement A snc, la projection de congruence pure A cgr(S) et la première réponse aux requêtes. Ae(S) . Nous pensons que ces automates pourraient devenir plus grands que A cgr(S)e(S) que nous avons construit d'un seul coup.

11.2. Outil d'évaluation en continu : AStream

Nous utilisons et développons l'outil AStream pour répondre aux requêtes monadiques dSHA sur les flux XML avec des restrictions de schéma. La version 1.01 d'AStream a été présentée dans [21] prend en charge la première réponse aux requêtes sans projection. Étant donné un dSHA A définissant une monadique requête, un ensemble d'états finaux de schéma FS et un flux XML w, il construit à la volée le

avait besoin d'une partie du SHA↓ Ae(S) lors de l'analyse de w et de l'évaluation de A sur la couverture codant w.

Pour la version conférence FCT du présent article [24], nous avons amélioré AStream avec

projection sûre et sans changement. Cela conduit à la version AStream 2.01. Il construit la première projection sûre et sans changement dSHA↓ A $_{e(S)}^{snc}$ à la volée lors de l'évaluation de la requête monadique définie par A sur la haie codant pour le flux XML w. La projection de sous-haie peut être activée ou

off afin de comparer les deux versions sans autres différences d'implémentation.

Pour cette version actuelle de la revue, nous avons ajouté la première projection de congruence et l'avons intégrée dans AStream menant à AStream 3.0.

AStream 3.0 est différent des deux versions précédentes dans la mesure où le dSHA↓ A cgr $_{e(S)}^{est}$ construit de manière statique et entièrement, indépendamment de la haie d'entrée. On utilise alors un générique le premier évaluateur de streaming pour les SHA↓ qui rejette dans les états de rejet, sélectionne dans la sélection États et projets dans tous les États de projection de sous-couverture. Cet évaluateur pourrait également être exécuté avec $_{UN}^{snc}{}_{e(S)}$ ou Ae(S) , à condition qu'ils ne deviennent pas trop gros.

Il convient de noter que A cgr $_{e(S)}$ s'est avéré bien petit pour l'ensemble de notre benchmark, avec les autres dSHA↓ s A $_{e(S)}^{snc}$ et Ae(S) risquent de devenir plus gros. Comme indiqué précédemment, nous n'avons pas construisez ces dSHA↓ jusqu'à présent pour nos requêtes de référence. C'est pourquoi nous avons continué à courir

Machine Translated by Google

le premier streaming avec projection sûre et sans changement avec AStream 2.01, tandis que nous avons utilisé AStream 3.0 pour le premier streaming avec projection de congruence.

Toutes les versions d'AStream s'appuient sur abc-Datalog de Java pour calculer les points les moins fixes des programmes Datalog de manière ascendante. Les programmes d'enregistrement de données sont nécessaires pour tout raisonnement logique : pour calculer des sous-ensembles d'états qui peuvent être rejetés, sélectionnés ou projetés en toute sécurité à tous les niveaux. De plus, les relations de différence sont calculées sur la base du Datalog. Notez que la première évaluation de requête à la volée nécessite l'exécution de Datalog pendant l'évaluation de la requête, tandis qu'avec une approche statique pour construire entièrement les dSHA↓ , Datalog n'est nécessaire qu'au moment du prétraitement pendant la construction de l'automate.

11.3. Mesures d'évaluation

Nous souhaitons mesurer le temps d'exécution des trois évaluateurs de requêtes en streaming pour tous dSHA obtenus à partir des expressions XPath de la collection des tableaux 1 et 2.

L'un des avantages du benchmark XPathMark est qu'il est livré avec un générateur de documents XML auxquels les requêtes peuvent être appliquées et dont la taille peut être mise à l'échelle. Nous avons créé un document XML d'une taille de 1,1 Go, suffisamment grande pour rendre le streaming obligatoire. Nos expériences montrent que l'efficacité de l'évaluation des requêtes croît linéairement avec la taille de la partie non projetée du document XML . Cela vaut pour chacun de nos trois évaluateurs. Par conséquent, mesurer le temps de l'évaluateur sur des documents XML d'autres tailles ne donnerait aucun nouvel aperçu, sauf que la consommation de mémoire reste également modérée.

Le gain de temps de projection pour une requête Q est le pourcentage de temps économisé par la projection lors de l'évaluation de la requête en ignorant le temps d'analyse du document pur en secondes :

$$\text{gain de temps}_Q = 100 \quad (1 - (t_{\text{noproject}Q} - t_{\text{parse}})/(t_{\text{project}Q} - t_{\text{parse}})).$$

Nous pouvons mesurer le gain de temps pour la première projection sûre sans changement par AStream 2.01 et pour la première projection de congruence par AStream 3.0, puisque la projection peut être désactivée dans les deux cas. L'inconvénient du gain de temps est qu'il dépend des détails de mise en œuvre.

Le gain d'événement est une meilleure mesure de la puissance de projection. L'analyseur envoie des événements nécessaires au SHA↓ , un flux tout en ignorant les parties du document d'entrée qui peuvent être projetées. Pour cela, il doit toujours être informé par l'automate du type d' événements qui peuvent être projetés dans l'état actuel. Le gain d'événement de la projection event-gain$_Q$ est alors le pourcentage d'événements reconnus comme non pertinents et donc non plus créés par un évaluateur de projection pour la requête Q. On pourrait s'attendre à ce que le gain de temps soit toujours inférieur au gain d'événement.

$$\text{time-gain}_Q \leq \text{event-gain}_Q$$

En effet, cela sera confirmé par nos expériences. Nous pensons que ces écarts entre ces deux mesures indiquent combien il reste de marge pour optimiser la mise en œuvre d'un évaluateur. De cette façon, nous pouvons constater qu'il reste encore une certaine marge pour d'autres optimisations d'AStream 3.0.

11.4. Expériences

Nous avons utilisé l'interface XMLStreamReader v1.0 de Java du package javax.xml.stream pour analyser et diffuser des fichiers XML dans Scala. L'analyse d'un document XML de 1,1 Go a nécessité $t_{\text{parse}} = 15$ s, tandis que l'interroger sans projection nous a pris en moyenne $t_{\text{noproject}_{\text{avg}}} = 752$ s, tout en variant de 600 à 900 s. Le temps de traitement moyen pour 1 Mo était ainsi de 0,72 s. Une fois que nous avons su cela, nous avons prédit le temps d'évaluation attendu à partir de la taille de la partie non projetée du document XML . C'est le suivant :

$$\text{event-gain}_Q \quad 1100 \quad 0{,}72 \text{ s} + t_{\text{parse}} \text{ s}.$$

Sans projection, l'analyseur génère 1 012 018 728 événements pour le document XML de 1,1 Go . Il s'agit de la base de calcul du gain d'événement Q pour toutes les requêtes Q.

Machine Translated by Google

Nous avons exécuté AStream avec Scala v2.13.3 sur une machine avec le système d'exploitation Ubuntu 20.04.06 LTS (64 bits). La machine est une machine Dell Precision-7750, équipée d'un Processeur Intel® Core™ i7-10875H à 2,30 GHz × 16 et 32 Go de RAM.

11.4.1. Première projection de congruence

Nous présentons les mesures de notre évaluateur avec la première projection de congruence en Tableau 4. Le gain d'événement est élevé, variant entre 75,7% pour A1_0c et 100% pour A0, A1_0a, A1_0b, A1_4, A1_5 et A4_1. Le gain d'événement pour les requêtes sans axe descendant est supérieur 98,2%. Pour ces requêtes, tous les sous-arbres qui ne se trouvent pas sur le chemin principal de la requête peuvent être projeté au loin. En particulier, seuls les sous-arbres doivent être inspectés jusqu'à une profondeur fixe. Ces sont les requêtes A1, A4, A6-A8 et A4_0 et les requêtes avec gain d'événement 100% listées ci-dessus. Le gain de temps pour toutes ces requêtes est un peu inférieur au gain événement mais tout de même supérieur 98,2%.

Tableau 4. Gain de temps et gain d'événements par première projection de congruence avec AStream3.01 sur un disque de 1,1 Go Flux XPathMark.

Requête IDENTIFIANT	#Ans- Wer en s	Heure Gain en s	Événement horaire Gagner	Requête IDENTIFIANT	#Ans- Nous	Heure	Événement horaire Gagner
A1	38 267	29,3	98,1% 98,9%	A1_0c	3 600 816	226,6	72,9% 75,7%
A2	117 389	171,8	77,9% 81,1%	A1_1a	2	187,8	78,1% 80,3%
A3	117 389	33,4	97,4% 97,8%	A1_1d	2	217,3	74,8% 80,3%
A4	24 959	33,2	97,8% 98,9%	A1_2	1 062 965	238,2	68,2% 76,0%
A5	50 186			A1_3	25074	16,2	99,1% 99,8%
A6	30 307	45,9	96,9% 98,2%	A1_4	5170	17,5	99,7% 100,0%
A7	179 889	30,5	98,0% 98,7%	A1_5	6	14,3	100,0% 100,0%
A8	67 688	37,8	97,3% 98,7%	A1_6	117 389	188,7	75,4% 81,1%
A0	1	21,3	99,1 % 100,0 %	A2_1	50 186	199,7	76,5% 81,1%
A1_0a	6	15,2	99,9 % 100,0 %	A4_0	91 650	21,8	99,0% 99,3%
A1_0b	0	0,0	100,0 % 100,0 %	A4_1	1	12,0	100,0 % 100,0 %

Nous considérons ensuite le deuxième type de requêtes ayant l'axe descendant, plus précisément A2, A1_0c, A1_1a, A1_1d, A1_2, A1_6 et A2_1. Intuitivement, un gain d'événement plus faible est attendu dans ces cas car la projection de sous-couverture avec l'axe descendant ne peut pas exclure directement un sous-arbre entier sous un élément donné ; tous les éléments descendants doivent être inspectés. Par exemple, la requête A1_0c égale à /site//@* doit sélectionner tous les attributs de tous les éléments sous le site de l'élément racine, nécessitant l'inspection de chaque élément XML du document, pour la présence d'attribut, à l'exception du site de l'élément racine .

Le gain d'événement inférieur signalé pour les requêtes susmentionnées correspond à ce qui précède. attente, affichant des pourcentages allant de 75,7% à 81,1%. De plus, ces requêtes présentent un écart plus important entre le gain de temps et le gain d'événement, qui varie de 2,2% à 5,7% et faire l'objet d'une optimisation future. Il convient de noter que certaines de ces requêtes combinent plusieurs fonctionnalités à la fois, comme les requêtes A1_1a et A1_1d, qui ont un axe descendant, filtre avec attributs et comparaison de chaînes.

La seule exception à cette tendance pour les requêtes ayant un axe descendant sont A3 et A1_3, avec des gains d'événements respectifs de 97,8% et 99,8%. La raison derrière ce gain élevé est que ces requêtes partent de la racine, avec un chemin d'axe enfant de profondeur 3 avant d'interroger le axe descendant. Ce chemin initial permet l'exclusion et la projection de la plupart des éléments sous le site qui ne correspond pas au chemin spécifié avec l'intégralité de leurs sous-arbres.

Machine Translated by Google

11.4.2. Première projection sûre et sans changement

Nous comparons la première projection sans changement avec la première projection de congruence dans le tableau 5. Pour cela, nous nous limitons aux requêtes A1 à A8 du XPathMark dans Tableau 1.

Les requêtes XPath A1, A4, A6, A7 et A8 n'ont pas d'axes descendants. Le temps le gain pour la projection sûre et sans changement sur ces requêtes se situe entre 92,3 et 98,9 %. Pour le plus tôt Selon la projection de congruence, le gain de temps était meilleur pour A1, A4 et A6 avec au moins 96,6 %. Pour A7 et A7, le gain de temps de la projection sans changement est proche du gain d'événement de projection de congruence mais le gain de temps de la projection de congruence est inférieur de 1,5 %. Alors, nous j'espère qu'une version optimisée de la projection de congruence pourrait surpasser la sécurité sans changement projection dans tous les cas.

Les requêtes XPath avec axe descendant sont A2, A3 et A5. Pour A2 et A3 en En particulier, le gain de temps de la projection sans changement est très faible, avec même pas 10,9 % alors que la projection de congruence donne au moins 77,9 %. Nous pensons qu'il s'agit d'un bug dans notre système actuel mise en œuvre d'une projection sûre et sans changement, qui peut interdire la projection de attributs et nœuds de texte pour ces requêtes (mais pas pour les autres). Projection de congruence sur A2 gagne 77,9% de temps et 81,1% d'événements. C'est bien mieux, mais c'est encore loin du puissance de projection pour les requêtes sans axes descendants. Pour A3, la projection de congruence gagne 97,4% du temps. C'est bien mieux que pour A2 puisque l'axe descendant en A3 est à la fin, tandis que pour A2 c'est au début du chemin. Mais même pour A2, la projection de congruence est très utile.

Tableau 5. Gain de temps pour la première projection de sécurité sans changement et de congruence la plus précoce, et gain d'événement de la première projection de congruence.

Temps de requête (s)	Temps (s)		Gain de temps		Gain d'événement
IDENTIFIANT	SNC	Congrès	SNC	Congrès	Congrès
A1	72,8	29.3	92,3%	98,1%	98,9%
A2	664,6	171,8	8,5%	77,9%	81,1%
A3	666.1	33.4	10,9%	97,4%	97,8%
A4	78,5	33.2	92,4%	97,8%	98,9%
A5	77,7		92,3%		
A6	65.2	45,9	95,1%	96,9%	98,2%
A7	24.6	30,5	98,8%	98,0%	98,7%
A8	24,8	37,5	98,9%	97,3%	98,7%

11.4.3. Comparaison avec des outils externes

QuiXPath s'est révélé être l'outil externe à large couverture le plus efficace pour les Évaluation XPath sur les flux XML [4]. Par conséquent, nous comparons l'efficacité de la congruence projection vers QuiXPath dans le tableau 6, et donc, par transitivité vers les nombreuses alternatives outils. Nous notons que QuiXPath effectue des réponses anticipées aux requêtes par compilation pour éventuellement RNF non déterministes avec des états de sélection, sans toujours être les plus précoces. En dehors de pour cela, il fonde son efficacité à la fois sur la projection de sous-arbre et sur la projection descendante.

Machine Translated by Google

Tableau 6. Comparaison des délais sans analyse en secondes pour les requêtes XPathMark avec QuiXPath et première projection de congruence.

Requête	Fract sans analyse	sans analyse.		Requête	Fract sans analyse	sans analyse.	
IDENTIFIANT	Temps (s)	Temps (s)		IDENTIFIANT	Temps (s)	Temps (s)	
	QuiXPath	Congrès.	Prj.		QuiXPath	Congrès.	Prj.
A1	11.0	14.3	1.3	A2	11.4	156,8	13.8
A4	11.6	18.2	1.6	A3	11.5	18.4	1.6
A6	10.7	30,9	2.9	A5	12,0		
A7	8.6	15,5	1.8				
A8	8.8	22,5	2.6				

Les expériences avec QuiXPath dans [4] utilisent le temps sans analyse. Par conséquent, nous faisons idem pour la projection de congruence en réduisant le temps global mesuré de 15 s de temps d'analyse.

Par rapport aux temps sans analyse pour les requêtes XPath sans axe descendant, la première projection de congruence démontre des améliorations significatives. En moyenne, notre projection est seulement plus lente d'un facteur de 1,3 à 2,9 que QuiXPath. Cela montre que notre la mise en œuvre est déjà sur le point d'être compétitive avec les outils de streaming XML existants , tout en étant le seul à garantir une réponse rapide aux requêtes.

Pour la requête A2 avec un axe descendant, la projection de congruence est 13,8 fois plus lente que QuiXPath. La raison est que QuiXPath supporte la projection descendante, alors que la projection de congruence ne concerne que les sous-couvertures. Pour la requête A3 avec un axe descendant à la fin, la projection de congruence n'est en retard que d'un facteur 1,6, donc la projection descendante semble moins pertinent ici.

12. Modèles d'automates associés

Un grand nombre de notions alternatives d'automates pour les arbres, les forêts et les haies ont été proposées dans la littérature. Nous les comparons aux SHA et aux SHA↓ dans la sous-section suivante et discuter des implications.

12.1. SHA contre automates d'arbres, de forêts et de haies

Automates de couverture standard. Les automates de couverture standard [9,10,32,33] fonctionnent sur des les couvertures de manière ascendante, à l'instar des SHA. Ils ont aussi la même expressivité, mais les langages horizontaux sont spécifiés différemment, ce qui conduit à une notion problématique de déterminisme [34]. Pour cette raison, la minimisation unique échoue pour la norme déterministe automates de haie. Pourtant, ils ont la même expressivité que les SHA lorsqu'ils sont limités à haies labellisées.

Automates d'arbres standard et minimisation SHA. Syntaxiquement, tout SHA A est un standard automate arborescent sur la signature classée suivante :

- un symbole unaire pour toute lettre a Σ ;
- un symbole binaire @ ;
- une constante .

Sémantiquement, il n'existe pas de correspondance générale parfaite. Uniquement pour les sous-classes, un parfait la correspondance peut être établie via un codage binaire. Cela a été montré dans [8] pour la sous-classe de dSHA avec I = Δ. Dans [35], cela pourrait également être montré pour la sous-classe des multi-modules dSHA. Cela conduit à des résultats de minimisation uniques pour les deux sous-classes de méthodes déterministes. SHA.

Les automates forestiers de Bojanczyk. Les automates de haie standards ont également la même expressivité comme les automates forestiers de Bojanczyk (voir la section 3.3 de [36]). Ceux-ci sont basés sur l'idée de monoïdes de transition, plutôt que sur l'idée de règles de transition telles que les SHA ou les

Machine Translated by Google

automates d'état ou automates forestiers de Neuman et Seidl. En conséquence, il existe une différence exponentielle de concision entre les SHA et les automates forestiers de Bojanczyk [35].

12.2. SHA↓ contre SHA

Nous commençons à expliquer pourquoi les SHA↓ ont la même expressivité que les SHA. Fondamentalement, c'est la même raison pour laquelle les NWA ont la même expressivité que les SHA (voir par exemple [8]). Pour une direction plus simple, nous avons déjà soutenu que tout SHA A = (Σ, Q, Δ, I, F) peut être converti

l'inverse, on peut convertir tout SHA↓ A en un SHA équivalent . A éventuellement le non-déterminisme comme suit :

$$\text{Élim}\downarrow = je \times Q \qquad F^{\text{élimine}\downarrow} = F \times Q$$

$$\frac{-\!\!\rightarrow qq \quad ' \text{ en } \Delta}{-\!\!\rightarrow (q, q \quad ' \text{elim}\downarrow) \text{ dans } \Delta}$$

$$\frac{q\ a\!\!-\!\!\rightarrow \quad ' \text{ dans } \Delta\ r \quad Q}{q\ (r, q)\ a\!\!-\!\!\rightarrow (r, q \quad ' \text{elim}\downarrow) \text{ dans } \Delta} \qquad \frac{q@p \rightarrow q \quad ' \text{ dans } \Delta\ r \quad Q}{(r, q)@(q, p) \rightarrow (r, q \quad ' \text{elim}\downarrow) \text{ dans } \Delta}$$

Proposition 10. L(A) = L(elim↓ (A)).

La construction est analogue à la conversion de RNF en SHA [8] ou à la couverture automates [26]. Les preuves d'exactitude de ces compilateurs sont standard.

Minimisation unique. La minimisation unique échoue pour les dSHA↓ comme d'habitude pour les automates multivoies déterministes. Cela se produit même pour les automates déterministes à états finis bidirectionnels sur les mots. En revanche, la sous-classe des dSHA avec la restriction que l'état initial et l' état initial de l'arbre sont identiques bénéficie d'une minimisation unique [8].

Automates forestiers Pushdown de Neuman et Seidl. Les automates de haie standard sont également connus pour avoir la même expressivité que les automates forestiers à poussée de Neumann et Seidl [25]. Ces dernières peuvent être identifiées par des SHA↓ pour les haies labellisées. Ce qui est nécessaire pour mapper les automates forestiers pushdown aux automates de haie standard est l'élimination vers le bas, comme pour mapper les SHA↓ aux SHA.

12.3. SHA↓ par rapport aux NWA

L'évaluateur de streaming pour les SHA↓ via une machine visiblement pushdown peut également être obtenu en compilant un SHA↓ dans un automate de mots imbriqué (NWA) [11] dont l'évaluateur de streaming est fourni par une machine visiblement pushdown [12], précédemment connue sous le nom d'entrée. automates pilotés [13–15].

Définition 16. Un automate de mots imbriqué (NWA) est un tuple (Σ, Q, Γ, Δ, I, F), où Σ, Γ et Q sont des ensembles, I, F Q et Δ = ((a Δ)a Σ, Δ Δ)Q contient les relations suivantes : a Δ Q × (Γ × Q), et Δ , Q × Γ × Q. Une NWA est déterministe ou de manière équivalente une dNWA si je contient au plus un élément et toutes les relations ci-dessus sont des fonctions partielles.

Les éléments de Γ sont appelés symboles de pile. Les règles de transition dans Δ ont encore trois formes : les règles internes q a−→ q ↑γ et les ' ↓γ dans Δ comme pour les SHA, règles d'ouverture q −−→ q' dans Δ si Δ (q) = (q ', y), règles de clôture q −−→ q Les ' dans Δ si Δ (q, y) = q '.

évaluateurs de streaming pour les NWA. L'évaluateur de streaming pour les NWA peut être considéré comme une machine déroulante permettant d'évaluer les mots imbriqués des haies de manière continue. Une configuration de la machine pushdown est une paire dans K = Q × Γ contenant un état et une pile.*

Machine Translated by Google

Pour tout mot $v \in \Sigma^*$, nous définissons une relation de transition streaming $v \longrightarrow_{str} Q \times K$, tel que, pour tout $q, q' \in Q$ et $\sigma \in \Gamma^*$, ce qui suit est applicable :

$$\frac{\text{vrai}}{(q,\sigma) \xrightarrow{\varepsilon}_{str} (q,\sigma) \text{ par rapport à } \Delta} \qquad \frac{(q,\sigma)\, v \longrightarrow_{str} (q',\sigma) \quad (q',\sigma) \xrightarrow{v'}_{str} (q'',\sigma) \text{ par rapport à } \Delta}{(q,\sigma) \xrightarrow{v \cdot v'}_{str} (q'',\sigma) \text{ par rapport à } \Delta}$$

$$\frac{q\, a \longrightarrow q' \in \Delta}{q\, a \longrightarrow_{str} q' \text{ par rapport à } \Delta} \qquad \frac{q \longrightarrow qq' \in \Delta}{(q,\sigma) \longrightarrow_{str} (q',\sigma) \text{ par rapport à } \Delta} \qquad \frac{q \longrightarrow_{\uparrow\gamma} qq' \in \Delta}{(q,\sigma \cdot \gamma) \longrightarrow_{str} (q',\sigma) \text{ par rapport à } \Delta}$$

Des SHA↓ aux RNF. Pour tout SHA↓ $A = (\Sigma, Q, \Delta, I, F)$, on peut définir la NWA $A_{nwa} = (\Sigma, Q, \Gamma, \Delta_{nwa}, I_{nwa}, F_{nwa})$ tout en préservant le déterminisme, tel que $\Gamma = Q$, tandis que Δ_{nwa} contient pour tous $a \in \Sigma$ et $q, p \in Q$ les règles de transition :

$$\frac{q\, a \longrightarrow q' \in \Delta}{qq\, a \longrightarrow_{nwa} q' \in \Delta_{nwa}} \qquad \frac{q \longrightarrow_{\downarrow q} q' \in \Delta}{q \longrightarrow_{nwa} q' \in \Delta_{nwa}} \qquad \frac{q@p \longrightarrow_{\uparrow q} q' \in \Delta}{p \longrightarrow_{nwa} q' \in \Delta_{nwa}}$$

Par exemple, le dNWA A_{nwa} du dSHA↓ A sur la figure 9 obtenu à partir du dSHA pour le [self::list/child::item] de l'introduction est donné dans la figure 24. Le dNWA A du dSHA↓ A_{nwa} dans la figure 21 est donné dans la figure 25.

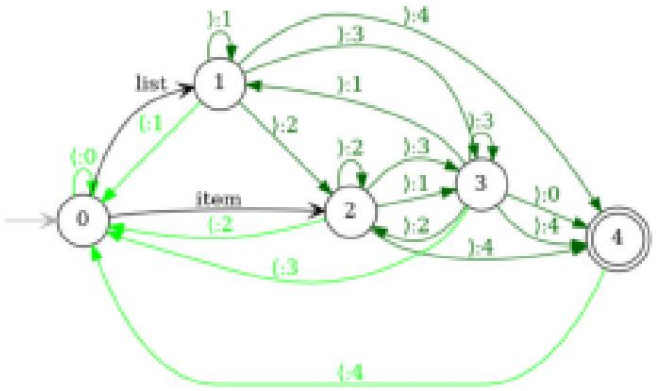
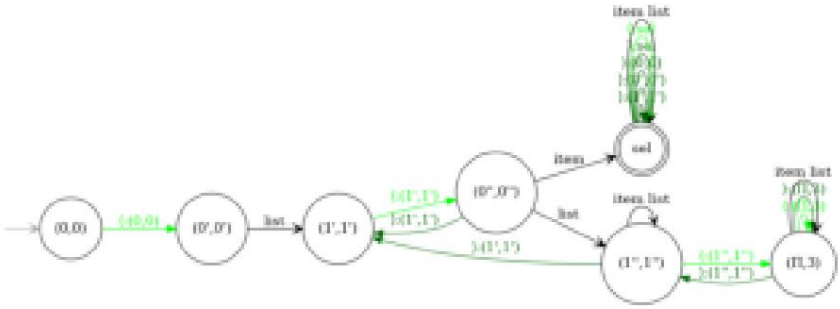


Figure 24. Le dNWA A_{nwa} pour le dSHA↓ A sur la figure 9 obtenu en appliquant l'opérateur down à dSHA de la figure 4 pour le filtre [self::list/child::item].



dNWA $(A_{snc})_{nwa}$ pour le premier dSHA↓ avec projection sûre sans changement dans la figure 25. Le e(doc) Graphique 21.

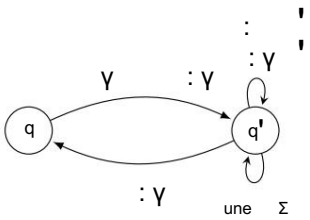
Machine Translated by Google

Lemme 21. L(A nwa) = L(A).

Les exécutions des SHA↓ s A et des NWA A^{nwa} peuvent être identifiés.

Projection pour les RNF. Des évaluateurs de projection pour les RNF ont été proposés dans le contexte de la projection des RNF [4]. Ils sont basés sur la notion suivante d'états de sous-arbres non pertinents pour les RNF.

Définition 17 (Variante de la Définition 3 de [4]). Nous appelons un état q d'une NWA un état de sous-arbres non pertinents s'il existe deux symboles de pile différents γ, γ tels que les transitions indiquées à droite existent mais aucune autre transition d'ouverture avec γ, aucune autre transition avec γ et aucune autre transition de fermeture dans q faisant éclater γ. Dans ce cas, on écrit q i-treeΣ\ .



Lemme 22. Tout état de projection de sous-couverture d'un SHA↓ A complet est un état de sous-arbres non pertinents d' Anwa .

Il a ensuite été montré dans [4] comment cartographier une NWA avec un sous-ensemble QΔ d'états de sous-arbres shp non pertinents pour une NWA en projection, dont la sémantique de streaming produit un évaluateur de streaming avec projection de sous-couverture. La présentation de la projection de sous-couverture pour les SHA↓ sans passer par les RNF donne le même résultat de manière plus directe.

Il convient également de noter que la projection des RNF prend en charge la projection descendante en plus de la projection de sous-couverture. Pour les SHA, cela est laissé à des recherches futures.

13. Conclusions et travaux futurs

Nous avons introduit la notion de sous-couvertures non pertinentes pour l'adhésion afin de protéger le langage sous les restrictions de schéma. Nous avons développé des algorithmes avec projection de sous-couverture qui sautent par-dessus les sous-couvertures non pertinentes pour tester l'appartenance aux langages de couverture habituels définis par les dSHA. Tous nos algorithmes peuvent être exécutés en mode mémoire et en mode streaming. La difficulté était de savoir comment transmettre les informations d'état fini nécessaires à la projection de sous-arbres de haut en bas, étant donné que les SHA fonctionnent de bas en haut. Nous l'avons résolu en nous basant sur des compilateurs allant des SHA aux SHA descendants .

Ce premier compilateur propage des informations de sécurité concernant les états non changeants. Le deuxième compilateur, que nous démontrons complet pour la projection de sous-couverture, propage les relations de différence. Nous avons ensuite combiné la projection de sous-couverture avec les premiers tests d'adhésion et la première réponse aux requêtes monadiques. Nous avons intégré à la fois la sécurité sans changement et la congruence complète dans notre outil AStream et avons confirmé leur utilité expérimentalement : nous avons montré qu'il peut effectivement accélérer l'outil AStream pour répondre aux requêtes XPath régulières sur les flux XML . De plus, la projection de congruence a montré des taux de gain de projection d'événements beaucoup plus élevés que la projection sans changement, compris entre 75,7 % et 100 % pour la classe de requêtes qui n'a pas d'axe descendant, devenant ainsi compétitive dans le temps avec les meilleurs outils de streaming existants pour XPath standard. requêtes sans axe récursif.

Une question théorique et pratique restant ouverte est de savoir comment obtenir une projection descendante pour les SHA, nécessaire pour accélérer l'évaluation des requêtes XPath avec un axe descendant. Du côté de la mise en œuvre et de l'expérimentation, il serait bien d'étudier également la version en mémoire de nos algorithmes de sous-couverture. Enfin, ce serait bien d'op

Machine Translated by Google

diffuser davantage les outils de réponse aux requêtes les plus précoces jusqu'à ce qu'ils surpassent les meilleurs outils de streaming dans tous les cas, tout en étant meilleurs pour les requêtes pour lesquelles les autres outils ne sont pas les plus précoces.

Contributions des auteurs : rédaction – version originale, AAS et JN. Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche n'a reçu aucun financement externe.

Déclaration de disponibilité des données : ensemble de données disponible sur demande auprès des auteurs.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Annexe A. Première adhésion Nous adaptons

le premier testeur d'appartenance pour les dSHA de [21] afin qu'il se compile en SHA↓ au lieu de NWA et qu'il prenne en compte les schémas.

Annexe A.1. La sécurité du schéma est un problème d'inclusion

L'idée pour tester certaines adhésions et certaines non-adhésions est de considérer que les états peuvent atteindre les états finaux en toute sécurité, sauf s'ils sortent du schéma. Soit A = (Σ, Q, Δ, I, F) un SHA de schéma S, tel qu'il existe F FS Q avec S = L(A[F/FS]). Notons que L(A) S HΣ en particulier. L'ensemble des états qui peuvent atteindre en toute sécurité un sous-ensemble P Q pour toutes les couvertures qui ne sortent pas du schéma S est le suivant :

$$\text{sûr}_{\Delta,S}(P) = \text{sûr}_{\Delta}(P \quad (Q \setminus FS)).$$

Si A est déterministe et schéma-complet pour S alors la sécurité d'un état de A par rapport au schéma S est un problème d'inclusion.

Lemme A1. Soit S = L(A[F/FS]). Si A est déterministe et schéma-complet pour S, alors pour tous les états q Q :

$$L(A[I/\{q\}, F/FS]) \quad L(A[I/\{q\}]) \quad q \quad \text{sûr}_{\Delta,S}(F).$$

Preuve. « ». Supposons que q $\text{safe}_{\Delta,S}(F)$. La définition de la sécurité basée sur un schéma donne $\text{acc}_{\Delta}(\{q\}) \cap (FS \setminus F) = \emptyset$. Il existe donc une couverture h HΣ, telle que $q \xrightarrow{h} FS \setminus F$ par rapport à Δ. En particulier, q L(A[I/{q}, F/FS]). Par déterminisme, il n'existe aucune autre transition sur h à partir de q, et donc, q L(A[I/{q}]). Par conséquent, L(A[I/{q}, F/FS]) L(A[I/{q}]. " ". Supposons que q $\text{safe}_{\Delta}(F)$.

Soit h L(A [I/{q}, F/FS]). Alors, il existe q $\text{acc}_{\Delta}(\{q\})$ tel que $\text{acc}_{\Delta}(\{q\})$ F $(Q \setminus FS)' \quad FS'$, h tel que $q \xrightarrow{h} q'$ F. Ainsi, q' par rapport à Δ. Par conséquent, q' h L(A[I/{q}]). implique q

□

Le lemme A1 avec FS = Q montre le lemme 5 de [21], qui indique pour tout dSHA complet Et que la sécurité est un problème d'universalité :

$$L(A[I/\{q\}]) = H\Sigma \quad q \quad \text{safe}_{\Delta}(F).$$

En effet, la complétude du schéma de A pour S = HΣ implique la complétude de A, et donc, L(A[I/{q}, F/Q]) = HΣ.

Annexe A.2. Algorithme Nous

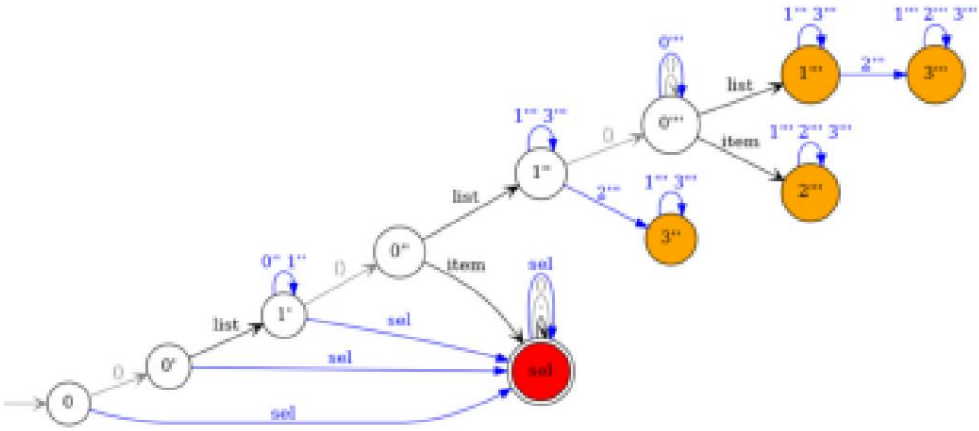
présentons maintenant notre premier testeur d'adhésion pour les dSHA avec des restrictions de schéma. Étant donné un SHA A, nous calculons un SHA↓ $A_e(S) = (\Sigma, Q_e, \Delta_e, I_{e(S)}, F_{e(S)})$ comme suit. Laissons- nous être un

Machine Translated by Google

nouveau symbole. Nous commençons par un ensemble d'états sûrs pour la sélection S0 = sûrΔ (F) et, respectivement, sûrs pour le rejet R0 = sûrΔ (Q \ F). Les ensembles d'états de Ae(S) sont alors les suivants :

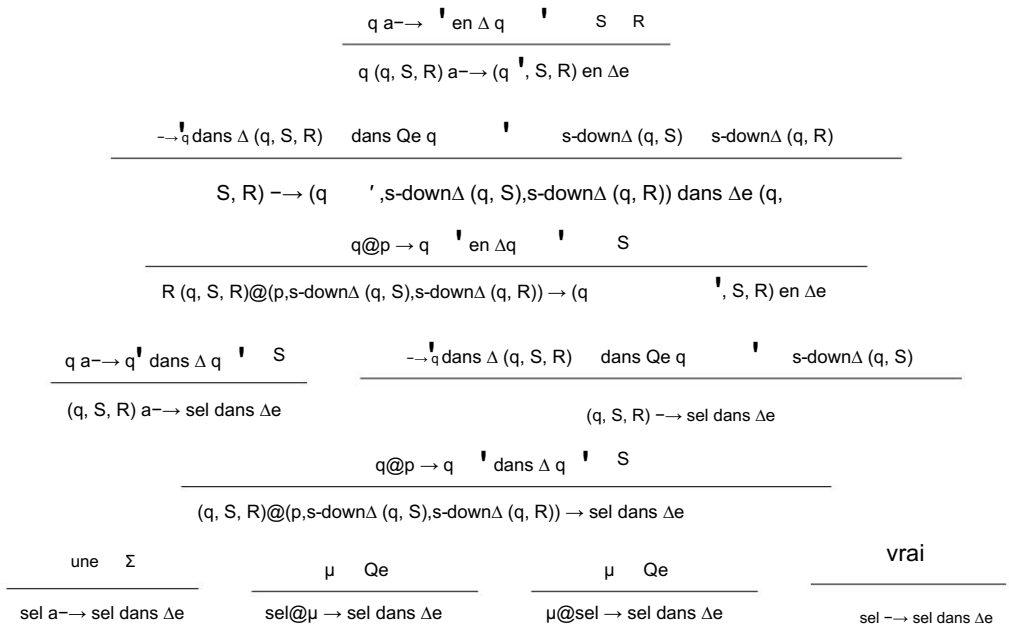
$$Qe = (Q \times 2^{Q} \times 2^{Q}) \quad \{sel\} =$$
$$Je_{e(S)} \{(q, R0, S0) \mid q \quad I, q \quad R0 \quad S0\} \quad \{sel \mid je \cap S0 \quad = \quad \}$$
$$Fe(S) = \{(q, R0, S0) \mid q \quad F\} \quad \{sel\}$$

À titre d'illustration, reconsidérez le dSHA de la figure 4. Ce dSHA n'est pas complet mais schématiquement complet pour le document de schéma. On obtient s-downΔ (0, {4}) = {3, 4}. Il peut sembler contre-intuitif que seul l'état 3, mais également l'état 4 vers le bas, restent sûrs pour la sélection. Cela reflète le fait qu'aucune sous-couverture appropriée d'une haie satisfaisant au schéma prévu ne pourra jamais entrer dans l'état 4. L'application de la première construction avec projection sûre sans changement au dSHA dans la figure 4 donne le dSHA↓ dans la figure A1. Les règles de transition en Δe sont données sur la figure A2.



Graphique A1. Le premier dSHA↓ Ae pour le dSHA A dans la figure 19 avec la documentation de schéma pour le filtre XPath [self::list/child::item]. Les états de Ae correspondent aux tuples suivants : 0 = (0, {1, 2, 3, 5}, {4}), 0′ = (0, {2, 4, 5}, {3, 4}) , 1′ = (1, {2, 4, 5}, {3, 4}), 0″ = (0, , {2, 1″ 4, 5}), = (1, , {2, 4 , 5}), 0‴ = (0, ,), 1‴ = (1, ,), 2‴ = (1, ,) et 3‴ = (1, ,).

Machine Translated by Google



Graphique A2. Les premières règles de transition Δe sont déduites des règles de transition Δ de certains SHA.

Annexe A.3. Solidité et exhaustivité

Nous fournissons ensuite des résultats d'exactitude et d'exhaustivité pour notre premier testeur d'adhésion en mémoire avec projection.

Proposition A1 (Finalité). Pour tout SHA A avec schéma S = L(A[F/FS]), couverture h S et exécution partielle v de Ae(S) sur h qui se termine par l'état sel, alors h L(A).

Ainsi, si l'exécution partielle atteint l'état sel, alors projΣ (v) est certain d'être membre de L(A) par rapport à S. L'évaluation d'un SHA↓ s Ae(S) donne un premier testeur d'appartenance en mémoire pour dSHA A. A L'adaptation unique est la suivante : chaque fois que sel est atteint, l'évaluation peut s'arrêter complètement et accepter la couverture d'entrée.

Théorème A1 (Exhaustivité). Soit A = (Σ, Q, Δ, I, F) un dSHA dont le schéma est complet pour le schéma S = L(A[F/FS]). Pour toute couverture h S et préfixe v de nw(h) : Si v est certain pour l'appartenance à L(A)

- avec S, alors il existe une exécution partielle w pour Ae(S) sur h, telle que v = projΣ (w) et w se termine par state sel.
- Si v est certain pour la non-appartenance à L(A) avec S, alors il existe une exécution partielle bloquante w pour Ae(S) sur h, telle que projΣ (w) est un préfixe de v.

Cela signifie que la certitude d'appartenance et de non-appartenance est détectée au premier préfixe lors de l'exécution de l'évaluateur pour Ae(S) .

Preuve. Nous remarquons que cet algorithme est fondamentalement le même que le premier testeur d'appartenance de la proposition 6 de [21], sauf qu'il vérifie également le rejet sûr et qu'il prend en compte les schémas. Le fait que des RNF soient utilisées ici à la place des SHA↓ n'est pas essentiel. Nous pouvons donc en tirer la preuve de solidité et d'exhaustivité sans problème.

Annexe B. Preuves pour la section 6 (Projection Safe-No-Change)

Preuve. Nous devons prouver que plus aucun changement d'état q P aucun changementΔ n'est acceptable.

Si q aucun changementΔ , cela découle de la complétude du schéma de Δ de sorte qu'on ne peut ni bloquer aucune haie du schéma ni changer l'état. Dans le cas q P, le

Machine Translated by Google

L'intuition est que l'état au niveau supérieur, disons r, ne peut ni changer, puisque P = s-pas de changementΔ (r), et l'automate ne peut pas non plus bloquer toute haie du schéma en raison de l'exhaustivité du schéma.

Nous prouvons d'abord l'inclusion L(A) L(A snc). Puisque L(A) S par définition des schémas, cela implique L(A) L(A snc) ∩ S. La preuve sera basée sur les trois revendications suivantes A1.1a, A1.2a et A1.3a. Notez que l'exhaustivité du schéma n'est pas nécessaire pour cette direction.

h Réclamation A1.1a. $\Pi \xrightarrow{h} \Pi$ par rapport à Δ snc pour toutes les haies h HΣ.

La preuve est simple par récurrence sur la structure de h. Il utilise les trois dernières règles de transition de Δ de la figure 10 permettant de toujours rester dans Π quelle que soit la haie qui suit.

Réclamation A1.2a. Pour tout h HΣ, q Q et P Q, tel que q P no-changeΔ ,

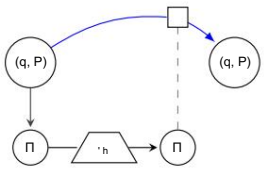
$$(q, P) \xrightarrow{h} (q, P) \text{ par rapport à } \Delta \text{ snc}.$$

Nous prouvons la revendication A1.2a par récurrence sur la structure de h.

Cas h = h' . Dans ce cas, nous pouvons utiliser la revendication A1.1a pour montrer $\Pi \xrightarrow{h} \Pi$ par rapport à Δ snc et les règles d'inférence

$$\frac{q \quad P \quad \text{aucun changement}\Delta}{(q, P) \xrightarrow{h} \Pi \text{ dans } \Delta \text{ snc}} \qquad \frac{q \quad P \quad \text{aucun}}{\text{changement}\Delta\ (q, P)@\Pi \rightarrow (q, P) \text{ dans } \Delta \text{ snc}}$$

afin de fermer le diagramme suivant par rapport à Δ snc :



Cela prouve $(q, P) \xrightarrow{h} (q, P)$ par rapport à Δ snc, comme l'exige la revendication.

Cas h = a. Puisque q P aucun changementΔ nous pouvons appliquer la règle d'inférence :

$$\frac{a \quad \Sigma\ q \quad P \quad \text{aucun changement}\Delta}{(q, P)\ a \xrightarrow{} (q, P) \text{ dans } \Delta \text{ snc}}$$

Cela prouve le cas de cette affirmation.

Cas h = ε. On a trivialement $(q, P) \xrightarrow{\varepsilon} (q, P)$ par rapport à Δ snc.

Cas h = $h' \cdot h''$. Par hypothèse d'induction appliquée à h' et h'' on a : $(q, P) \xrightarrow{h'} (q, P)$ et $(q, P) \xrightarrow{h''} (q, P)$ par rapport à Δ snc. Par conséquent, $(q, P) \xrightarrow{h' \cdot h''} (q, P)$ par rapport à Δ snc.

Ceci met fin à la preuve de la revendication A1.2a. L'affirmation suivante, dans laquelle l'étape d'induction est un peu plus fastidieuse à prouver, est la clé de la preuve de solidité. Nous définissons le prédicat suivant pour tout q', q'' Q et P Q :

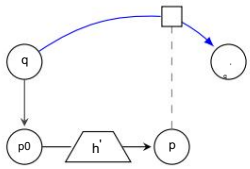
$$P(q', q'') \text{ si et seulement si } (q' = q'' \text{ ou } q', q'' \quad P)$$

Réclamation A1.3a. Soit h HΣ une haie, q, q' h qui accΔ(P) P et Q états et P Q un sous-ensemble d'états, tel que q P no-changeΔ . Si $q \xrightarrow{h} q'$ par rapport à Δ, alors il existe q '', tel que $(q, P) \xrightarrow{h} (q'', P)$ par rapport à Δ snc et $P(q', q'')$.

Machine Translated by Google

Preuve. Induction sur la structure de h.

Cas h = h ' h . L'hypothèse q −→ qp Q clôturant ' par rapport à ∆ montre qu'il existe des états p0 ∆ et le schéma suivant :

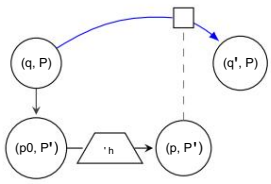


Soit P ' = s-pas de changement∆ (q) et notons que acc∆(P ′) P '. Puisque q P aucun changement∆ nous peut déduire ce qui suit :

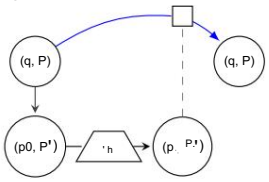
$$\frac{-\!\!\rightarrow p0 \text{ dans } \Delta \; q \qquad P \quad \text{aucun changement}\Delta}{(q, P) - \!\!\rightarrow (p0, P') \text{ en } \Delta^{\text{snc}}} \qquad \frac{q@p \rightarrow q' \text{ dans } \Delta \; q \qquad P \quad \text{aucun changement}\Delta}{(q, P)@(p, P') \rightarrow (q', P) \text{ en } \Delta^{\text{snc}}}$$

Sous-cas p0 P ' aucun changement∆ . L'hypothèse d'induction appliquée à h telle que ' montre que il existe p −→ (p les (p0, P ') pour cas justifiant ce dernier P.') par rapport à ∆ snc et p P p'. Nous distinguons prédicat :

Diagramme du ' = p. Donc (p0, P ') −→ (p, P ') par rapport à ∆ snc, nous pouvons donc fermer le sous-sous-cas p comme suit :



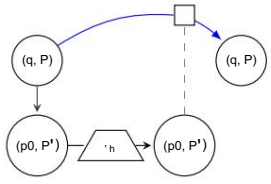
Cela montre que (q, P) −→ ', P) par rapport à ∆ snc, et donc, l'affirmation puisque q ' Pq →'. Sous-sous-cas p, pq ' P '(q . Puisque p P ' et P ' = s-pas de changement∆ (q), nous avons q@p dans ∆. On peut donc clôturer le diagramme comme suit :



Puisque p P ' et q@p → q ' en ∆, on a q = q ' par définition de P ' = s-aucun changement∆ (q). Cela montre que (q, P) −→ (q, P) par rapport à ∆ snc. Puisque q P q, le la réclamation suit.

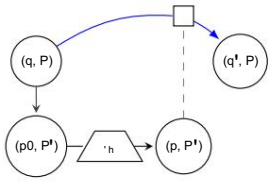
Sous-cas p0 P ' aucun changement∆ . La revendication A1.2a montre alors que (p0, P ') −→ (p0, P ') par rapport ∆ snc.

Sous-sous-cas p0 P '. Puisque p acc∆(p0) et acc∆(P ′) P = s-no- ', il s'ensuit que P les aussi. Par définition P p ' change∆ (q) et la complétude de ∆, impliquent que appartenances p0 P peuvent ' et p P ' q@∆p0 = {q} = q@∆p. Nous désormais clôturer le diagramme ci-dessous comme suit :



Sous-sous-cas p0 aucun changement∆ . Dans ce cas, p0 = p donc q ' q @∆p0. Ainsi,

Machine Translated by Google



Cas h = a. Puisque q P no-changeΔ , nous pouvons appliquer la règle d'inférence :

$$\frac{q \xrightarrow{a} q' \text{ dans } \Delta \quad q \quad P \quad \text{aucun changement}_\Delta(q, P)}{q' \quad P \quad q'} \text{ snc}$$

Cela montre que (q, P) −→ (q h ', P) valide la revendication puisque q ' P q '.

Cas h = ε. Dans ce cas, on a q = q ' et (q, P) ε−→ (q, P) donc la revendication est valable.

h Cas h = h1 · h2. Puisque q −→ q ' h1 par rapport à Δ, il existe q1 Q, tel que q −→ q1 par rapport à Δ et h2 q1 −→ q ' par rapport à Δ. Puisque q P aucun changementΔ , nous appliquons l'hypothèse d'induction sur h1. Cela implique qu'il existe q ' 1 tel que,

$$h1 \ (q, P) \longrightarrow (q'_{1}, P) \text{ par rapport à } \Delta^{\text{snc}} \text{ et } q1 \quad P \ q \quad '_{1}.$$

On distingue les deux cas de q1 P q ' : 1

Sous-cas q1 = q ' 1 . Nous distinguons également ici deux sous-cas : Sous-sous-cas q1 P. L'hypothèse d'induction appliquée à h2 donne la suite-ing:

$$q\text{»}. \ (q1, P) \xrightarrow{h2} (q \quad \text{»}, P) \text{ par rapport à } \Delta^{\text{snc}} \quad q' \quad P \ q\text{»},$$

Ainsi

$$q\text{»}. \ (q, P) \xrightarrow{h} (q \quad \text{»}, P) \text{ par rapport à } \Delta^{\text{snc}} \quad q' \quad P \ q\text{»}$$

.

Sous-sous-sous-cas q1 P. D'après la revendication A1.2a, nous avons (q1, P) h2 −→ (q1, P). Nous avons aussi P. q ' acc({q1}), et puisque nous supposons acc(P) P, cela implique q ' Par conséquent, (q, P) h −→ (q1, P) et q ', q1 P, ce qui implique que la revendication est q ' Pq1 .

Sous-cas q1, q ' 1 P. Puisque q ' 1 P, la revendication A1.2a implique (q ' 1 , P) h2 −→ ' 1 , P) par rapport à Δ snc. P) accΔ(P) h −→ (q 1 ', P) par rapport à Δ snc. Depuis q ' (q accΔ({q1}) et q1 P, il s'ensuit que Ainsi, (q, q ' P. Ici, nous avons utilisé comme dans le sous-sous-cas précédent que acc(P) P est assumé par la réclamation. Soient q et q» = q ' 1 . Alors, on a (q, P) −→ (q h », P) par rapport à Δ snc ', q » P montrant la revendication.

Ceci met fin à la preuve de la revendication A1.3a. □

Preuve d'inclusion L(A) L(A snc). Soit h L(A). Alors, il existe q0 I et q F, tels que q0 −→ q. Nous distinguons deux cas : h

Cas q0 aucun changementΔ . Par définition de non-changementΔ et puisque q accΔ(q0) nous avons q0 = q. La revendication A1.2a montre que (q0,) −→ (q0,) par rapport à Δ snc et donc (q0,) −→ (q,) de sorte que h L(A snc). h h

Cas q0 aucun changementΔ . La revendication A1.3a avec P = montre que (q0,) h −→ (q,) par rapport à Δ snc et donc h L(A snc).

Machine Translated by Google

Ceci termine la preuve de la première inclusion. Nous voulons ensuite montrer l'inclusion inverse L(A snc) ∩ S L(A). Cela découlera finalement des trois revendications suivantes A1.1b, A1.2b et A1.3b.

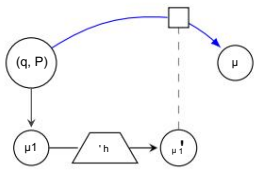
Réclamation A1.1b. Pour toute couverture h et état μ Qsnc, si Π −→ μ par rapport à Δ snc alors μ = Π.

La preuve est simple par récurrence sur la structure de h : les seules règles de transitions de Δ snc avec Π du côté gauche sont déduites par les trois dernières règles de la figure 10. Celles-ci nécessitent de rester dans Π quelle que soit la haie qui suit.

Réclamation A1.2b. Pour toute couverture h, posez P Q, l'état q P no-changeΔ et l'état μ Qsnc : si (q, P) −→ μ par rapport à Δ snc, alors μ = (q, P).

Preuve. Par induction sur la structure de h. Supposons que (q, P) −→ μ par rapport à Δ snc :

Cas h = h ′ . Il doit exister des états μ1, μ ′ 1 Qsnc fermant le schéma suivant :



Puisque q P no-changeΔ , la règle suivante doit avoir été appliquée pour déduire (q, P) μ1 par rapport à Δ snc :

$$\frac{q \quad P \quad \text{aucun changement}\Delta}{(q, P) \quad \longrightarrow \Pi \text{ dans } \Delta^{snc}}$$

Donc μ1 = Π. La revendication A1.1b montre que ′ 1 = Π aussi. Donc μ devait être μ déduit en appliquant la règle :

$$\frac{q \quad P \quad \text{aucun}}{\text{changement}\Delta\ (q, P)@\Pi \rightarrow (q, P)^{snc} \text{ dans } \Delta}$$

Donc, μ = (q, P), comme requis.

Cas h = a. La règle suivante doit avoir été appliquée :

$$\frac{q \quad P \quad \text{aucun}}{\text{changement}\Delta\ (q, P)\ a \longrightarrow (q, P)^{snc} \text{ dans } \Delta}$$

Par conséquent, μ = (q, P).

Cas h = ε. Évident.

h = h1 · h2. Il doit exister μ1, tel que (q, P) −→ μ par rapport à Δ snc. Par hypothèse d'induction −→ μ1 appliquée à h1, = (q, P). On peut ainsi appliquer l' hypothèse d'induction à h2 pour obtenir μ2 = (q, P).

Ceci termine la preuve de la revendication A1.2b. Nous avons ensuite besoin d'un inverse de la revendication A1.3a. □

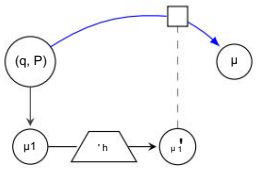
Réclamation A1.3b. Soit q Q, P Q, tel que q P aucun changementΔ et accΔ(P) P. Pour tout h HΣ, tel que (q, P) −→ μ par rapport à tout blocage partiel snc pour certains μ Qsnc et tels que Δ n'a pas exécuté sur h à partir de q, il existe q ′, q ", tel que

$$\mu = (q' , P), q \xrightarrow{h} q'' \text{ par rapport à } \Delta, \text{ et } q' \quad P\ q''.$$

Machine Translated by Google

Preuve. Par récurrence sur la structure de h HΣ, on distingue des cas pour toutes les formes possibles h.

Cas h = h −→ ' μ par définition de (q, P) il existe μ1, μ pour que le schéma suivant puisse être ' 1 Qsnc, tel fermé :



Puisque q P aucun changementΔ , la règle suivante a été appliquée pour déduire (q, P) −→ μ1 où = s-P.' aucun changementΔ (q) :

$$\frac{-\!\!\rightarrow p \text{ dans } \Delta \; q \qquad P \quad \text{aucun changement}\Delta}{(q, P) - \!\!\rightarrow (p, P') \text{ en } \Delta \quad \text{snc}}$$

Il existe donc p Δ, tel que μ1 = (p, P '). Nous corrigeons un tel état de manière arbitraire.

Puisque Δ n'a pas d'exécutions bloquantes sur h à partir de q, il existe p ', q » Q, tel que h' p p → q à partir de p, ' » par rapport à Δ. De plus, Δ n'a pas de course partielle bloquante sur h '

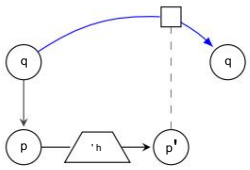
Sous-cas p P '. Dans ce cas, nous pouvons appliquer la revendication A1.2b ') h ' par rapport à Δ dans l'ordre −→ μ 1 que μ il s'ensuit que ' 1 = (p, P ' à (p, P). Puisque p acc∆({p}) et p P = s-no-changeΔ (q) pour montrer q@Δp = {q}. Par conséquent, la règle suivante a été appliquée pour déduire (q, P) −→ μ :

h

$$\frac{q@p \rightarrow q \text{ dans } \Delta \; q \qquad P \quad \text{aucun changement}\Delta \; (q, P)@(p, P') \text{ dans } \Delta \quad \text{snc}}{(q, P) \; ,}$$

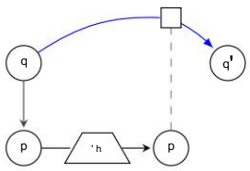
Cela montre que μ = (q, P). Soit q ' = q de sorte que μ = (q ', P). Puisque p par rapport à Δ, nous → q avoir p ' acc∆({p}) tel que q@p ' par rapport à Δ par définition de s-pas de changementΔ ({q}).
On peut donc clôturer le schéma suivant :

h '



Soit q ⇒q, de sorte que q ' = q ''. On considère alors que q ' P q » h et q −−→ q ' par rapport à Δ, comme requis par la revendication.

h Sous-cas p aucun changementΔ . Dans ce cas p −→ p ' peut ' par rapport à Δ implique que p ' = p. Par conséquent, nous fermer le diagramme suivant :



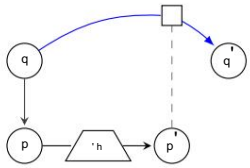
Cela montre que q −−→ q h ' ' par rapport à Δ.

Machine Translated by Google

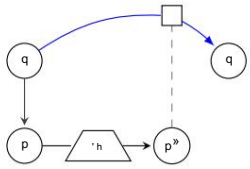
Sous-cas p P aucun changementΔ . Par hypothèse d'induction appliquée à (p, P h) par rapport à −→ μ 1
Δ, il existe p ", tel que

μ 1 = (p , P. h), p −→ p » par rapport à Δ, et p P′p ».

Puisque Δ n'a pas d'exécutions bloquantes sur h commençant par q, il existe q P′ p " : q@p » tel que
» → q » par rapport à Δ. Il existe deux manières de satisfaire p
Sous-sous-cas p = p ». On a alors ce qui suit :



Cela montre q, h − par rapport à Δ.
Sous-sous-cas p , p » −→ q P. Par définition de P , il s'ensuit que q = q = q ". Ainsi,



h Cela montre q −−→ q par rapport à Δ.

Cas h = a. Puisque q P no-changeΔ , la règle d'inférence suivante doit être utilisée :

$$\frac{q\ a{-\!\!\rightarrow}\ q' \text{ dans } \Delta \quad q \quad P \quad \text{aucun changement}\Delta\ (q, P)}{P\ a\text{-en-}q'}\ \text{snc}$$

Donc μ = (q , P) et q a−→ q Cas par rapport à Δ.
h = ε. Évident.

Cas h = h1 · h2. Le jugement (q, P) −→ μ par rapport à Δ h2 snc montre qu'il existe μ1, tel que
h1 (q, P) −→ μ1 −→ μ par rapport à Δ snc. Puisque q P aucun changementΔ , on peut appliquer l'induction h1 −→
hypothèse à h1. Cela montre qu'il existe q 1 et q » 1 , tel que μ1 = (q 1 , P), q
q » 1 par rapport à Δ, et q1 Pq » 1 .

Sous-cas q 1 = q » h21 . Par conséquent, (q1, P) −→ μ par rapport à Δ snc.
Sous-sous-cas q 1 P aucun changementΔ . Dans ce cas, on peut appliquer l'hy-
pothèse à (q μ = (q 1 , P) h2 −→ μ par rapport à Δ snc, montrant l'existence de q , tel que
, P) et énoncer q ", tel que q 1 h2 −→ q » et q donc P q '. Ainsi,
q » q . q h −→ » par rapport à Δ et q P q » la revendication est valable.
Le sous-sous-cas 1 P aucun changementΔ . La revendication A1.2b s'applique à h2 snc
1 , P) −→ μ par rapport à Δ
q montre que μ = (q (q 1 , P).

Sous-cas q 1 , q 1 » P. Rappelons que (q » 1 , P) h2 −→ μ par rapport à Δ » 1 P. Réclamation A1.2b
montre que μ = (q » snc et q h1 1 , P) par rapport à Δ snc. Nous avons » 1 par rapport à Δ. Depuis là
aussi q −→ q il n'y a pas d'exécutions partielles bloquantes sur h à partir de q il existe un état q » tel
qui q » 1 h2 −→ » q par rapport à Δ. Depuis q1 P et P est fermé par accessibilité, on a
q » acc({q » 1 }) acc(P) P. De q h1 −→ » q par rapport à Δ, nous obtenons h » par rapport à Δ. Depuis
q » 1 , q » P, il s'ensuit que q » q −→ q 1 P q ", et donc, l'affirmation est valable.

Ceci met fin à la preuve de la revendication A1.3b.

Machine Translated by Google

□

Preuve d'inclusion L(A snc) ∩ S L(A). Soit h L(A snc) ∩ S. Puisque h L(A snc), alors il existe q0 I et q F, tels que (q0,) −→ (q,) par rapport à ∆ snc. En utilisant le lemme 7, nous avons −→ q par rapport à ∆. q0 On
h h
distingue
deux cas :

Cas q0 aucun changement∆ . La revendication A1.2b montre que q = q0. Puisque A est schéma-complet pour S, h
S et q0 I ∩ no-change∆ , le lemme 7 montre que q0 −→ q0 par rapport à ∆. Puisque q = q0, cela donne h
h
L(A).

Cas q0 aucun changement∆ . Puisque A est un schéma complet pour S et h S, il n'existe aucune exécution bloquante sur h commençant en q0. Par conséquent, nous pouvons appliquer la revendication A1.3b avec P =
h h
à (q0,) −→ (q,) par rapport à ∆ snc. Cela montre que q0 −→ q par rapport à ∆, et donc h L(A).

Ceci termine la preuve de l'inclusion inverse, et donc de L(A) = L(A snc). □

Références

1. Marian, A. ; Siméon, J. Projection de documents XML. Dans les Actes du VLDB ; Morgan Kaufmann : San Francisco, Californie, États-Unis, 2003 ; pp. 213-224.
2. Frisch, A. Reconnaissance régulière du langage des arbres avec informations statiques. Dans Actes de l'exploration de nouvelles frontières de l'informatique théorique, 18e Congrès mondial de l'informatique de l'IFIP, 3e Conférence internationale du TCS sur l'informatique théorique, Toulouse, France, 22-27 août 2004 ; pp. 661-674.
3. Maneth, S. ; Nguyen, K. Optimisation complète des requêtes XPath. VLPB J. 2010, 3, 882-893.
4. Sébastien, T. ; Niehren, J. La projection pour les automates de mots imbriqués accélère l'évaluation XPath sur les flux XML. Dans Actes de la Conférence internationale sur les tendances actuelles de la théorie et de la pratique de l'informatique (SOFSEM), Harrachov, République tchèque, 23-28 janvier 2016.
5. Kay, M. Le processeur Saxon XSLT et XQuery. 2004. Disponible en ligne : https://www.saxonica.com (consulté le 29 juillet 2024).
6. Gienieczko, M. ; Murlak, F. ; Paperman, C. Prise en charge des descendants dans JSONPath accéléré par SIMD. Dans les actes de la 29e conférence internationale ACM sur le support architectural pour les langages de programmation et les systèmes d'exploitation, La Jolla, Californie, États-Unis, 27 avril-1er mai 2024 ; Volume 1, p. 15-29.
7. Al Serhali, A. ; Niehren, J. Une collection de référence d'automates déterministes pour les requêtes XPath. Dans Actes du XML Prague 2022, Prague, République tchèque, 9-11 juin 2022.
8. Niehren, J. ; Sakho, M. Détermination et minimisation des automates pour les mots imbriqués revisités. Algorithmes 2021, 14, 68. https://doi.org/10.3390/a14030068.
9. Comon, H. ; Dauchet, M. ; Gilleron, R. ; Loding, C. ; Jacquemard, F. ; Lugiez, D. ; Tison, S. ; Tommasi, M. Techniques des automates d'arbres et Applications. Disponible en ligne : http://tata.gforge.inria.fr (consulté le 29 juillet 2024).
10. Thatcher, JW Caractérisation des arbres de dérivation de grammaires hors contexte grâce à une généralisation de la théorie des automates. J. Informatique. Système. Sci. 1967 : 1 : 317-322. https://doi.org/10.1016/S0022-0000(67)80022-9.
11. Alur, R. Marier les mots et les arbres. Dans les actes du 26e symposium ACM SIGMOD-SIGACT-SIGART sur les principes de Database Systems, Pékin, Chine, 11-14 juin 2007 ; ACM-Press : New York, NY, États-Unis, 2007, pp. 233-242.
12. Alur, R. ; Madhusudan, P. Langues visiblement pushdown. Dans Actes du 36e Symposium annuel de l'ACM sur la théorie de l'informatique, Chicago, IL, États-Unis, 13-16 juin 2004 ; Babai, L., éd.; ACM : New York, NY, États-Unis, 2004 ; pp. 202-211. https : //doi.org/10.1145/1007352.1007390.
13. Mehlhorn, K. Pebbling Mountain Ranges et son application de la reconnaissance DCFL. Dans Proceedings of the Automata, Languages and Programming, 7e colloque, Noordweijkerhout, Pays-Bas, 14-18 juillet 1980 ; Procédure; de Bakker, JW, van Leeuwen, J., éd. ; Springer : Berlin/Heidelberg, Allemagne, 1980 ; Volume 85, pages 422 à 435. https://doi.org/10.1007/3-540-10003- 2_89.
14. von Braunmühl, B. ; Verbeek, R. Les langages pilotés par les entrées sont reconnus dans l'espace log n. Dans Thèmes de théorie du calcul ; Études mathématiques de Hollande du Nord ; Karpinski, M., van Leeuwen, J., éd. ; Hollande du Nord : 1985 ; Volume 102, pages 1 à 19. https://doi.org/10.1016/S0304-0208(08)73072-X.
15. Okhotine, A. ; Salomaa, K. Complexité des automates pushdown pilotés par les entrées. Actualités SIGACT 2014, 45, 47-67. https://doi.org/10.1145/2636805.2636821.
16. Niehren, J. ; Sakho, M. ; Al Serhali, A. Détermination des automates basés sur des schémas. Dans Actes du 13e Symposium international sur les jeux, les automates, la logique et la vérification formelle, GandALF 2022, Madrid, Espagne, 21-23 septembre 2022 ; Ganty, P., Monica, DD, éd.; EPTCS : 2022 ; Volume 370, p. 49-65. https://doi.org/10.4204/EPTCS.370.4.
17. Lécher, A. ; Schmitz, S. Référence XPath. Disponible en ligne : https://archive.softwareheritage.org/browse/directory/1ea68cf5bb3f9f3f2fe8c7995f1802ebadf17fb5 (consulté le 29 juillet 2024).

Machine Translated by Google

18. Mozafari, B. ; Zeng, K. ; Zaniolo, C. Traitement d'événements complexes hautes performances sur des flux XML. Dans les actes de la conférence SIGMOD ; Candan, KS, Chen, Y., Snodgrass, RT, Gravano, L., Fuxman, A., Candan, KS, Chen, Y., Snodgrass, RT, Gravano, L., Fuxman, A., Eds.; ACM : New York, NY, États-Unis, 2012 ; pp. 253-264. https://doi.org/10.1145/2213836.2213866.
19. Grez, A. ; Riveros, C. ; Ugarte, M. Un cadre formel pour le traitement des événements complexes. Dans Actes de la 22e Conférence internationale sur la théorie des bases de données (CIDC 2019), Lisbonne, Portugal, 26-28 mars 2019 ; pp. 5 : 1 à 5 : 18. https://doi.org/10.4230/LIPIcs.CIDC.2019.5.
20. Muñoz, M. ; Riveros, C. Énumération en streaming sur des documents imbriqués. Dans Actes de la 25e Conférence internationale sur la théorie des bases de données, CIDC 2022, Édimbourg, Royaume-Uni, 29 mars-1er avril 2022 ; Olteanu, D., Vortmeier, N., éd.; Château Dagstuhl - Leibniz-Zentrum für Informatik : 2022 ; Volume 220, p. 19 :1-19 :18. https://doi.org/10.4230/LIPIcs.ICDT.2022.19.
21. Al Serhali, A. ; Niehren, J. Première réponse aux requêtes pour les automates déterministes de couverture par étapes. Dans Actes de la mise en œuvre et de l'application des automates — 27e Conférence internationale, CIAA 2023, Famagouste, Chypre du Nord, 19-22 septembre 2023 ; Procédure; Nagy, B., éd.; Springer : Berlin/Heidelberg, Allemagne, 2023 ; Volume 14151, p. 53-65. https : //doi.org/10.1007/978-3-031-40247-0_3.
22. Gauwin, O. ; Niehren, J. ; Tison, S. Première réponse aux requêtes pour les automates de mots imbriqués déterministes. Dans Actes du 17e Symposium international sur les principes fondamentaux de la théorie informatique, Wroclaw, Pologne, 2-4 septembre 2009 ; Springer : Berlin/Heidelberg, Allemagne, 2009 ; Volume 5699, p. 121-132.
23. Debarbieux, D. ; Gauwin, O. ; Niehren, J. ; Sébastien, T. ; Zergaoui, M. Premiers automates de mots imbriqués pour répondre aux requêtes XPath sur les flux XML.Theor. Calculer. Sci. 2015, 578, 100-125. https://doi.org/10.1016/j.tcs.2015.01.017.
24. Al Serhali, A. ; Niehren, J. Projection Subhedge pour les automates Stepwise Hedge. Dans Actes des principes fondamentaux de la théorie du calcul — 24e Symposium international, FCT 2023, Trèves, Allemagne, 18-21 septembre 2023 ; Procédure; Fernau, H., Jansen, K., éd. ; Springer : Berlin/Heidelberg, Allemagne, 2023 ; Volume 14292, p. 16-31. https://doi.org/10.1007/978-3-031-43587-4_2.
25. Neumann, A. ; Seidl, H. Localisation des correspondances de modèles d'arbres dans les forêts. Dans Proceedings of the Foundations of Software Technology and Theoretical Computer Science, Chennai, Inde, 17-19 décembre 1998 ; Springer : Berlin/Heidelberg, Allemagne, 1998 ; Volume 1530, p. 134-145.
26. Gauwin, O. ; Niehren, J. ; Roos, Y. Automates d'arbres en streaming. Inf. Processus. Lett. 2008, 109, 13-17. https://doi.org/10.1016/j.ipl.2008.08.002.
27. Franceschet, M. XPathMark Test de performances. Disponible en ligne : https://users.dimi.uniud.it/~massimo.franceschet/xpathmark/PTbench.html (consulté le 25 octobre 2020).
28. Hosoya, H. ; Pierce, BC XDuce : un langage de traitement XML typé statiquement. ACM Trans. Technologie Internet. 2003, 3, 117-148. https://doi.org/10.1145/767193.767195.
29. Gecseg, F. ; Steinby, M. Automates d'arbres ; Akadémiai Kiadó : Budapest, Hongrie, 1984.
30. Gold, identification du langage EM dans la limite. Inf. Contrôle. 1967, 10, 447-474. https://doi.org/10.1016/S0019-9958(67)91165-5.
31. Carme, J. ; Niehren, J. ; Tommasi, M. Interrogation d'arbres non classés avec des automates d'arbres pas à pas. Dans Proceedings of the Rewriting Techniques and Applications, 15e Conférence internationale, RTA 2004, Aix-la-Chapelle, Allemagne, 3-5 juin 2004 ; Procédure; van Oostrom, V., éd.; Springer : Berlin/Heidelberg, Allemagne, 2004 ; Volume 3091, pages 105 à 118. https://doi.org/10.1007/978-3-540-25 979-4_8.
32. Paire, C. ; Quéré, A. Définition et étude des bilans réguliers. Inf. Contrôle 1968, 13, 565-593.
33. Murata, M. Hedge Automata : Un modèle formel pour les schémas XML. 2000. Disponible en ligne : http://www.xml.gr.jp/relax/hedge_ sympa.html (consulté le 29 juillet 2024).
34. Martens, W. ; Niehren, J. Sur la minimisation des schémas XML et des automates d'arbres pour les arbres non classés. J. Informatique. Système. Sci. 2007, 73, 550-583.
35. Niehren, J. Les automates Stepwise Hedge sont exponentiellement plus succincts que les automates forestiers. 2024, en préparation.
36. Bojanczyk, M. ; Walukiewicz, I. Algèbres forestières. Dans Actes de la logique et des automates : histoire et perspectives [en l'honneur de Wolfgang Thomas] ; Flum, J., Grädel, E., Wilke, T., éd.; Presses universitaires d'Amsterdam : Amsterdam, Pays-Bas, 2008 ; Volume 2, p. 107-132.

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Cadre d'optimisation hiérarchique pour la conception de la configuration de Réseau de pipelines de collecte de gaz Star-Tree dans des espaces discrets

Yu Lin [1], Yanhua Qiu [1], Hao Chen [2], Jun Zhou 3,*, Jiayi He [3], Penghua Du [3] et Dafan Liu [4]

1 Institut de recherche sur les technologies d'ingénierie de collecte et de transport, PetroChina Southwest Oil & Gasfield Company, Chengdu 610041, Chine ; linyu223@petrochina.com.cn (YL); qiuyan0320@petrochina.com.cn (YQ)
2 Sichuan Shale Gas Exploration and Development Co., Ltd., Chengdu 610051, Chine ; chenhao05@petrochina.com.cn
3 École d'ingénierie pétrolière, Southwest Petroleum University, Chengdu 610500, Chine ; hjywyyx@163.com (JH); scdupenghua@163.com (PD)
4 Département d'exploration et de développement du gaz de schiste, CNPC Chuanqing Drilling Engineering Co., Ltd., Chengdu 610066, Chine ; liudf_cqyyq@cnpc.com.cn
* Correspondance : zhoujunswpu@163.com

Résumé : Le réseau de gazoducs de collecte est une infrastructure essentielle pour la collecte et le transport du gaz naturel du site d'extraction jusqu'à l'installation de traitement. Cet article présente un modèle d'optimisation de la conception d'un réseau de gazoducs en étoile dans un espace discret, visant à déterminer la configuration optimale de cette infrastructure. L'objectif est de réduire les investissements nécessaires à la construction du réseau. Les principales variables décisionnelles comprennent l'emplacement des stations, l' emplacement de l'usine, les connexions entre les puits et les stations, ainsi que les interconnexions entre les stations. Plusieurs contraintes d'égalité et d'inégalité sont formulées, portant principalement sur l'affiliation entre puits et stations, le rayon de transmission et la capacité des stations. La conception d'un réseau de pipelines en étoile et en arbre représente un problème complexe et non déterministe d'optimisation combinatoire polynomiale (NP) . Pour relever ce défi, un cadre d'optimisation hiérarchique couplé à un algorithme génétique amélioré (IGA) est proposé. L'efficacité de l'algorithme génétique est validée par des tests et des comparaisons avec d'autres algorithmes traditionnels. Par la suite, le modèle d'optimisation et la méthodologie de solution sont appliqués à la conception d'un réseau de pipelines. Les résultats révèlent que la configuration réseau optimisée réduit les coûts d'investissement de 16 % par rapport à la conception originale. De plus, lorsque l'on compare la disposition optimale sous une topologie étoile-étoile, on observe que l'investissement nécessaire pour la topologie étoile-étoile est 4 % plus élevé que celui nécessaire pour la topologie étoile-arbre.

Mots-clés : structure topologique ; conception de la mise en page ; optimisation hiérarchique ; algorithme génétique

check for updates

Référence : Lin, Y. ; Qiu, Y. ; Chen, H. ; Zhou, J. ; Lui, J. ; Du, P. ; Liu, D. Cadre d'optimisation hiérarchique pour la conception de l'agencement du réseau de pipelines de collecte de gaz Star-Tree dans des espaces discrets. Algorithmes 2024, 17, 340. https://doi.org/10.3390/a17080340

Rédacteurs académiques : Frank Werner, Ming-Feng Ge et Chen Liu

Reçu : 3 juillet 2024
Révisé : 27 juillet 2024
Accepté : 3 août 2024
Publié : 5 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Motivation

Actuellement, la communauté mondiale est en transition vers une ère caractérisée par une énergie propre et efficace, le gaz naturel devenant la source d'énergie prédominante. Selon le Global Energy Outlook, la consommation mondiale de gaz naturel devrait augmenter continuellement jusqu'en 2050, avec un taux de croissance annuel moyen supérieur à 1,4 % (Richard, 2020). L'augmentation significative de la demande de gaz naturel nécessite des efforts accrus en faveur de la construction de réseaux de gazoducs de surface. L'investissement requis pour les réseaux de gazoducs de collecte est important, représentant souvent plus de 30 % des coûts totaux de développement d'un champ gazier [1]. Par conséquent, optimiser de manière rationnelle la conception de ces réseaux de pipelines peut générer des avantages économiques considérab

Le système comprend généralement des puits, des stations de collecte de gaz, des installations centrales de traitement et des pipelines (Figure 1). Dans un premier temps, le gaz naturel extrait des puits est acheminé vers

Machine Translated by Google

la station de collecte de gaz via des conduites de production. Une fois collecté à la station de collecte de gaz, le gaz est ensuite transporté via des canalisations de collecte jusqu'à l'usine centrale de traitement pour être purifié, garantissant ainsi qu'il répond aux normes requises pour la distribution externe.
Sur la base de la topologie du réseau, les réseaux de gazoducs de collecte peuvent être classés en trois types fondamentaux : les topologies en étoile, en arbre et cyclique [2]. Dans une topologie en étoile, plusieurs puits sont directement connectés à une seule station, comme illustré sur la figure 2a. Dans une topologie arborescente, les puits sont reliés à la station via des puits adjacents, comme le montre la figure 2b. Une topologie cyclique implique des puits interconnectés formant une boucle, comme le montre la figure 2c.

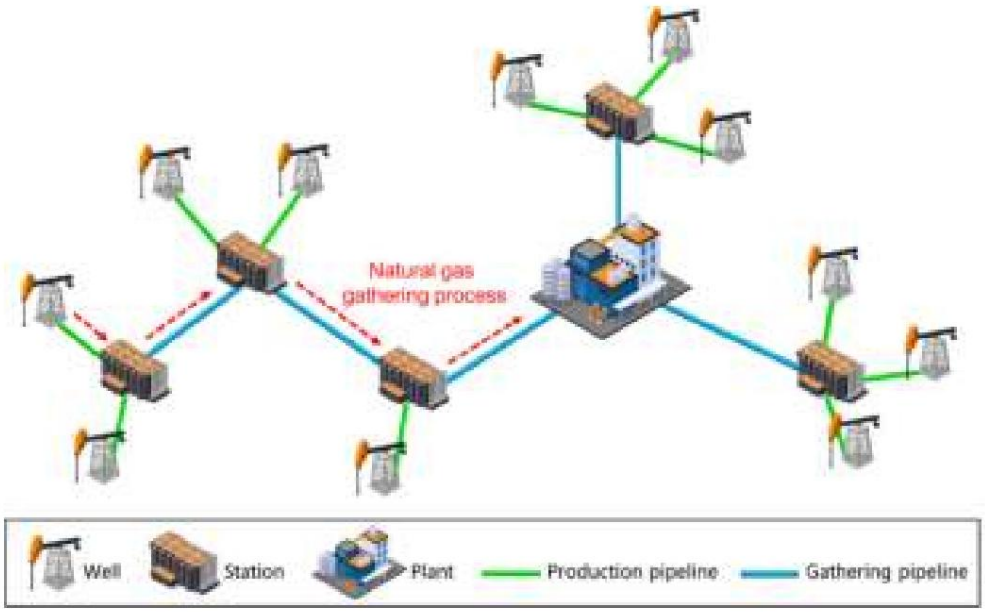


Figure 1. Réseau typique de gazoducs de collecte.

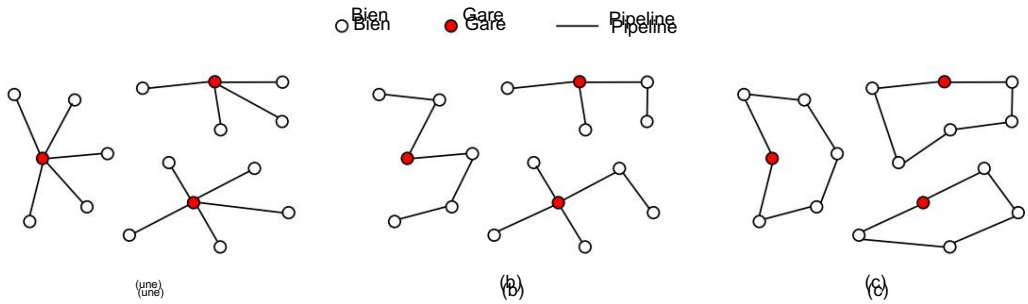


Figure 2. Structure du réseau de pipelines : (a) topologie en étoile, (b) topologie arborescente, (c) topologie cyclique.

Dans les réseaux de gazoducs de collecte de gaz à grande échelle, diverses combinaisons topologiques sont souvent observées. À l'heure actuelle, dans la construction réelle de gisements de gaz, en fonction de divers facteurs, notamment la distribution des puits de gaz, le coût de construction, la commodité de gestion, l'efficacité opérationnelle, etc., les structures de réseau à topologie étoile-étoile et les structures de réseau à topologie étoile-arbre ont été largement utilisé dans le développement des champs de gaz. La structure du réseau à topologie étoile-étoile composée de plusieurs puits de gaz directement connectés à un site central de collecte de gaz rend la configuration simple et facile à gérer et à contrôler. Ses avantages sont que la longueur totale du gazoduc peut être réduite, le coût de construction initial peut être réduit et il est pratique de traiter et de transporter de manière centralisée le gaz naturel produit par le gisement de gaz, ce qui convient à un gisement de gaz avec un petit zone gazière, un grand nombre de têtes de puits et un faible rendement d'un seul puits. La structure du réseau topologique étoile-arbre combine les caractéristiques des structures en étoile et dendritiques. Dans cette structure, chaque puits de gaz et ses stations de collecte de gaz sont reliés par une forme en étoile, et les stations de collecte de gaz sont reliées par des conduites principales de collecte de gaz pour former une structu

Machine Translated by Google

out équilibre le coût et l'efficacité de la ligne de collecte longue distance, tout en offrant une certaine flexibilité pour s'ajuster et s'étendre en fonction du stade de développement du domaine. La figure 3 illustre un réseau typique de gazoducs de collecte de gaz en étoile. Cette structure de réseau est particulièrement adaptée aux scénarios dans lesquels les puits au sein de blocs locaux sont concentrés, mais ces blocs sont largement dispersés avec des distances importantes entre eux.

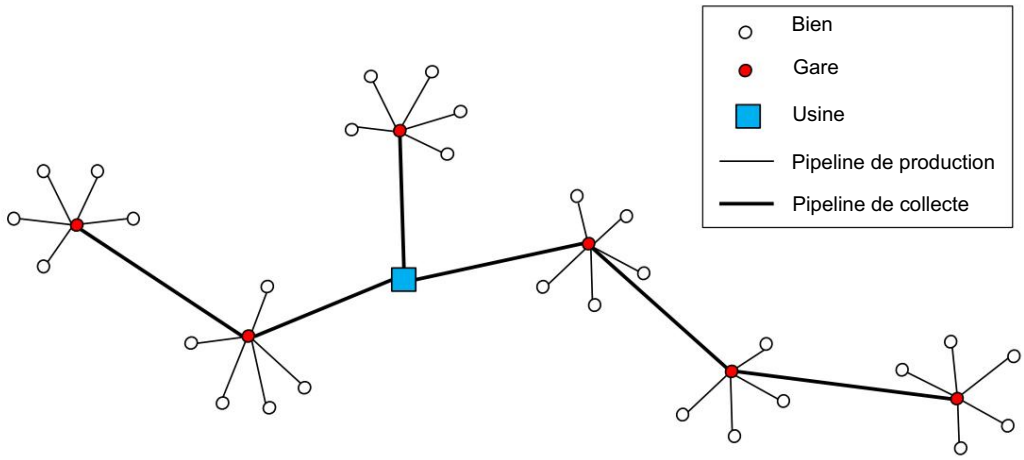


Figure 3. Réseau de pipelines en étoile.

En général, le problème d'optimisation peut être classé en optimisation de conception et optimisation opérationnelle [3]. L'optimisation de la conception implique la sélection de la configuration d'implantation, de l'emplacement des stations, de la taille des canalisations et des échelles d'équipement les plus appropriées pour un nouveau réseau de pipelines afin d'obtenir des coûts d'investissement de construction optimaux. L'optimisation opérationnelle, quant à elle, se concentre sur l'affinement des stratégies opérationnelles du réseau de pipelines établi, ce qui comprend l'optimisation des flux de pipelines et des paramètres de pression des nœuds pour atteindre les objectifs techniques ou économiques.

Le défi d'optimisation de la disposition des réseaux de gazoducs de collecte englobe plusieurs sous-processus, tels que le regroupement des puits, la détermination de l'emplacement des stations et des usines et l'établissement de la connectivité entre les installations [4]. La figure 4 illustre le processus d'optimisation de la mise en page. De plus, la détermination de l'emplacement de la station (ou de l'usine) est divisée en deux scénarios. Le premier scénario est que la station peut être construite à n'importe quel endroit du champ gazier. Le deuxième scénario est que la station doit être construite conjointement avec le puits pour économiser sur les coûts d'investissement dans la construction. En d'autres termes, l'emplacement de la station ne peut être sélectionné qu'à partir de l'emplacement du puits existant, qui est défini comme la sélection de l'emplacement de la station dans un espace discret. Puisque ce scénario est plus largement utilisé dans la pratique de l'ingénierie, cet article se concentrera sur ce problème.

1.2. Revue de littérature

Sur la base de recherches documentaires, il est évident que la communauté universitaire a mené des études approfondies sur l'optimisation de la conception des réseaux de gazoducs. Ces études se concentrent principalement sur la modélisation et la résolution des problèmes d'optimisation liés à l'agencement du réseau . En fonction de la topologie du réseau de pipelines (telle que les topologies en étoile, en arbre et cyclique), différents problèmes d'optimisation surviennent. Sanaye et Mahmoudimehr [5] ont développé un modèle d'optimisation pour les réseaux de gazoducs à topologie arborescente, qui prend en compte des paramètres tels que la disposition du réseau, le diamètre des tuyaux et les pressions des nœuds. Rodríguez et coll. [6] ont étudié la conception de réseaux en étoile à plusieurs niveaux, en abordant les configurations optimales et l'emplacement des installations. Fu [7] a exploré la conception d'optimisation multi-objectifs des réseaux d'oléoducs de type anneau . Wang et coll. [8] ont formulé un modèle MILP (Mixed Integer Linear Programming) multipériode pour optimiser les réseaux de topologie arborescente, en tenant compte du terrain et des obstacles. Zhou et coll. [9] ont établi un MINLP (Mixed Integer Nonlin-

Machine Translated by Google

ear Programming) modèle pour optimiser les réseaux de pipelines à topologie en étoile à deux niveaux pour stockage souterrain de gaz naturel.

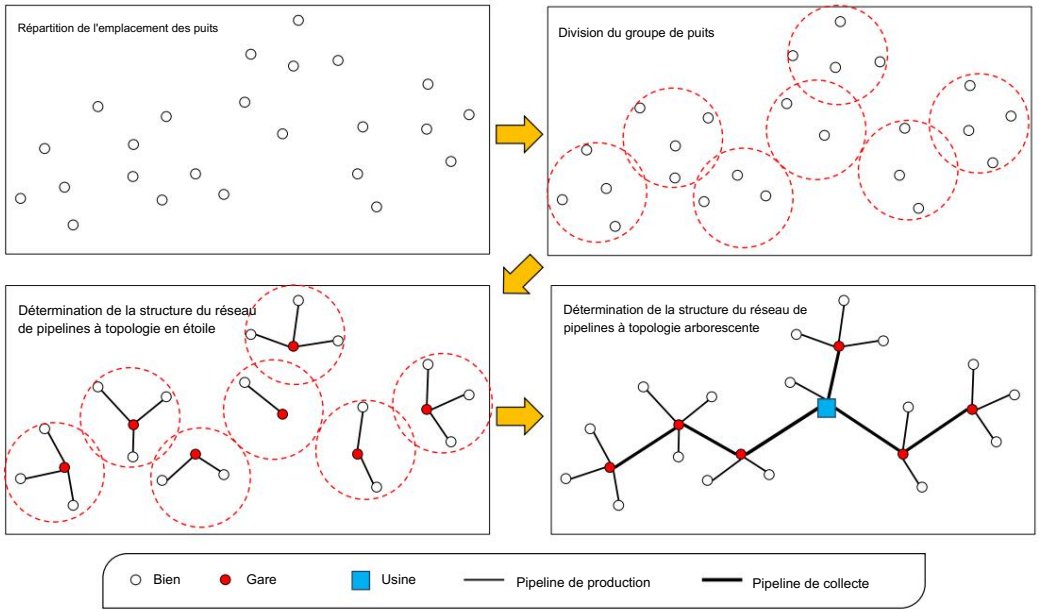


Figure 4. Diagramme d'optimisation hiérarchique.

Concernant les méthodes de résolution d'optimisation, les premières recherches utilisaient principalement des techniques de recherche opérationnelle, en particulier des algorithmes heuristiques, pour aborder le problème de la collecte de gaz. modèles d'optimisation du réseau de pipelines. D'une optimisation et d'une opération mathématiques Du point de vue de la recherche, la conception de réseaux optimaux de gazoducs de collecte est classée comme une tâche importante. Problème d'optimisation combinatoire NP-difficile, qui pose des défis et des découvertes importantes application dans divers domaines industriels [10,11]. Balinski [12] a formulé un modèle d'optimisation pour le placement de sites dans des espaces discrets avec des capacités de site illimitées, qui a été principalement résolu à l'aide d'algorithmes heuristiques. Les algorithmes exacts courants incluent la branche et méthode liée, algorithme de plan de coupe et algorithme de génération de colonnes [13,14]. Cependant, les théories d'optimisation traditionnelles ont souvent du mal à résoudre efficacement les problèmes liés aux modèles de réseaux de gazoducs en raison des temps de calcul longs et des inadéquation entre les résultats théoriques et les applications pratiques.

Ces dernières années, les solveurs commerciaux et les algorithmes heuristiques sont progressivement devenus la manière la plus courante de gérer l'optimisation de la disposition du réseau de pipelines. Une publicité Le solveur peut éliminer le travail fastidieux de programmation d'algorithmes, mais la méthode de résolution est limité et il est difficile de modifier les paramètres intégrés de l'algorithme. L'heuristique L'algorithme a une charge de travail de programmation importante, mais l'algorithme a une bonne évolutivité et peut même combiner les avantages de plusieurs algorithmes pour obtenir les résultats d'optimisation. Uester et Dilaveroglu [15] ont résolu un problème de conception et d'optimisation d'un réseau de gazoducs à l'aide du solveur en ligne Bonmin. Zhang et coll. [16] ont développé une étude complète Modèle d'optimisation de la disposition MILP qui intègre des considérations de terrain et d'obstacles. Certains chercheurs utilisent également des méthodes de théorie des graphes et des méthodes d'apprentissage automatique pour résoudre des problèmes d'optimisation [17,18]. Ils ont utilisé le solveur Gurobi pour résoudre globalement le modèle. Rosa et coll. [19] ont introduit une approche d'optimisation pratique et efficace pour les systèmes sous-marins. réseaux de production à l'aide du solveur GAMS. Liu et coll. [20] proposent une amélioration algorithme d'optimisation des essaims de particules avec des capacités de recherche globales améliorées pour lutter contre le modèle MINLP pour optimiser la disposition des réseaux de pipelines. Hong et coll. [21] ont proposé un modèle d'optimisation de la configuration du réseau de pipelines qui tient compte des caractéristiques hydrauliques et a utilisé un algorithme de colonie de fourmis pour sa solution.

Machine Translated by Google

1.3. Cotisations

L'investissement en surface des champs de gaz représente environ 30 à 40 % du coût total d'investissement en développement, et la proportion des investissements dans la construction de réseaux de gazoducs de collecte et de transport en surface représente souvent 60 à 70 % du coût d'investissement en surface. On peut constater qu'un réseau de gazoducs de collecte et de transport de gaz naturel occupe une place importante dans la construction de l'ingénierie du sol des champs de gaz, et que son optimisation de la conception peut améliorer considérablement les avantages économiques
À l'heure actuelle, certaines méthodes d'optimisation traditionnelles ne suffisent pas pour étudier les structures multi-niveaux des réseaux de pipelines et découvrir une solution efficace. Par conséquent, cet article propose un ensemble de méthodes d'optimisation pour le réseau de gazoducs de collecte de gaz en étoile plus complexe dans des espaces discrets.

Sur la base de la littérature examinée, les chercheurs ont réalisé des progrès significatifs dans l'optimisation de la conception des réseaux de gazoducs. Cependant, ils se sont principalement concentrés sur l'optimisation des topologies à réseau unique, négligeant la complexité introduite par de multiples combinaisons topologiques, en particulier les configurations multi-niveaux en étoile et en arbre qui prévalent dans les champs de pétrole et de gaz. L'inclusion de divers types de réseaux augmente considérablement la complexité de la résolution de ces modèles. Nous abordons ce problème comme suit :

(1) Nous établissons un modèle MINLP pour la conception optimale de réseaux étoile-arbre dans des environnements d espace.

(2) Un cadre d'optimisation hiérarchique est proposé. La première étape porte sur l'optimisation de la conception du réseau de pipelines à topologie en étoile reliant les puits aux stations. La deuxième étape consiste à optimiser la conception du réseau de canalisations à topologie arborescente reliant les stations aux usines.

(3) Un algorithme génétique amélioré (IGA) est proposé et l'algorithme de Kruskal est utilisé pour obtenir des solutions d'aménagement optimales pour le réseau arborescent.

La section 2 délimite le modèle mathématique. La section 3 introduit une hiérarchie cadre d'optimisation. La section 4 présente un cas pratique et examine le résultat de l'optimisation. Enfin, la section 5 fournit un résumé.

2. Modèle mathématique

Dans le réseau de gazoducs de collecte et de transport de gaz naturel, il existe deux principaux types de modèles : l'un est un modèle MIP et l'autre est un modèle MINLP. Dans le modèle MIP, différentes distances point à point sont calculées à l'avance. Pour le développement des gisements de gaz, l'échelle du nombre de puits est très grande, et le calcul de la distance à l'avance est très grand ; par conséquent, dans le modèle de cet article, la longueur du pipeline entre les puits et les stations Lij et la longueur du pipeline entre les stations Ljj′ sont toutes deux euclidienn distances, qui sont respectivement des équations non linéaires, $L_{ij} = \left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2)^{0,5}$ et $L_{jj'} = \left(x_j - x_{j'}\right)^2 + \left(y_j - y_{j'}\right)^2)^{0,5}$, ce qui peut améliorer l'efficacité de la résolution. Dans addition, la variable de décision d'emplacement de la station (δj), la variable de décision d'emplacement de l'usine (δk) et la variable de décision de connexion au pipeline (δij,δjj′) sont des variables binaires. Par conséquent, le modèle présenté dans cet article est un modèle MINLP. Le modèle a une bonne évolutivité, ce qui permettrait d' augmenter le nombre de nœuds à l'avenir.

2.1. Objectif Fonction Un

modèle MINLP est proposé. L'investissement minimum est utilisé comme fonction objectif du modèle, comme le montre l'équation (1).

$$\text{Min } F = FPP + FGP + FS + FC \quad (1)$$

où F est l'investissement dans le réseau de pipelines, CNY. FPP est l'investissement dans le pipeline de production, CNY. FGP est l'investissement dans le pipeline de collecte, CNY. FS est l'investissement de la station, CNY. FC est l'investissement de l'usine, CNY.

Machine Translated by Google

(1) Investissement dans le pipeline de production.

L'investissement dans les pipelines de production correspond au coût de connexion de tous les puits et stations dans le réseau, comme le montre l'équation (2).

$$FPP = \sum_{je} \sum_{I_j} \delta ijcijLij \quad (2)$$

où δij est la variable de décision de connexion entre les puits et les stations, 0 ou 1. cij est le prix du pipeline entre les puits et les stations, CNY/m. Lij est la longueur du pipeline des puits aux stations, en m. Lij = $xi - xj\ (xj, yj)$ est la coordonnée de la station. $\quad + yi - yj \quad ^{0,5}$, (xi , yi) est la coordonnée du puits, et

(2) Rassembler les investissements dans les pipelines.

L'investissement dans les canalisations de collecte correspond au coût de connexion des canalisations entre toutes les stations du réseau, comme le montre l'équation (3).

$$FGP = \sum_{j} \sum_{j'} \delta jj' cjj' Ljj' \quad (3)$$

où δjj′ est la variable de décision de connexion entre les stations, 0 ou 1. cjj′ est le prix du gazoduc entre les stations, CNY/m. Ljj′ est la longueur du pipeline entre les stations, m. Ljj′ = coordonnée $xj - \quad + yj - yj' \quad ^{0,5}$, (xj , yj), et (xj ′ , yj ′) sont les stations ou usines xj ′ .

(3) Investissement en gare.

L'investissement dans la station est la somme des coûts d'investissement de toutes les stations du réseau, comme le montre l'équation (4).

$$FS = \sum_{j} \delta j\ fS \quad (4)$$

où δj est la variable de décision d'emplacement de la station, 0 ou 1. fS est le coût de la station, CNY.

(4) Investissement en usine.

L'investissement dans l'usine est la somme des coûts d'investissement de toutes les usines du réseau de pipelines, comme le montre l'équation (5).

$$FC = \sum_{k} \delta kfP \quad (5)$$

où δk est la variable de décision d'emplacement de l'usine, 0 ou 1. fP est le coût de l'usine, CNY.

2.2. Contraintes

(1) Contrainte d'unicité des connexions de puits.

Chaque puits est exclusivement associé à une station, comme le montre l'équation (6).

$$\sum_{j} \delta ij = 1 \quad i \quad I \quad (6)$$

(2) Contrainte du nombre d'usines.

Il existe une seule usine au sein du réseau de pipelines, comme le montre l'équation (7).

$$\sum_{k} \delta k = 1 \quad (7)$$

(3) Contrainte d'affiliation aux puits et stations.

Eh bien, je peux être affecté à la station j lorsque la station j existe, comme le montre l'équation (8).

$$\delta ij \le \delta j \quad i \quad I, \quad j \quad J \quad (8)$$

Machine Translated by Google

(4) Contrainte de rayon de transmission.

La pression du gaz peut diminuer pendant le transport en raison du frottement des tuyaux. Afin de maintenir la chute de pression dans une plage raisonnable, la distance entre le puits et la station doit être inférieure à une valeur spécifique, appelée rayon de transmission, comme le montre la figure 5.

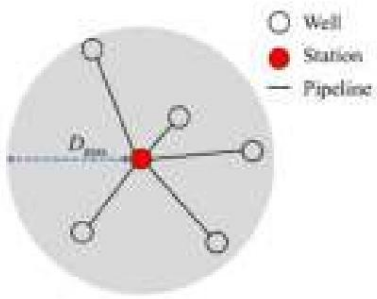


Figure 5. Diagramme schématique du rayon de transmission.

En tant que telle, la contrainte de rayon de transmission est calculée comme indiqué dans l'équation (9).

$$\delta ijLij \leq Dmax_{maximum} \quad I, \quad j \quad ع J \tag{9}$$

où Dmax est la longueur maximale autorisée du pipeline entre le puits et la station, en m.
maximum

(5) Contrainte de capacité de la station.

Généralement, la contrainte de rayon de rassemblement n'est pas prise en compte lors de la résolution de la contrainte de capacité de la station, qui relève du problème de localisation des installations capacitaires (CFLP). Dans cet article, la contrainte de rayon et la contrainte de capacité sont considé
Lorsque la contrainte de rayon de collecte est satisfaite mais que la contrainte de capacité ne l'est pas, la configuration du réseau de pipelines sera repensée, comme le montre la figure 6.

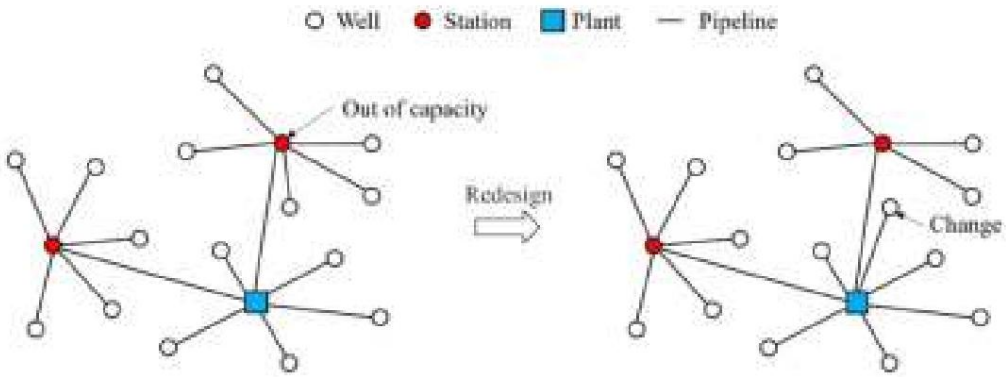


Figure 6. Reconception du réseau de pipelines en violation des contraintes de capacité.

Ainsi, la capacité d'une station ne doit pas dépasser celle indiquée dans l'équation (10).

$$\delta jqj \leq qmax \quad j \quad J \tag{10}$$

maximum

où qj est le débit gazeux du nœud de la station j, m3/d. qmax est la capacité de la station, m3/j.

(6) Contrainte de structure arborescente.
maximum

La contrainte de connexion arborescente garantit que la topologie générée est la plus courte et sans boucle, comme le montre l'équation (11).

$$\sum_{j \ J} \sum_{j' \ J} \delta jj = 2 \sum_{j \ J} \delta j \tag{11}$$

$$2$$

Machine Translated by Google

(7) Contrainte de flux de nœuds.

La contrainte de débit du nœud garantit que le volume total de gaz reste constant, comme le montre l'équation (12).

$$\sum_{je} qi - \sum_{k} \delta kqk = 0 \quad (12)$$

où qi est le débit de gaz du puits i, m3/d. qk est le flux de nœuds de l'usine k, m3/d.

2.3. Variables de décision

Les variables de décision comprennent la variable de décision d'emplacement de station (δj), la variable de décision d'emplacement d'usine (δk) et la variable de décision de connexion de pipeline (δij, δjj'). Toutes ces variables de décision sont des variables binaires, comme le montre l'équation (13).

$$\delta j , \delta k , \delta ij, \delta jj' = \{0, 1\} \quad i \in I, j \in J, j' \in J, k \in K \quad (13)$$

3. Méthode de résolution

L'optimisation de la conception d'un réseau de pipelines en étoile est reconnue comme NP-difficile, ce qui présente des défis pour trouver une solution globale [22]. Pour résoudre ce problème, un cadre d'optimisation hiérarchique est proposé dans cette étude. Le cadre se compose de deux étapes : la première étape consiste à déterminer la division des groupes de puits et l'emplacement des stations au sein du réseau topologique en étoile, tandis que la deuxième étape se concentre sur l'établissement des connexions des stations et des emplacements des installations dans le réseau topologique arborescent. Notamment, l'optimisation du réseau topologique en étoile dans la première étape constitue l'élément fondamental et essentiel de l'ensemble du problème d'optimisation. Sa complexité et son ampleur dépassent celles de l'optimisation du réseau à topologie arborescente de deuxième étape. Par conséquent, cet article se concentre sur l'amélioration de l' optimisation de la conception et présente un IGA adapté pour relever ce défi. De plus, l'algorithme de Kruskal, exploitant l'approche de l'arbre couvrant minimum, est utilisé pour l'optimisation de la conception du réseau de pipelines à topologie arborescente de deuxième étape.

3.1. Algorithme génétique de

base Une GA est une méthode de recherche heuristique qui s'inspire des principes de la sélection naturelle et de la génétique. Il vise à optimiser l'attribution des installations en simulant le processus évolutif des chromosomes biologiques. Le processus de résolution de base de GA [23] se déroule comme suit :

(1) Codage : coder les informations sur la répartition des puits et des stations dans la séquence chromosomique.

(2) Génération de la population initiale : une taille fixe de chromosomes (individus) est générée aléatoirement pour former la population initiale P(t) (dans cet article, 200 groupes sont générés). Ici, t désigne le compteur de génération évolutive et T représente le nombre maximum de cycles d'évolution défini pour l'algorithme.

(3) Fonction de condition physique : chaque individu de la population est évalué sur la base d'une fonction de condition physique, en utilisant généralement la valeur totale de l'investissement comme critère d'évaluation.

(4) Sélection et héritage : la population initiale est sélectionnée et croisée en simulant le processus d'évolution biologique, et la population de génération suivante P(t + 1) est obtenue via l'étape 3.

(5) Jugement des conditions de terminaison : lorsque t ≤ T, alors t + 1 remplace t, et on passe à l'étape 2 ; si t > T, nous considérons l'individu ayant une forme physique optimale tout au long du processus d'évolution comme la solution optimale et mettons fin à l'opération [24].

Le processus ci-dessus est également décrit par l'organigramme illustré à la figure 7.

Machine Translated by Google

≤

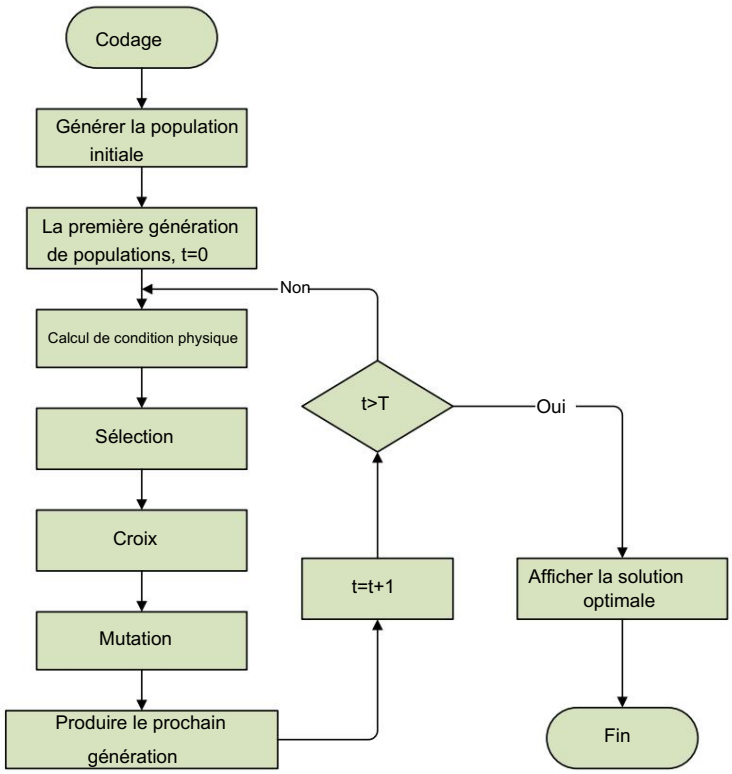


Figure 7. Organigramme du GA de base.

3.2. Amélioration de l'algorithme

Compte tenu de la nature discrète de l'espace et de la structure spécifique des réseaux de gazoducs, l'AG traditionnelle et ses paramètres associés peuvent ne pas s'aligner entièrement sur le modèle mathématique proposé dans cet article. Il est donc essentiel d'améliorer l'AG de base pour mieux répondre à ces caractéristiques et optimiser efficacement la conception du réseau de pipelines. L'amélioration comprend le codage global de la correspondance puits-station, la formulation de la fonction d'aptitude de la solution réseau et la refonte de l' opérateur de croisement et de l'opérateur de mutation.

(1) Conception du codage.

En tenant compte des caractéristiques des contraintes, en utilisant la relation de connexion entre les puits et les stations comme base de codage, un codage en nombres réels est utilisé pour améliorer l'efficacité du GA. La séquence codant pour les chromosomes est structurée de telle sorte qu'elle a un impact direct sur la fonction objectif, qui dans ce cas est l'investissement total.
La longueur de chaque chromosome correspond au nombre total de puits, et au sein de la séquence, l'identifiant de chaque puits est systématiquement représenté par ordre croissant. Cette méthode de codage garantit que le GA répond de manière optimale aux contraintes de conception et cible efficacement l'optimisation de la configuration du réseau de gazoducs. Le numéro de station correspondant à chaque puits est généré de manière stochastique à chaque position dans le chromosome pour indiquer le lien entre le puits et sa station respective. Dans le schéma de codage chromosomique proposé, chaque chiffre d'un chromosome = (C1, C2, . . . Cm) représente le numéro de la station correspondant à chaque puits.

(2) Conception de l'opérateur croisé.

Pour le GA, la méthode d'opération de croisement et la probabilité de croisement sont des paramètres extrêmement importants qui affectent directement la qualité d'optimisation de l'algorithme. À l'aide des données CAP101 basées sur OR-LIBRARY (50 puits, 25 stations) [25], la méthode d'opération de croisement et la probabilité de croisement sont testées pour leur influence sur la solution, et le meilleur mode de croisement et la meilleure probabilité de croisement sont déterminés.

Chacune des techniques croisées a été testée 50 fois et leurs résultats ont été moyennés à des fins de comparaison. Au fur et à mesure que le nombre de générations itératives augmentait, l'algorithme con-

Machine Translated by Google

La méthode de croisement multipoint a démontré des performances supérieures, conduisant à sa sélection. La figure 8 illustre les résultats moyens des tests.

La probabilité de croisement varie entre 0,5 et 0,9. Une comparaison des effets de ces différentes probabilités de croisement sur les résultats a été réalisée en faisant la moyenne des résultats. provient de 50 tests. Lorsque la probabilité de croisement est comprise entre 0,7 et 0,9, les résultats sont similaires. Augmenter correctement la probabilité de croisement peut améliorer la convergence de la solution, c'est pourquoi la probabilité de croisement doit être maintenue à une valeur élevée dans la conception. Le chiffre 9 montre l'effet de différentes probabilités de croisement sur les résultats dans la méthode de croisement multipoint. La méthode de croisement multipoint a démontré une performance supérieure, conduisant à sa sélection. La figure 8 illustre les résultats moyens des tests.

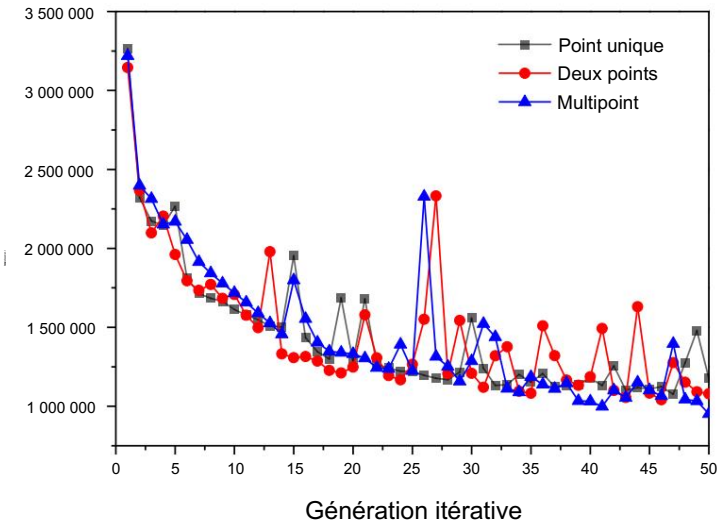


Figure 8. Courbe de convergence des différentes méthodes de fonctionnement du crossover.

La probabilité de croisement varie entre 0,5 et 0,9. Une comparaison des effets de ces différentes probabilités de croisement sur les résultats a été réalisée en faisant la moyenne des résultats de 50 tests. Lorsque la probabilité de croisement est comprise entre 0,7 et 0,9, les résultats sont semblables. Augmenter correctement la probabilité de croisement peut améliorer la convergence de la solution, la probabilité de croisement doit donc être maintenue à une valeur élevée dans la conception. La figure 9 montre l'effet de différentes probabilités de croisement sur les résultats de l'analyse multipoint. méthode de croisement.

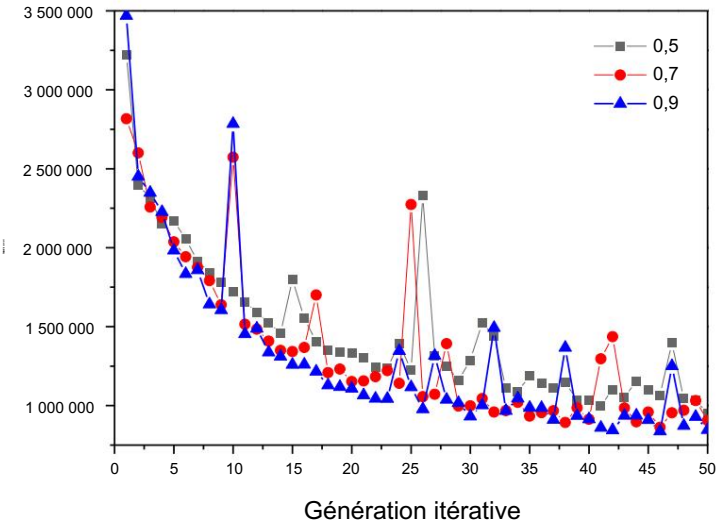


Figure 9. Courbe de convergence des différentes probabilités de croisement.

(3) Conception de l'opérateur de mutation

L'opérateur de mutation est une mesure qui empêche l'AG de tomber dans le local optimum local en simulant les valeurs génétiques et évolutives de l'allèle de rupture [26]. Dans le processus de débogage de l'opérateur de mutation, une fonction de variation est testée pour le GA afin d'atténuer le risque de tomber dans des optima locaux dans les étapes ultérieures. Les opérateurs de mutation emploient une approche adaptative, comme indiqué dans la formule (14).

$$= \begin{cases} - & + \\ & - \\ = & + \end{cases} \quad (14)$$

Machine Translated by Google

atténuer le risque de tomber dans des optima locaux à des stades ultérieurs. Les opérateurs de mutation emploient une approche adaptative, comme indiqué dans la formule (14).

$$\begin{cases} MP = \varepsilon \cdot \dfrac{Fitmax - Fit}{Fitmax - Fitmin} & \text{Ajustement} \geq \dfrac{Fitmax + Fitmin}{2} \\ MP = \varepsilon & \text{Ajuster} < \dfrac{Fitmax + Fitmin}{2} \end{cases} \quad (14)$$

où MP est la probabilité de mutation ; L'ajustement est la valeur de la condition physique ; Fitmax et Fitmin sont les valeur de condition physique maximale et minimale de chaque génération ; et ε est le coefficient de mutation.

Grâce à des tests, il a été constaté que lorsque le coefficient de mutation est inférieur à 0,003, le la stabilité itérative et la convergence de l'ensemble du processus de solution sont meilleures, et la solution le temps est plus raisonnable ; cependant, si le coefficient de mutation a une valeur supérieure à 0,1, la solution déterminée peut être très éloignée de la solution optimale, affectant la stabilité de l'itération et de la vitesse de convergence et d'autres problèmes. Ainsi, la mutation Le coefficient a été choisi inférieur à 0,003.

L'impact de la probabilité de mutation sur les investissements est décrit dans le tableau 1. Ces les expériences sur CAP101 démontrent que lorsque la probabilité de mutation est définie sur un valeur, il garantit la diversité de la population sans compromettre la convergence de l'AG vers la solution optimale globale. La raison principale en est que lorsque le nombre réel le codage est utilisé, chaque position sur chaque chromosome contient plus d'informations que lorsque d'autres méthodes de codage sont utilisées, et lorsque la probabilité de mutation est élevée, il est facile de détruire directement l'individu dominant.

Tableau 1. Impact des différentes probabilités de mutation sur les performances de GA.

Taille du problème	Nombre total de calculs	ε	Résultat optimal moyen
50 × 25	10	0,0001	861 682,74
		0,0002	830 511,15
		0,0003	818 015,08
		0,0004	796 668,16
		0,0005	806 771,78
		0,001	809 352,78
		0,002	855 745,99
		0,003	834 490,13

3.3. Tests IGA

Pour mieux aborder la conception optimale des réseaux de pipelines en étoile dans des espaces discrets, cet article met l'accent sur les améliorations apportées au schéma de codage, à l'opérateur de croisement et à l'opérateur de mutation de l'AG fondateur. Les étapes de résolution de l'IGA sont illustrées à la figure 10.

Valider l'efficacité de l'algorithme génétique modifié dans la résolution du pipeline problème de conception de configuration du réseau, une série de tests d'algorithmes ont été effectués. Tous les algorithmes ont été implémentés dans le logiciel MATLAB 2022. Les expériences numériques ont été exécutées sur un ordinateur équipé d'un processeur Intel(R) Core (TM) i5-2450M à 2,50 GHz, 4 Go de RAM et système d'exploitation Windows 7.

Deux algorithmes classiques ont été sélectionnés pour comparaison avec le GA amélioré, à savoir optimisation des essaims de particules avec facteur de constriction (CPSO) et recherche tabou (TABU). Plusieurs les problèmes de OR-LIBRARY ont été utilisés pour des tests algorithmiques, et la solution les résultats sont présentés dans le tableau 2. LOCAL est la solution optimale réelle. n signifie bien nombre. m représente le numéro de la station. Pour un plus petit nombre de puits et de stations, comme dans le cas du CAP101, les avantages de l'utilisation de l'IGA ne sont pas évidents. En fait, le IGA peut nécessiter plus de temps de calcul que d'autres algorithmes. Cependant, comme le le nombre de n (puits) et m (stations) augmente, les avantages de l'efficacité de l'IGA dans

Machine Translated by Google

tt

la résolution des problèmes devient de plus en plus évidente. Il tend à converger vers la solution optimale plus rapidement que les autres algorithmes à mesure que n et m grandissent.

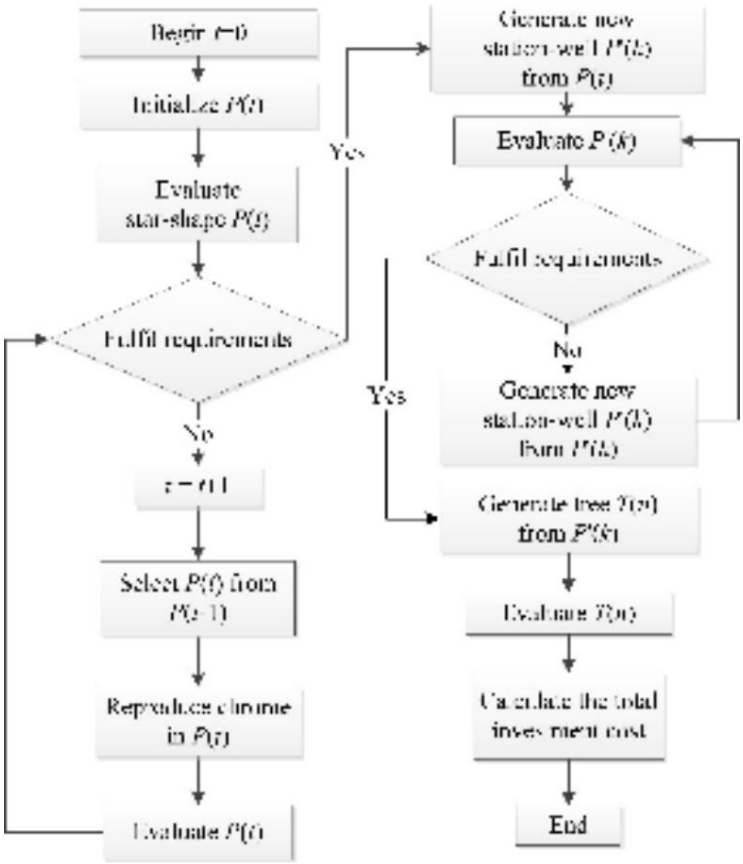


Figure 10. Organigramme de l'IGA.

ff

Tableau 2. Comparaison avec d'autres algorithmes.

Problème	Taille (n × m)	Algorithme	Solution optimale (moy.)	Temps (s)
CAP101	50 × 25	LOCALE	796 648.437	-
		TABU	797 508,70	0,19
		OMCO	796 648,44	0,88
		IGA	796 668,16	0,37
CAP103	50 × 25	LOCALE	893 782,11	-
		TABU	894 008,10	0,24
		OMCO	893 782,11	0,99
		IGA	895 192,66	0,17
CAP104	50 × 25	LOCALE	928 941,75	-
		TABU	928 941,80	0,15
		OMCO	928 941,75	0,60
		IGA	928 946,48	0,17

Machine Translated by Google

Tableau 2. Suite

Problème	Taille (n × m)	Algorithme	Solution optimale (moy.)	Temps (s)
CAP131	50 × 50	LOCALE	793 439,56	-
		TABU	794 299,90	0,20
		OMCO	793 439,56	3,62
		IGA	793 902,48	0,37
CAP133	50 × 50	LOCALE	893 076,71	-
		TABU	893 782,10	0,30
		OMCO	893 076,71	3,78
		IGA	893 994,01	0,22
CAPB	1000 × 100	LOCALE	13 198 556,43	-
		TABU	13 214 718,1	3,91
		IGA	13 211 519,82	3.26

3.4. Algorithme de Kruskal

Lors de la connexion de chaque station à l'aide d'une topologie arborescente, la tâche fondamentale est de répondre le problème de l'arbre couvrant minimum. Dans la conception de l'arbre couvrant minimum, les algorithmes couramment utilisés incluent l'algorithme Prim et l'algorithme Kruskal, dont le noyau est basé sur le principe gourmand. En comparant les deux algorithmes, il a été constaté que L'algorithme de Kruskal est toujours supérieur à l'algorithme de Prim en termes de temps de résolution, nous avons donc choisi d'utiliser l'algorithme de Kruskal comme algorithme de conception de réseau à topologie arborescente. Le Kruskal L'algorithme est un algorithme Spanning Tree basé sur des arêtes pondérées. En cours de résolution, le plus petit bord pondéré est successivement sélectionné et ajouté à l'ensemble d'arbres. Quand le l'arête ajoutée génère une topologie cyclique, l'arête pondérée est supprimée, jusqu'à ce que toutes les arêtes aient été sélectionné [27].

Le processus de résolution de l'algorithme de Kruskal est illustré à la figure 11. Tous les nœuds sont divisé en l'ensemble de nœuds de jointure U et l'ensemble de nœuds de dispersion non joints V. Au début, V = {v1, v2, v3, v4, v5} et U = {φ}. Premièrement, les nœuds qui forment l'arête la plus courte sont sélectionné et ajouté à l'ensemble de nœuds de jointure, c'est-à-dire U = {v4, v5} et V = {v1, v2, v3}. Puis, dans de la même manière, les arêtes v1, v2, v1, v3, v4, v2 et v4, v3 sont sélectionnées dans l'ensemble de nœuds de jointure. Quand un la typologie cyclique est créée, l'arête qui l'a produite est supprimée et la solution est complète. Enfin, V = {φ} et U = {v1, v2, v3, v4, v5}.

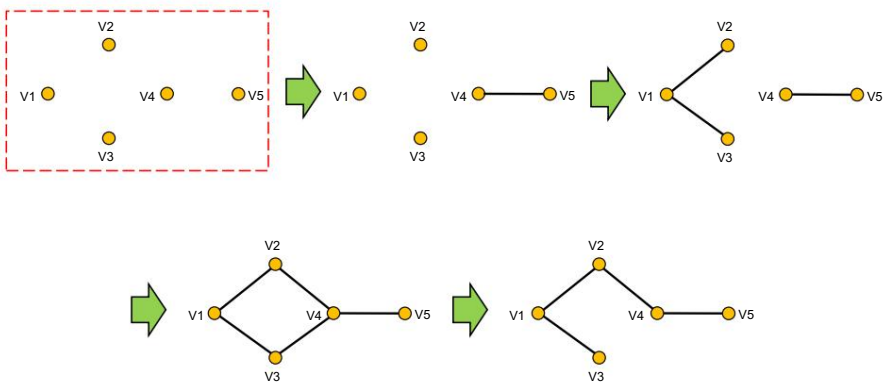


Figure 11. Diagramme schématique du processus de résolution de l'algorithme de Kruskal.

4. Résultats et discussion

Pour un système de réseau de gazoducs de collecte et de transport de gisements de gaz naturel, le tracé la conception est un problème NP, qui n'est pas propice au calcul. Afin d'améliorer le l'efficacité de la solution, les quatre hypothèses suivantes sont faites : (1) Pour le

Machine Translated by Google

conditions de calcul dans l'étude d'implantation, les conditions standards deviennent les conditions par défaut pour le calcul (pression de 0,1 MPa, température de 20 ◦C). (2) Le débit dans le gazoduc est instable. Afin de résoudre le problème de commodité, le flux de gaz dans le pipeline est simplifié en un flux stable. (3) Dans le flux du pipeline, la température du pipeline enterré change généralement avec la longueur du pipeline, et lorsque le pipeline est trop long, le gaz finira par se rapprocher de la température ambiante. Afin de faciliter les calculs, et puisque le réseau de canalisations de collecte est généralement situé dans la même zone, la température de la canalisation est supposée constante dans cet article. (4) Pour le sens d'écoulement du gaz dans le réseau, il est stipulé que chaque nœud ne peut transporter le gaz que d'un sens à l'autre.

Un véritable réseau de gazoducs en Chine a été sélectionné pour analyse et recherche d'une solution [28]. La disposition originale du réseau de gazoducs dans le champ gazier est représentée à la figure 12. Ce champ gazier comprend 93 puits, chacun produisant entre 2 000 et 5 000 m3/j de gaz, avec des pressions en tête de puits allant de 0,2 à 1,0 MPa et une production moyenne. température de 20 ◦C. La spécification des canalisations de production primaires entre les puits et les stations était de De160 × 9,5 mm, tandis que la spécification des canalisations de collecte secondaires entre les stations (et l'usine) était de De225 × 12,8 mm. Les coûts détaillés sont fournis dans le tableau 3.

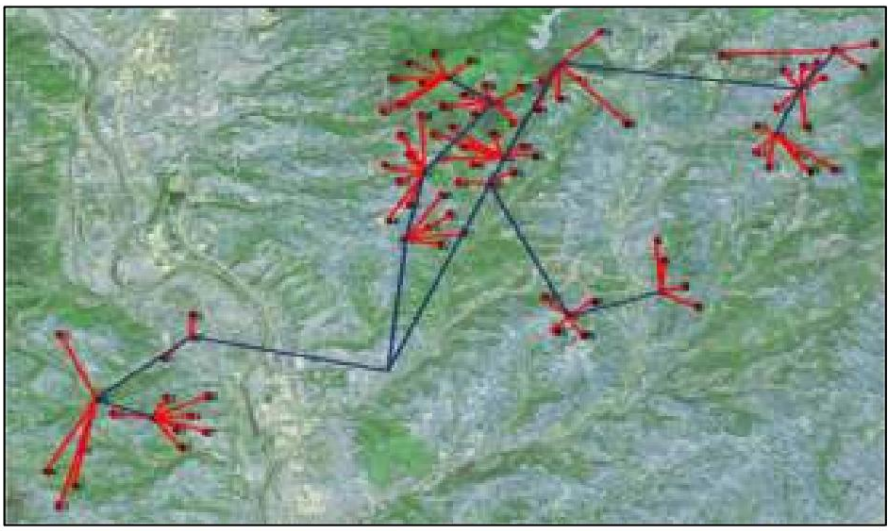


Figure 12. Croquis de la disposition originale du réseau de pipelines.

Tableau 3. Coûts projetés du réseau de pipelines.

Projet	Coût
Gare	52,17 CNY × 104
Usine	1725,93 CNY × 104
Coût du pipeline (De160 × 9,5 mm)	CNY/km 15,636 × 104
Coût du pipeline (De225 × 12,8 mm)	CNY/km 25,406 × 104

La contrainte de rayon maximum était de 1,5 km, et la capacité de traitement d'une seule station était de 20 000 m3/jour. Puisque les données provenaient d'un champ de gaz à basse pression, les contraintes de pression n'ont pas été prises en compte dans ce cas. Dans l'algorithme original (GA), le calcul a prenait plus de 2 heures la fois. Après avoir optimisé la structure de l'algorithme (IGA), le temps de calcul a été réduit à 15 à 30 minutes, augmentant ainsi l'efficacité du calcul de plus de 400 %. L' IGA a été utilisé pour résoudre le tracé du réseau de pipelines.

En utilisant le cadre d'optimisation hiérarchique établi, nous avons dérivé un schéma de mise en page optimisé. La disposition initiale de la topologie en étoile dans la première étape est illustrée dans

Machine Translated by Google

Figure 13. La figure 14 affiche les résultats du regroupement des puits et de la détermination de l'emplacement des stations . La configuration finale optimisée est illustrée à la figure 15. Ce réseau se compose de quatorze stations de collecte de gaz et d'une usine de traitement, soit une station de moins que la conception originale du champ de gaz.



Figure 13. Relation de connexion entre les puits et les stations.

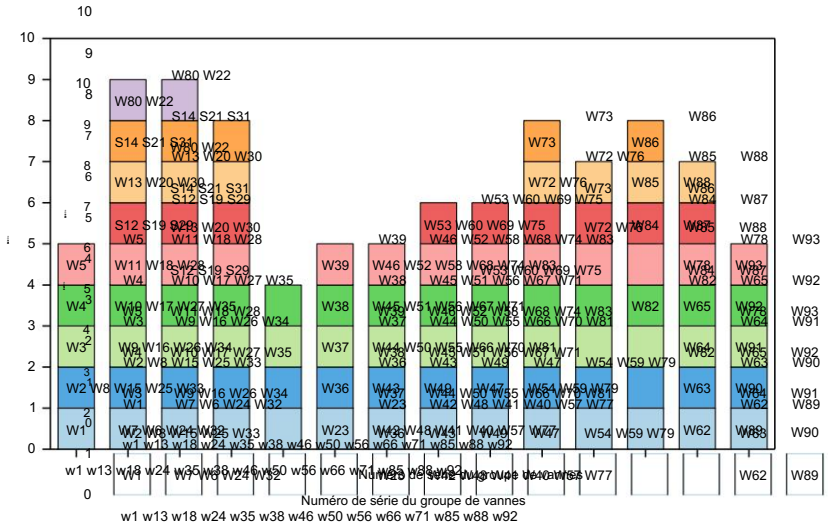


Figure 14. Résultats pour l'attribution des puits, l'emplacement de la station et l'emplacement de l'usine.

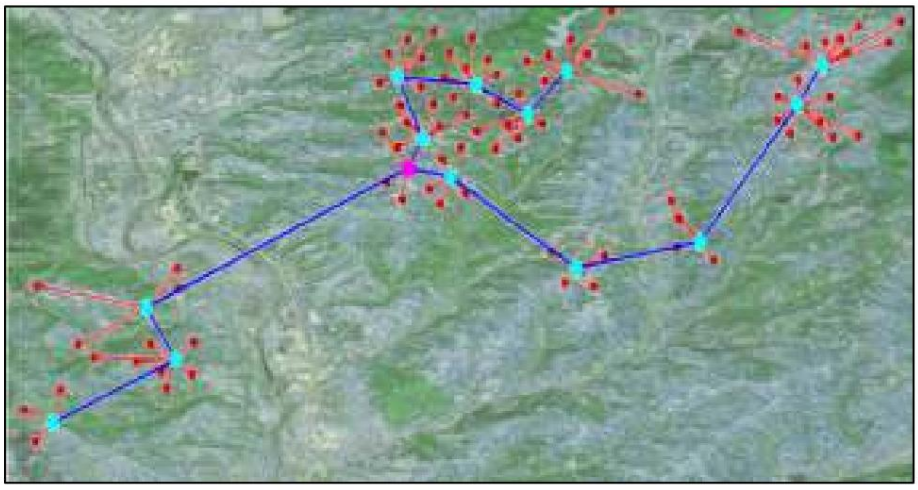


Figure 15. Résultat de conception pour la disposition du réseau de pipelines en étoile.

Machine Translated by Google

En utilisant le logiciel de simulation pour vérifier la faisabilité de la structure du réseau de pipelines et en considérant que le nombre de puits dans le réseau de pipelines était trop grand, le nombre de puits a été déterminé comme la somme du volume de gaz et du nœud du collecteur de vannes. pour la vérification de la connexion. Le schéma de connexion est présenté à la figure 16. L'étalonnage de la chute de pression correspondant est illustré à la figure 17. La chute de pression se situait dans la plage de chute de pression admissible, répondant aux exigences du transport de gaz.

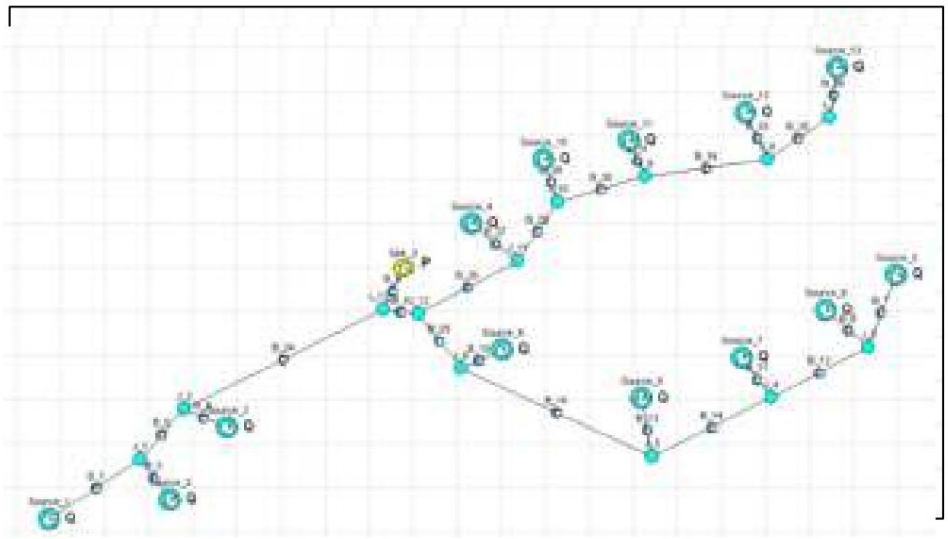


Figure 16. Illustration d'une simulation de connexion à un réseau de pipelines.

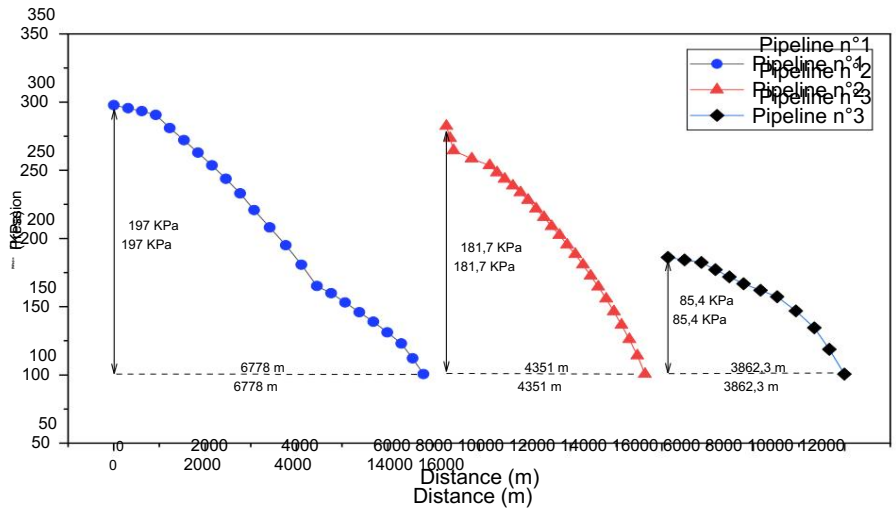
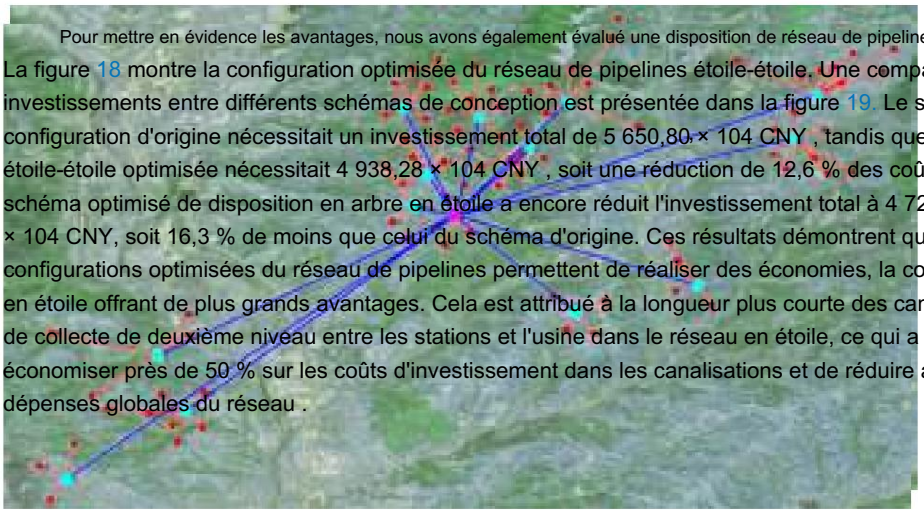


Figure 17. Résultats d'étalonnage de la chute de pression de l'agencement du réseau de pipelines.

Pour mettre en évidence les avantages, nous avons également évalué une disposition de réseau de pipelines étoile à étoile. La figure 18 montre la configuration optimisée du réseau de pipelines étoile-étoile. Une comparaison des investissements entre différents schémas de conception est présentée dans la figure 19. Le schéma de configuration d'origine nécessitait un investissement total de 5 650,80 × 104 CNY , tandis que la disposition étoile-étoile optimisée nécessitait 4 938,28 × 104 CNY , soit une réduction de 12,6 % des coûts. Le schéma optimisé de disposition en arbre en étoile a encore réduit l'investissement total à 4 728,28 lequel × 104 CNY, soit 16,3 % de moins que celui du schéma d'origine. Ces résultats démontrent que les deux configurations optimisées du réseau de pipelines permettent de réaliser des économies, la configuration en étoile offrant de plus grands avantages. Cela est attribué à la longueur plus courte des canalisations de collecte de deuxième niveau entre les stations et l'usine dans le réseau en étoile, ce qui a permis d' économiser près de 50 % sur les coûts d'investissement dans les canalisations et de réduire ainsi les dépenses globales du réseau .



Machine Translated by Google

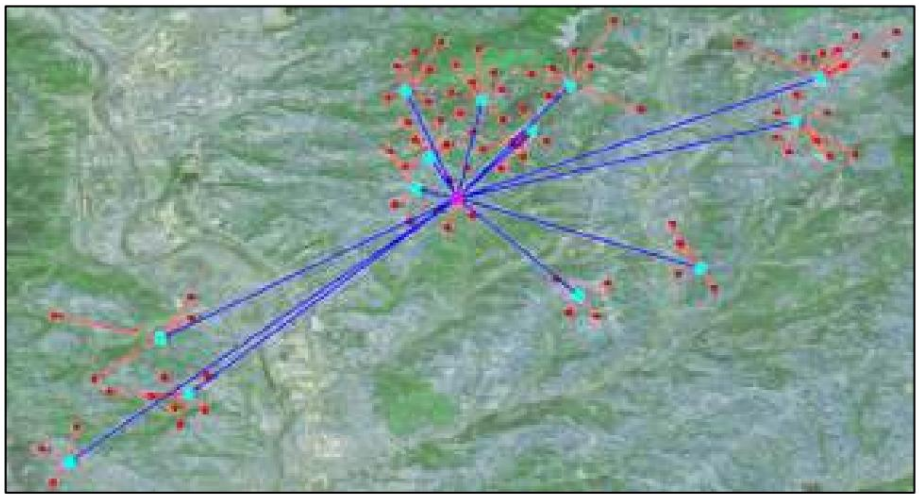


Figure 18. Résultat de conception pour la disposition du réseau de pipelines étoile-étoile.

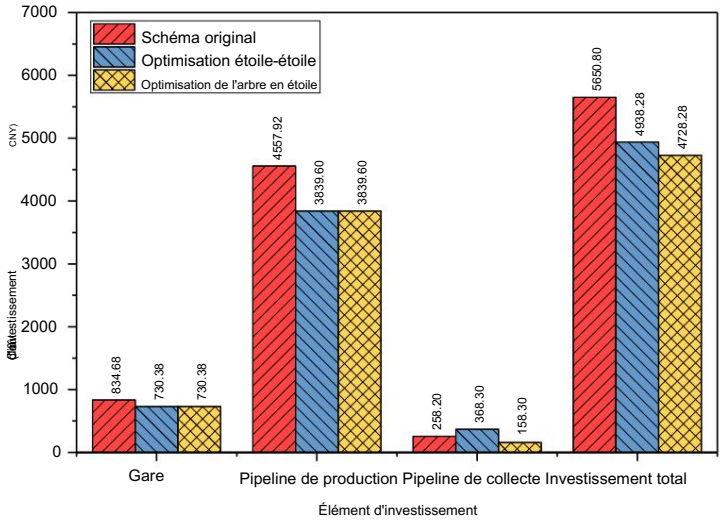


Figure 19. Comparaison des coûts d'investissement.

5. Conclusions

La recherche universitaire sur l'optimisation de l'agencement des réseaux de gazoducs s'est principalement concentrée sur les réseaux à un seul niveau et à structure unique, avec une attention limitée accordée aux réseaux de gazoducs combinés à plusieurs niveaux et multitypes. Pour combler cette lacune, cet article établit un modèle d'optimisation de la conception pour un réseau de gazoducs de collecte de gaz en étoile à deux étages dans des espaces discrets. Reconnaissant la complexité du problème et la nécessité d' une efficacité informatique, un cadre d'optimisation hiérarchique est développé parallèlement à un IGA pour résoudre efficacement ce problème d'optimisation. Sur la base de ce modèle,

nous avons déterminé la conception optimisée d'un réseau de pipelines en étoile pour un champ de gaz naturel en Chine, ce qui a abouti à des résultats de conception optimaux. Les résultats révèlent que le schéma d'aménagement optimisé du réseau permet d'obtenir une réduction de 16 % des coûts d'investissement par rapport au schéma d'origine, ce qui représente une économie de 922,52 x 104 CNY . De plus , nous avons comparé cette disposition optimisée avec une topologie étoile-étoile. Il a été observé que l'adoption de la disposition étoile-étoile augmenterait le coût d'investissement de 4 % par rapport à la disposition étoile-arbre. Cette comparaison souligne l'applicabilité supérieure de la topologie en étoile pour ce système de réseau particulier.

Bien que le modèle établi dans cet article résolve efficacement le problème d'optimisation de la conception du réseau de gazoducs en étoile , il présente toujours les inconvénients suivants : Le modèle ne prend pas en compte les caractéristiques d'écoulement hydraulique du gaz naturel, il ne peut donc pas efficacement

Machine Translated by Google

juger si les changements de pression du fluide dans le réseau de canalisations répondent aux exigences technologiques exigences. De plus, la stratégie d'optimisation hiérarchique utilisée dans cette étude ne garantit pas que la solution obtenue soit la solution globale optimale. Par conséquent, le futur la recherche devrait intégrer pleinement les caractéristiques de l'exploitation du gaz dans le gazoduc modèle d'optimisation de la conception du réseau. De plus, l'adoption d'une approche mondiale plus systématique La méthode d'optimisation sera essentielle pour relever efficacement ce défi.

Contributions des auteurs : Validation, YQ ; analyse formelle, HC; enquête, JH ; logiciel, PD; données curation, DL; rédaction – ébauche originale, YL et JZ ; rédaction – révision et édition, YL et JZ Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche a été financée par une subvention de la [Fondation nationale des sciences naturelles de Chine] numéro [51704253].

Déclaration de disponibilité des données : Les données qui soutiennent les conclusions de cette étude sont disponibles auprès du auteurs correspondants sur demande raisonnable.

Remerciements : Les auteurs remercient chaleureusement le soutien financier fourni par le Fondation nationale des sciences naturelles de Chine (51704253) pour ce projet.

Conflits d'intérêts : les auteurs Yu Lin et Yanhua Qiu étaient employés par la société PetroChina Société pétrolière et gazière du sud-ouest. L'auteur Hao Chen était employé par la société Sichuan Shale Gas Exploration and Development Co., Ltd. L'auteur Dafan Liu était employé par l'entreprise CNPC Chuanqing Drilling Engineering Co., Ltd. Les autres auteurs déclarent que la recherche a été menée en l'absence de toute relation commerciale ou financière pouvant être interprétée comme un conflit d'intérêts potentiel.

Nomenclature

Ensembles	
je	Bien défini, i I = {1, 2, . . . , NI}
J.	Ensemble de positions candidates de station, j J = 1, 2, . . . , New Jersey
K	Ensemble de positions de candidats végétaux, k K = {1, 2, . . . , NK}
Paramètres	
F	Investissement dans le réseau de pipelines, CNY
cij	Prix du pipeline entre les puits et les stations, CNY/m
$c_{jj'}$	Prix du pipeline entre les stations, CNY/m
Dmax	Longueur maximale autorisée du pipeline du puits à la station, m
fP	Coût de l'usine, CNY
fS	Coût de la gare, CNY
FC	Investissement dans l'usine, CNY
FS	Investissement dans la gare, CNY
FPP	Investissement dans le pipeline de production, CNY
FGP	Rassemblement des investissements dans le pipeline, CNY
Lij	Longueur du pipeline des puits aux stations, m
Ljj′	Longueur du pipeline entre les stations, m
qj	Débit de gaz du nœud de la station j, m3/j
qmax	Capacité de la station, m3/j
qi	Débit de gaz du puits i, m3/j
qk	Débit de gaz nœud de l'installation k, m3/j
δij	Variable de décision de connexion entre puits et stations, 0 ou 1
δj	Variable de décision d'emplacement de la station, 0 ou 1
δjj′	Variable de décision de connexion entre stations, 0 ou 1
δk	Variable de décision d'emplacement de l'usine, 0 ou 1

Machine Translated by Google

Références

1. Zhou, J. ; Fu, T. ; Chen, Y. ; Xiao, Y. ; Peng, J. ; Liang, G. Un modèle de programmation non linéaire en nombres entiers mixtes pour la conception optimale de réseau bitube de surface de stockage de gaz. J. Stockage d'énergie 2021, 44, 103379. [CrossRef]
2. Rios-Mercado, R. ; Borraz-Sanchez, C. Problèmes d'optimisation dans les systèmes de transport de gaz naturel : un état de l'art. Appl. Énergie 2015, 147, 536-555. [Référence croisée]
3. Demissie, A. ; Zhu, W. ; Belachew, CT Un modèle d'optimisation multi-objectifs pour l'exploitation des gazoducs. Calculer. Chimique. Ing. 2017, 100, 94-103. [Référence croisée]
4. Hong, B. ; Li, X. ; Creuser.; Chanson, S. ; Yu, W. ; Chen, S. ; Li, Y. ; Gong, J. Un modèle MILP intégré pour une planification optimale de plusieurs périodes système de pipelines de collecte de champs de gaz terrestres. Calculer. Ind. Ing. 2020, 146, 106479. [Réf. croisée]
5. Sanaye, S. ; Mahmoudimehr, J. Conception optimale d'un réseau de transport de gaz naturel. Chimique. Ing. Rés. Des. 2013, 91, 2465-2476. [Référence croisée]
6. Rodriguez, DA ; Oteiza, PP; Brignole, NB Optimisation du recuit simulé pour les réseaux de pipelines d'hydrocarbures. Ind. Ing. Chimique. Rés. 2013, 52, 8579-8588. [Référence croisée]
7. Fu, YX Conception d'optimisation des paramètres multi-objectifs d'un réseau de canalisations de collecte d'huile et d'eau de mélange de type anneau unique. Clé Ing. Maître. 2011, 467, 1285-1290. [Référence croisée]
8. Wang, B. ; Yuan, M. ; Zhang, H. ; Zhao, W. ; Liang, Y. Un modèle MILP pour la conception optimale d'un réseau de transport de gaz naturel multipériode . Chimique. Ing. Rés. Des. 2018, 129, 122-131. [Référence croisée]
9. Zhou, J. ; Zhou, X. ; Liang, G. ; Peng, J. Un modèle MINLP pour l'aménagement du réseau de stockage souterrain de gaz naturel. J. Intell. Système flou . 2020, 38, 4619-4642. [Référence croisée]
10. Daskin, M. Modèles, algorithmes et applications de réseaux et de localisation discrète. J. Opér. Rés. Soc. 1997, 48, 763-764. [Référence croisée]
11. Posta, M. ; Ferland, JA; Michelon, P. Une méthode coopérative exacte pour le problème de localisation des installations sans capacité. Mathématiques. Programme. Calcul 2014, 6, 199-231. [Référence croisée]
12. Balinski, ML Programmation entière : méthodes, utilisations, calculs. Gérer. Sci. 1965, 12, 253-313. [Référence croisée]
13. Klincewicz, JG; Luss, H. ; Rosenberg, E. Algorithmes optimaux et heuristiques pour la localisation d'installations multiproduits sans capacité. EUR. J. Opéra. Rés. 1986, 26, 251-258. [Référence croisée]
14. Efroymson, MA; Ray, TL Un algorithme lié aux branches pour l'emplacement des usines. Opéra. Rés. 1966, 14, 361-368. [Référence croisée]
15. Uester, H. ; Dilaveroglu, S. Optimisation pour la conception et l'exploitation des réseaux de transport de gaz naturel. Appl. Énergie 2014, 133, 56-69. [Référence croisée]
16. Zhang, H. ; Liang, Y. ; Zhang, W. ; Wang, B. ; Yan, X. ; Liao, Q. Un modèle MILP unifié pour la structure topologique du puits de production réseau de pipelines de collecte. J. Pet. Sci. Ing. 2017, 152, 284-293. [Référence croisée]
17. Bosikov, II ; Klyuev, RV; Silaev, IV ; Stas, GV Évaluation complète de la difficulté et de la technologie de la ventilation formalisée procédés dans les mines de charbon. Soutenir. Dév. Mont Territ. 2023, 15, 516-527. [Référence croisée]
18. Nguyen, H. ; Bui, XN ; Choi, Y. ; Topal, E. Application de l'intelligence artificielle à la prévision de la stabilité des pentes dans les mines à ciel ouvert : une étude de cas avec un nouveau réseau neuronal à fonction de base radiale basé sur un algorithme compétitif impérialiste. Dans Applications de l'intelligence artificielle dans les mines, la géotechnique et la géo-ingénierie ; Elsevier : Amsterdam, Pays-Bas, 2024 ; pp. 97-111.
19. Rosa, VR ; Camponogara, E. ; Ferreira Filho, VJM Conception d'optimisation d'infrastructures sous-marines de champs pétrolifères avec placement de collecteurs et la disposition des pipelines. Calculer. Chimique. Ing. 2018, 108, 163-178. [Référence croisée]
20. Liu, Y. ; Chen, S. ; Guan, B. ; Xu, P. Optimisation de la configuration d'un système de collecte de pétrole et de gaz à grande échelle basée sur une stratégie d'optimisation combinée. Neuroinformatique 2019, 332, 159-183. [Référence croisée]
21. Hong, B. ; Li, X. ; Creuser.; Li, Y. ; Liu, X. ; Chen, S. ; Gong, J. Une méthode MILP intégrée pour rassembler les réseaux de pipelines en tenant compte caractéristiques hydrauliques. Chimique. Ing. Rés. Des. 2019, 152, 320-335. [Référence croisée]
22. Montagna, AF; Cafaro, DC ; Grossmann, IE ; Burch, D. ; Shao, Y. ; Wu, XH ; Furman, K. Conception d'un réseau de pipelines pour rassembler la production pétrolière et gazière non conventionnelle à l'aide d'une optimisation mathématique. Optimale. Ing. 2023, 24, 539-589. [Référence croisée]
23. Slowik, A. ; Kwasnicka, H. Algorithmes évolutifs et leurs applications aux problèmes d'ingénierie. Calcul neuronal. Appl. 2020, 32, 12363-12379. [Référence croisée]
24. Tohyama, H. ; Ida, K. ; Matsueda, J. Un algorithme génétique pour le problème de localisation des installations sans capacité. Electr. Commun. Jpn. 2011, 94, 47-54. [Référence croisée]
25. Beasley, JE OR-Library : Distribution des problèmes de test par courrier électronique. J. Opér. Rés. Soc. 1990, 41, 1069-1072. [Référence croisée]
26. Al-Sultan, KS ; Hussein, MF; Nizami, JS Un algorithme génétique pour le problème de couverture d'ensemble. J. Opér. Rés. Soc. 1996, 47, 702-709. [Référence croisée]
27. Pettie, S. ; Ramachandran, V. Un algorithme d'arbre couvrant minimum optimal. J.ACM 2002, 49, 16-34. [Référence croisée]
28. Zhou, J. ; Peng, J. ; Liang, G. ; Deng, T. Optimisation de l'aménagement du réseau de gazoducs arbre-arbre. J. Pet. Sci. Ing. 2019, 173, 666-680. [Référence croisée]

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Entropie d'erreur minimale rapide pour la régression linéaire

Qian Li [1], Xiao Liao [1], Wei Cui [1], Ying Wang [1], Hui Cao 2* et Qingshu Guan [2]

[1] State Grid Information & Telecommunication Group Co., Ltd., Pékin 102209, Chine ; qiangli_sgcc@163.com (QL); xiaoliao_sgcc@163.com (XL); weicui_sgcc@163.com (WC); yingwang_sgcc@163.com (YW)

[2] École de génie électrique, Université Xi'an Jiaotong, Xi'an 710049, Chine ; qsguan@stu.xjtu.edu.cn

* Correspondance : huicao@mail.xjtu.edu.cn

Résumé : Le critère d'entropie d'erreur minimale (MEE) trouve une grande utilité dans diverses applications , en particulier dans des contextes caractérisés par un bruit non gaussien. Cependant, ses exigences de calcul sont notables et sont principalement attribuables à l'opération de double sommation impliquée dans le calcul de la fonction de densité de probabilité (PDF) de l'erreur. Pour résoudre ce problème, notre étude introduit une nouvelle approche, appelée algorithme d'entropie d'erreur minimale rapide (FMEE), visant à atténuer la complexité informatique grâce à l'utilisation d'expansions polynomiales de la PDF d'erreur. Initialement, l' approximation PDF d'une variable aléatoire est dérivée via le développement de Gram – Charlier. Par la suite, nous procédons à la vérification et à la rationalisation de l'entropie de la variable aléatoire. Ensuite, l' entropie d'erreur inhérente au modèle de régression linéaire est délimitée et exprimée en fonction du vecteur de coefficient de régression. Enfin, en utilisant l'algorithme de descente de gradient, nous calculons le vecteur de coefficient de régression correspondant à l'entropie d'erreur minimale. Un examen théorique révèle que la complexité temporelle du FMEE s'élève à O(n), ce qui contraste fortement ^{2}) complexité associée à avec le O(n MEE. Expérimentalement, nos résultats soulignent les gains d'efficacité remarquables apportés par FMEE, avec une consommation de temps inférieure à 1‰ de celle observée. Avec MEE, il est encourageant de constater que ce gain d'efficacité est réalisé sans compromettre la précision, comme en témoignent les différences négligeables observées entre les précisions du FMEE et du MEE. De plus, des expériences de régression complètes sur des ensembles de données électriques réelles dans le nord-ouest de la Chine démontrent que notre FMEE surpasse la référence . méthodes avec une marge nette.

Mots-clés : entropie d'erreur minimale ; régression linéaire; Extension Gram-Charlier ; complexité informatique

check for updates

Citation : Li, Q. ; Liao, X. ; Cui, W. ; Wang, Y. ; Cao, H. ; Guan, Q. Entropie d'erreur minimale rapide pour la régression linéaire. Algorithmes 2024, 17, 341. https://doi.org/10.3390/a17080341

Rédacteur académique : Frank Werner

Reçu : 23 juin 2024
Révisé : 1er août 2024
Accepté : 3 août 2024
Publié : 6 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.

Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

L'entropie sert de mesure essentielle pour quantifier l'incertitude inhérente à un système, trouvant une application omniprésente dans une multitude de domaines, notamment la réduction de dimension [1,2], l'estimation des paramètres [3,4], l'identification [5,6], le contrôle par rétroaction [7,8] et détection anormale [9–11]. Le critère d'entropie d'erreur minimale (MEE) se distingue comme un cadre notable visant à minimiser l'entropie associée aux erreurs d'estimation, atténuant ainsi l'incertitude au sein du modèle d'estimation. Particulièrement pertinent pour les dilemmes d'estimation par régression linéaire, MEE s'écarte des méthodologies conventionnelles des moindres carrés non seulement en considérant la variance des erreurs de prédiction, mais également en incorporant des cumulants d'ordre supérieur, ce qui le rend apte à gérer les distributions de bruit non gaussiennes [12,13] . Compte tenu de la prévalence du bruit non gaussien dans des scénarios réels, l'efficacité du MEE a été dûment validée dans diverses applications, notamment le filtrage adaptatif [14-16], la reconnaissance faciale [17,18], l'identification de systèmes clairsemés [19-21].], les systèmes de contrôle stochastiques [22,23] et la communication par lum

De plus, les propriétés de convergence du MEE ont été méticuleusement examinées dans la littérature antérieure [25-27]. Des analyses approfondies ont porté sur la dynamique de convergence des méthodologies MEE à virgule fixe [28], tandis que des recherches sur l'interaction entre

Machine Translated by Google

Le MEE et l'erreur quadratique moyenne minimale (MMSE) [29], ainsi que sa relation avec la correntropie maximale [30], ont été largement explorés. Les extensions notables du cadre MEE conventionnel incluent la proposition de variantes du noyau MEE [31], des explications sur les stratégies de régularisation pour les implémentations MEE [32] et le développement d' adaptations semi-supervisées et distribuées [33]. De plus, dans le domaine de la modélisation des erreurs dans les variables (EIV), des analyses de robustesse des méthodologies MEE ont été exposées [34], parallèlement à la proposition de nouvelles méthodologies telles que l' approche d'entropie d'erreur totale minimale [35].

Les exigences de calcul associées au critère MEE proviennent de son époque . Au cœur plexité, qui évolue quadratiquement comme ^{2}de la charge de calcul de MEE se trouve le O(n nécessité d'une opération de double sommation lors du calcul du gradient relatif à la fonction de densité de probabilité d'erreur (PDF), en particulier lors de l'utilisation du fenêtrage de Parzen [36] et de certains types de fonctions du noyau. De plus , la sélection d'une fonction de noyau inappropriée ou de ses paramètres peut avoir un impact négatif sur la précision du PDF résultant. Pour résoudre le problème, des efforts ont été déployés pour optimiser l' approche MEE de descente de gradient, comme en témoigne la normalisation de la taille du pas dans. [37], exploitant la puissance de l'entropie d'entrée. De plus, des méthodologies innovantes telles que l'utilisation d'un opérateur de quantification, comme proposé dans [38], ont joué un rôle déterminant dans l'atténuation de la complexité informatique en mappant l'erreur sur une série de. Avec des mots de code à valeur réelle , cette approche transforme efficacement l'opération de double sommation en une forme singulière, réduisant ainsi la charge de calcu

Cet article présente une nouvelle approche, l'algorithme d'estimation rapide de l'entropie d'erreur minimale (FMEE), spécialement conçu pour les tâches de régression linéaire, exploitant la technique d'expansion polynomiale appliquée à la fonction de densité de probabilité d'erreur (PDF). Contrairement à la méthodologie traditionnelle d'entropie d'erreur minimale (MEE) employant le fenêtrage de Parzen pour la dérivation de PDF d'erreur, FMEE adopte la méthode d'expansion de Gram – Charlier pour approximer la PDF d'erreur, ce qui entraîne une réduction notable de la complexité temporelle à partir des paramètres O (n) . L'algorithme comporte plusieurs étapes clés : d'une part, la dérivation de l' approximation PDF pour une variable aléatoire via le développement de Gram – Charlier, suivie de la simplification de l'entropie de la variable aléatoire facilitée par l'orthogonalité des polynômes d'Hermite. le modèle de régression linéaire est déterminé, avec l'entropie d'erreur exprimée en fonction du vecteur de coefficient de régression. Finalement, la technique de descente de gradient est utilisée pour déduire le vecteur de coefficient de régression correspondant à l'entropie d'erreur minimale. La validation expérimentale souligne l'efficacité du FMEE, révélant une réduction remarquable de la consommation de temps, avec un écart de moins de 1‰ par rapport à celui de l'approche MEE pour des problématiques identiques . Il est important de noter que des écarts minimes sont observés entre les niveaux de précision des méthodologies FMEE et MEE.

^{2}) à O(n). Notamment, FMEE évite le besoin de fonctions du noyau ou de leurs fonctions associées.

Le reste de l'article est organisé comme suit : la section 2 présente les algorithmes associés , le FMEE est proposé en détail dans la section 3, la section 4 contient les expériences et les conclusions sont dans la section 5.

2. Les algorithmes associés 2.1. MEE

MEE minimise les informations contenues dans l'erreur et maximise les informations capturées par le modèle estimé. Habituellement, l'entropie Renyi de second ordre [39] est utilisée pour mesurer l'entropie contenue dans l'erreur du modèle.

$$H(E) = - \log E_p(pE) = - \log p \int_E^2 (e)de, \quad (1)$$

Machine Translated by Google

où E est la variable aléatoire d'erreur de MEE, et H(E) est l'entropie de E et pE(e) donne la PDF de E. Dans cet article, Ep(·) fait don de l'espérance afin d'éviter tout malentendu. Ensuite, la PDF de E peut être estimée par fenêtrage de Parzen :

$$\hat{p}E(e) = \frac{1}{n}\sum_{je=1}^{n} Kh\,(e - ei), \tag{2}$$

où n est le numéro des données observées, K est la fonction noyau et h est la bande passante. Habituellement, le noyau gaussien est choisi, et $Kh(z) = \frac{1}{\sqrt{2\pi h}} \exp(-z\,2h\,\frac{2}{2})$. Ensuite, l'empirique information d'erreur $\hat{H}(f)$ peut être calculée comme suit :

$$\hat{H}(f) = -\log \frac{1}{2n}\sum_{je=1}^{n}\sum_{j=1}^{n} Kh\,(ei - ej). \tag{3}$$

Comme la fonction logarithmique est monotone croissante, elle peut être supprimée sans aucune influence sur le processus de minimisation. Ensuite, les informations d'erreur empirique transformées peuvent être obtenues sous la forme de l'équation (4) :

$$R(e) = -\frac{1}{\sum_{je=1}^{n}\sum_{j=1}^{n} Kh\,(ei - ej).\,2n} \tag{4}$$

Pour la régression linéaire, y = wTx + e, l'objectif du modèle est d'estimer w à partir des échantillons limités. Comme ei = yi − wTxi , les informations d'erreur empirique correspondantes sont calculées comme l'équation (5) :

$$R(w) = -\frac{1}{2n}\sum_{je=1}^{n}\sum_{j=1}^{n} Kh\,(yi - wT\,xi) - (yj - wT\,xj)\,. \tag{5}$$

Ensuite, l'estimateur MEE $\hat{w}$ peut être obtenu en minimisant R(w) avec descente de gradient. La complexité temporelle de MEE est $O(n^2)$ en raison de l'opération de double sommation, comme le montre équation (5). Comme la consommation de temps de MEE est quadratique par rapport au nombre de données observées, la durée d'exécution de MEE augmenterait considérablement avec l'augmentation de la quantité de données observées.

2.2. Les polynômes d'Hermite

Les polynômes d'Hermite [40] sont une séquence polynomiale orthogonale classique. Ils sont définis par les dérivées de la PDF gaussienne standardisée φ(ξ) comme l'équation (6) :

$$\frac{\partial\, i\varphi(\xi)}{\partial \xi i} \equiv (-1)\,iHi(\xi)\varphi(\xi), \tag{6}$$

où H(i) donne le polynôme Hermite d'ordre ie.

Les polynômes Hermite forment un système orthogonal, comme dans l'équation (7) :

$$\varphi(\xi)Hi(\xi)Hj(\xi)d\xi = \begin{matrix} \text{si } je = j \\ je!,\ 0, \quad \text{si } je \neq j \end{matrix}, \tag{7}$$

Cet article utilise les cinq premiers polynômes d'Hermite, présentés dans l'équation (8) :

$$\begin{aligned} H0(x) &= 1, \\ H1(x) &= x, \\ H2(x) &= x^2 - 1, \\ H3(x) &= x^{3-3x}, \\ H4(x) &= x^{4\,2} - 6x + 3. \end{aligned} \tag{8}$$

Machine Translated by Google

2.3. L'extension Gram-Charlier

Le développement de Gram – Charlier obtient l'approximation de la PDF d'une variable aléatoire, et il n'y a pas de perte de précision significative avec ce développement, ce qui est démontré dans [41]. En supposant que la variable aléatoire est proche de la distribution normale avec la même moyenne et la même variance qu'elle-même, le développement de Gram – Charlier de la PDF de x peut être exprimé comme suit :

$$p_x(\xi) \approx \hat{p}_x(\xi) = \exp\left(-\frac{(\xi-\mu)^2}{2\sigma^2}\right)\left(1 + \frac{\kappa_3(x)}{3!\sigma^3}H_3\left(\frac{\xi-\mu}{\sigma}\right) + \frac{\kappa_4(x)}{4!\sigma^4}H_4\left(\frac{\xi-\mu}{\sigma}\right)\right), \quad (9)$$

où μ et σ sont la moyenne et l'écart type de x. ξ est la variable aléatoire gaussienne dont la moyenne et la variable sont les mêmes que celles de x. κ3(x) et κ4(x) sont les cumulants de troisième et quatrième ordre de x, qui peuvent également être appelés asymétrie et aplatissement. H3 et H4 sont les polynômes Hermite du troisième et du quatrième ordre.

3. Méthodologie

Cette section dérive la relation entre l'entropie différentielle et le développement de Gram – Charlier de la PDF de l'erreur pour la régression linéaire. De plus, l'hypothèse selon laquelle la PDF de l'erreur de régression est proche de la distribution gaussienne φ(ξ) qui a la même moyenne et la même variance que l'erreur est mise en œuvre.

L'entropie différentielle [42] d'une variable aléatoire X est présentée dans l'équation (10) :

$$H(x) = -\int p(x)\log p(x)dx. \quad (10)$$

Remplacez l'équation (9) par l'équation (10) ; le résultat peut être dérivé de l'équation (11) :

$$\begin{aligned} H(x) &\approx -\int \hat{p}_x(\xi)\log \hat{p}_x(\xi)d\xi \\ &= -\int \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(\xi-\mu)^2}{2\sigma^2}\right)\left(1 + \frac{\kappa_3(x)}{3!\sigma^3}H_3\left(\frac{\xi-\mu}{\sigma}\right) + \frac{\kappa_4(x)}{4!\sigma^4}H_4\left(\frac{\xi-\mu}{\sigma}\right)\right) * \\ &\quad \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(\xi-\mu)^2}{2\sigma^2}\right)\left(1 + \frac{\kappa_3(x)}{3!\sigma^3}H_3\left(\frac{\xi-\mu}{\sigma}\right) + \frac{\kappa_4(x)}{4!\sigma^4}H_4\left(\frac{\xi-\mu}{\sigma}\right)\right)\right) * \\ &\quad \left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(\xi-\mu)^2}{2\sigma^2}\right)\left(1 + \frac{\kappa_3(x)}{3!\sigma^3}H_3\left(\frac{\xi-\mu}{\sigma}\right) + \frac{\kappa_4(x)}{4!\sigma^4}H_4\left(\frac{\xi-\mu}{\sigma}\right)\right)\right)d\xi. \end{aligned} \quad (11)$$

Cette intégrale est assez difficile à calculer. Cependant, notez que si la PDF de x est proche de la densité normale comme supposé, on peut en déduire que κ3(x) et κ4(x) sont très petits. Ensuite, l'approximation de l'équation (12) peut être utilisée :

$$\log(1+\varepsilon) \approx \varepsilon - \frac{\varepsilon^2}{2}, \quad (12)$$

où ε désigne l'infinitésimal.

Ensuite, l'équation (11) peut être transformée en équation (13) :

Machine Translated by Google

$$H(x) = -\varphi(\xi) \quad \kappa 3(x)\ 1 \\ 3!\sigma^3 + H3 \quad \frac{\xi-\mu}{\sigma} + \frac{\kappa 4(x)}{4!\,\sigma^4} H4 \quad \frac{\xi-\mu}{\sigma} \quad * \\ \log \varphi(\xi) + H3\ \frac{\kappa 3(x)}{3!\sigma^3} \quad \frac{\xi-\mu}{\sigma} + \\ \frac{\kappa 4(x)}{4!\,\sigma^4} H4 \quad \frac{\xi-\mu}{\sigma} - \frac{1}{2} \quad \frac{\kappa 3(x)}{3!\,\sigma^3} H3 \quad \frac{\xi-\mu}{\sigma} + \\ \frac{\kappa 4(x)}{4!\,\sigma^4} H4 \quad \frac{\xi-\mu}{\sigma} \quad {}^2 d\xi. \tag{13}$$

Comme Hi forme un système orthogonal, l'équation (13) peut être simplifiée comme l'équation (14),

$$H(x) = \quad + \log 2\pi e\sigma\ 2^{\ 2} - \frac{^2\kappa 3(x)}{2 \quad 3!\sigma\ 6} - \frac{^2\kappa 4(x)}{2 \quad 4!\sigma\ 8}. \tag{14}$$

Les détails du processus de simplification de l'équation (14) sont illustrés comme suit. Cette partie consiste à dériver l'équation (14). Il faut d'abord montrer que

$$2H3(\ \varphi(\xi)(\ )d\xi = 0 \quad \frac{\xi-\mu}{\sigma} \tag{15}$$

et

$$2H4(\ \varphi(\xi)(\ )d\xi = 0 \quad \frac{\xi-\mu}{\sigma} \tag{16}$$

Pour l'équation (15), supposons que t = (ξ − μ)/σ,

$$2H3(\ \varphi(\xi)(\ )d\xi\ ) \quad \frac{\xi-\mu}{\sigma} \\ = \quad \frac{1}{} \exp - 2\sigma \frac{(\xi-\mu)^{\ 2}}{2\sqrt{2\pi\sigma}} \quad (\ )\frac{\xi-\mu}{\sigma}^{\ 3} - 3(\ )\frac{\xi-\mu}{\sigma} d\xi \\ = \quad \frac{1}{\sqrt{2\pi}} \ 2\ t \exp(-\ )(t\ 2 \quad {}^3 - 3t)dt \tag{17}$$

Comme $\exp(-\frac{t^2}{2})$ est une fonction paire et $t^{\ 5}\ 3 - 3t$ est une fonction étrange, alors

$$\frac{1}{2\ t\sqrt{2\pi}} \quad \exp(- \quad \frac{^2}{2})(t^{\ 3} - 3t)dt = 0 \tag{18}$$

et l'équation (15) est vraie.

Pour l'équation (16), supposons que t = (ξ − μ)/σ,

$$2H4(\ \varphi(\xi)(\ )d\xi\ ) \quad \frac{\xi-\mu}{\sigma} \\ = \quad \frac{1}{} \exp - \sqrt{\ } \quad \frac{(\xi-\mu)^{\ 2}}{2\pi\sigma\ 2\sigma\ 2\ \xi-\mu\ (\ )} \quad * \\ \frac{}{\sigma}^{\ 6} - 6(\ )\frac{\xi-\mu}{\sigma}^{\ 4} + 3(\ )\frac{\xi-\mu}{\sigma}^{\ 2} d\xi \\ = \quad \frac{1}{2\pi} \exp(-\sqrt{\ }\frac{2t}{2})(t^{\ 6\ 4\ 2} - 6t + 3t \quad)dt \\ =3 \quad \frac{1}{2\pi} \exp(-\sqrt{\ }\frac{^2}{2})t\ 2dt - 6 \quad \frac{1}{2\pi} \exp(-\sqrt{\ }\frac{^2}{2})t\ 4dt \\ + \quad \frac{1}{2\pi} \exp(-\sqrt{\ }\frac{^2}{2})t\ 6dt \tag{19}$$

Machine Translated by Google

Le premier terme de l'équation (19) peut être calculé comme l'équation (20),

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)t\,2dt \\
&= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)d\left(-t\right) \\
&= \frac{1}{\sqrt{2\pi}}\left(-t\right)\exp\left(-\frac{t^2}{2}\right)\Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}}\int\exp\left(-\frac{t^2}{2}\right)dt \\
&= 1
\end{aligned} \tag{20}$$

Le deuxième terme de l'équation (19) peut être calculé comme l'équation (21),

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)t\,4dt \\
&= \frac{1}{\sqrt{2\pi}}\left(-t^3\right)d\exp\left(-\frac{t^2}{2}\right) \\
&= \frac{1}{\sqrt{2\pi}}\left(-t^3\right)\exp\left(-\frac{t^2}{2}\right)\Big|_{-\infty}^{+\infty} + 3\int\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)t\,2dt \\
&= 3
\end{aligned} \tag{21}$$

Le dernier terme de l'équation (19) peut être calculé comme l'équation (22),

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}^2}\exp\left(-\sqrt{\frac{t^2}{2}}\right)t\,6dt \\
&= \frac{1}{\sqrt{2\pi}}\left(-t^5\right)d\exp\left(-\frac{t^2}{2}\right) \\
&= \frac{1}{\sqrt{2\pi}}\left(-t^5\right)\exp\left(-\frac{t^2}{2}\right)\Big|_{-\infty}^{+\infty} + 5\int\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)t\,4dt \\
&= 15
\end{aligned} \tag{22}$$

D'après les équations (20) à (22),

$$2H4\left(\varphi\left(\frac{\xi-\mu}{\sigma}\right)\left(\frac{\xi-\mu}{\sigma}\right)d\xi\right) = 15 - 6\frac{\xi-\mu}{\sigma} + 3 = 0 \tag{23}$$

et l'équation (16) est vraie.

Machine Translated by Google

Alors, pour l'équation (13),

$$H(x) = -\varphi(\xi)\log\varphi(\xi)d\xi - \varphi(\xi)\frac{\kappa 3(x)}{\sigma^3 4!}\left(\frac{\xi-\mu}{\sigma}\right)^4 + H3\left(\frac{\xi-\mu}{\sigma}\right) + H4\left(\frac{\xi-\mu}{\sigma}\right)d\xi + \frac{3!}{\sigma} - \frac{\kappa 3(x)\,\kappa 4(x)}{H3\left(\frac{\xi-\mu}{\sigma}\right) + H4\left(\frac{\xi-\mu}{\sigma}\right)3!\sigma^3 4!\sigma^4 2}\left(\frac{\xi-\mu}{\sigma}\right)^2 d\xi - \varphi(\xi)\frac{\kappa 3(x)}{H3\left(\frac{\xi-\mu}{\sigma}\right)+3!\sigma^3}\left(\frac{\xi-\mu}{\sigma}\right)^3\log\varphi(\xi)\frac{\kappa 4(x)}{4!\sigma^4}\left(\frac{\xi-\mu}{\sigma}\right)^4 H4\left(\frac{\xi-\mu}{\sigma}\right)d\xi - \frac{\kappa 3(x)}{H3\left(\frac{\xi-\mu}{\sigma}\right)+3!\sigma^3}\left(\frac{\xi-\mu}{\sigma}\right)^3 - \varphi(\xi)\frac{\kappa 4(x)}{4!\sigma^4}\left(\frac{\xi-\mu}{\sigma}\right)^4 H4\left(\frac{\xi-\mu}{\sigma}\right)^2 d\xi + \frac{1}{2}\frac{\kappa 3(x)\,\kappa 4(x)}{H3\left(\frac{\xi-\mu}{\sigma}\right) + H4\left(\frac{\xi-\mu}{\sigma}\right)3!\sigma^3 4!\sigma^4}\left(\frac{\xi-\mu}{\sigma}\right)^3 d\xi \quad (24)$$

Notez que H0 = 1, alors le deuxième terme de l'équation (24) est 0. Comme on suppose que x est proche d'une variable aléatoire normale, κ3(x) et κ4(x) sont très petits. Ensuite, le dernier terme de l'équation (24) peut être omis comme 0, car il implique des monômes de troisième ordre, qui sont infiniment plus petits que les termes contenant uniquement des monômes de deuxième ordre. Ensuite, pour l'équation (24), prenez en considération les équations (15) et (16) , et notez à nouveau que H0 = 1,

$$H(x) = H(\xi) - \frac{1}{2}\left[H3\left(\frac{\xi-\mu}{\sigma}\right)\frac{\kappa 3(x)}{3!\sigma^3}(\xi-\mu)\right]^2\varphi(\xi) + \frac{\kappa 4(x)}{\sigma^4}\left[H4\left(\frac{\xi-\mu}{\sigma}\right)\frac{\kappa 3(x)}{4!}(\xi-\mu)\right]^2 + 2\frac{\xi-\mu}{\sigma^4}\left(\frac{\xi-\mu}{\sigma}\right)H3\left(\frac{\xi-\mu}{\sigma}\right)H4\left(\frac{\xi-\mu}{\sigma}\right)d\xi\frac{3!\sigma^3 4!}{\sigma} - \varphi(\xi)\log + \log\exp\left(-\frac{1}{\sqrt{2\pi\sigma^2\kappa 3(x)}}\frac{(\xi-\mu)^3\kappa 4(x)(\xi-\mu)}{H3\left(\frac{\xi-\mu}{\sigma}\right)+H4\left(\frac{\xi-\mu}{\sigma}\right)d\xi\,3!\sigma^3 4!\sigma^4}\frac{\xi^2}{\sigma}\right) \quad (25)$$

$$= H(\xi) - \frac{^2\kappa 3(x)}{18\sigma^6} - \frac{^2\kappa 3(x)}{32\sigma^8} - \log\frac{1}{\sqrt{2\pi\sigma}}\varphi(\xi)\frac{8\,\kappa 3(x)}{3!\sigma^3}\left(\frac{\xi-\mu}{\sigma}\right)^3 + H3\left(\frac{\xi-\mu}{\sigma}\right)\frac{\kappa 4(x)}{4!\sigma^4}\left(\frac{\xi-\mu}{\sigma}\right)^4 + H4\left(\frac{\xi-\mu}{\sigma}\right)d\xi + \frac{1}{2\,l\,en\,2}\varphi(\xi)\xi^{2*}\frac{\kappa 3(x)\,\kappa 4(x)}{3!\sigma^3 4!\sigma^4}\left(\frac{\xi-\mu}{\sigma}\right)^3 + H3\left(\frac{\xi-\mu}{\sigma}\right)H4\left(\frac{\xi-\mu}{\sigma}\right)d\xi\frac{3!\sigma^3 4!\sigma^4}{\sigma} = \frac{1}{2}\log 2\pi e\sigma^2 - \frac{^2\kappa 3(x)}{18\sigma^6} - \frac{^2\kappa 4(x)}{32\sigma^8}$$

Ainsi, l'équation (14) est vraie.

Pour la régression linéaire, ei = yi − w Txi , l'écart type de l'erreur ew est dans l'équation (26) :

$$^2\sigma(ew) = Ep(E\,^2w) - [Ep(Ew)]^2 = Ep(E\,^2w) - (\bar{y} - w^T\bar{x})^2, \quad (26)$$

Machine Translated by Google

où $\bar{y}$ est la moyenne de y et $\bar{x}$ donne la moyenne de x.

Pour simplifier le calcul de κ4(ew), soit ew une variable aléatoire de moyenne 0 :

$$ewi = yi - wT xi - \bar{y} + wT \bar{x}. \quad (27)$$

Notez que la transformation dans l'équation (27) ne modifie pas l'écart type et l'entropie de ew comme $\sigma^2(e + c) = \sigma^2(e)$ et H(e + c) = H(e) lorsque c est un nombre constant. Dans le reste de l'article, pour rester clair, $\sigma^2(\cdot)$ est remplacé par var(·).

Ensuite, pour la régression linéaire, son entropie de l'erreur peut être exprimée par l'équation (28) :

$$H(ew) = \frac{1}{2}\log 2\pi e + \log \mathrm{var}(ew) - \frac{2\kappa 3}{3!\mathrm{var}3(ew)} - \frac{2\kappa 4(ew)}{4!\mathrm{var}4(ew)}, \quad (28)$$

où

$$\kappa 3(ew) = Ep[(y - wT x - \bar{y} + wT \bar{x})^3], \quad (29)$$

$$\kappa 4(ew) = Ep[(y - wT x - \bar{y} + wT \bar{x})^4] - 3. \quad (30)$$

Pour minimiser H(ew), la dérivée de H(ew) par rapport à w est calculée en Équation (31) :

$$\frac{\partial H(ew)}{\partial w} = \frac{1}{2\mathrm{var}(ew)\ln 2}\frac{\partial \mathrm{var}(ew)}{\partial w} - \frac{1}{2}\left[\frac{1}{3!\mathrm{var}6(ew)}\left(2\mathrm{var}3(ew)\kappa 3(ew)\frac{\partial \kappa 3(ew)}{\partial w} - 3\kappa 3(ew)\mathrm{var}2(ew)\frac{\partial \mathrm{var}(ew)}{\partial w}\right) - \frac{1}{4!\mathrm{var}8(ew)}\left(2\mathrm{var}4(ew)\kappa 4(ew)2\frac{\partial \kappa 4(ew)}{\partial w} - 4\kappa 4(ew)\mathrm{var}3(ew)\frac{\partial \mathrm{var}(ew)}{\partial w}\right)\right], \quad (31)$$

où

$$\frac{\partial \mathrm{var}(ew)}{\partial w} = -2Ep[(y - wT x)x] + 2(\bar{y} - wT \bar{x})\bar{x}, \quad (32)$$

$$\frac{\partial \kappa 3(ew)}{\partial w} = 3Ep[(y - wT x - \bar{y} + wT \bar{x})^2(-x + \bar{x})], \quad (33)$$

$$\frac{\partial \kappa 4(ew)}{\partial w} = 4Ep[(y - wT x - \bar{y} + wT \bar{x})^3(-x + \bar{x})]. \quad (34)$$

Ensuite, le w optimal pour le minimum de H(ew) peut être obtenu par le schéma d'itération dans l'équation (35) avec descente de gradient :

$$\hat{w}_{k+1} = \hat{w}_k - \alpha \frac{\partial H(ewk)}{\partial \hat{w}k}, \quad (35)$$

où α est la taille du pas déterminée par la condition d'Armijo [43].

Pour FMEE à chaque itération, var(ewk), κ3(ewk), κ4(ewk), $\frac{\partial H(ewk)}{\partial wk}$, $\frac{\partial \mathrm{var}(ewk)}{\partial wk}$, $\frac{\partial \kappa 3(ewk)}{\partial wk}$, $\frac{\partial \kappa 4(ewk)}{\partial wk}$ et $\frac{\partial var(ewk)}{\partial wk}$ sont nécessaire. D'après les équations (26) et (29) – (34), la complexité temporelle de calculer var(ewk), κ3(ewk), κ4(ewk) et $\frac{\partial var(ewk)}{\partial wk}$, la complexité temporelle de $\frac{\partial H(ewk)}{\partial wk}$, $\frac{\partial \kappa 3(ewk)}{\partial wk}$, $\frac{\partial \kappa 4(ewk)}{\partial wk}$ est O(n), respectivement. Ensuite, le complexité de calcul du FMEE est O(n). Comme la complexité temporelle de MEE est $O(n^2)$ pour l'opération dû de double sommation, FMEE s'exécute plus rapidement que MEE.

Machine Translated by Google

4. Expériences

Dans cette section, une gamme complète d'expériences est entreprise, englobant à la fois des simulations numériques et des scénarios du monde réel. Les simulations numériques servent à valider l'efficacité de l'approche FMEE et examiner l'efficacité temporelle des deux Méthodologies FMEE et MEE. Parallèlement, des expérimentations pratiques sont menées dans le but à prévoir les pannes de courant dans une ville située au nord-ouest de la Chine.

Concernant les simulations numériques, nous considérons un scénario où x R10, avec le modèle défini comme y = w ^{T} x + e, où w = [1, −1, 1, −1, 1 , −1, 1, −1, 1, −1] T et x suit une distribution normale N (0, I10). Deux types distincts de bruit sont examinés dans nos expériences. Premièrement, le bruit gaussien e N (0, 1) est utilisé, tandis que le deuxième type comprend un bruit gaussien généralisé caractérisé par une fonction de densité de probabilité f(e) exp(−α|e|) 0,3. Cette distribution à queue lourde, utilisée ici, s'aligne sur les précédentes ouvrages, tels que les références [27,35]. Bien que le FMEE soit dérivé sous l'hypothèse de bruit quasi gaussien, il est prévu que le FMEE reste efficace en présence de bruits non gaussiens à queue lourde. Ici, α désigne une constante ajustée pour maintenir une variance de 1 pour e. Pour MEE, la taille du pas est définie sur √ 0,005π et le paramètre d'échelle pour le modèle gaussien le noyau est 10, reflétant les paramètres adoptés dans la référence [27]. Au fil des expériences, la taille des échantillons varie de 100 à 500, avec chaque combinaison de taille d'échantillon et de type de bruit répété 100 fois. Les résultats expérimentaux qui en résultent sont détaillés ci-dessous.

Le tableau 1 présente une analyse comparative de la consommation de temps entre MEE et FMEE en présence de bruit gaussien. Le FMEE démontre une accélération notable efficacité de calcul par rapport à MEE. Plus précisément, alors que le temps de calcul de MEE présente une croissance quadratique par rapport à la taille de l'échantillon, ce qui représente la charge de calcul du FMEE évolue linéairement avec le nombre d'échantillons. FMEE atteint notamment une vitesse remarquable, traiter 500 objets en 0,1 s environ, le positionnant ainsi comme un outil très prometteur candidat pour les candidatures urgentes.

Tableau 1. Le temps d'exécution de MEE et FMEE pour le bruit gaussien.

Taille de l'échantillon	Durée moyenne de fonctionnement		Durée de fonctionnement optimale	
	MEE	FMEE	MEE	FMEE
100	103,89	0,0401	77,77	0,0156
200	384.31	0,0580	310.05	0,0312
300	801.52	0,0769	654,21	0,0468
400	1430.61	0,1000	1254,12	0,0468
500	2306.24	0,1104	2072,50	0,0624

Le tableau 2 compare l'erreur quadratique moyenne entre MEE et FMEE dans le contexte de Bruit gaussien. Dans l'ensemble, MEE présente des performances supérieures à FMEE en termes de de MSE. Il convient particulièrement de noter l'avantage substantiel du MEE par rapport au FMEE, en particulier évident lorsque la taille de l'échantillon est de 100. Cependant, à mesure que la taille de l'échantillon augmente, l'écart en MSE entre les deux algorithmes diminue. Au moment où la taille de l'échantillon atteint 500, la différence se réduit à seulement 2 %. Cette observation suggère que même si le FMEE peut être à la traîne légèrement en retard sur MEE, en particulier dans les ensembles de données à plus grande échelle, une telle disparité reste acceptable dans des applications pratiques, étant donné le temps de calcul considérablement réduit du FMEE par rapport à MEE.

Machine Translated by Google

Tableau 2. Erreur quadratique moyenne de MEE et FMEE pour le bruit gaussien.

Taille de l'échantillon	MSE moyen		MSE optimale	
	MEE	FMEE	MEE	FMEE
100	0,9108	1.4027	0,5697	0,6542
200	0,9311	1,0528	0,6295	0,7258
300	0,9573	1,0190	0,7600	0,7462
400	0,9696	0,9769	0,8024	0,7588
500	0,9604	0,9938	0,8024	0,8302

Le tableau 3 analyse le nombre d'itérations requis pour la convergence entre MEE et FMEE dans le contexte du bruit gaussien. Le FMEE réalise notamment une convergence avec des moins d'itérations par rapport à MEE. De plus, on observe que le nombre d'itérations de Le MEE diminue progressivement à mesure que la taille de l'échantillon augmente. À l'inverse, le nombre d'itérations de Le FMEE reste relativement constant, quelles que soient les variations de la taille de l'échantillon.

Tableau 3. Le nombre d'itérations de MEE et FMEE pour le bruit gaussien.

Taille de l'échantillon	Nombre d'itérations moyen		Nombre d'itérations optimal	
	MEE	FMEE	MEE	FMEE
100	797.01	38.21	592	6
200	697.09	38.29	592	7
300	656.61	36,88	542	7
400	650.03	40,67	571	9
500	637.03	36.17	586	7

Le tableau 4 fournit une évaluation comparative de la consommation de temps entre MEE et FMEE dans le contexte de bruit non gaussien. Dans les cinq groupes expérimentaux, FMEE démontre systématiquement une réduction remarquable de la consommation de temps, de manière constante représentant moins de 1‰ du temps passé par le MEE. De plus, conformément à prédictions théoriques, le coût de calcul du MEE présente une croissance quadratique relative à la taille de l'échantillon, alors que les frais généraux de temps du FMEE évoluent linéairement avec la taille de l'échantillon.

Tableau 4. Temps d'exécution de MEE et FMEE pour le bruit non gaussien.

Taille de l'échantillon	Durée moyenne de fonctionnement		Durée de fonctionnement optimale	
	MEE	FMEE	MEE	FMEE
100	107,96	0,0413	85,24	0,0156
200	348,66	0,0635	270,26	0,0312
300	800,50	0,0819	727,45	0,0468
400	1431,34	0,1027	1274,87	0,0624
500	2179,33	0,1178	1952,27	0,0624

Le tableau 5 effectue une évaluation comparative de l'erreur quadratique moyenne (MSE) entre MEE et FMEE dans le contexte de bruit non gaussien. D'une manière générale, la disparité entre les valeurs MSE de FMEE et MEE sont relativement mineures. Notamment, FMEE surpasse MEE lorsque la taille de l'échantillon est de 100 ou 400, tandis que MEE présente des performances supérieures pour l'échantillon tailles de 200, 300 et 500. De plus, FMEE atteint le MSE optimal sur tous les échantillons tailles, sauf pour n = 100. De plus, on observe que les valeurs MSE pour les le bruit, comme illustré dans le tableau 5, couvre une plage plus large que celle du bruit gaussien, indiquant une plus grande variabilité dans les résultats pour le MEE et le FMEE.

Machine Translated by Google

Tableau 5. Erreur quadratique moyenne de MEE et FMEE pour le bruit non gaussien.

Taille de l'échantillon	MSE moyen		MSE optimale	
	MEE	FMEE	MEE	FMEE
100	1,0844	0,9604	0,1584	0,1630
200	0,8485	1,1554	0,2654	0,2607
300	0,9110	1.2314	0,3140	0,2659
400	1.1037	1,0550	0,3672	0,2891
500	0,9536	1,0491	0,4995	0,4668

Le tableau 6 compare le nombre d'itérations nécessaires à la convergence entre MEE et FMEE dans le contexte de bruit non gaussien. FMEE démontre systématiquement plus rapidement convergence par rapport au MEE. De plus, on observe que le nombre d'itérations de MEE diminue à mesure que la taille de l'échantillon augmente. En revanche, le nombre d'itérations de FMEE reste relativement stable malgré les variations de la taille de l'échantillon.

Tableau 6. Le nombre d'itérations de MEE et FMEE pour le bruit non gaussien.

Taille de l'échantillon	Nombre d'itérations moyen		Nombre d'itérations optimal	
	MEE	FMEE	MEE	FMEE
100	782,88	34.33	618	8
200	689,67	35,72	563	9
300	644,95	36,72	584	10
400	648,70	37.55	580	12
500	633.25	35h35	571	12

Sur la base des résultats présentés dans les tableaux 1 à 6, il est évident que le FMEE est systématiquement donne des résultats comparables à ceux du MEE, mais avec une consommation de temps considérablement réduite et l'itération compte. FMEE se distingue notamment par son efficacité remarquable, étant environ 1 000 fois plus rapide que MEE et parvenant à une convergence dans un délai d'environ 0,1 s pour 500 instances. Compte tenu de sa complexité temporelle linéaire de O(n), FMEE possède des propriétés considérables promesse de déploiement dans des scénarios en temps réel caractérisés par un bruit non gaussien.

Par la suite, nous nous lançons dans des expérimentations pratiques utilisant les données caractéristiques des transformateurs pour prévoir les pannes des réseaux de distribution. Les transformateurs sont classés en deux groupes en fonction de leur situation géographique au sein de niveaux distincts d'activité urbaine dans une ville spécifique du nord-ouest de la Chine. Dans un premier temps, notre attention se porte vers un sous-ensemble de 1 265 transformateurs situés dans une zone plus petite et moins densément peuplée du ville susmentionnée. Nous entreprenons un partitionnement aléatoire de cet ensemble de données en deux segments : un ensemble d'entraînement comprenant 80 % des données, totalisant 1012 instances, et un ensemble de vérification englobant les 20 % restants, soit 253 instances. Pour être plus précis, l' ensemble de données sur les transformateurs comprend diverses variables caractéristiques, impliquant les charges lourdes standardisées. durée de surcharge, rapport de charge actif maximum, rapport de charge actif moyen, triphasé moyen déséquilibre, durée de déséquilibre triphasé lourde standardisée, etc.

Nous effectuons une analyse comparative, comparant notre FMEE proposée à trois méthodologies de base établies : régression logistique, réseau neuronal et vecteur de support. machine. L'ensemble de données comprend 1265 variables caractéristiques du transformateur associées à niveaux de charge importants, soumis à 30 cycles de partitionnement aléatoire. Par la suite, chacun L'ensemble de données divisé aléatoirement est soumis à quatre algorithmes distincts. L'efficacité prédictive de ces algorithmes sur l'ensemble de vérification est évaluée sur la base de la mesure F et de l'erreur. mesures de taux.

La figure 1 représente les évaluations de mesure F issues de 30 instances de prédiction menée par quatre algorithmes distincts. Notamment, chaque évaluation implique un partitionnement cohérent de l'ensemble de données entre les quatre algorithmes. Une tendance perceptible émerge selon laquelle tous les algorithmes donnent des valeurs de mesure F supérieures à 0,8. Notamment, notre projet FMEE ex-

Machine Translated by Google

hibits présente les performances les plus prometteuses, avec une mesure F moyenne de 0,904, surpassant ainsi largement les trois autres méthodologies de base.

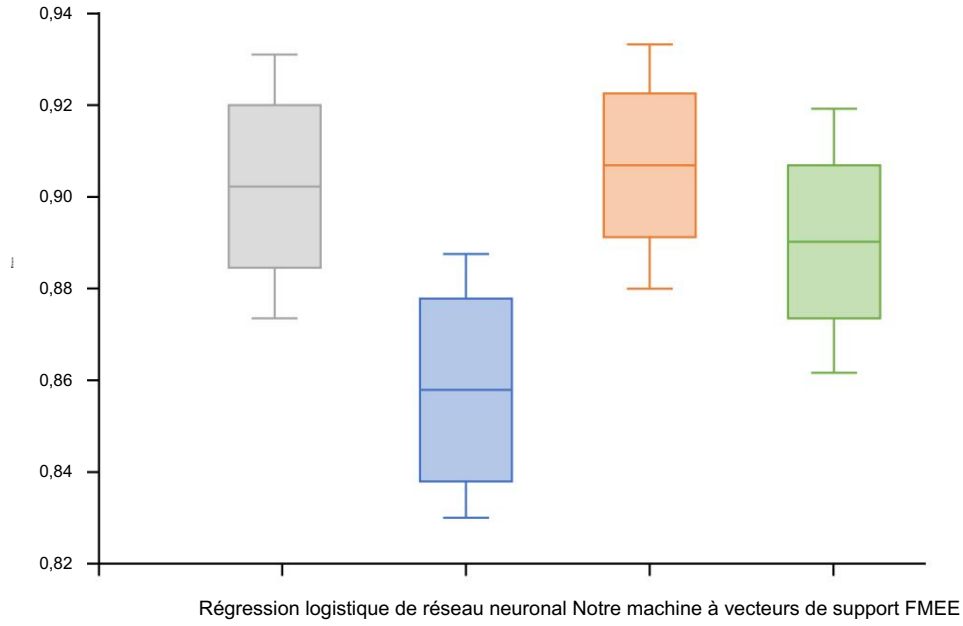


Figure 1. Diagramme en boîte de la mesure F parmi quatre algorithmes pour les résultats de prédiction des pannes.

Pour faciliter une comparaison complète entre les quatre algorithmes, la figure 2 représente le taux d'erreur et la mesure F des résultats de prévision des pannes sous forme de graphiques linéaires et rectangulaires. Notamment, FMEE apparaît comme un favori, présentant un avantage substantiel par rapport à la machine à vecteurs de support sophistiquée en termes de taux d'erreur. Plus précisément, notre FMEE atteint un taux d'erreur moyen de 9,1 %, dépassant le réseau neuronal, la régression logistique et la machine vectorielle de support avec des marges de 0,3 %, 4,2 % et 4,3 %, respectivement. Ces résultats soulignent la nette supériorité de notre méthodologie FMEE proposée.

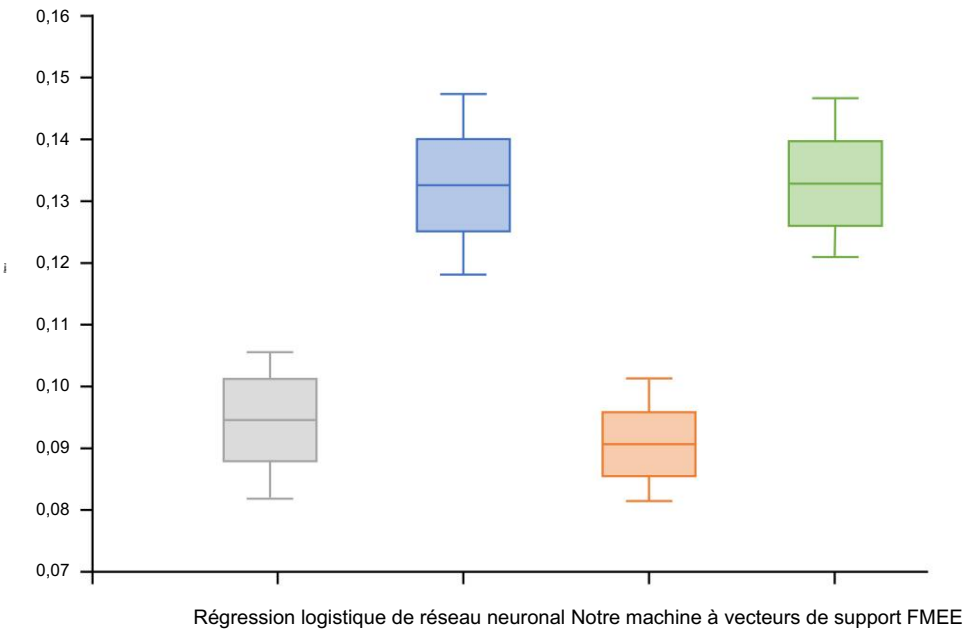


Figure 2. Diagramme en boîte du taux d'erreur parmi quatre algorithmes pour les résultats de prédiction des pannes.

Machine Translated by Google

De plus, nous étendons notre expérimentation à une région urbaine plus densément peuplée au sein de la ville du nord-ouest de la Chine. La figure 3 illustre les évaluations de mesure F dérivées de 30 tests indépendants menés sur les quatre algorithmes comparatifs. À une échelle globale, notre FMEE proposé atteint une mesure F de 0,891, dépassant les méthodologies de base telles que les réseaux de neurones, la régression logistique et les machines vectorielles de support avec des marges de 0,013, 0,023 et 0,028, respectivement.

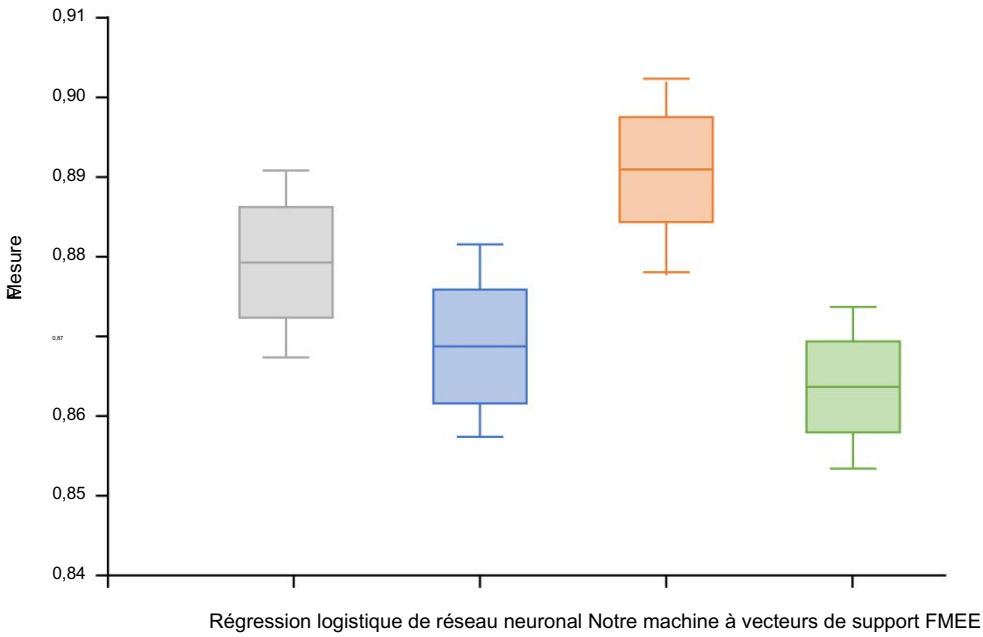


Figure 3. Diagramme en boîte de la mesure F de quatre algorithmes pour les résultats de prédiction des pannes.

Enfin, nous rapportons le taux d'erreur de la mesure F concernant les résultats de prédiction des pannes dérivés de quatre algorithmes comparatifs, représentés sous forme de graphiques linéaires et rectangulaires. Comme le montre la figure 4, notre FMEE proposée atteint un taux d'erreur moyen de 11,2 %, démontrant une supériorité notable sur trois méthodologies de base. Ces résultats soulignent de manière robuste l'efficacité de notre algorithme pour résoudre les problèmes de régression.

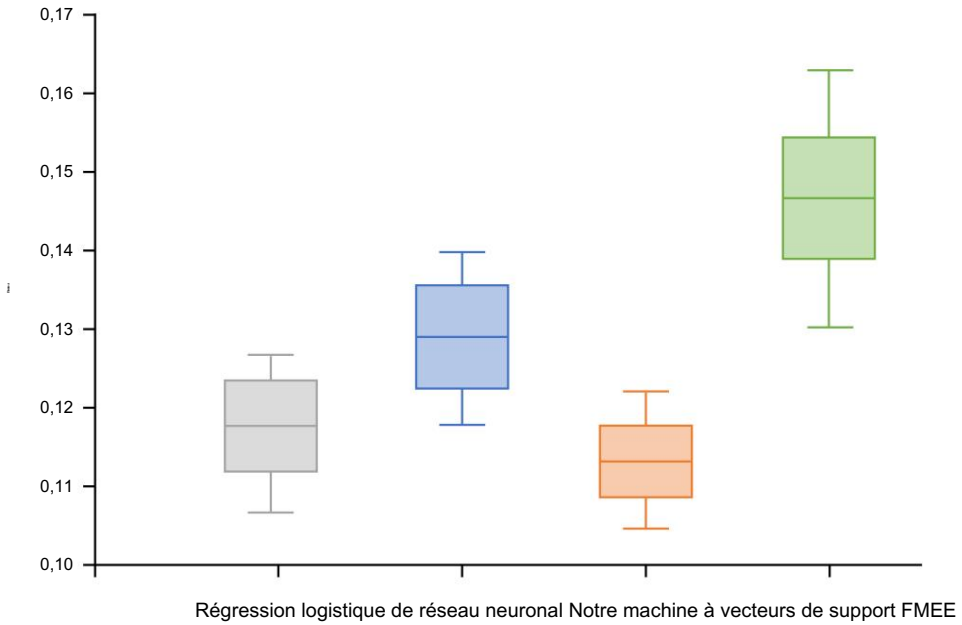


Figure 4. Diagramme en boîte du taux d'erreur de quatre algorithmes pour les résultats de prédiction des pannes.

Machine Translated by Google

5. Conclusions

Cette étude présente un algorithme rapide d'entropie d'erreur minimale adapté à la régression linéaire. Plusieurs conclusions notables ressortent de notre enquête. Premièrement, le FMEE proposé présente une convergence beaucoup plus rapide, nécessitant moins d'itérations que le MEE. Deuxièmement, bien que FMEE ne surpasse pas le MEE traditionnel en termes de performances pour le bruit gaussien et donne des résultats comparables pour le bruit non gaussien mesuré par l'erreur quadratique moyenne, son efficacité de calcul est étonnamment supérieure, consommant moins de 1 ‰ du temps requis par MEE. Troisièmement, bien qu'il soit dérivé de l'hypothèse d'une proximité du bruit avec la distribution gaussienne, le FMEE démontre son efficacité dans la gestion des distributions de bruit non gaussiennes à queue lourde . Quatrièmement, en raison de son efficacité temporelle exceptionnelle, FMEE est prometteur pour les applications en temps réel dans des scénarios comportant du bruit non gaussien. À l'avenir, une analyse plus approfondie des conditions de convergence du FMEE et l'exploration de méthodes d'approximation du bruit pour des distributions arbitraires représentent des pistes de recherche futures.

Contributions des auteurs : Conceptualisation, QL et XL ; méthodologie, QL; logiciel, WC ; validation , QL, XL et WC ; analyse formelle, YW ; enquête, XL; ressources, qualité de vie ; conservation des données, YW ; rédaction – préparation du projet original, QG ; rédaction – révision et édition, HC ; visualisation, QG ; supervision, SC; administration de projet, WC; acquisition de financement, QL et HC Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Ce travail est soutenu par les projets d'innovation scientifique et technologique du State Grid Information and Telecommunication Group « Recherche sur le système de technologie spatiale numérique de puissance et les technologies clés » (numéro de subvention : SGIT0000XMJS2310456).

Déclaration du comité d'examen institutionnel : sans objet.

Déclaration de consentement éclairé : sans objet.

Déclaration de disponibilité des données : Les données présentées dans cette étude sont disponibles sur demande auprès de l'auteur correspondant. Les données ne sont pas accessibles au public en raison de l'importance particulière du système d'alimentation électrique utilisé dans le processus expérimental, qui implique la confidentialité dans certains domaines de recherche, il n'est pas pratique de divulguer entièrement leurs données.

Conflits d'intérêts : les auteurs (Qiang Li, Xiao Liao, Wei Cui et Ying Wang) étaient employés par State Grid Information & Telecommunication Group Co., Ltd. Les autres auteurs déclarent que la recherche a été menée en l'absence de tout contact commercial ou commercial. relations financières qui pourraient être interprétées comme un conflit d'intérêts potentiel.

Références

1. Baggenstoss, PM Conception de PDF à entropie maximale utilisant des contraintes de densité de fonctionnalités : applications dans le traitement du signal. IEEE Trans. Processus de signal. 2015, 11, 2815-2825. [Référence croisée]
2. Li, Y. ; Bai, X. ; Yuan, D. ; Yu, C. ; San, X. ; Guo, Y. ; Zhang, L. ; Ye, J. Oxyde bidimensionnel à haute entropie à base de Cu comme catalyseur photothermique stable et actif. Nat. Commun. 2023, 14, 3171. [Réf. croisée] [Pub Med]
3. Bisikalo, O. ; Kharchenko, V. ; Kovtun, V. ; Krak, moi; Pavlov, S. Paramétrage du modèle stochastique pour évaluer de petites données variables dans la base d'entropie de Shannon. Entropie 2023, 25, 184. [CrossRef] [Pub Med]
4. Chen, Q. ; Zhang, F. ; Su, L. ; Lin, B. ; Chen, S. ; Zhang, Y. Estimation de fusion des paramètres d'état pour les véhicules intelligents basée sur IMM-MCCKF. Appl. Sci. 2024, 14, 4495. [Réf. croisée]
5. Carli, FP; Chen, T. ; Ljung, L. Noyaux d'entropie maximale pour l'identification du système. IEEETrans. Automatique. Contrôle 2016, 62, 1471-1477. [Référence croisée]
6. Ventilateur, X. ; Chen, L. ; Huang, D. ; Tian, Y. ; Zhang, X. ; Jiao, M. ; Zhou, Z. Des métaux uniques aux alliages à haute entropie : comment l'apprentissage automatique accélère le développement d'électrocatalyseurs métalliques. Av. Fonction. Maître. 2024, 1, 2401887. [Réf. croisée]
7. Han, TT; Ge, SS ; Lee, TH Systèmes non linéaires à commutation de temps de séjour persistant : paradigme de variation et conception de jauge. IEEE Trans. Automatique. Contrôle 2010, 55, 321-337.
8. Chesi, G. Stabilisation et réduction de l'entropie via la conception basée sur SDP de contrôleurs de rétroaction de sortie à ordre fixe et de paramètres de réglage. IEEETrans. Automatique. Contrôle 2017, 62, 1094-1108. [Référence croisée]
9. Wei, P. ; Li, HX Entropie spatio-temporelle pour la détection et la localisation d'anomalies des batteries Li-ion. IEEETrans. Indiana Électron. 2023, 70, 12851-12859. [Référence croisée]
10. Wang, Z. ; Li, G. ; Yao, L. ; Cai, Y. ; Lin, T. ; Zhang, J. ; Dong, H. Schéma intelligent de détection de défauts pour les éoliennes à vitesse constante basé sur une entropie floue multi-échelle améliorée et une machine à vecteurs de support adaptative chaotique basée sur l'optimisation Aquila. ISA Trans. 2023, 138, 582-602. [Référence croisée] [Pub Med]

Machine Translated by Google

11. Daneshpazhouh, A. ; Sami, A. Détection des valeurs aberrantes basée sur l'entropie utilisant une approche semi-supervisée avec quelques exemples positifs. Reconnaissance de formes. Lett. 2014, 49, 77-84. [Référence croisée]
12. Qu, B. ; Wang, Z. ; Shen, B. ; Dong, H. Estimation d'état dynamique décentralisée pour les systèmes électriques multi-machines avec bruits non gaussiens : détection et localisation des valeurs aberrantes. Automatique 2023, 153, 111010. [CrossRef]
13. Feng, Z. ; Wang, G. ; Peng, B. ; Lui, J. ; Zhang, K. Nouveau filtre de Kalman à distribution de Wasserstein à entropie d'erreur minimale robuste sous incertitude du modèle et bruit non gaussien. Processus de signal. 2023, 203, 108806. [Réf. croisée]
14. Soleil, M. ; Davies, MOI ; Proudler, IK ; Hopgood, JR Filtre de Kalman à noyau adaptatif. IEEETrans. Processus de signal. 2023, 71, 713-726. [Référence croisée]
15. Wu, L. ; Lin, H. ; Moyeu.; Tan, C. ; Gao, Z. ; Liu, Z. ; Li, SZ Au-delà de l'hypothèse d'homophilie et d'homogénéité : réseaux neuronaux à graphes adaptatifs de fréquence basés sur les relations. IEEETrans. Réseau neuronal. Apprendre. Système. 2023, 35, 8497-8509. [Référence croisée] [Pub Med]
16. Guo, L. ; Wang, H. Filtrage d'entropie minimale pour les systèmes stochastiques multivariés avec des bruits non gaussiens. IEEETrans. Automatique. Contrôle 2006, 51, 695-700.
17. Wang, Y. ; Tang, YY ; Li, L. Représentation clairsemée basée sur l'entropie d'erreur minimale pour un clustering de sous-espace robuste. IEEETrans. Processus de signal . 2015, 63, 4010-4021. [Référence croisée]
18. Wang, M. ; Deng, W. Reconnaissance faciale profonde : une enquête. Neuroinformatique 2021, 429, 215-244. [Référence croisée]
19. Lin, M. ; Cheng, C. ; Peng, Z. ; Dong, X. ; Qu, Y. ; Meng, G. Identification de systèmes dynamiques non linéaires à l'aide des méthodes de régression clairsemée et des moindres carrés séparables. J. Son Vib. 2021, 505, 116141. [Réf. croisée]
20. Wu, Z. ; Peng, S. ; Maman, W. ; Chen, B. ; Principe, JC Algorithmes d'entropie d'erreur minimale avec contraintes de pénalité de parcimonie. Entropie 2015, 17, 3419-3437. [Référence croisée]
21. Huang, S. ; Tran, TD Récupération du signal clairsemé via la minimisation des fonctions d'entropie généralisées. IEEETrans. Processus de signal. 2018, 67, 1322-1337. [Référence croisée]
22. Yue, H. ; Wang, H. Contrôle d'entropie minimum des erreurs de suivi en boucle fermée pour les systèmes stochastiques dynamiques. IEEETrans. Automatique. Contrôle 2003, 48, 118-122.
23. Wang, H. Contrôle d'entropie minimum des systèmes stochastiques dynamiques non gaussiens. IEEETrans. Automatique. Contrôle 2002, 47, 398-403. [Référence croisée]
24. Mitra, R. ; Bhatia, V. Estimation de canal basée sur le critère d'entropie d'erreur minimale pour le MIMO massif dans VLC. IEEETrans. Véh. Technologie. 2018, 68, 1014-1018. [Référence croisée]
25. Erdogmus, D. ; Principe, JC Propriétés de convergence et efficacité des données du critère d'entropie d'erreur minimale dans la formation adaline. IEEETrans. Processus de signal. 2003, 51, 1966-1978.
26. Chen, B. ; Yuan, Z. ; Zheng, N. ; Principe, algorithme d'entropie d'erreur minimale du noyau JC. Neuroinformatique 2013, 121, 160-169. [Référence croisée]
27. Hu, T. ; Wu, Q. ; Zhou, DX Convergence de la descente de gradient pour le principe d'entropie d'erreur minimale dans la régression linéaire. IEEETrans. Processus de signal. 2016, 64, 6571-6579. [Référence croisée]
28. Zhang, Y. ; Chen, B. ; Liu, X. ; Yuan, Z. ; Principe, JC Convergence d'un algorithme d'entropie d'erreur minimale à virgule fixe. Entropie 2015, 17, 5549-5560. [Référence croisée]
29. Chen, B. ; Zhu, Y. ; Hu, J. ; Zhang, M. Sur les estimations optimales avec critère d'entropie d'erreur minimale. J. Frankl. Inst. 2010, 347, 545-558.
30. Heravi, AR ; Hodtani, GA Une nouvelle relation théorique de l'information entre l'entropie d'erreur minimale et la correntropie maximale. Processus de signal IEEE. Lett. 2018, 25, 921-925.
31. Yang, S. ; Tan, J. ; Chen, B. Méta-apprentissage continu robuste basé sur des pointes amélioré par un critère d'entropie d'erreur minimale restreint. Entropie 2022, 24, 455. [CrossRef] [Pub Med]
32. Dan, L. ; Chen, B. ; Wang, S. ; Maman, W. ; Ren, P. Estimation robuste de l'état du système électrique avec une entropie d'erreur minimale sans parfum Filtre de Kalman. IEEETrans. Instrument. Mesures. 2020, 69, 8797-8808. [Référence croisée]
33. Wang, B. ; Hu, T. Principe d'entropie d'erreur minimale semi-supervisée avec méthode distribuée. Entropie 2018, 20, 968. [CrossRef] [Pub Med]
34. Chen, B. ; Xing, L. ; Xu, B. ; Zhao, H. ; Principe, JC Aperçu de la robustesse de l'estimation de l'entropie d'erreur minimale. IEEETrans. Réseau neuronal. Apprendre. Système. 2016, 29, 731-737. [Référence croisée] [Pub Med]
35. Shen, P. ; Li, C. Méthode d'entropie d'erreur totale minimale pour l'estimation des paramètres. IEEETrans. Processus de signal. 2015, 63, 4079-4090. [Référence croisée]
36. Kothari, P. ; Kreiss, S. ; Alahi, A. Prévision de la trajectoire humaine dans les foules : une perspective d'apprentissage en profondeur. IEEETrans. Intell. Transp. Système. 2021, 23, 7386-7400. [Référence croisée]
37. Kim, N. ; Kwon, K. Algorithme d'entropie d'erreur minimale normalisé avec estimation de puissance récursive. Entropie 2016, 18, 239. [Référence croisée]
38. Chen, B. ; Mar.; Yu, S. ; Du, S. ; Qin, J. Granger analyse de causalité basée sur le critère d'entropie d'erreur minimale quantifiée. IEEE Processus de signal. Lett. 2019, 26, 347-351. [Référence croisée]
39. Niknam, M. ; Santos, LF ; Cory, DG Détection expérimentale de l'entropie de Rényi de corrélation dans le modèle de spin central. Phys. Tour. Lett. 2021, 127, 080401. [Réf. croisée] [Pub Med]
40. Kim, T. ; San Kim, D. ; Jang, LC; Lee, H. ; Kim, H. Représentations de polynômes Hermite dégénérés. Av. Appl. Mathématiques. Mécanique. 2022, 139, 102359. [Réf. croisée]

Machine Translated by Google

41. Lin, W. ; Zhang, JE Les régions valides des densités de Gram – Charlier avec des cumulants d'ordre élevé. J. Informatique. Appl. Mathématiques. 2022, 407, 113945. [Référence croisée]
42. Matsushita, R. ; Brandão, H. ; Nobre, moi; Da Silva, S. Estimation différentielle de l'entropie avec un noyau parétien : lourdeur de la queue et lissage. Phys. Une statistique. Mécanique. Appl. 2024, 1, 129850. [Réf. croisée]
43. Canales, C. ; Galarcé, C. ; Rubio, F. ; Pineda, F. ; Anguita, J. ; Barros, R. ; Parragué, M. ; Daille, LK; Aguirre, J. ; Armijo, F. ; et coll. Test du test : Une étude comparative de la corrosion microbienne marine dans des conditions de laboratoire et sur le terrain. ACS Omega 2021, 6, 13496-13507. [Référence croisée] [Pub Med]

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Optimisation de la sélection génétique pour la classification du cancer Données de grande dimension utilisant un Africain amélioré Algorithme des vautours

Mona G. Gafar 1,2,*, Amr A. Abohany 3,†, Ahmed E. Elkhouli 4,† et Amr A. Abd El-Mageed 5,†

1 Département de génie informatique et d'information, Collège d'ingénierie de Wadi Alddawasir, Université Prince Sattam bin Abdulaziz, Kharj 16278, Arabie Saoudite Département
2 d'apprentissage automatique et de recherche d'informations, Intelligence artificielle, Université Kafrelsheikh, Kafrelsheikh 33511, Égypte
3 Faculté d'informatique et d'information, Université Kafrelsheikh , Kafrelsheikh 33511, Égypte ; amrabohany8@gmail.com
4 Département de génie biomédical, Faculté de génie électrique, Université de Menofia, Menofia 32951, Égypte ; eng.elkholy.ahmed@gmail.com
5 Département des systèmes d'information, Université de Sohag, Sohag 82511, Égypte ; amr.atef@commerce.sohag.edu.eg
* Correspondance : m.gafar@psau.edu.sa ou dr.mona.gamal85@gmail.com Ces
† auteurs ont contribué à parts égales à ce travail.

Résumé : Cette étude présente une nouvelle méthode, appelée RBAVO-DE (Relief Binary African Vultures Optimization based on Differential Evolution), visant à relever le défi de la sélection génétique (GS) dans les données RNA-Seq de grande dimension, en particulier le rnaseqv2 lluminaHiSeq rnaseqv2 un edu Ensemble de données normalisé sur les gènes RSEM de niveau 3, qui contient plus de 20 000 gènes. Le séquençage d'ARN (RNA-Seq) est une approche transformatrice qui permet la quantification et la caractérisation complètes des expressions géniques, dépassant les capacités des technologies de micro-réseaux en offrant une vue plus détaillée des données d'expression génique d'ARN-Seq. L'analyse quantitative de l'expression des gènes peut jouer un rôle essentiel dans l'identification des gènes qui différencient les tissus normaux des tissus malins. Cependant, la gestion de ces données matricielles denses de haute dimension présente des défis importants. L'algorithme RBAVO-DE est conçu pour sélectionner méticuleusement les gènes les plus informatifs à partir d'un ensemble de données comprenant plus de 20 000 gènes et évaluer leur pertinence dans vingt-deux ensembles de données sur le cancer. Pour déterminer l'efficacité des gènes sélectionnés, cette étude utilise les classificateurs Support Vector Machine (SVM) et k-Nearest Neighbour (k-NN). Comparé aux versions binaires d'algorithmes méta-heuristiques largement reconnus, RBAVO-DE démontre des performances supérieures. Selon le test de somme de rangs de Wilcoxon, avec un niveau de signification de 5 %, RBAVO-DE atteint une précision de classification allant jusqu'à 100 % et réduit la taille des caractéristiques jusqu'à 98 % dans la plupart des vingt-deux ensembles de données sur le cancer examinés. Cette avancée souligne le potentiel de RBAVO-DE à améliorer la précision de la sélection génétique pour la recherche sur le cancer, facilitant ainsi une identification plus précise et plus efficace des marqueurs génétiq

Mots clés : séquençage d'ARN ; sélection génétique; haute dimensionnalité; algorithmes méta-heuristiques ; Algorithme des vautours africains ; évolution différentielle; technique de soulagement

check for updates

Référence : Gafar, MG ; Abohany, AA; Elkhouli, AE; Abd El-Mageed, AA
Optimisation de la sélection des gènes pour Classification du cancer dans
Données de grande dimension utilisant un algorithme amélioré pour les vautours africains. Algorithmes 2024, 17, 342. https:// doi.org/ 10.3390/a17080342

Rédacteurs académiques : Muhammad Adnan Khan et Frank Werner

Reçu : 27 avril 2024
Révisé : 16 juillet 2024
Accepté : 1er août 2024
Publié : 6 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

L'acide désoxyribonucléique (ADN) constitue notre code génétique et contient la recette de notre existence. Bien que chaque cellule possède le même ADN, chaque structure tissulaire est unique et répond à un objectif distinct. Le mécanisme de transcription de l'ARN détermine quels gènes d' une cellule sont actifs, permettant à l'ARN d'être transformé en protéines responsables de la structure et de la fonctionnalité de la cellule [1]. Nous analysons les données d'expression génique RNA-Seq (GED) pour déterminer les changements génétiques et évaluer les biomarqueurs de la maladie. L'analyse de l'expression différentielle identifie les distinctions quantitatives dans l'expression des gènes, nous permettant de catégoriser les gènes dont l'expression change dans diverses conditions. Cette méthode nous aide

Machine Translated by Google

maladies et trouver des moyens de les gérer. Les technologies de profilage de l'expression génique se sont considérablement développées au fil des années [2].

Deux techniques populaires sont les micro-réseaux, qui utilisent une approche basée sur l'hybridation, et le RNA-Seq, qui est basé sur le séquençage de nouvelle génération. Ces technologies ont démontré leur valeur pour faire progresser notre compréhension de l'expression des gènes [3]. Des études notables , comme celles de Chen et al. [4] et Nunez et al. [5], ont mis en évidence le potentiel de ces techniques, tandis que Wang et al. [6] ont fourni des informations précieuses sur l' approche RNA-Seq.

Les deux méthodes considérées servent à quantifier l'expression des gènes à des fins de classification et d'investigation statistique. Cet article a choisi les données de quantification obtenues grâce à l'approche RNA-Seq basée sur le séquençage de nouvelle génération (NGS) car elles offrent plus de précision dans la détection des niveaux de quantification de l'ARN par rapport aux données de micropuces [7]. RNA-Seq a surmonté plusieurs limitations de l'analyse par micropuces, telles que la dépendance à l'égard des connaissances préalables en matière de séquençage, qui limitaient la plage de détection (8,9). Contrairement aux micro-arrays, RNA-Seq ne nécessite aucune connaissance préalable et élargit considérablement la plage de détection dynamique. Cette amélioration améliore la précision des résultats et permet l'identification d'un ensemble complet de gènes, fournissant ainsi une compréhension précise des biomarqueurs de la maladie. En conséquence, RNA-Seq est devenu un outil puissant pour les chercheurs, offrant des informations précieuses sur les mécanismes moléculaires sous-jacents à diverses maladies (10).

La « malédiction de la dimensionnalité » [11] est devenue un défi courant à l'époque contemporaine en raison de l'abondance des données disponibles. Cela a provoqué une augmentation du développement de techniques et d'algorithmes de sélection de fonctionnalités (FS). Les algorithmes FS peuvent être classés en quatre méthodes distinctes : filtre, wrapper, intégré et hybride [12,13]. Ces approches visent à identifier les caractéristiques les plus informatives permettant de distinguer les différentes classes.
Dans notre cas, il s'agit d'identifier les gènes liés aux tumeurs.

L'approche par filtre est une technique largement utilisée en FS qui consiste à évaluer la pertinence de gènes individuels sur la base de scores statistiques. Cette méthode est connue pour sa grande précision dans la sélection du groupe de gènes le plus satisfaisant. Néanmoins, l'approche par filtre présente certaines limites, car fonctionner sur chaque gène individuellement ignore les interrelations entre les gènes, ce qui peut entraîner un problème d'optima local (14). Il convient de noter que l' approche par filtre peut être classée en deux sous-types : univariée et multivariée, cette dernière considérant les corrélations entre les gènes. Quelques exemples d'approche de filtre incluent le soulagement [15], le score de Fisher [16], le test t [17] et le gain d'information [18].

La méthode wrapper étudie tous les sous-ensembles de gènes réalisables pour tester et créer un sous-ensemble de gènes afin de déterminer leurs implications. Un classificateur spécifique est utilisé pour déterminer le résultat de chaque sous-ensemble, et la technique de catégorisation est utilisée plusieurs fois pour chaque évaluation. Par rapport à l'approche filtre, l'approche wrapper offre des performances supérieures en utilisant une technique de catégorisation qui dirige la procédure d'apprentissage. Cependant , cette méthode nécessite beaucoup de temps et de ressources de calcul, en particulier lorsqu'il s'agit de données à grande échelle [19].

Les méthodes méta-heuristiques (MHM) sont des heuristiques de haut niveau utilisées en optimisation mathématique et en informatique pour trouver des solutions satisfaisantes aux problèmes d'optimisation lorsque l'information est imparfaite ou que les ressources sont limitées [20]. Les MHM échantillonnent un sous-ensemble de solutions, ce qui les rend utiles pour divers problèmes d'optimisation en raison de leurs rares hypothèses sur le problème [21]. Cependant, ils ne garantissent pas des solutions globalement optimales. De nombreux MHM utilisent l'optimisation stochastique, ce qui signifie que la solution est basée sur des variables aléatoires générées au cours du processus [22].

Les méthodes méta-heuristiques (MHM) sont plus pratiques que les méthodes itératives et les algorithmes d'optimisation traditionnels, car elles peuvent explorer un plus large éventail de solutions possibles. Par conséquent, les MHM sont devenus une stratégie privilégiée pour résoudre les problèmes d'optimisation [23]. De nombreux articles de recherche ont démontré que parmi les différentes méthodes de wrapper, les MHM sont bien adaptées pour résoudre le problème de sélection de fonctionnalités (FS). Les méthodes stochastiques, y compris les MHM, peuvent produire rapidement des résultats optimaux ou quasi optimaux. Ils

Machine Translated by Google

offrent des avantages tels que la flexibilité, l'autogestion sans propriétés mathématiques détaillées et la capacité d'évaluer de nombreux résultats simultanément. Divers MHM ont récemment été développés pour résoudre le problème FS, fournissant des solutions fiables presque optimales à des coûts de calcul considérablement réduits [24].

Les méthodes FS intégrées utilisent une technique d'apprentissage pour sélectionner les gènes appropriés qui interagissent avec le processus de classification. Cette méthode combine la technique FS dans le cadre du modèle d'apprentissage. L'algorithme d'apprentissage est entraîné avec un sous-ensemble d'attributs initial pour estimer une mesure permettant d'évaluer les valeurs de rang des attributs [25]. Le but ultime est de réduire le temps de calcul nécessaire à la reclassification des différents sous-ensembles en intégrant l'étape FS dans le processus de formation. Certaines techniques de cette approche effectuent une pondération des caractéristiques sur la base de modèles de régularisation avec des fonctions objectives qui minimisent les erreurs d'ajustement tout en imposant des coefficients de caractéristiques petits ou précisément nuls. Des exemples de méthodes intégrées sont l'algorithme de sélection de sous-ensembles de fonctionnalités basé sur des règles FOIL (First-Order Inductive Learner) et le SVM basé sur l'élimination récursive de fonctionnalités (SVM-RFE) [26].

La méthode hybride est une technique bien conçue qui fusionne les méthodes de filtrage et d'emballage pour exploiter les atouts de chacune. Cette approche commence par un filtre qui réduit la dimensionnalité de l'espace des fonctionnalités, générant plusieurs sous-ensembles de complexité intermédiaire. Par la suite, un wrapper est utilisé comme technique d'apprentissage pour choisir le sous-ensemble candidat le plus approprié [27]. L'approche hybride intègre la précision des wrappers et l' efficacité des filtres, aboutissant à une méthodologie optimale [28,29].

1.1. Motivation et contribution

Un algorithme d'optimisation méta-heuristique globale efficace est présenté dans cet article, appelé algorithme d'optimisation des vautours africains (AVO) [30], qui imite les habitudes de vie et d'alimentation des vautours africains. L'algorithme proposé se compose de quatre phases de base : division de la population en trois groupes : la meilleure solution, la deuxième meilleure solution et les solutions restantes ; mesure du niveau de famine pour formuler un modèle mathématique pour l'exploitation, l'exploration et le transfert des vautours ; exploration des vautours en utilisant deux stratégies pour permettre aux vautours de parcourir de grandes distances sur de longues périodes pour trouver de la nourriture sur des sites aléatoires ; et l'exploitation des vautours impliquant deux sous-phases. Deux stratégies distinctes sont utilisées dans la première sous-phase : le combat de siège pour les vautours puissants et le vol en rotation, qui forme une activité en spirale entre le vautour exceptionnel et les autres. Deux stratégies sont utilisées dans la deuxième sous-phase : rassembler les vautours autour de la source de nourriture et un combat de siège hostile, au cours duquel les vautours deviennent plus agressifs et tentent de voler la nourriture laissée par les vautours en bonne santé.

Cette étude propose une variante binaire améliorée de la technique AVO, à savoir la technique RBAVO-DE, qui est une approche productive qui démontre des performances précises dans la gestion du problème GS. Au début, il y a de fortes chances que la technique recommandée évite les optima locaux et permette d'obtenir une précision d'examen suffisante, une convergence rapide et une stabilité améliorée. Par rapport aux MHM récents, la technique RBAVO-DE proposée atteint une efficacité accrue en produisant des solutions optimales ou substantiellement optimales pour de nombreuses situations analysées. L'algorithme Relief est utilisé pour identifier uniquement les entités associées afin de produire l'ensemble de données de classification final. RBAVO-DE combine l'algorithme Relief avec l'approche DE pour augmenter la capacité d'exploration et obtenir les meilleurs résultats dans l'espace de solution via des répétitions. Il utilise également une fonction de transfert pour transformer les valeurs de position réelles en valeurs binaires. L'approche RBAVO-DE a du sens dans GS car elle est simple à comprendre et à mettre en œuvre, peut traiter divers problèmes d'optimisation, produit des résultats précieux dans un laps de temps acceptable, nécessite moins de puissance de calcul et utilise un petit nombre de paramètres de contrôle. Les

principales contributions de cet article sont les suivantes : • Les données de niveau 3 basées sur le séquençage de nouvelle génération (RNA-Seq) ont été prétraitées. • La capacité de

l'algorithme méta-heuristique AVO proposé à résoudre le problème GS n'a pas été étudiée, car RNA-Seq GED n'a jamais été utilisé avec l' algorithme AVO.

Machine Translated by Google

• Pour construire une version binaire connue sous le nom d'algorithme RBAVO-DE, l'algorithme AVO est modifié et recréé. • L'algorithme

RBAVO-DE proposé combine la variante binaire de l'AVO avec une approche Relief et DE pour améliorer la capacité d'exploration de l'espace de recherche et améliorer les résultats optimaux obtenus. • L'algorithme RBAVO-DE proposé est appliqué pour la

première fois au RNA-Seq GED. • Plusieurs indicateurs de performance, notamment la condition physique moyenne, l'exactitude de la classification, le nombre de gènes sélectionnés, la précision, le rappel et le score F1, sont utilisés pour évaluer les résultats. • Une comparaison est faite entre l'impact de l'algorithme RBAVO-DE présenté, employant les deux classificateurs ML recommandés (SVM et k-NN), et d'autres algorithmes de la littérature.

• Vingt-deux ensembles de données distincts sur le cancer sont utilisés pour évaluer la méthode RBAVO-DE proposée, et les résultats sont présentés. •

Les gènes choisis sont étudiés à l'aide de biomarqueurs associés au cancer.

1.2. Structure

Cet article est structuré en cinq sections principales. La section 2 passe en revue les recherches antérieures sur la FS utilisant RNA-Seq GED. Vient ensuite la section 3, qui propose une discussion approfondie de l'algorithme RBAVO-DE, une version améliorée d'AVO, y compris ses paramètres d'adressage de GS. La section 4 présente les résultats expérimentaux par rapport à plusieurs MHM récents. Enfin, la section 6 conclut cette étude et propose des pistes d' investigation future.

2. Travaux connexes

Dans cette section, nous discutons de la littérature qui se concentre sur les méthodes utilisées par les chercheurs pour classer le RNA-Seq GED, qui présente généralement une dimensionnalité élevée. Pour obtenir des performances optimales des méthodes de classification, il est crucial d'ignorer les gènes non pertinents et sans rapport ; par conséquent, la sélection des gènes appropriés est une étape vitale avant d'utiliser les techniques de ML et d'apprentissage profond (DL) [31] ou toute autre technique de classification. À cet égard, nous avons exploré certains articles pertinents dans ce domaine pour atteindre l'objectif de catégorisation RNA-Seq pour l'identification du cancer.

Yaqoob et coll. [32] ont introduit une technique de pointe pour GS connue sous le nom d' algorithme de recherche sinus-cosine-coucou (SCCSA), une méthode hybride conçue pour fonctionner aux côtés de modèles de ML établis tels que SVM. Cet algorithme GS innovant a été évalué à l'aide d'un ensemble de données de référence sur le cancer du sein, où ses performances ont été méticuleusement analysées et comparées à d'autres méthodologies GS. Pour affiner la sélection des fonctionnalités, la méthode de pertinence maximale de redondance minimale a été initialement appliquée comme étape de filtrage préliminaire. Par la suite, l'approche hybride SCCSA a été déployée pour améliorer et affiner davantage le processus GS. La dernière étape consistait à utiliser le classificateur SVM pour classer l' ensemble de données en fonction des gènes sélectionnés. Compte tenu de l'importance cruciale de GS dans le décodage d'ensembles de données biologiques complexes, SCCSA apparaît comme un atout crucial pour la classification des ensembles de données liés au cancer.

Joshi et coll. [33] ont introduit une stratégie d'optimisation innovante nommée PSO-CS intégrée à DL pour la classification des tumeurs cérébrales. Cette méthode améliore l'efficacité de la technique d'optimisation des essaims de particules (PSO) en incorporant l' algorithme Cuckoo Search (CS), optimisant ainsi le processus de classification. Suite à cette optimisation, PSO-CS utilise DL pour classer les GED liés aux tumeurs cérébrales, identifiant diverses classes associées à des tumeurs spécifiques parallèlement à la méthode d'optimisation PSO-CS. En intégrant la technique PSOC-CS à DL, il a largement surpassé les autres modèles DL et ML en termes de précision de classification, comme en témoignent diverses mesures de performance.

Mahto et coll. [34] ont dévoilé une approche révolutionnaire pour la classification des cancers grâce à une méthode intégrée basée sur CS-SMO (CS et Spider Monkey Optimization) pour GS. Initialement, la fonction de fitness de l'algorithme Spider Monkey Optimization (SMO) est modifiée à l'aide de l'algorithme CS. Cette modification exploite les atouts des deux MHM

Machine Translated by Google

identifier un sous-ensemble de gènes capables de prédire le cancer à un stade précoce. Pour affiner davantage la précision de l'algorithme CS-SMO, une étape de prétraitement, MRMR, est utilisée pour réduire la complexité des ensembles de données sur l'expression des gènes du cancer. Par la suite, ces sous-ensembles de gènes sont traités à l'aide de DL pour classer différents types de cancer. L'efficacité de l' approche CS-SMO couplée à la DL a été évaluée à l'aide de huit ensembles de données de référence sur l'expression génique de micro-réseaux pour le cancer, en examinant ses performances selon diverses mesures. La méthode CS-SMO intégrée à DL a démontré une précision de classification supérieure dans tous les ensembles de données d'expression génique à grande échelle examinés pour le cancer, surpassant les modèles DL et ML existants.

Neggaz et coll. [35] ont introduit une version améliorée de l'optimisation de la recherche de nourriture des raies manta , appelée MRFO-SC, qui utilisait des opérateurs trigonométriques inspirés de l' algorithme sinus-cosinus (SC) pour gérer le problème GS. Le modèle k-NN a été utilisé pour la sélection d'ensembles de gènes. De plus, la signification statistique du MRFO-SC a été évaluée à l'aide du test de somme des rangs de Wilcoxon à un niveau de signification de 5 %. Les résultats ont été évalués et comparés à certaines MHM récentes. La comparaison et les résultats expérimentaux ont confirmé les performances efficaces du MRFO-SC proposé sur des ensembles de données de référence de haute et basse dimensions en obtenant la plus grande précision de classification sur 85 % des ensembles de données de référence GS.

Lyu et coll. [36] ont exploré les biomarqueurs du cancer en se concentrant sur l'importance des gènes dans leur impact sur la classification. Ils ont suivi une approche en deux phases : prétraitement des données et utilisation d'un réseau neuronal convolutionnel pour classer le type de tumeur. Dans la deuxième phase, ils ont créé des cartes thermiques pour chaque catégorie afin d'identifier les gènes liés aux pixels présentant les intensités les plus élevées dans les cartes thermiques. Ils ont ensuite évalué les voies des gènes sélectionnés. Au cours du prétraitement, ils ont supprimé les niveaux d'expression génique qui n'avaient pas été modifiés au cours de la phase GS, en utilisant un seuil de variance de 1,19, ce qui a réduit le nombre de gènes de 19 531 à 10 381. La précision de classification finale obtenue était de 95,59, ce qui est bon mais pourrait encore être amélioré en utilisant une méthodologie FS plus efficace pour réduire davantage la dimensionnalité des données.

Khalifa et coll. [37] construit sur l'article susmentionné [36]. Leur étude portait sur cinq types de données sur le cancer : le carcinome de l'endomètre du corps utérin (UCEC), le carcinome épidermoïde du poumon (LUSC), l'adénocarcinome du poumon (LUAD), le carcinome rénal à cellules claires du rein (KIRC) et le carcinome invasif du sein (BRCA).). L'ensemble de données de référence utilisé pour l'étude comprenait 2 086 enregistrements et 972 attributs. Chaque enregistrement fournissait des informations détaillées sur l'échantillon, tandis que chaque attribut incluait les valeurs RNA-Seq pour un gène spécifique, représenté par RPKM (Reads Per Kilobase per Million) (38). Les chercheurs ont utilisé une approche mixte utilisant le PSO binaire avec la méthode des arbres de décision (BPSO-DT) pour prétraiter les données. Sur 971 attributs, 615 ont été sélectionnés comme les meilleurs attributs de RNA-SEquation. La méthode proposée a atteint une précision globale de classification des tests de 96,90 %, comme le démontrent les résultats suggérés et les mesures d'évaluation utilisées.

Xiao et coll. [39] ont évalué leur méthodologie à l'aide de trois ensembles de données RNA-Seq : Stomach Adenocarcinome (STAD), BRCA et LUAD. Leur approche reposait sur des techniques DL, dans lesquelles ils utilisaient cinq classificateurs et utilisaient ensuite la technique DL pour regrouper chaque sortie des cinq classificateurs. Cela a conduit à une amélioration de la précision de la classification de toutes les prédictions, l'ensemble de données BRCA atteignant une précision de 98,4 %, STAD une précision de 98,78 % et LUAD une précision de 99,20 %.

Liu et coll. [40] ont utilisé des données de micro-réseaux et une approche hybride pour résoudre le problème. Contrairement aux articles mentionnés précédemment, ils ont étudié chaque classe de cancer séparément. Pour évaluer les performances, ils ont utilisé quatre références génétiques liées aux petites tumeurs à cellules bleues rondes, au cancer du côlon, au cancer du poumon et à la leucémie. Leur méthode proposée utilisait Relief comme technique de pré-traitement pour éliminer les gènes présentant une corrélation plus faible avec la classe spécifique de cancer, suivi par PSO comme technique de recherche. Enfin, ils ont utilisé le modèle SVM pour évaluer l'exactitude de la classification du sous-ensemble de gènes sélectionné et obtenir le sous-ensemble de gènes optimal concluant pour chaque type de cancer.

Sur la base des recherches existantes, il semble que la plupart des études utilisant RNA-Seq GED en sont encore à leurs débuts, les chercheurs tentant de mettre en œuvre et de tester diverses idées.

Machine Translated by Google

dans ce domaine prometteur. Bien que la littérature contienne une pléthore d'expériences employant plusieurs techniques, telles que les méthodes récentes FS et DL, aucune technique n'est parfaite en raison de la haute dimensionnalité du RNA-Seq GED. FS de RNA-Seq GED joue un rôle crucial dans la détermination de la relation entre un gène et sa catégorie. Il s'agit d'une tâche de prétraitement essentielle pour valider les biomarqueurs génétiques du cancer et surmonter la malédiction de la dimensionnalité Par conséquent, cette étude vise à introduire une nouvelle approche wrapper, l' algorithme RBAVO-DE, et à l'appliquer pour la première fois aux données RNA-Seq. Il compare également l' efficacité de l'algorithme proposé avec celle d'autres techniques FS.

3. RBAVO-DE proposé pour GS

Une variante améliorée d'AVO connue sous le nom de RBAVO-DE est proposée dans cet article pour découvrir les plus petits sous-ensembles de gènes pertinents issus du processus de classification et ignorer les gènes non pertinents. RBAVO-DE utilise un algorithme AVO binaire basé sur Relief combiné à la technique DE. La principale caractéristique de RBAVO-DE est sa capacité à maximiser la précision tout en utilisant le moins de fonctionnalités possible. Le RBAVO-DE proposé comprend deux étapes principales. Tout d'abord, il existe une étape de prétraitement dans laquelle l'algorithme Relief est utilisé pour déterminer quelles caractéristiques sont significatives en attribuant à chaque caractéristique une pondération, puis en supprimant les caractéristiques qui ne sont pas pertinentes et qui ont les pondérations les plus faibles. Dans la deuxième étape, l' algorithme binaire AVO et la technique DE sont appliqués pour identifier les caractéristiques les plus pertinentes et uniques. L'algorithme AVO est sujet au piège optimal local lors du traitement de problèmes à grande échelle. L'algorithme AVO intègre la technique DE pour éviter cela.

L'algorithme RBAVO-DE proposé pour aborder la stratégie GS nécessite l'application de l'algorithme Relief, l'initialisation, l'amélioration de la position à l'aide de l'algorithme AVO, la conversion binaire, l'évaluation de la condition physique et l'intégration avec DE. Les sous-sections suivantes expliquent ces étapes.

3.1. Application de l'algorithme Relief pour la filtration des

fonctionnalités Cette étape vise à prétraiter la population à l'aide de l'algorithme Relief [41], qui est considéré comme une technique de filtrage rapide, simple et efficace pour trouver des fonctionnalités liées les unes aux autres. L'objectif principal de cet algorithme est de trouver des caractéristiques qui différencient les valeurs des échantillons et de regrouper les échantillons similaires à proximité. En conséquence, la méthode dépend du classement pondéré des caractéristiques, où une caractéristique avec un poids plus élevé indique de meilleures performances de classification.

Après avoir choisi un échantillon de manière aléatoire, l'algorithme Relief examine deux types différents d'échantillons les plus proches : les échantillons quasi-touchés, qui sont associés à des échantillons de la même classe, et les échantillons quasi-accidentés, qui sont associés à des échantillons d'autres classes. Les valeurs de quasi-impact et de quasi-accident peuvent être utilisées pour déterminer les pondérations des entités. Pour évaluer l'importance dans le processus de classification, les pondérations des caractéristiques sont classées du plus significatif au moins significatif. Enfin, les caractéristiques ayant les poids les plus significatifs sont sélectionnées. La formule suivante peut être utilisée pour déterminer le poids W pour la caractéristique A

$$WA = \sum_{j=1}^{N} {}^{jx}{}_{UN} - NM(xj)\,A \quad {}^{2} \quad - x \quad {}^{j}{}_{UN} - NH(xj)\,A \quad {}^{2}. \qquad (1)$$

où WA désigne le poids de la caractéristique A, x j j désigne la valeur de la caractéristique A pour les données x , $_{UN}$ identique ou différent et N désigne le nombre d'échantillons. Les points de données les plus proches des classes x sont NH(x j) et NM(x j).

L'algorithme Relief se concentre uniquement sur les fonctionnalités nécessaires et minimise la zone de recherche pour aider l'algorithme AVO à trouver plus rapidement de meilleures fonctionnalités.

3.2. Initialisation de la population

L'algorithme BAVO proposé commence par générer aléatoirement une population de N positions. Un vecteur de dimensions D équivalent au nombre d'entités dans l'ensemble de données d'origine décrit chaque position comme une solution possible à ses contraintes inférieure et supérieure.

Machine Translated by Google

bornes. Cette étape initialisée aléatoirement, limitée dans la plage [−1, 1] pour la variable de chaque vecteur de position, utilise la position produite aléatoirement.

3.3. Booster les positions via l'algorithme AVO

Cet article présente l'algorithme AVO [30], un algorithme d'optimisation méta-heuristique inspiré des vautours d'Afrique et de leurs comportements de vie et d'alimentation. Basé sur des idées fondamentales sur les vautours, l'algorithme AVO est configuré comme suit : La taille de la population de vautours africains est initialement présumée par l'algorithme AVO comme étant composée de N vautours. La population de vautours africains est ensuite divisée en trois groupes pour refléter la fonction naturelle principale des vautours, sur la base de la fonction de fitness calculée. Le premier groupe est constitué du vautour le plus fort, qui constitue la meilleure solution ; le deuxième groupe contient un vautour plus faible que le premier, ce qui constitue la deuxième meilleure solution ; et le dernier groupe comprend les vautours restants, les plus faibles, qui constituent les pires solutions.

Sur la base de ce qui précède, l'algorithme AVO proposé est composé de quatre phases fondamentales pour simuler le comportement de différents types de vautours. Les sous-sections suivantes clarifient ces phases.

3.3.1. Phase de division de la population

La population initiale est divisée en groupes en évaluant la fonction de fitness de la solution. La meilleure solution est choisie comme étant le meilleur vautour du premier groupe, tandis que le deuxième groupe contient la deuxième meilleure solution. Le troisième groupe contient les solutions restantes. Comme les solutions s'efforcent constamment de se rapprocher des meilleures et des deuxièmes meilleures solutions, la population doit être réévaluée à chaque itération, comme suit :

$$Rg = \begin{cases} \text{BestVultureg si } prg = L1, \\ \text{SecondBestVultureg 2 si } prg = L2, \end{cases} \quad (2)$$

$$prg = \sum_{u\,g\,g=1}^{N} \frac{ug}{} . \quad (3)$$

Le meilleur vautour du premier groupe et le deuxième meilleur vautour de la deuxième itération du groupe sont désignés par les expressions BestVultureg au $g^{ème}$ et SecondBestVultureg 2 , itération 1 respectivement. La probabilité de sélectionner la meilleure solution pour chaque groupe au ^{ème} niveau g est définie à l'aide de l'approche de la roulette, comme le montre l'équation (3), prg . Deux paramètres aléatoires compris dans la plage [0, 1] sont L1 et L2.

3.3.2. Phase de mesure du niveau de famine

Cette phase est utilisée pour formuler un modèle mathématique pour l'exploitation, l'exploration et le transfert entre les vautours. Lorsqu'ils ne sont pas affamés, les vautours ont une plus grande capacité à rechercher de la nourriture et peuvent voler plus loin. De plus, lorsqu'ils sont affamés, les vautours ne peuvent pas parcourir de longues distances à la recherche de nourriture et peuvent devenir violents. Le niveau de famine du vautour (F_i ^{ème}) au cours de l' ^{ème} itération g peut s'écrire comme suit, qui est utilisé pour développer un modèle mathématique pour l'exploitation, l'exploration et le transfert des vautours :

$$F_{je}^{g} = (2 \times rand + 1) \times z \times 1 - \frac{gi}{Gmax} + t, \quad (4)$$

$$t = h \times péchéw \frac{\pi}{2} \times \frac{gi}{Gmax} + parce\ que \frac{\pi}{2} \times \frac{gi}{Gmax} - 1 . \quad (5)$$

Les vautours sont vraisemblablement rassasiés depuis $^{g}_{je}$ montre le passage des vautours de l'exploration variable F jusqu'à l'exploitation. Un nombre aléatoire compris entre 0 et 1 s'appelle un rand. z désigne une valeur aléatoire dans l'intervalle [−1, 1]. Le numéro d'itération actuel est noté tandis que le numéro gi , d'itération maximum est noté Gmax. L'équation (5) calcule la valeur t pour aider à résoudre plus efficacement des problèmes d'optimisation complexes et éviter d'atteindre un optimal local. Une valeur arbitraire dans l'intervalle [−2, 2] est représentée par h. La probabilité de réaliser le processus d'exploration est régulée par le prédéfini

Machine Translated by Google

paramètre constant w ; la probabilité d'exploration augmente à mesure que sa valeur augmente. L'exploration est moins probable à mesure que sa valeur diminue.

L'équation (4) indique qu'à mesure que le nombre d'itérations augmente, F un $_{„}^{g}$ diminue progressivement. Comme résultat, l'étape suivante de l'algorithme AVO proposé peut être définie comme suit :

$$\begin{array}{ll} \text{Phase d'exploration (Recherche de nourriture dans diverses zones),} & \text{si } |F_{„}^{g}| \geq 1, \\ \text{Phase d'exploitation (Recherche de nourriture aux alentours),} & \text{si } |F_{„}^{g}| < 1. \end{array} \quad (6)$$

3.3.3. Phase d'exploration Les

vautours se caractérisent par leur grande capacité oculaire à localiser la nourriture appropriée pendant cette phase. Deux stratégies différentes sont utilisées dans cette phase pour permettre aux vautours de parcourir de grandes distances pendant de longues périodes pour rechercher de la nourriture dans des endroits aléatoires. Ces emplacements sont sélectionnés à l'aide d'un nombre aléatoire randP1 et d'un paramètre prédéfini P1, qui ont tous deux des valeurs comprises dans la plage [0, 1]. En phase d'exploration, le niveau de famine |F est supérieur $_{„}^{g}$ | est ou égal à 1. Les techniques d'exploration sont décrites comme suit :

$$X_{je}^{g+1} = \begin{array}{ll} R_{g} - D_{gg} \times F_{je}, & \text{si randP1} \leq P1, \\ R_{g} - F_{je}^{ig} + \text{rand} \times ((UB - LB) \times \text{rand} + LB), & \text{si randP1} > P1, \end{array} \quad \text{si } |F_{„}^{g}| \geq 1, \quad (7)$$

$$ré_{„}g = |(2 \times \text{rand}) \times R_{g} - X_{„}^{g}|. \quad (8)$$

où la position mise à jour à venir à la prochaine itération (g + 1) est notée X. Le meilleur vautour choisi pour $_{„}^{g+1}$. l'itération actuelle g est noté R g . Ceci est déterminé à l'aide de l'équation (2). Les vautours se déplacent de manière aléatoire pour protéger la nourriture des autres vautours et pour offrir un degré élevé de caractère aléatoire dans leur comportement de recherche. rand est une valeur aléatoire comprise entre zéro et un. La limite supérieure des variables est notée UB ; leur borne inférieure par LB ; et la position actuelle au g ^{ème} la ème itération g est X i .

3.3.4. Phase d'exploitation

Dans cette phase, $|F_{„}^{g}|$ est inférieur à 1. L'efficacité de l'algorithme AVO proposé est évaluée en deux sous-phases qui comprennent la phase d'exploitation. Chacune de ces sous-phases utilise deux stratégies différentes. Pour chaque sous-phase interne, la stratégie appropriée est choisie à l'aide de deux paramètres prédéfinis : P2 pour la première sous-phase et P3 pour la seconde, avec des valeurs allant de 0 à 1. Ces deux sous-phases internes s'expliquent comme suit :

- Première sous-phase d'exploitation : Cette sous-phase applique deux stratégies différentes quand |F $_{„}^{g}$ | est inférieur à 1 et supérieur ou égal à 0,5. La sélection de l'une de ces deux stratégies se fait à l'aide d'une valeur aléatoire randP2 comprise dans l'intervalle [0, 1] et d'un paramètre P2 spécifié .

  La stratégie initiale de cette sous-phase est appelée combat de siège, qui implique des vautours suffisamment puissants et quelque peu rassasiés. Les vautours plus robustes et en meilleure santé essaient de ne pas partager de nourriture avec d'autres car ils se rassemblent autour d'une seule source de nourriture. En grouillant près des vautours en bonne santé et en se livrant à de petits combats, les vautours les plus faibles tentent de leur prendre de la nourriture. A l'inverse, la deuxième stratégie est appelée vol en rotation ; cela crée un mouvement en spirale entre un vautour supérieur et les autres. Ce qui suit illustre les stratégies pour la première sous-phase d'exploitation :

$$X_{je}^{g+1} = \begin{array}{ll} D_{gi} \times (F + \text{rand}) - d_{t}^{g}, & \text{si randP2} \leq P2, \text{ si} \\ R_{g} - (S_{\text{concert 1}} + S_{2}^{g}), & \text{randP2} > P2, \end{array} \quad \text{si } 1 > |F_{„}^{g}| \geq 0{,}5, \quad (9)$$

$$d_{t}^{g} = R_{g} - X_{je,}^{g} \quad (10)$$

$$S^{g} = R_{g} \times 1 \quad \frac{\text{grand rand} \times X_{je}}{2\pi} \times \cos(X_{„}^{g}), \quad (11)$$

Machine Translated by Google

$$S2_{i}^{g} = R_{i}^{g} \times \frac{\text{grand} \times X_{i}^{g}}{2\pi} \times \text{péché}(X_{i}^{g}). \quad (12)$$

où la prochaine position mise à jour du vautour au prochain (g + 1) g+1 par X $_{i}^{g+1}$ est définie à l'aide de l'équation (8), et rand est une valeur arbitraire comprise entre 0 et 1. La distance d g entre le vautour et l'un des meilleurs deux vautours est estimée à l'aide de l'équation (10) ; R g désigne le meilleur vautour approprié dans le calcul indique la position actuelle le courant $^{\text{ème}}$ itération, qui est calculée à l'aide de l'équation (2) ; S $_{i1}^{g}$ et S 2 $_{i}^{g}$ g utilisant les équations (11) et (12), respectivement ; et X en g $_{i}^{g}$ $^{\text{ème}}$ itération.

• Deuxième sous-phase d'exploitation : Cette sous-phase est réalisée lorsque |F $_{i}^{gi}$ | est inférieur à 0,5. Diverses espèces de vautours se rassemblent autour de la réserve de nourriture et se livrent à de nombreux sièges et bagarres au cours de cette sous-phase. Cette sous-phase utilise deux stratégies différentes. Un paramètre prédéfini P3 et une valeur arbitraire randP3 , avec des valeurs allant de 0 à 1, sont utilisés pour décider laquelle de ces deux stratégies utiliser.
La stratégie initiale de cette sous-phase s'appelle rassembler les vautours autour de la source de nourriture. Dans cette stratégie, différents types de vautours recherchent de la nourriture et peuvent rivaliser à proximité d'une seule source. La deuxième stratégie est connue sous le nom de combat de siège hostile. Dans cette stratégie, les vautours deviennent plus agressifs et tentent de piller les restes de nourriture des vautours en bonne santé en se rassemblant autour d'eux de différentes manières. Les vautours en bonne santé, en revanche, se détériorent et sont incapables de repousser les autres vautours. Ce qui suit illustre les stratégies pour la deuxième sous-phase d'exploitation :

$$X_{i}^{g+1} = \begin{cases} \frac{\text{Un}_{1}^{g} + A_{2}^{g}}{2}, & \text{si randP3} \le \text{P3, si randP3} > \text{P3,} \\ R_{i}^{g} - |d_{t}^{g}| \times F_{i}^{g} \times \text{Lévyd}, & \end{cases} \quad \text{si } |F_{i}^{g}| < 0,5, \quad (13)$$

$$\text{Un}_{1}^{g} = \text{MeilleurVultureg}_{1} - \frac{\text{MeilleurVultureg}_{1} \times X_{i}^{g}}{\text{MeilleurVultureg}_{1} - (X_{i}^{g})^{2}} \times F_{i}^{g}, \quad (14)$$

$$\text{Un}_{2}^{g} = \text{DeuxièmeMeilleurVultureg}_{2} - \frac{\text{DeuxièmeMeilleurVultureg}_{2} \times X_{i}^{g}}{\text{SecondBestVultureg}_{2} - (X_{i}^{g})^{2}} \times F_{i}^{g}, \quad (15)$$

$$\text{Levyd} = 0{,}01 \times \frac{\mu \times \sigma}{|v|^{\frac{1}{\beta}}}, \quad (16)$$

$$\text{sigma} = \frac{\Gamma(1+\beta) \times \text{péché}(\frac{\Pi\beta}{2})}{\Gamma(1+\beta 2) \times \beta \times 2^{(\frac{\beta-1}{2})}}^{\frac{1}{\beta}}. \quad (17)$$

où la prochaine position mise à jour du vautour au prochain (g + 1) g+1 A g et A g l'assemblage de vautours, est $^{\text{ème}}$ itération, reflétant notée X sont évaluées en utilisant, respectivement, 1 2 , les équations (14) et (15). Pour augmenter l'efficacité de la stratégie AVO, Levyd est la fonction de répartition des vols obtenue à l'aide de l'équation (16). d est l'espace dimensionnel ; σ est défini par l'équation (17), où β = 1,5 est une valeur constante ; et μ et v sont des nombres aléatoires répartis également dans la plage [0, 1].

3.4. Conversion en nature binaire

Dans l'algorithme AVO présenté, les positions sont affichées sous forme de valeurs réelles. En tant que tels, ils ne sont pas immédiatement applicables au problème binaire GS. Par conséquent, il est nécessaire de convertir ces valeurs de position réelles en valeurs binaires pour se conformer à la nature binaire de GS et conserver la structure de l'algorithme d'origine. Dans cette conversion, les 1 représentent les valeurs réelles des gènes sélectionnés pertinents dans le vecteur de binarisation, tandis que les 0 représentent les valeurs réelles des gènes non sélectionnés, qui ne sont pas pertinents. A chaque itération g, l' expression mathématique suivante peut être utilisée pour convertir la position réelle X en une position binaire (X $_{i}^{g}$ ^{g})bac :

Machine Translated by Google

$$(X^{g}_{je})bin = \begin{cases} 1 \text{ si } X^{g}_{je} > \delta, \\ 0 \text{ sinon.} \end{cases} \quad (18)$$

où un point seuil arbitraire dans [0, 1] est représenté par δ. Selon cette approche fondamentale de conversion binaire, le « 1 » binaire (caractéristique sélectionnée) remplace sa valeur réelle si $(X^{g}_{je})bin$ est plus petit que δ. Dans le cas contraire, si $(X^{g}_{je})bin$ est supérieur à δ, la valeur binaire est fixée à 0 (caractéristique qui n'a pas été choisie) .

3.5. Évaluation de la valeur de la fonction fitness

Trouver le plus petit nombre de fonctionnalités sélectionnées et optimiser la précision de classification des classificateurs disponibles (modèles k-NN et SVM) sont deux objectifs contradictoires qui doivent être équilibrés pour obtenir la meilleure solution et déterminer sa qualité. Étant donné que la précision des classificateurs k-NN et SVM peut être entravée si le nombre de fonctionnalités sélectionnées est inférieur au nombre optimal, la fonction de fitness équilibre la taille et la précision des fonctionnalités sélectionnées. La fonction d'aptitude se concentre sur la réduction du taux d'erreur de classification plutôt que sur l'exactitude, comme suit :

$$\text{mangécueilli| f it} = w1 \times \frac{|f}{\text{Errrate} + w2}, \quad \times |D| \; w1 \quad [0, 1], w2 = 1 - w1. \quad (19)$$

où f eatpicked désigne le nombre d'entités sélectionnées, D désigne le nombre total d' entités dans l'ensemble de données et Errrate représente le taux d'erreur de classification des classificateurs k-NN et SVM. L'importance de la précision de la classification et le nombre de caractéristiques sélectionnées sont indiqués respectivement par les paramètres de pondération w1 et w2 . Les valeurs de w1 (0,99) et w2 (0,01) sont déterminées grâce à des essais approfondis menés dans des études antérieures (21, 42, 43).

3.6. Intégration de la technique DE [44]

se caractérise par sa grande efficacité et sa simplicité dans la recherche d'une solution appropriée à des problèmes d'optimisation complexes. Cela peut rapidement produire des résultats à valeur ajoutée. Trois processus principaux – mutation, croisement et sélection – sont nécessaires pour la DE. Le processus de mutation différentielle cherche à produire un vecteur modifié ui pour le vecteur solution de chaque itération . Cela peut être calculé mathématiquement comme suit :

$$\vec{u}_{je} = \vec{X}_{r1} + WM(\vec{X}_{r2} \vec{X}_{r3}) \quad (20)$$

Dans la plage [1, taille de la population], trois vecteurs asymétriques sélectionnés au hasard sont désignés par les symboles Xr1 , Xr2 et Xr3 . WM désigne le facteur de pondération de la mutation dans l'intervalle [0, 1].

Après l'opération de mutation, DE effectue une opération de croisement pour augmenter la variété de la population. Un vecteur de progéniture ui est produit en combinant les valeurs du vecteur cible Xi et du vecteur modifié ui . Le facteur de base le plus couramment utilisé dans la recherche croisée est le croisement binaire, qui a l'expression mathématique suivante :

$$u_{i,d} = \begin{cases} u_{i,d}, & \text{si rand} \le CR \text{ ou } d = j\text{rand, sinon.} \\ X_{i,d}, & \end{cases} \quad (21)$$

où un nombre aléatoire uniformément distribué jrand [1, 2, ..., DX] est utilisé pour garantir que le vecteur muté a au moins une dimension. La possibilité de croiser chaque élément est calculée à l'aide du taux de croisement CR, souvent déterminé comme étant un montant important (CR = 0,9).

L'opération de sélection est ensuite effectuée, comme le montre l'équation (22). Ici, une comparaison entre la fonction de fitness f(Xi) du vecteur cible et le vecteur de progéniture correspondant f(ui) est effectuée, et la fonction de fitness de valeur la plus basse est conservée pour l' itération à venir.

Machine Translated by Google

$$Xi = \begin{cases} \overline{Xl}, & \text{si } f(ui) < f(Xi) \\ & \text{sinon.} \end{cases} \quad (22)$$

3.7. L'algorithme RBAVO-DE complet

Pour gérer la stratégie GS, les étapes de l'algorithme RBAVO-DE recommandé sont décrites dans les sous-sections suivantes. Le pseudo-code de l' algorithme RBAVO-DE proposé est fourni dans l'algorithme 1. Un organigramme de l'algorithme RBAVO-DE proposé est présenté à la figure 1, illustrant ses principales étapes.

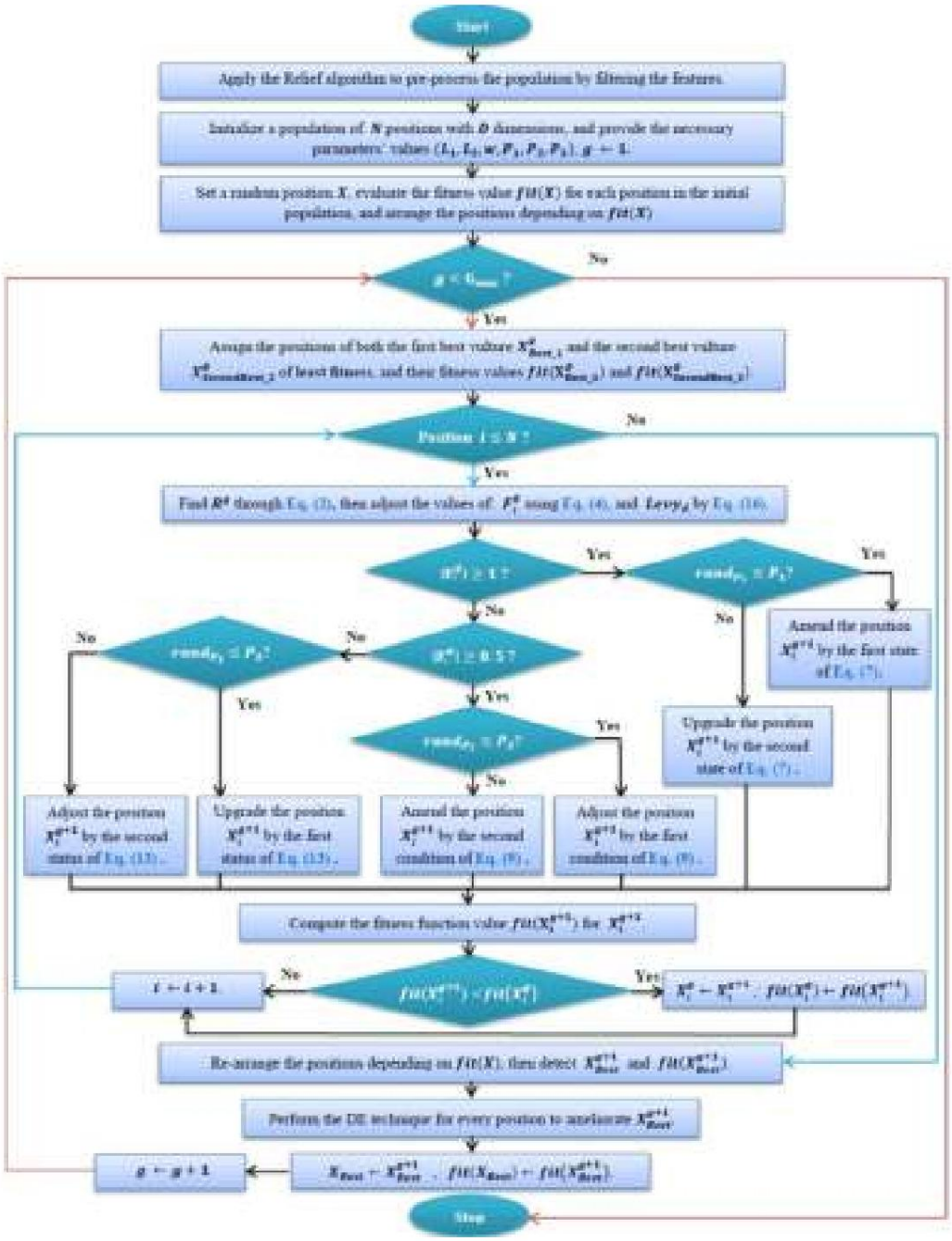


Figure 1. Organigramme de l'algorithme RBAVO-DE.

Machine Translated by Google

Algorithme 1 L'algorithme RBAVO-DE proposé.

Saisir:
N – nombre total de postes (taille de la population)
Gmax : nombre maximum d'itérations autorisées
D—espace dimensionnel du problème
LB – limites inférieures des variables
UB – limites supérieures des variables
CR – taux de croisement
WM – facteur de pondération des mutations
Sortir:
XBest : la meilleure position mondiale du vautour trouvée lors d'une recherche
f it(XBest) : la meilleure valeur globale de fonction de condition physique trouvée, qui doit être diminuée
1 : Démarrer
2 : Appliquer l'approche Relief pour sélectionner les fonctionnalités associées et les filtrer, comme démontré dans la section 3.1 ;
3 : Initialiser une population de N positions, et fournir les valeurs des paramètres nécessaires (L1, L2, w, P1, P2 et P3) ;
4 : Définir une position aléatoire X dans la population initiale ;
5 : Évaluer les valeurs de fitness f it(X) pour chaque position dans la population initiale ;
6 : Organiser les positions par ordre croissant en fonction de leur fonction fitness f it(X) ;
7 : g ← 1 ; 8 : Nombre actuel d'itérations
tandis que g < Gmax faire
9 : Attribuez les positions du premier meilleur vautour X $^{g}_{Meilleur1}$ et le deuxième meilleur vautour X $^{g}_{DeuxièmeMeilleur2}$, ainsi que leur forme physique valeurs f it (X $^{g}_{Meilleur1}$) et l'ajuster (X $^{g}_{DeuxièmeMeilleur2}$), parmi toutes les positions de la population ;
pour la position du vautour i = 1 : N do
Trouvez le meilleur vautour R g à l'aide de l'équation (2) ;
Ajuster le niveau de famine du vautour F $^{g}_{i}$ en utilisant l'équation (4) ;
Modifier la fonction de répartition des vols de prélèvement Levyd à l'aide de l'équation (16) ;
si |F $^{g}_{i}$ | ≥ 1 alors
10 : 11 : 12 : 13 : 14 : 15 : si randP1 ≤ P1 alors
Modifier la position du vautour X $^{g+1}_{i}$ sur la base de la première étape de la phase d'exploration utilisant l'équation (7) ;
16 : 17 : sinon si randP1 > P1 alors
Améliorez la position du vautour X $^{g+1}_{i}$ sur la base de la deuxième étape de la phase d'exploration utilisant l'équation (7) ;
finir si
sinon si |F $^{g}_{i}$ | < 1 alors
si |F $^{g}_{i}$ | ≥ 0,5 alors
18 : 19 : 20 : 21 : 22 : si randP2 ≤ P2 alors
23 : Ajuster la position du vautour X $^{g+1}_{i}$ basé sur la première condition de la première sous-phase d'exploitation en utilisant Équation (9);
24 : sinon si randP2 > P2 alors
25 : Modifier la position du vautour X $^{g+1}_{i}$ basé sur la deuxième condition de la première sous-phase d'exploitation en utilisant l'équation (9) ;
26 : fin si
sinon si |F $^{g}_{i}$ | < 0,5 alors
27 : 28 : si randP3 ≤ P3 alors
29 : Améliorez la position du vautour X $^{g+1}_{i}$ en fonction du premier statut de la deuxième sous-phase d'exploitation en utilisant Équation (13);
30 : sinon si randP3 > P3 alors
31 : Ajuster la position du vautour X $^{g+1}_{i}$ sur la base du deuxième statut de la deuxième sous-phase d'exploitation en utilisant Équation (13);
finir si
finir si
finir si
32 : 33 : 34 : 35 : ça va (X $^{g+1}_{i}$) ← Calculer la valeur de la fonction fitness pour X $^{g+1}_{i}$;
36 : si c'est le cas (X $^{g+1}_{i}$) < s'adapter (X $^{g}_{i}$) alors
37 : X $^{g}_{i}$ ← X $^{g+1}_{i}$;
ça va (X $^{g}_{i}$) ← s'adapter (X $^{g+1}_{i}$);
finir si
fin pour
38 : 39 : 40 : 41 Réorganiser les positions par ordre croissant en fonction de leur fonction de fitness f it(X) ;
Détecter la meilleure position globale X $^{g+1}_{Meilleur}$ et sa meilleure valeur de remise en forme au monde ($^{g+1}$Best) lorsque l'itération en cours g + 1 est terminée ;
42 : 43 : Effectuez la technique DE pour chaque position pour améliorer X $^{g+1}_{Meilleur}$, comme indiqué à la section 3.6 ;
44 : XMeilleur ← X$^{g+1}_{Meilleur}$;
45 : f it (XBest) ← f it (X $^{g+1}_{Meilleur}$);
46 : g ← g + 1 ;
47 : fin pendant
48 : Fin

4. Résultats expérimentaux et discussion

Cette section présente les résultats empiriques du projet RBAVO-DE proposé et ses homologues : colonie d'abeilles artificielles binaires (BABC) [45], algorithme d'essaim de salpes binaires (BSSA) [46], PSO binaire (BPSO) [47], algorithme binaire de chauve-souris (BBA) [48], loup gris binaire Optimisation (BGWO) [49], Algorithme d'optimisation Binary Grasshopper (BGOA) [50], Algorithme d'optimisation de baleine binaire (BWOA) [51], ASO binaire (BASO) [52], oiseau binaire Algorithme d'essaim (BBSA) [53], HGSO binaire (BHGSO) [54] et binaire Harris Hawks Op-

Machine Translated by Google

timisation (BHHO) [55]. Les optimiseurs sont évalués par le biais de formations et de tests des critères de référence, avec des résultats concluants dérivés des valeurs moyennes des paramètres d'évaluation . Les critères utilisés pour évaluer les performances du modèle proposé sont détaillé à la section 4.1. Les paramètres utilisés dans les environnements opérationnels sont décrits à la section 4.2. Les paramètres utilisés pour l'évaluation sont décrits à la section 4.3. L'analyse de les résultats expérimentaux sont discutés à la section 4.4.

4.1. Description de l'ensemble de données

Des approches expérimentales approfondies et divers algorithmes wrapper ont été appliqués à vingt-deux ensembles de données de descriptions de gènes ; les données comprennent un RNA-Seq normalisé de niveau 3 données d'expression génique pour vingt-deux types de tumeurs du Broad Institute. Ces données sont accessibles au public et peuvent être trouvés dans [56]. Nous avons respecté la méthodologie décrite dans [36] et des écarts observés entre les données référencées depuis GitHub dans l'article et les chiffres rapportés dans le document, qui proviennent du site Web. Le les données du site Web comprenaient un mélange d'échantillons de tumeurs et d'échantillons standards, alors que l'article traitait les données uniformément sous forme d'échantillons de tumeurs. Nous avons donc procédé à un examen détaillé des données. Initialement, le site Web proposait différents formats de l'ensemble de données identique à celui que nous souhaitions à analyser. En examinant les données, nous avons rencontré les défis suivants :

- Certains gènes sont identifiés par ID mais n'ont pas de symbole associé.
- Certains gènes sont absents du fichier d'annotation.
- Il existe une confusion des échantillons, y compris les types normaux et tumoraux.

Par conséquent, un prétraitement était nécessaire pour séparer et identifier les échantillons, distinguer les échantillons normaux et tumoraux à utiliser dans la classification binaire et rationaliser le processus GS. Nous avons relevé les défis ci-dessus de la manière suivante :

- Nous avons recherché le symbole du gène correspondant pour chaque ID trouvé dans le fichier d'annotation.
- Après croisement avec le fichier d'annotation, plus d'une centaine de gènes ont été éliminés.
- Sur la base de l'exemple de rapport, chaque ligne de la feuille Excel a été organisée selon le sorte d'échantillon pour la classification binaire.

De plus, l'approche Relief, telle que détaillée dans la section 3.1, a été utilisée lors de l' étape de pré-traitement pour calculer le poids de chaque gène dans le référentiel. Ces poids ont ensuite été classés du plus élevé au plus bas. Les gènes avec des poids inférieurs étaient alors supprimé. L'approche Relief s'est avérée utile pour éliminer les gènes qui ne contribuent pas au classement.

Après la phase de prétraitement, les données ont été affinées et prêtes pour le processus GS. Contrairement à l'approche utilisée dans [36], qui abordait la multi-classification de tous les types de ensemble, nous avons choisi d'analyser chaque type de cancer individuellement pour une plus grande spécificité. Le tableau 1 [57] présente un aperçu complet des vingt-deux types de tumeurs et de leurs comptes d'échantillons respectifs. Le nombre de fonctionnalités utilisées dans les ensembles de données de référence est 32 000 fonctionnalités (gènes).

Tableau 1. Ensembles de données utilisés dans cet article [57].

# Cancer	Normale Enregistrements	Tumeur Enregistrements
1 Carcinome urothélial de la vessie (BLCA)	19	408
2 Cancer de la thyroïde (THCA)	59	501
3 cancers du col de l'utérus et endocervicaux (CESC)	3	304
4 Cholangiocarcinome (CHOL)	9	36
5 Adénocarcinome du côlon (COAD)	41	458
6 Cancer de l'œsophage (ESCA)	13	184
7 Glioblastome Multiforme (GBM)	5	153

Machine Translated by Google

Tableau 1. Suite

# Cancer	Normale Enregistrements	Tumeur Enregistrements
8 Thymome (THYM)	2	120
9 Carcinome épidermoïde de la tête et du cou (HNSC)	44	520
dixChromophobe rénal (KICH)	25	66
11KIRC	72	533
12 Rein Rénal Papillaire (KIRP)	32	290
13 Carcinome hépatocellulaire du foie (LIHC)	50	371
14LUAD	59	515
15 LUSC	51	501
16 Adénocarcinome pancréatique (PAAD)	4	178
17 UCEC	35	176
18 Phéochromocytome et Paragangliome (PCPG)	3	179
19 Adénocarcinome du rectum (LIRE)	10	94
20 Sarcome (SARC)	2	259
21 Mélanome cutané (SKCM)	1	103
22 STADE	37	415

4.2. Paramétrage

L'algorithme RBAVO-DE proposé a été comparé à des variantes binaires de différents des optimiseurs méta-heuristiques, tels que BABC, BSSA, BPSO, BBA, BGWO, BWOA, BBSA, BGOA, BHHO, BASO et BHGSO. Les paramètres critiques pour les modèles ML utilisés dans ce document l'étude sont détaillées dans le tableau 2.

Tableau 2. Les paramètres clés des modèles ML utilisés.

Modèle	Paramètres
SVM	Noyau polynomial = 2
k-NN	Métrique de distance euclidienne k = 5 [21,58,59]

Dans le cadre proposé, la validité des résultats a été vérifiée à l'aide d'un facteur 10 approche de validation croisée pour garantir la fiabilité des résultats. Cela impliquait au hasard diviser chaque ensemble de données en deux sous-ensembles distincts : 80 % pour la formation et les 20 % restants pour les tests. La partie formation a été utilisée pour former les classificateurs présentés en utilisant techniques d'optimisation, tandis que la partie test a été utilisée pour évaluer les gènes sélectionnés efficacité. Les paramètres les plus couramment appliqués à tous les algorithmes comparés sont résumé dans le tableau 3. Pour exécuter toutes les expériences de cet article, Python a été utilisé dans un environnement informatique avec un processeur Dual Intel® Xeon® Gold 5115 2,4 GHz et 128 Go de RAM sur le système d'exploitation Microsoft Windows Server 2019.

Tableau 3. Configurations de paramètres de tous les optimiseurs utilisés.

Optimiseur	Paramètres
Tous les optimiseurs	Nombre d'exécutions = 30
	Nombre maximum d'itérations Gmax = 100
	Taille de la population N = 10

Machine Translated by Google

Tableau 3. Suite

Optimiseur	Paramètres
RBAVO-DE proposé	Dimensionnalité D = Le nombre de gènes dans les ensembles de données utilisés Limites supérieures UB Limites inférieures LB Paramètre L1 = 0,7 Paramètre L2 = 0,2 Paramètre w = 2
BSSA	Paramètre P1 = 0,6 Paramètre P2 = 0,6 Paramètre P3 = 0,5 Nombre de voleurs SD = 0,1 × N Nombre de producteurs PD = 0,2 × N Seuil de sécurité ST = 0,8 Nombre
BABC	d'abeilles employées = 16 Nombre d'abeilles éclaireuses = 3 Nombre d'abeilles observatrices =
BPSO	4 Masse d'inertie (ωmax = 0,9ωmin = 0,4) Coefficients d'accélération (c2 = c1 = 1,2)
BBA	Intensité sonore A = 0,8 Taux d'émission d'impulsions r = 0,95
BWOA	Fréquences d'impulsion inférieure et supérieure = 0, 10 a est linéairement diminuée de 2 à 0 p = 0,5 b = 1,0
BHHO	Énergie du lapin E [−1, 1] a est linéairement diminuée
BGWO BGOA BBSA	de 2 à 0 Cmin = 0,00004 et Cmax = 1 Fréquence de vol ff = 10 Effet sur la vigilance des oiseaux Coefficient suivi fl = 0,5 Probabilité de recherche de nourriture p = 0,8 Coefficients d'accélération (c1 = c2 = 1,5) comportements (a1 = a2 = 1,0)
BASO	Poids de profondeur α = 50 Poids multiplicateur β = 0,2
BHGSO	Nombre de grappes = 2 α = β = 0,1 K = 1,0 et l3 = 1E − 02 l1 = 5E − 03, l2 = 1E + 02

4.3. Critères d'évaluation

Pour évaluer l'efficacité du RBAVO-DE proposé par rapport à d'autres méthodes, chaque stratégie a été testée indépendamment trente fois sur chaque référence pour garantir la validation statistique des résultats. À cette fin, les mesures de performance établies suivantes pour les problèmes GS ont été utilisées :

- Précision moyenne de la classification (MeanAC) : cette mesure représente la précision de la classification correcte des données, calculée en exécutant l'algorithme indépendamment trente fois. Il est déterminé de la manière suivante :

$$\text{MoyenneAC} = \sum_{r=1}^{30} \sum^{m} \text{correspondance}(PLr, ALr) \quad (23)$$

Dans cette formule, m désigne le nombre d'échantillons dans l'ensemble de données de test, tandis que PLr et ALr désignent respectivement les étiquettes prédites du classificateur et les étiquettes de classe réelles pour l'échantillon r . La fonction match(PLr , ALr) sert de mécanisme de comparaison, où si PLr est égal à ALr , alors match(PLr , ALr) se voit attribuer la valeur 1 ; s'ils ne correspondent pas, la valeur 0 lui est attribuée.

Machine Translated by Google

• Condition physique moyenne (MeanFi) : cette mesure calcule la moyenne Fi, obtenue en exécutant l'approche proposée trente fois de manière indépendante. Il évalue l'équilibre entre la réduction du taux d'erreur de classification et la minimisation du nombre de gènes sélectionnés. Une valeur inférieure indique une solution supérieure, évaluée en fonction de l'adéquation :

$$\text{MoyenneFi} = 30 \frac{1}{k=1} \sum^{30} f^{k}_{*} \tag{24}$$

où f^{k}_{*} montre la valeur de condition physique optimale obtenue lors de la k − ème exécution.

• Nombre moyen de gènes choisis (MeanGe) : Cette métrique calcule le nombre moyen de gènes choisis (ou ratio GS) en exécutant l'approche proposée trente fois de manière indépendante, et elle est définie comme suit :

$$\text{MoyenneFe} = 30 \frac{1}{k=1} \sum^{30} \frac{|d^{k}|}{|D|} \tag{25}$$

Ici, $|d^{k}|$ désigne le nombre total de gènes choisis dans la solution optimale pour le k ème exécution et |D| représente le nombre total de gènes présents dans le benchmark initial. • Écart type (STDE) : sur la base des résultats ci-dessus, la stabilité des résultats moyens globaux dérivés de trente exécutions distinctes de chaque optimiseur sur chaque test de référence a été évaluée de la manière suivante :

$$\text{STDE} = \sqrt{\frac{1}{29} \sum_{k=1}^{30} (Yk - \text{MoyenneY})^2} \tag{26}$$

où Y montre la mesure à utiliser, Y k désigne la valeur de mesure Y dans la kème exécution et MeanY est la moyenne de la mesure de trente exécutions indépendantes.

Les données des tableaux suivants représentent les valeurs moyennes obtenues à partir de trente exécutions indépendantes, en mettant l'accent sur l'exactitude de la classification, le nombre de gènes choisis, la condition physique moyenne, la précision, le rappel et le score F1. Les sous-sections suivantes examinent et discutent en profondeur ces résultats expérimentaux, avec des chiffres en gras mettant en évidence les résultats optimaux.

4.4. Résultats de la comparaison du RBAVO-DE proposé avec divers algorithmes méta-heuristiques

L'approche RBAVO-DE proposée avec les modèles SVM et k-NN a été comparée à d'autres méthodes méta-heuristiques récentes exécutées dans les mêmes conditions pour montrer sa supériorité sur ses homologues. L'algorithme RBAVO-DE proposé a été comparé aux variantes binaires de plusieurs optimiseurs, notamment BABC, BSSA, BPSO, BBA, BGWO, BWOA, BBSA, BGOA, BHHO, BASO et BHGSO.

4.4.1. Résultats utilisant le modèle k-NN

Le tableau 4 montre les résultats de l'algorithme RBAVO-DE proposé et d'autres techniques d'optimisation utilisant le modèle k-NN, basés sur des mesures de précision comparées dans des conditions similaires. Les résultats expérimentaux révèlent que le RBAVO-DE proposé a obtenu les résultats les plus prometteurs dans quatre critères. Il convient de noter que toutes les méthodes concurrentes, y compris la proposition RBAVO-DE utilisant le modèle k-NN, ont produit des résultats comparables sur dix-huit critères.

Le reste des métriques du modèle k-NN sont présentés à l'annexe A.1. En termes de valeurs moyennes de fitness, l'algorithme RBAVO-DE proposé a démontré une efficacité supérieure à celle de ses homologues utilisant le modèle k-NN dans des conditions similaires.

L'algorithme RBAVO-DE a donné les résultats de fitness les plus bas et le STDE le plus compétitif parmi tous les benchmarks. En outre, tous les benchmarks utilisés sont de grande dimension, démontrant que le RBAVO-DE proposé peut fonctionner efficacement sur tous les benchmarks, quelle que soit leur taille. L' algorithme RBAVO-DE s'avère prometteur en équilibrant efficacement l'exploitation et l'exploration de l'espace de recherche, évitant ainsi de se retrouver piégé dans des optima locaux lors des itérations. Contrairement à

Machine Translated by Google

de nombreux autres algorithmes qui ont tendance à se laisser piéger, il a démontré leur capacité à s'échapper de tels pièges.

En ce qui concerne les résultats moyens pour les gènes sélectionnés par l'algorithme RBAVO-DE proposé et ses homologues utilisant k-NN, l'algorithme RBAVO-DE proposé a surpassé les autres méthodes sur tous les benchmarks concernant le nombre de gènes choisis. De plus, la capacité du RBAVO-DE à identifier des gènes significatifs est due à son efficacité exploration des zones possibles tout en améliorant la précision.

En termes de précision moyenne de l'algorithme RBAVO-DE proposé et de ses homologues avec k-NN, l'algorithme RBAVO-DE proposé a surpassé les approches alternatives pour trois des vingt-deux ensembles de données. Pour dix-neuf ensembles de données, BWOA a obtenu des résultats similaires, tandis que BABC, BPSO, BGWO, BGOA et BHHO ont obtenu des résultats similaires à ceux du proposé RBAVO-DE. BASO et BHGSO, qui ont produit les mêmes résultats que le a proposé RBAVO-DE sur dix-sept ensembles de données, classé quatrième. En fin de compte, l'approche BBA s'est classée au dernier rang parmi toutes les approches, produisant des résultats similaires à ceux de l'approche RBAVO-DE proposée pour les quinze jeux de données.

En ce qui concerne le rappel moyen du RBAVO-DE proposé et de ses homologues utilisant k -NN, RBAVO-DE a surpassé les approches alternatives pour trois des vingt- deux ensembles de données. D'autre part, BSSA, BABC, BPSO, BGWO, BWOA, BGOA et BBSA a produit les mêmes résultats que RBAVO-DE pour 19 ensembles de données, tandis que BHHO a obtenu des résultats similaires pour 18 ensembles de données. BHGSO, qui a obtenu les mêmes résultats que le projet proposé RBAVO-DE pour dix-sept ensembles de données, classé quatrième. Au final, BASO et BBA se sont classés la plus basse de toutes les approches, produisant des résultats similaires à ceux de RBAVO-DE pour seize ensembles de données.

En ce qui concerne le score F1 moyen du RBAVO-DE proposé et de ses homologues en utilisant k-NN, le RBAVO-DE proposé a surpassé les approches alternatives pour trois des vingt-deux ensembles de données. En revanche, pour dix-huit jeux de données, BSSA, BABC, BPSO, BGWO, BGOA et BBSA ont produit les mêmes résultats que RBAVO-DE, tandis que BWOA et BHHO a produit des résultats comparables pour dix-sept ensembles de données. BASO et BHGSO, qui a obtenu les mêmes résultats pour 16 ensembles de données que le RBAVO-DE proposé, classé quatrième. En fin de compte, l'approche BBA s'est classée au dernier rang parmi toutes les approches, produisant des résultats similaires à ceux de l'approche RBAVO-DE proposé pour quatorze ensembles de données.

Tableau 4. Les résultats de l'algorithme RBAVO-DE proposé et de ses homologues utilisant k-NN concernant résultats moyens d'exactitude de classification.

Ensemble de données	Métrique	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA	Moyenne	0,998	0,993	0,998	0,993	0,989	0,995	0,994	0,993	0,991	0,994	0,994	0,989
	STDE	0,004	0,006	0,005	0,006	0,003	0,006	0,006	0,006	0,005	0,006	0,006	0,002
CESC	Moyenne	0,984	0,984	0,984	0,984	0,984	0,984	0,984	0,984	0,984	0,984	0,984	0,984
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
CHOL	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
COAD	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
ESCA	Moyenne	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
GBM	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
HNSC	Moyenne	0,993	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974
	STDE	0,004	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
KICH	Moyenne	1 .000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0.000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
KIRC	Moyenne	0.998	0,991	0,991	0,991	0,991	0,992	0,991	0,991	0,991	0,991	0,991	0,991
	STDE	0.003	0,002	0,002	0,000	0,000	0,002	0,000	0,000	0,000	0,002	0,000	0,000
KIRP	Moyenne	1.000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0.000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
LIHC	Moyenne	1.000	1. 000	1. 000	1. 000	1. 000	1. 000	1. 000	1. 000	1. 000	1. 000	1. 000	1.000
	STDE	0.000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
LUAD	Moyenne	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
PAAD	Moyenne	1,000	1.000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
STDE	LUSC Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
STDE	Moyenne	0,993	0,994	0,992	0,992	0,994	0,992	0,994	0,993	0,993	0,993	0,992	
	STDE	0,003	0,004	0,002	0,000	0,004	0,002	0,004	0,003	0,003	0,003	0,000	
	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
	Moyenne	1. 000	1. 000	1. 000	1. 000	1. 000	1. 000	1. 000	1. 000	1. 000	1. 000	1.000	
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
	Moyenne	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991	
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
	Moyenne	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
	Moyenne	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	

Machine Translated by Google

Tableau 4. Suite

Ensemble de données	Métrique	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
PCPG	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
READ	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SARC	Moyenne	0,981	0,981	0,981	0,981	0,981	0,981	0,981	0,981	0,981	0,981	0,981	0,981
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SKCM	Moyenne	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	STDE	0,999	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
STADE	STDE	0,003	0,996	0,999	0,996	0,989	0,996	0,995	0,997	0,992	0,995	0,996	0,989
THCA	Moyenne	1,000	0,005	0,004	0,005	0,000	0,005	0,006	0,005	0,005	0,006	0,005	0,000
THYM	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
UCEC	Moyenne	0,998	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,008	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
			0,985	0,988	0,977	0,969	0,984	0,977	0,978	0,978	0,979	0,976	0,968
			0,012	0,013	0,006	0,011	0,012	0,006	0,008	0,009	0,009	0,005	0,011
Classement W|T|L		4|18|0	0|18|4	0|18|4	0|18|4	Les valeurs en gras		0|18|4	0|18|4	0|18|4	0|18|4	0|18|4	0|18|4	0|18|4

indiquent les meilleurs résultats.

4.4.2. Résultats utilisant le modèle SVM

Le tableau 5 présente les résultats du RBAVO-DE proposé et de diverses optimisations. techniques utilisant le modèle SVM concernant les résultats de précision de la classification évalués dans des conditions de fonctionnement identiques. Les résultats expérimentaux démontrent que L'algorithme RBAVO-DE proposé a surpassé les autres approches en obtenant le plus valeurs prometteuses pour quatre ensembles de données. Toutes les méthodes concurrentielles, y compris la méthode proposée RBAVO-DE employant SVM, a obtenu des valeurs équivalentes sur 18 benchmarks.

Le reste des métriques utilisant SVM sont présentés à l'Annexe A.2. Concernant le valeurs moyennes de fitness du RBAVO-DE proposé et de ses homologues utilisant le SVM modèle dans des conditions de fonctionnement équivalentes, le RBAVO-DE proposé s'est avéré plus efficace que les autres méthodes. Le RBAVO-DE proposé utilisant le modèle SVM a donné les valeurs de condition physique les plus basses et le STDE le plus compétitif sur tous les points de référence. De plus, tous les benchmarks utilisés sont de grande dimension, démontrant que le modèle proposé RBAVO-DE peut fonctionner efficacement sur tous les benchmarks, quelle que soit leur taille. Le RBAVO-DE l'algorithme est prometteur en équilibrant efficacement l'exploitation et l'exploration de la recherche l'espace, en évitant de se retrouver piégé dans des optima locaux lors des itérations. Contrairement à beaucoup d'autres algorithmes qui ont tendance à se laisser piéger, il a démontré la capacité d'échapper à de tels pièges.

Concernant les résultats moyens pour les gènes sélectionnés par le RBAVO-DE proposé et ses homologues employant SVM, le RBAVO-DE proposé a obtenu des résultats plus prometteurs que les autres techniques dans tous les critères utilisés dans cette étude. Aussi, la supériorité du RBAVO-DE proposé utilisant SVM dans ce contexte démontre sa capacité pour explorer efficacement les régions précieuses de l'espace de recherche tout en évitant les régions avec des solutions non réalisables.

En ce qui concerne les valeurs de précision moyennes du RBAVO-DE proposé et de ses homologues employant SVM, le RBAVO-DE proposé a surpassé l'alternative approches pour trois des vingt-deux ensembles de données. En revanche, pour dix-huit jeux de données, BABC, BGWO, BWOA, BBSA et BASO ont donné des résultats comparables. Pour dix-sept ensembles de données, BSSA, BPSO, BGOA et BHHO ont donné des résultats similaires à ceux de RBAVO-DE, se classant quatrième. En fin de compte, BBA s'est classée au dernier rang parmi toutes les approches, produisant des résultats similaires. résultats à ceux du RBAVO-DE proposé pour douze ensembles de données.

En termes de rappel moyen de la méthode RBAVO-DE proposée et de ses homologues en utilisant SVM, le RBAVO-DE proposé a surpassé les approches alternatives pour trois des les vingt-deux ensembles de données. En revanche, BPSO et BHHO ont réalisé des performances similaires pendant dix-huit ensembles de données, tandis que BSSA, BABC, BGWO, BWOA, BGOA et BBSA ont produit les mêmes résultats comme le RBAVO-DE proposé pour dix-neuf ensembles de données. BHGSO, qui a obtenu le même résultats comme RBAVO-DE pour dix-sept ensembles de données, classés quatrièmes. Au final, BASO et BBA classé au dernier rang parmi toutes les approches, produisant des résultats similaires à ceux de RBAVO-DE pour seize ensembles de données.

Machine Translated by Google

En ce qui concerne le score F1 moyen de l'algorithme RBAVO-DE proposé et de ses homologues utilisant SVM, l'algorithme RBAVO-DE proposé a surpassé les approches alternatives. pour cinq des vingt-deux ensembles de données. D'un autre côté, BWOA a produit des résultats similaires à ceux du RBAVO-DE proposé pour dix-sept ensembles de données, tandis que BSSA, BABC, BPSO, BGWO, BGOA et BBSA ont produit des résultats comparables pour seize ensembles de données. BHHO, qui a obtenu les mêmes résultats que le RBAVO-DE proposé pour quinze ensembles de données, classé quatrième. En fin de compte, BBA s'est classée au dernier rang parmi toutes les approches, produisant des résultats similaires à ceux de RBAVO-DE pour 12 jeux de données.

Tableau 5. Les résultats de l'algorithme RBAVO-DE proposé et de ses homologues utilisant SVM concernant les valeurs moyennes de précision de classification.

Ensemble de données	Métrique	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA	Moyenne	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
CESC	Moyenne	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
CHOL	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
COAD	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
GBM	Moyenne	0,986	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,013	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
ESCA	Moyenne	1,000	0,975	0,977	0,974	0,974	0,977	0,974	0,975	0,975	0,975	0,974	0,974
	STDE	0,000	0,004	0,008	0,000	0,000	0,008	0,000	0,005	0,005	0,005	0,000	0,000
HNSC	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
KICH	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
LUAD	Moyenne	0,993	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,003	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
KIRC	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
PAAD	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
KIRP	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
PCPG	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
LIHC	Moyenne	1,000	1,000	1,000	0,9992	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
READ	Moyenne	1,000	0,000	0,000	0,000	0,0029	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	STDE		0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,992	0,991
			0. 000	0. 000	0. 000	0,000	0. 000	0. 000	0. 000	0. 000	0. 000	0. 002	0,000
LUSC			1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
			0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
			1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
			0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
			1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SARC	Moyenne	0,994	0,986	0,989	0,986	0,981	0,989	0,987	0,986	0,984	0,986	0,984	0,981
	STDE	0,008	0,008	0,009	0,008	0,000	0,009	0,009	0,008	0,006	0,008	0,007	0,000
SKCM	Moyenne	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
STDE	Moyenne	0,990	0,989	0,989	0,989	0,989	0,989	0,989	0,989	0,989	0,989	0,989	0,989
	STDE	0,003	0,000	0,000	0,000	0,002	0,000	0,000	0,000	0,002	0,000	0,000	0,000
THCA	Moyenne	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
STADE	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
THYM	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
UCEC	Moyenne	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
	STDE	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
			1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
			0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
			1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
			0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Classement W|T|L		4|18|0	0|18|4	0|18|4	0|18|4	0|18|4	0|18|4	Les chiffres en gras indiquent les meilleurs résultats.		0|18|4	0|18|4	0|18|4	0|18|4	0|18|4

4.5. Analyse de convergence

Il est évident dans l'Annexe B [57] que le RBAVO-DE proposé utilisant le SVM et les modèles k-NN ont atteint un comportement de convergence optimal sur tous les benchmarks. Donc, les performances de convergence du RBAVO-DE proposé utilisant le SVM et le k-NN modèles démontre sa capacité à atteindre rapidement des résultats optimaux tout en maintenant un équilibre efficace entre recherche et exploitation.

Machine Translated by Google

4.6. Test de somme de rangs de Wilcoxon

Cet article compare les valeurs de fitness obtenues par le RBAVO-DE proposé et ses homologues par paires à l'aide du test de somme de rangs de Wilcoxon [60], qui vise à déterminer s'il existe une différence statistiquement significative entre les différentes approches. Ce test est un outil crucial pour évaluer le succès de l'algorithme proposé. Le test de Wilcoxon est utilisé dans les tests d'hypothèses pour comparer des données appariées. Cela implique de trier les différences absolues dans les résultats de procédures appariées sur les $j^{ème}$ des N questions. Ensuite, la valeur la plus basse est trouvée en additionnant les classements positifs (R +) et né
L'hypothèse nulle est rejetée si le niveau de signification résultant est inférieur à 5 %, et n'est pas rejetée dans le cas contraire.

En examinant les résultats de la comparaison effectuée entre le RBAVO-DE et ses homologues par paires à l'aide du test de Wilcoxon, nous avons constaté que toutes les valeurs R + sont égales à 253, toutes les valeurs R − sont égales à 0 et toutes les valeurs p sont égales à 4,768 × 10−7. qui sont inférieurs à 0,05 (niveau de signification de 5 %). Par conséquent, on peut conclure que le RBAVO-DE proposé utilisant SVM et k-NN fonctionne mieux que tout autre algorithme dans tous les cas. Par conséquent, toutes les valeurs p inférieures à 0,05 (niveau de signification de 5 %) offrent des preuves irréfutables que les résultats de la stratégie proposée sont statistiquement significatifs et ne sont pas une simple coïncidence.

4.7. Complexité informatique de l'algorithme RBAVO-DE et de divers algorithmes méta-heuristiques 4.7.1. Complexité informatique du temps d'exécution de l'algorithme RBAVO-DE

Chacune des cinq étapes principales de RBAVO-DE peut être analysée séparément pour déterminer sa complexité informatique. Ces étapes comprennent les fonctionnalités de filtrage, l'initialisation de la population, l'amélioration et la modification de la position, l'évaluation de la fonction de remise en forme et l'utilisation de DE. Ensuite, Oexecute_time(RBAVO − DE) peut être utilisé pour représenter la complexité informatique globale de l'algorithme RBAVO-DE proposé. Cela peut être calculé à l'aide des formules de notation big-O suivantes :

$$\begin{aligned} \text{Oexecute_time (RBAVO} - \text{DE)} = \text{Oexecute_time (}& \text{Fonctionnalités de filtrage)} + \\ & \text{Oexecute_time (Initialisation de la population)} \\ & + \text{Oexecute_time (Amélioration et modification de la position)} \\ & + \text{Oexecute_time (Évaluation de la fonction de fitness)} \\ & + \text{Oexecute_time (Incorporation de la technique DE).} \end{aligned} \quad (27)$$

Puisque N détermine la taille de la population, Gmax est le nombre maximum d'itérations autorisé, et D est l'espace dimensionnel du problème. Par conséquent,

$$\text{Oexecute_time (}\text{Fonctionnalités de filtrage)} = \text{Oexecute_time (D).}$$
$$\text{Oexecute_time (Initialisation de la population)} = \text{Oexecute_time (N).}$$
$$\text{Oexecute_time (}\text{Position de renforcement et de modification)} = \text{Oexecute_time (Gmax} \times \text{N} \times \text{D).}$$
$$\text{Oexecute_time (}\text{fonction d'évaluation de la condition physique)} = \text{Oexecute_time (Gmax} \times \text{N).}$$
$$\text{Oexecute_time (Incorporant la technique DE)} = \text{Oexecute_time (N} \times \text{D).}$$

Pour cela,

$$\begin{aligned} \text{Oexecute_time(RBAVO} - \text{DE)} &= \text{Oexecute_time(D)} + \\ &\text{Oexecute_time(N)} + \text{Oexecute_time(Gmax} \times \text{N} \times \text{D)} + \\ &\text{Oexecute_time(Gmax} \times \text{N)} + \text{Oexecute_time(N} \times \text{D)} = \\ &\text{Oexecute_time(Gmax} \times \text{N} \times \text{D).} \end{aligned}$$

Machine Translated by Google

4.7.2. Complexité informatique de l'utilisation de la mémoire de l'algorithme RBAVO-DE

Cela implique de mesurer la quantité de mémoire requise par un algorithme pour résoudre un problème à mesure que la quantité d'entrée augmente. Elle est fréquemment exprimée comme la mémoire supplémentaire nécessaire à l'algorithme au-delà de l'entrée. Cela implique la fusion des deux éléments principaux suivants : 1. Utilisation de la mémoire des variables d'entrée : il s'agit de la quantité de mémoire requise pour que l'

algorithme stocke les données d'entrée. Il existe 13 variables d'entrée liées à l' algorithme RBAVO-DE proposé, comme suit : N, Gmax, D, LB, UB, CR, WM, L1, L2, w, P1, P2 et P3. Étant donné que chaque variable stocke des valeurs numériques, 4 octets de mémoire sont utilisés par chacune . Par conséquent, la complexité totale d'utilisation de la mémoire de ces 13 variables d'entrée est de 52 octets (13 × 4 octets = 52 octets). La complexité d'utilisation de la mémoire des valeurs d'entrée est constante.

2. Utilisation de la mémoire supplémentaire : cela indique la quantité de mémoire supplémentaire dont l'algorithme a besoin en plus de l'entrée. Il comprend la mémoire requise pour les structures de données, les variables internes et d'autres composants de l'algorithme. Quelle que soit la taille de l'entrée, l'algorithme RBAVO-DE nécessite une quantité spécifique de mémoire supplémentaire. Ce qui suit les variables sont impliquées:

- La complexité d'utilisation de la mémoire consommée par Xinitial est de (4 × N × D) octets. En effet , chaque position dans le vecteur de positions Xinitial nécessite 4 octets de mémoire et sa taille est (N × D), proportionnelle à la population initiale de N positions de dimension D. Parce que la quantité de mémoire requise augmente linéairement avec la valeur (N × D), sa complexité d'utilisation de la mémoire est linéaire. • Les variables pr, g, Fi , t, R, Di , dt , S1, S2, A1, A2, sigma, f it(Xi), f

  it(XBest1), fit (XSecondBest2), et f it (XBest) nécessitent chacun 4 octets d'espace mémoire car ils ne représentent que des valeurs numériques. En conséquence, la complexité totale d'utilisation de la mémoire de ces 16 variables est de 64 octets (16 × 4 octets = 64 octets). Cette complexité d'utilisation de la mémoire est constante.

- Les vecteurs positions BestVulture1, SecondBestVulture2, Xi , Levyd , μ, σ, X $_{bin}$, ui ,

  Xr1 , Xr2 , Xr3 , ui et XBest nécessitent chacun 4 octets d'espace mémoire et la taille de chaque vecteur est D, proportionnelle à la taille des positions acquises.

  En conséquence, les vecteurs en ces 13 points ont une complexité totale d'utilisation de la mémoire de (52 × D) octets (13 × 4 × D octets). Puisque la mémoire nécessaire augmente linéairement avec la valeur D, sa complexité de consommation de mémoire est linéaire.

Par conséquent, la complexité globale de l'utilisation de la mémoire pour toutes les variables supplémentaires répertoriées ci-dessus est de (4 × N × D) + 64 + (52 × D) octets.

Enfin, la formule suivante peut être utilisée pour déterminer la complexité de calcul de l'utilisation globale de la mémoire de l'algorithme RBAVO-DE proposé :

Complexité d'utilisation de la mémoire (RBAVO-DE) = Valeurs d'entrée utilisation de la mémoire +

Utilisation de la mémoire supplémentaire = 52 + (4 × N × D) + 64 + (52 × D) octets.

Gardez à l'esprit qu'il existe des octets constants qui ne sont pas pris en compte. En notation big-O , la complexité de calcul de la consommation totale de mémoire de RBAVO-DE peut être représentée par Omemory_usage (RBAVO − DE). Cela peut ensuite être calculé en notation big-O après avoir éliminé toutes les constantes de la manière suivante :

Omemory_usage (RBAVO − DE) = Omemory_usage (valeurs d'entrée utilisation de la mémoire)

+ Omemory_usage (utilisation de la mémoire

supplémentaire) = Omemory_usage (1) + Omemory_usage (N × D) + Omemory_usage (1) +

Omemory_usage (D) = Omemory_usage

Machine Translated by Google

Il peut être difficile de compiler une comparaison approfondie de l'utilisation de la mémoire et de la complexité du temps d'exécution de divers algorithmes méta-heuristiques, car ces complexités peuvent différer en fonction de l'implémentation particulière, de la taille du problème et d'autres opérateurs. De plus, tous les algorithmes répertoriés ne disposent pas d'évaluations complètes du temps d'exécution et de la complexité de l'utilisation de la mémoire, et leurs propriétés peuvent varier en fonction du problème à résoudre.

5. Avantages et inconvénients de l'algorithme RBAVO-DE

Cette partie présente une discussion équilibrée des avantages et des inconvénients du RBAVO-Algorithme DE. Les avantages de l'algorithme RBAVO-DE peuvent être répertoriés comme suit :

- Haute précision de classification : L'algorithme RBAVO-DE a démontré une précision de classification élevée, atteignant jusqu'à 100 % dans certains cas. Il s'agit d'une réalisation remarquable, en particulier en ce qui concerne les ensembles de données sur le cancer, dans la mesure où une sélection génétique précise peut avoir un impact direct sur les résultats diagnostiques et thérapeutiques.
- Réduction efficace de la taille des entités : l'algorithme a démontré une capacité remarquable à réduire la taille des entités jusqu'à 98 % tout en maintenant ou en améliorant la précision de la classification. Cette réduction de dimensionnalité est cruciale pour traiter efficacement des ensembles de données de grande dimension .
- Robustesse sur divers ensembles de données : L'efficacité de l'algorithme RBAVO-DE sur vingt-deux ensembles de données sur le cancer indique sa robustesse et son adaptabilité à diverses caractéristiques des données génétiques. Cette polyvalence est bénéfique pour des applications plus larges dans la recherche en génomique.
- Performance supérieure à ses concurrents : comparé aux variantes binaires d'algorithmes méta-heuristiques largement reconnus, RBAVO-DE se distingue par ses performances exceptionnelles en termes de précision et de réduction des caractéristiques, mettant en évidence son approche innovante de la sélection des gènes.

D'un autre côté, les inconvénients de la méthodologie proposée peuvent être répertoriés comme suit :

- Complexité informatique : La haute dimensionnalité des ensembles de données et la nature itérative des algorithmes méta-heuristiques suggèrent que RBAVO-DE peut avoir un coût de calcul important. Cet aspect pourrait limiter son applicabilité dans des environnements aux ressources informatiques limitées.
- Problèmes de généralisabilité : malgré son succès dans divers ensembles de données sur le cancer, la généralisabilité de l'algorithme à d'autres types de données biologiques ou de maladies reste à faire l'objet d'études approfondies. Il est crucial de tester RBAVO-DE dans des contextes plus larges pour confirmer son applicabilité au-delà des ensembles de données examinés.
- Sensibilité des paramètres : comme de nombreux algorithmes méta-heuristiques, RBAVO-DE peut être sensible aux paramètres définis, affectant les performances et l'efficacité. Des études détaillées sur l'impact de différentes configurations de paramètres et stratégies pour leur optimisation pourraient améliorer l'utilité de l'algorithme.

En résumé, l'algorithme RBAVO-DE représente une avancée significative dans la sélection de gènes pour la classification du cancer, avec des atouts notables en termes de précision et d'efficacité. Cependant, remédier à ses inconvénients potentiels grâce à des recherches plus approfondies pourrait élargir son applicabilité et améliorer ses performances, ce qui en ferait un outil encore plus précieux en génomique et en médecine personnalisée.

6. Conclusions et orientations futures

L'approche RBAVO-DE proposée dans cet article a été la première à être mise en œuvre pour gérer les problèmes de GS dans les données d'expression génique RNA-Seq et déterminer des biomarqueurs potentiels pour différentes classes de tumeurs afin d'améliorer la meilleure solution découverte. Les résultats étaient prometteurs, révélant que l'efficacité et les capacités de l'algorithme RBAVO-DE proposé ont été considérablement améliorées. SVM et k-NN, deux modèles de classification largement utilisés, ont été utilisés pour évaluer l'efficacité de chaque ensemble de gènes sélectionnés. La capacité de l'algorithme RBAVO-DE proposé a été comparée à onze variantes binaires largement utilisées.

Machine Translated by Google

techniques méta-heuristiques pour l'évaluer sur différentes classes de tumeurs avec diverses instances. L'évaluation a été réalisée en utilisant une combinaison de mesures d'évaluation, telles que la condition physique moyenne, l'exactitude de la classification, le nombre de gènes sélectionnés, la précision, le rappel et le score F1. L'algorithme RBAVO-DE proposé utilisant les classificateurs SVM et k-NN a obtenu des résultats plus prometteurs que les autres optimiseurs dans la gestion des problèmes GS. Malgré les résultats prometteurs , cette recherche ouvre plusieurs pistes d'exploration future pour faire progresser davantage le domaine :

- Améliorations algorithmiques : amélioration de l'efficacité de l'algorithme et réduction de la complexité des calculs.
- Applicabilité multi-maladies : tester l'algorithme proposé sur des données génétiques provenant de diverses maladies au-delà du cancer.
- Analyses comparatives avec des modèles d'apprentissage profond : évaluation de RBAVO-DE par rapport modèles avancés d'apprentissage profond pour l'analyse des données génétiques.
- Validation clinique en situation réelle : collaborer avec des experts cliniques pour valider l' utilité pratique de gènes sélectionnés dans le traitement du cancer.
- Évolutivité et parallélisation : amélioration de l'évolutivité de l'algorithme grâce au calcul parallèle pour gérer efficacement des ensembles de données génétiques plus importants.
- Applications interdisciplinaires : explorer le potentiel de l'algorithme dans d'autres domaines traitant de données de grande dimension, tels que la modélisation financière et environnementale.

En conclusion, bien que l'algorithme RBAVO-DE représente une avancée significative dans le domaine de la sélection génétique pour la classification du cancer, les voies tracées pour les recherches futures mettent en évidence le potentiel de progrès supplémentaires et d'applications plus larges de ce travail. La collaboration interdisciplinaire et l'innovation continues seront cruciales pour libérer toute la capacité des méthodologies de sélection génétique afin d'améliorer les résultats des soins de santé et de faire progresser notre compréhension des maladies complexes.

Contributions de l'auteur : MGG : Ressources, validation et rédaction – révision et édition. AAA : ressources, analyse formelle, conservation des données, validation, rédaction (ébauche originale) et rédaction (révision et édition). AEE : investigation, visualisation, conservation des données, validation, rédaction (ébauche originale) et rédaction (révision et édition). AAAE-M. : Conceptualisation, méthodologie, logiciels, analyse formelle, enquête, conservation des données, validation, rédaction (ébauche originale) et rédaction (révision et édition). Tous les auteurs ont lu et accepté la version publiée du manu

Financement : L'Université Prince Sattam bin Abdulaziz a financé ce travail de recherche sous le numéro de projet PSAU/2023/01/25612.

Déclaration de disponibilité des données : pour améliorer l'applicabilité et la reproductibilité de cette recherche, le logiciel développé et le code Python pertinent de cette recherche sont accessibles au public et peuvent être obtenus dans [61].

Remerciements : Les auteurs expriment leur gratitude à l'Université Prince Sattam bin Abdulaziz pour avoir apporté son soutien financier à ce projet de recherche.

Conflits d'intérêts : Les auteurs déclarent qu'ils n'ont aucun intérêt financier concurrent connu ni aucune relation personnelle qui aurait pu sembler influencer le travail rapporté dans cet article.

Abréviations

Les abréviations suivantes sont utilisées dans cette recherche :

AVO	Optimisation des vautours africains
CNSS	Carcinome épidermoïde de la tête et du cou
DE	Evolution différentielle
KICH	Rein chromophobe
RBAVO-DE Relief	Optimisation binaire des vautours africains basée sur l'évolution différentielle
KIRP	Rein Rénal Papillaire
GS	Sélection de gènes
LIHC	Carcinome hépatocellulaire du foie
Séquençage d'ARN RNA-Seq	

Machine Translated by Google

PAAD	Adénocarcinome pancréatique
ADN	Acide désoxyribonucléique
PCPG	Phéochromocytome et paragangliome
SVM	Machine à vecteurs de support
LIRE	Adénocarcinome du rectum
k-NN	k-Voisin le plus proche
SARC	Sarcome
FS	Sélection des fonctionnalités
SKCM	Mélanome cutané
MHM	Méthodes méta-heuristiques
MeanFi	Forme physique moyenne
ML	Apprentissage automatique
Moyenne	Taille moyenne des gènes choisis
DL	Apprentissage profond
STDE	Écart type
SA	Recuit simulé
RF	Forêt aléatoire
DÉJOUER	Apprenti inductif de premier ordre
BSG	Optimiseur gravitationnel
SCCSA	Algorithme de recherche sinus-cosinus – coucou
SMO	Optimisation du singe araignée
BHGSO	Optimisation de la solubilité du gaz Henry binaire
BPSO	Optimisation des essaims de particules binaires
BABC	Colonie d'abeilles artificielles binaires
BSSA	Algorithme d'essaim de salpes binaires
BBA	Algorithme binaire de chauve-souris
BGWO	Optimisation binaire Grey-Wolf
BGOA	Algorithme d'optimisation de sauterelle binaire
BWOA	Algorithme d'optimisation de baleine binaire
BASO	Optimisation de la recherche d'atomes binaires
BBSA	Algorithme binaire d'essaim d'oiseaux
BHHO	Optimisation binaire Harris Hawks
BBBO	Optimisation binaire de l'ours brun
BAO	Optimisation binaire Aquila
BMOA	Algorithme d'optimisation binaire suricate
LUAD	Adénocarcinome du poumon
LUSC	Carcinome épidermoïde du poumon
BRCA	Carcinome invasif du sein
KIRC	Carcinome rénal à cellules claires du rein
UCEC	Carcinome de l'endomètre du corps utérin
RPKM	Lectures par kilobase par million de lectures mappées
STADE	Adénocarcinome de l'estomac
SDAE	Encodeur automatique de débruitage empilé
CGA	Atlas du génome du cancer
BLCA	Carcinome urothélial de la vessie
THCA	Cancer de la thyroïde
CESC	Cancers du col de l'utérus et endocervicaux
CHOL	Cholangiocarcinome
COAD	Adénocarcinome du côlon
ESCA	Cancer de l'œsophage
GBM	Glioblastome Multiforme
THYM	Thymome

Machine Translated by Google

Annexe A. Résultats de comparaison du RBAVO-DE proposé et de ses homologues

Annexe A.1. Résultats de comparaison basés sur le modèle k-NN

Tableau A1. Les résultats de l'algorithme RBAVO-DE proposé et de ses homologues utilisant k-NN concernant les valeurs moyennes de condition physique.

Ensemble de données	Métrique	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA	Moyenne	0,00190	STDE	0,00430		0,01110	0,00710	0,01120	0,01470	0,00940	0,01070	0,01140	0,01310	0,01030	0,01080	0,01590
	Moyenne	0,01610	STDE			0,00530	0,00440	0,00540	0,00240	0,00550	0,00540	0,00540	0,00440	0,00550	0,00540	0,00180
CESC		0,00000	CHOL	Moyenne		0,01990	0,02020	0,02050	0,01980	0,02000	0,02020	0,02020	0,01970	0,02020	0,02030	0,02060
		0,00000	STDE			0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
						0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
				0,00000		0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
COAD	Moyenne	0,00000	STDE	Moyenne		0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
	0,02560		0,00000			0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
ESCA		STDE				0,02930	0,02960	0,02990	0,02930	0,02940	0,02970	0,02960	0,02910	0,02960	0,02980	0,03000
			0,00000			0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00020
GBM	Moyenne	0,00000	STDE			0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
			0,00000			0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
HNSC	Moyenne	0,00710	STDE	0,00350		0,01250	0,01290	0,01330	0,01280	0,01240	0,01310	0,01320	0,01270	0,01290	0,01320	0,01360
	Moyenne	0,00000	STDE			0,00160	0,00140	0,00010	0,00030	0,00210	0,00010	0,00010	0,00020	0,00150	0,00010	0,00010
KICH		0,00000	Moyenne	0,00250		0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
	STDE	0,00380	Moyenne			0,00020	0. 00010	0. 00010	0,00020	0,00020	0,00020	0. 00010	0. 00010	0. 00010	0. 00010	0,00010
KIRC		0,00000	STDE	0,00000		0,01110	0,01040	0,01220	0,01200	0,01020	0,01200	0,01070	0,01040	0,01120	0,01170	0,01290
	Moyenne	0,00000	STDE			0,00270	0,00340	0,00190	0,00020	0,00340	0,00190	0,00320	0,00300	0,00280	0,00220	0,00010
KIRP		0,00000	LUAD	Moyenne		0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
	0,00870	STDE	0,00000			0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
LIHC		Moyenne	0 .00000	STDE		0. 00390	0,00430	0,00450	0. 00390	0. 00400	0. 00420	0,00430	0,00380	0,00430	0,00420	0,00460
	0,00000	PAAD	Moyenne			0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
0,00000	STDE	0,00000	PCPG	Moyenne		0,01260	0,01290	0,01310	0,01240	0,01270	0,01290	0,01290	0,01240	0,01290	0,01300	0,01340
	0,00000	STDE	0,00000			0,00020	0,00010	0,00010	0,00020	0,00020	0,00010	0,00010	0,00020	0,00010	0,00010	0,00010
LUSC		READ	Moyenne	0,00000		0,00400	0,00430	0,00450	0,00380	0,00410	0,00430	0,00430	0,00390	0,00430	0,00430	0,00480
	STDE	0,00000	SARC			0,00020	0,00010	0,00010	0,00020	0,00010	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
Moyenne	0,01890	STDE	0,00000	SKCM		0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
	Moyenne	0,00000	STDE			0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
0,00000	Moyenne	0,00070	STDE	0,00280		0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
						0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
						0,00390	0,0043 0	0,0045 0	0,00390	0,00400	0,00430	0,0043 0	0,0037 0	0,0042 0	0,0042 0	0,00470
						0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
						0,02260	0,02290	0,02320	0,02260	0,02270	0,02300	0,02290	0,02240	0,02290	0,02300	0,02330
						0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
						0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
						0,0002 0	0,00010	0,00010	0,0002 0	0,0002 0	0,0002 0	0,00010	0,00010	0,00010	0,00010	0,00010
STADE						0,00830	0,00620	0,00930	0,01470	0,00810	0,00950	0,00800	0,01170	0,00940	0,00840	0,01570
						0,00500	0,00360	0,00510	0,00020	0,00500	0,00510	0,00480	0,00480	0,00520	0,00490	0,00010
THCA	Moyenne	0,00000	STDE			0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00460
			0,00000			0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
THYM	Moyenne	0,00000	STDE			0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
			0,00000			0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
UCEC	Moyenne	0,00000	STDE			0,01920	0,01710	0,02780	0,03540	0,02010	0,02770	0,02690	0,02560	0,02520	0,02860	0,03670
			0,00250			0,01200	0,01220	0,00610	0,01000	0,01190	0,00620	0,00720	0,00830	0,00910	0,00450	0,01050
Classement W|T|L		22|0|0				0|0|22	0|0|22	0 |0|22	Les valeurs en		0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22

gras indiquent les meilleurs résultats.

Tableau A2. Les résultats de l'algorithme RBAVO-DE proposé et de ses homologues utilisant k-NN concernant les valeurs moyennes des gènes sélectionnés.

Ensemble de données	Métrique	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA	Moyenne	142,03	STDE	017,18		210,17	237,70	235,00	196,03	218,17	228,50	226,30	195,23	228,27	232,03	237.13
	Moyenne	122,50	STDE			020,50	019,07	017,12	023,83	018,27	019,43	014,81	013,17	014,60	018,90	014.14
CESC		005,45	CHOL	Moyenne		195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	213,33	215,90	233.43
		122,13	STDE	006,18		011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	005,88	006,02	005.64
COAD	Moyenne	120,30	STDE	005,82		195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63
	Moyenne	120,17	STDE			011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36
006,81	GBM	Moyenne	120,73	STDE		195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63
						011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36
ESCA						195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,57	218,87	230.07
						011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	006,94	004,68	007.64
						195,13	212,77	225,30	194,10	199,73	213,17	212,80	185,43	211,00	211,83	232,63
				006.54		011,17	005.40	003,11	009,66	009,05	008,51	005,72	006,40	006,44	004,99	004.39
HNSC	Moyenne	133,43	STDE	017,63		202,87	222,67	227,43	203.00	211,37	217,97	220,87	195,93	219,83	223,63	241.27
	Moyenne	121,43	STDE			009,26	011,96	003.91	013.24	007,65	007,41	005,32	008,26	007,71	005,68	006.21
KICH		006,58				195,13	213,03	225.30	194.10	199,50	213,63	212,67	185,60	211,00	211,57	232,63
						011,17	005,31	003.11	009.66	009,30	008,35	005,66	006,49	006,44	005,08	004.36

Machine Translated by Google

Tableau A2. Suite.

Ensemble de données Métrique	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
KIRC Moyenne 137,57 STDE 017,10 Moyenne 122,33	200,83	220,17	227.70	192.17	209,87	216.13	222.30	193,27	218,73	218,60	234.10	
	009,63	011,74	008.27	008.73	014,08	010.11	015.18	017,36	013,82	012,78	004.21	
KIRP STDE Moyenne 123,40 STDE 009.08	195,13	213,03	225.30	194.10	199,50	213.63	212.67	185,60	211,00	211,57	232,63	
	011,17	005,31	003.11	009,66	009h30	008,35	005.66	006,49	006,44	005,08	004.36	
LIHC 006,59 LUAD Moyenne 127,03 STDE 008,62	197,00	214,03	225,67	193,33	200,63	212,33	213.63	189,30	213,10	212,03	232.43	
	007,78	005,22	003.36	009,17	010,02	009,35	004.89	006,89	005,69	004,81	004.97	
Moyenne 125,73 STDE 006,35 PAAD Moyenne 121,10 STDE	197,47	216,13	222.93	190,83	205,33	212,57	215.67	191,80	215,27	217,83	237,90	
	008,98	004,74	006.16	008,61	010,57	006,15	004,43	008,50	006,13	004,71	004.94	
LUSC 007,71 PCPG Moyenne 122,17 STDE 006,42	198,87	215,63	223.00	189,13	203,60	213,23	216,37	194,67	215,53	215,23	241.23	
	008,77	003,54	004.89	010,69	007,20	009,70	003,44	006,40	005,54	006,16	004.66	
READ Moyenne 120,13 STDE 007,72 Moyenne SARC 122,47	195,13	213,03	225.30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63	
	011,17	005,31	003.11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36	
STDE 005,79 SKCM Moyenne 123,03 STDE 008,83 Moyenne	195,13	213,03	225.30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63	
	011,17	005,31	003.11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36	
142,60 STDE 013,31 THCA Moyenne 117,97 STDE 006,67	195,13	213,03	225.30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63	
	011,17	005,31	003.11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36	
THYM Moyenne 120,37 STDE 004,89 UCEC Moyenne 143,93	195,13	213,03	225.30	194,10	199,50	213,63	212,67	185,60	212,90	216,43	231,80	
	.17	.31	003.11	009,66	009,30	008,35	005,66	.49	.30	004,98	003.40	
STDE 013,18	195,13	213,03	225.30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63	
	011,17	005,31	003.11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36	
STADE	215,00	236,17	243.33	187 .50	221,17	.87	232,77	201,73	231,13	.10	236,97	
	018,43	013,94	016.29	010,27	.24	022,63	014,25	014,00	014,04	015,35	005.86	
	195,43	213,40	.13	194,10	199,63	214,13	212,80	186,53	211,00	211,57	231,90	
	011,20	005,29	003.10	009,66	009,81	008,59	005,69	005,88	006,44	005,08	004.78	
	195,13	213,03	225.30	194,10	199,50	213,63	.67	185,60	211,00	211,57	232,63	
	011,17	005,31	003.11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36	
	216,33	237,87	236.00	242,37	220,07	230,80	229,17	206,77	228,40	235,00	266,67	
	015,41	015,82	010.11	034,13	011,26	011,27	011,66	010,84	010,32	006,14	030.55	
Classement W|T|L	22|0|0	0|0|22	0|0|22 Les	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22

valeurs en gras indiquent les meilleurs résultats.

Tableau A3. Les résultats de l'algorithme RBAVO-DE proposé et de ses homologues utilisant k-NN concernant les valeurs de précision moyennes.

Jeu de données	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA 1,00000 CESC	1,00000	1,00000	1,00000	0,99960	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
0,98390 CHOL 1,00000	0,98390	0,98390	0,98390	0,98390	0,98390	0,98390	0,98390	0,98390	0,98390	0,98390	0,98390	
COAD 1,00000 ESCA	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
0,97370 GBM 1,00000	1,00000	1,00000	1,00000	0,99890	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
HNSC 0,99180 KICH	0,97370	0,97370	0,97370	0,97370	0,97370	0,97370	0,97370	0,97370	0,97370	0,97370	0,97370	
1,00000 KIRC 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
KIRP 1,00000 LIHC	0,96470	0,96540	0,96500	0,96350	0,96690	0,96660	0,96570	0,96540	0,96540	0,96100	0,96290	
1,00000 LUAD 1,0000 0	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
LUSC 1,00000 PAAD	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
1,00000 PCPG 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
READ 1,00000 SARC	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
0,98110 SKCM 1,00000	0,99770	0,99740	0,99840	0,99870	0,99770	0,99800	0,99540	0,99710	0,99740	0,99670	0,99800	
STAD 0,99920 THCA	1,00000	1,00000	1,00000	0,99790	1,00000	0,99970	1,00000	1,00000	1,00000	0,97900	0,99930	
1,00000 THYM 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
UCEC 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1. 00000	1.00000	
	0,98110	0,98110	0,98110	0,98110	0,98110	0,98110	0,98110	0,98110	0,98110	0,98110	0,98110	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	0,95180	0,95180	0,95180	0,94870	0,95180	0,95180	0,95180	0,95180	0,95180	0,99000	0,95090	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	0,99910	1,00000	1,00000	1,00000	1,00000	1,00000	0,93330	0,99910	
Classement (W|T|L)	3|19|0	0|19|3	0|19|3	0|19|3	0|15|7	0|19|3	0|18|4	0|19|3	0|19|3	0|19|3	0|17|5	0|17|5

Les valeurs en gras indiquent les meilleurs résultats.

Tableau A4. Les résultats de l'algorithme RBAVO-DE proposé et de ses homologues utilisant k-NN concernant les valeurs de rappel moyennes.

Jeu de données	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA 1,00000 CESC	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
1,00000 CHOL 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
COAD 1,00000 ESCA	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
1,00000 GBM 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	

Machine Translated by Google

Tableau A4. Suite.

Jeu de données RBAVO-DE BSSA		BABC	BPSO	BBA	BGWO BWOA BGOA BHHO BBSA					BASO BHGSO	
HNSC 1,00000 KICH	0,99520	0,99450	0,99480	0,99590	0,99280	0,99310	0,99420	0,99450	0,99450	0,99790	0,99730
1,00000 KIRC 0,99710	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
	0,99040	0,99040	0,99040	0,98970	0,99040	0,99040	0,99040	0,99040	0,99040	0,99010	0,99040
KIRP 1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
LIHC 1,00000 LUAD 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
LUSC 1,00000 PAAD	0,99240	0,99270	0,99170	0,99140	0,99240	0,99210	0,99470	0,99310	0,99270	0,99340	0,99210
1,00000 PCPG 1,00000	1.00000	1.00000	1.00000	0,99580	1.00000	1.00000	1.00000	0,99930	1.00000	0,99310	0,99270
READ 1,00000 SARC	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1,00000 SKCM 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
STAD 1,00000 THCA	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
1,00000 THYM 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
UCEC 1,00 000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
	1,00000	1,00000	1,00000	0,99750	1,00000	1,00000	1,00000	1,00000	1,00000	0,99620	0,99750
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
	1,00000	1,00000	1,00000	0,99370	1,00000	1,00000	1,00000	1,00000	1,00000	0,99640	1.00000
Classement (W|T|L) 3|19|0	0|19|3	0|19|3	0|19|3	0|16|6	0|19|3	0|19|3	0|19|3	0|18|4	0|19|3	0|16|6	0|17|5

Les valeurs en gras indiquent les meilleurs résultats.

Tableau A5. Les résultats de l'algorithme RBAVO-DE proposé et de ses homologues utilisant k-NN

concernant les valeurs moyennes du score F1.

Jeu de données RBAVO-DE BSSA		BABC	BPSO	BBA	BGWO BWOA BGOA BHHO BBSA					BASO BHGSO	
BLCA 1,00000 CESC	1,00000	1,00000	1,00000	0,99980	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
0,99190 CHOL 1,00000	0,99190	0,99190	0,99190	0,99190	0,99190	0,99190	0,99190	0,99190	0,99190	0,99190	0,99190
COAD 1,00000 ESCA GBM	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
1,00000 HNSC 0,99590	1,00000	1,00000	1,00000	0,99940	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
KICH 0,98670	0,98670	0,98670	0,98670	0,98670	0,98670	0,98670	0,98670	0,98670	0,98670	0,98670	0,98670
KIRC KIRP 1,00000 LIHC	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
1,00000 LUAD 0,99500	0,97970	0,97970	0,97970	0,97940	0,97970	0,97970	0,97970	0,97970	0,97970	0,97980	0,97970
LUSC 1.00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
1,00000 0,99860	0,99520	0,99520	0,99520	0,99480	0,99520	0,99520	0,99520	0,99520	0,99520	0,99500	0,99520
PAAD 1,00000 PCPG	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
1,00000 LIRE 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
SARC 0,99050 SKCM	0,99500	0,99500	0,99500	0,99500	0,99500	0,99500	0,99500	0,99500	0,99500	0,99500	0,99500
1,00000 STAD 0,99960	1,00000	1,00000	1,00000	0,99690	1,00000	0,99980	1,00000	0,99970	1,00000	0,99630	0,99600
THCA 1,00000 THYM	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
1,00000 UCEC 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
	0,99050	0,99050	0,99050	0,99050	0,99050	0,99050	0,99050	0,99050	0,99050	0,99050	0,99050
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
	0,97530	0,97530	0,97530	0,97240	0,97530	0,97530	0,97530	0,97530	0,97530	0,97180	0,97360
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
	1,00000	1,00000	1,00000	0,99640	1,00000	1,00000	1,00000	1,00000	1,00000	0,99640	0,99960
Classement (W|T|L) 3|19|0	0|18|4	0|18|4	0|18|4	0|14|8	0|18|4	0|17|5	0|18|4	0|17|5	0|18|4	0|16|6	0|16|6

Les valeurs en gras indiquent les meilleurs résultats.

Annexe A.2. Résultats de comparaison basés sur le modèle SVM

Tableau A6. Les résultats de l'algorithme RBAVO-DE proposé et de ses homologues utilisant SVM

concernant les valeurs moyennes de condition physique.

Ensemble de données	Métrique	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA Moyenne 0,00000 STDE 0,00000		0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00460
	Moyenne 0,00000 STDE	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
CESC	0,00000 CHOL Moyenne	0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
	0,00000 STDE 0,00000	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
COAD Moyenne 0,00000 STDE Moyenne		0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
	0,01370 STDE	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
		0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
	0,00000	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
ESCA		0,02840	0,02720	0,02990	0,02930	0,02690	0,02970	0,02880	0,02830	0,02880	0,02980	0,03000
	0,01280	0,00460	0,00750	0,00010	0,00020	0,00750	0,00020	0,00450	0,00460	0,00440	0,00010	0,00020
GBM Moyenne 0,00000 STDE		0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00460
	0,00000	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
Moyenne HNSC 0,00000 STDE		0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00460
	0,00000	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010

Machine Translated by Google

Tableau A6. Suite.

Ensemble de données Métrique	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
KICH Moyenne 0,00000 STDE	0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470	
0,00000 Moyenne 0,00000	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010	
KIRC STDE Moyenne 0,00000	0,00390	0,00430	0,00450	0,00380	0,00400	0,00430	0,00430	0,00380	0,00430	0,00430	0,00460	
STDE 0,00000	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00020	
KIRP Moyenne 0,00000 STDE	0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470	
0,00000	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010	
LIHC	0,00410	0,00440	0,00450	0,00500	0,00430	0,00440	0,00450	0,00400	0,00450	0,00450	0,00500	
0,00000	0,00010	0,00010	0,00010	0,00280	0,00010	0,00020	0,00010	0,00020	0,00010	0,00010	0,00020	
LUAD Moyenne 0,00840 STDE 0,00160	0,01250	0,01290	0,01310	0,01250	0,01260	0,01290	0,01290	0,01230	0,01290	0,01260	0,01330	
Moyenne 0,00000 STDE	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00150	0,00010	
LUSC 0,00000 PAAD Moyenne	0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470	
0,00000 STDE 0,00000	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010	
PCPG Moyenne 0,00000 STDE 0,00000	0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470	
READ Moyenne 0,00000	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010	
STDE 0,00000 SARC Moyenne 0,00570	0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470	
STDE 0,008 60 SKCM	0,00020	0. 00010	0. 00010	0,00020	0,00020	0,00020	0. 00010	0,00010	0. 00010	0. 00010	0,00010	
Moyenne 0,00000 STDE 0,00000 Moyenne	0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470	
0,01040 STDE 0,00280	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010	
THCA Moyenne 0,00000 STDE 0,00000	0,01780	0,01570	0,01840	0,02260	0. 01480	0. 01760	0,01870	0,02000	0,01870	0,02000	0,02330	
THYM Moyenne 0,00000	0. 00800	0,00880	0,00800	0. 00020	0,00900	0,00820	0,00770	0. 00610	0,00770	0,00670	0,00010	
STDE 0,00000 UCEC Moyenne 0,00000	0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470	
STDE 0,00000	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010	
STADE	0,01490	0,01530	0,01550	0,01450	0,01500	0,01530	0,01530	0,01470	0,01520	0,01530	0,01560	
	0,00020	0,00010	0,00010	0,00200	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010	
	0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470	
	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010	
	0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470	
	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010	
	0,00390	0,00430	0,00450	0,00380	0,00400	0,00430	0,00430	0,00380	0,00420	0,00430	0,00470	
	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00020	0,00010	0,00010	0,00010	
Classement W|T|L 22|0|0	0|0|22	0|0|22	0 |0|22	Les valeurs en	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22

gras indiquent les meilleurs résultats.

Tableau A7. Les résultats de l'algorithme RBAVO-DE proposé et de ses homologues utilisant SVM concernant les valeurs moyennes des gènes choisis.

Ensemble de données Métrique	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA Moyenne 121,37 STDE 007,49	194,70	213,03	225,30	194,10	199,50	213,40	212,67	186,27	211,00	211,83	232.37	
Moyenne 119,83 STDE	011,44	005,18	003,11	009,66	009,30	008,44	005,66	005,77	006,44	005,25	004.42	
CESC 007,02 CHOL Moyenne	195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63	
125,57 STDE 007,14	011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36	
COAD Moyenne 122,63 STDE 007,73	195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63	
Moyenne 125,50 STDE	011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36	
011,18 GBM Moyenne 119,73 STDE	195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63	
008,07 HNSC Moyenne	011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36	
ESCA 122,17 008,29 Moyenne	195,13	215,40	225,30	194,10	201,77	213,63	213,13	185,70	212,40	218,87	230.07	
121,73 STDE 008,49	011,17	009,58	003,11	009,66	012,33	008,35	006,44	006,47	008,98	004,68	007.64	
Moyenne 123,87 STDE 007,57 Moyenne	194,93	213,03	225,30	193,90	198,77	213,70	212,83	185,50	211,00	211,00	232.33	
120,07 STDE 007,87	011,27	005,31	003,11	009,47	010,28	008,36	005,69	006,51	006,44	004,92	004.15	
Moyenne 122,63 STDE 008,01 LUAD	196,40	213,90	225,13	194,03	202,10	213,83	213,50	186,93	211,47	210,87	232.17	
Moyenne 121,17 STDE	008,49	005,22	003,19	009,56	010,76	009,34	005,22	005,74	005,66	004,79	005.02	
KICH 007,56 Moyenne 122,67	195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63	
STDE	011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36	
KIRC	197,10	214,70	225,00	191,93	200,30	212,97	213,13	187,57	213,83	214,63	231.27	
	.42	.09	.23	.70	.59	.64	.94	0 06h49	.81	.20	008.02	
KIRP	195,13	213,03	225,30	194,10	199,50	213,63	212,67		211,00	211,57	232,63	
	011,17	005,31	003,11	009,66	009,30	008,35	005,66		006,44	005,08	004.36	
LIHC	204,27	220,37	226,40	212,83	214,07	219,07	223,70		223,03	225,17	249,90	
	006,99	006,29	006,80	017,60	007,06	007,88	004,13		006,48	004,81	007.97	
	195,13	213,03	225,30	194,10	199,43	213,63	212,67		212,40	214,50	233,00	
	011,17	005,31	003,11	009,66	009,37	008,35	005,66		006,10	005,98	004.34	
LUSC	195,13	213,03	225h30	194,10	199,50	213,63	212,67		211,00	211,57	232,63	
006.28	011,17	005,31	003.11	009,66	009.30	008,35	005,66		006,44	005,08	004.36	
PAAD Moyenne 119,23 STDE 006,90	195,13	213,03	225.30	194,10	199.50	213,63	212,67		211,00	211,57	232,63	
PCPG Moyenne 122,43	011,17	005,31	003.11	009,66	009.30	008,35	005,66		006,44	005,08	004.36	
STDE 006,68 READ Moyenne 123,93	195,13	213,03	225.30	194,10	199.50	213,63	212,67		211,00	211,57	232,63	
STDE 007,45 SARC	011,17	005,31	003.11	009,66	009.30	008,35	005,66		006,44	005,08	004.36	
Moyenne 135,27 STDE 018,57 SKCM	195,13	213,03	225.30	194,10	199.50	213,63	212,67		211,00	211,57	232,63	
Moyenne 122,47 STDE	011,17	005,31	003.11	009,66	009.30	008,35	005,66		006,44	005,08	004.36	
007,66	202,90	226,90	233.77	194,10	210.53	223,87	217,80		220,57	220,87	231.53	
	020,86	020,25	014.44	009,66	017.74	017,39	013,90		016,40	011,76	003.70	
	195,13	213,03	225.30	194,10	199.50	213,63	212,67		211,00	211,57	232,63	
	0 11.17	0 31/05	003.11	009,66	009.30	008,35	005,66		0 06h44	005,08	004.36	

Machine Translated by Google

Tableau A7. Suite.

Ensemble de données Métrique	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
STADE Moyenne 123,53 STDE 010,65 THCA Moyenne 119,47 STDE 005.58 THYM Moyenne 121,80 STDE 007,21 UCEC Moyenne 122,23 STDE 005,93		194,60	213,40	225.03	194.30	198.80	214.10	212.70	185,70	211,97	215.07	231,77
		010,22	005,38	003.29	009.79	009.09	008.68	005.54	006,92	005,88	005.12	004.58
		195,13	213,03	225.30	194.10	199.50	213.63	212.67	185,60	211,00	211.57	232,63
		011,17	005,31	003.11	009.66	009.30	008.35	005.66	006,49	006,44	005.08	004.36
		195,13	213,03	225h30	194.10	199.50	213.63	212.67	185,60	211,00	211.57	232,63
		011,17	005,31	003.11	009.66	009.30	008.35	005.66	006,49	006,44	005.08	004.36
		196,00	213,27	224.83	192.03	200.97	212.60	213.03	189,23	212,23	212.93	233,93
		009,06	006,31	004.27	007.85	008.24	007.74	005.67	008,78	004,98	004.97	003.97
Classement W|T|L	22|0|0	0|0|22	0|0|22	0 |0|22	Les valeurs en	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22

gras indiquent les meilleurs résultats.

Tableau A8. Les résultats de l'algorithme RBAVO-DE proposé et de ses homologues utilisant SVM concernant les valeurs de précision moyennes.

Jeu de données	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA 1,00000 CESC	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
1,00000 CHOL 1,00000	1,00000	1,00000	1,00000	0,99890	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
COAD 1,00000 ESCA	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
1,00000 GBM 1,00000	1,00000	1,00000	1,00000	0,99890	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
HNSC 1,00000 KICH	1,00000	1,00000	1,00000	0,99820	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	0,99210
1,00000 KIRC 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
KIRP 1,00000 1,00000	0,94140	0,94170	0,93750	0,92840	0,94140	0,93990	0,94170	0,94110	0,94050	0,84550	0,93150	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	0,99940	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
LIHC	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
LUAD 1,00000 LUSC	1,00000	1,00000	1,00000	0,99870	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	0,99930	
1,00000 PAAD 1,00000	0,99860	0,99860	0,99860	0,99830	0,99660	0,99860	0,99900	0,99930	0,99860	0,81400	0,99970	
PCPG 1,00000 READ	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	
1,00000 SARC 1,00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	
SKCM 1,00000 STAD	1,00000	1.00000	1,00000	1,00000	1,00000	1.00000	1,00000	1,00000	1,00000	1.00000	1.00000	
0,98830 THCA 1,00000	0,99940	1.00000	0,99940	0,99500	0,99940	1.00000	0,99940	0,99620	0,99940	1.00000	0,99120	
THYM 1,00000 UCEC	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
1,00000	0,94840	0,95110	0,94170	0,93610	0,94730	0,94690	0,94690	0,93910	0,94650	0,95200	0,93500	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	0,99910	0,99820	0,99820	0,99300	1,00000	0,99820	0,99820	0,99910	1,00000	1,00000	0,99390	
Classement (W|T|L)	3|19|0	0|17|5	0|18|4	0|17|5	0|12|10	0|18|4	0|18|4	0|17|5	0|17|5	0|18|4	0|19|3	0|15|7

Les valeurs en gras indiquent les meilleurs résultats.

Tableau A9. Les résultats de l'algorithme RBAVO-DE proposé et de ses homologues utilisant SVM concernant les valeurs de rappel moyennes.

Jeu de données	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA 1,00000 CESC	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
1,00000 CHOL 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
COAD 1,00000 ESCA	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
1,00000 GBM 1,00000	1,00000	1,00000	1,00000	0,99940	1,00000	1,00000	1,00000	1,00000	1,00000	0,99940	1.00000	
HNSC 1,00000 KICH KIRC	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
KIRP LIHC 1,00000 LUAD	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
1,00000 LUSC 1,00000	1,00000	1,00000	0,99970	0,99930	1,00000	1,00000	1,00000	1,00000	1,00000	0,99930	0,99930	
PAAD 1.00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
1,00000 1.00000	0,99040	0,99040	0,99040	0,99010	0,99040	0,99040	0,99040	0,99040	0,99040	0,98780	0,99040	
PCPG 1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	
1,00000 READ 1,00000	1.00000	1.00000	1.00000	1,00000	1.00000	1.00000	1.00000	1,00000	1.00000	1,00000	1.00000	
SARC 1,00000 SKCM	1,00000	1,00000	1,00000	0,99500	1,00000	1,00000	1,00000	0,99930	1,00000	0,99830	0,99770	
1,00000 STAD 1,00000	0,99100	0,99100	0,99100	0,99130	0,98960	0,99100	0,99060	0,99030	0,99100	0,99100	0,98990	
THCA 1,00000 THYM	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
1,00000 UCEC 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	
	0,99550	0,99640	0,98380	0,98200	0,99460	0,99100	0,99010	0,98470	0,99370	0,97840	0,97930	
Classement (W|T|L)	3|19|0	0|19|3	0|19|3	0|18|4	0|16|6	0|19|3	0|19|3	0|19|3	0|18|4	0|19|3	0|16|6	0|17|5

Les valeurs en gras indiquent les meilleurs résultats.

Machine Translated by Google

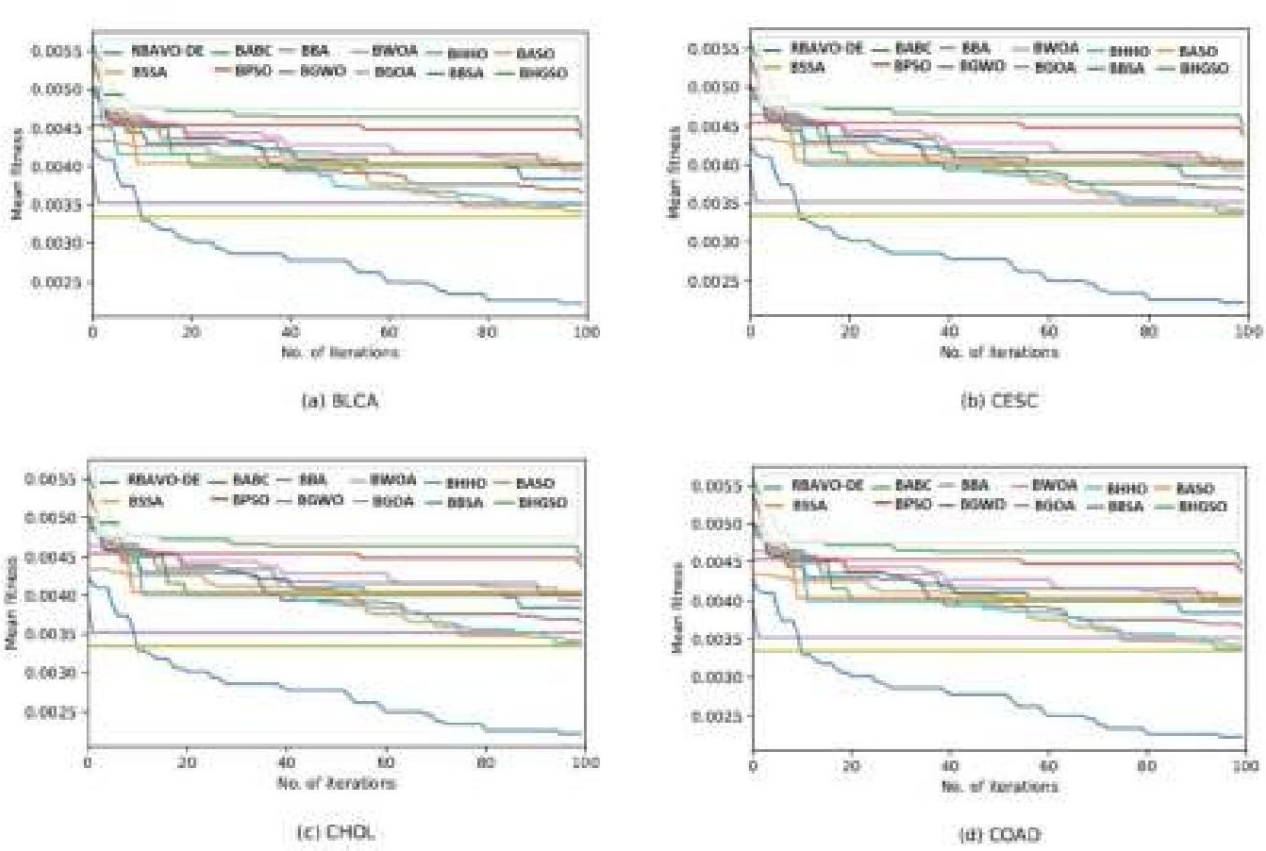
Tableau A10. Les résultats de l'algorithme RBAVO-DE proposé et de ses homologues utilisant SVM concernant les valeurs moyennes du score F1.

Jeu de données RBAVO-DE BSSA		BABC	BPSO	BBA	BGWO BWOA BGOA BHHO BBSA					BASO BHGSO	
BLCA 1,00000 CESC	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
1,00000 CHOL 1,00000	1,00000	1,00000	1,00000	0,99950	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
COAD 1,00000 ESCA	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
1,00000 GBM 1,00000	1,00000	1,00000	1,00000	0,99910	1,00000	1,00000	1,00000	1,00000	1,00000	0,99970	1.00000
HNSC 1,00000 KICH	1,00000	1,00000	1,00000	0,99910	1,00000	1,00000	1,00000	1,00000	1,00000	0,99820	0,99600
1,00000 KIRC 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
KIRP 1,00000 LIHC	0,96980	0,97000	0,96760	0,96250	0,96980	0,96900	0,96970	0,96970	0,96940	0,95730	0,96420
1,00000 LUAD 1,0000 0	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
LUSC 1,00000 PAAD	0,99520	0,99520	0,99520	0,99470	0,99520	0,99520	0,99520	0,99520	0,99520	0,99320	0,99520
1,00000 PCPG 1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
READ 1,00000 SARC	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
0,99900 SKCM 1,00000	1,00000	1,00000	1,00000	0,99690	1,00000	1,00000	1,00000	0,99970	1,00000	0,99880	0,99850
STAD 0,99410 THCA	0,99480	0,99480	0,99480	0,99480	0,99480	0,99480	0,99480	0,99480	0,99480	0,99480	0,99480
1,00000 THYM 1,00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
UCEC 1,00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000	1,00000	1,00000	1,00000	1,00000	1.00000
	0,99970	0,99810	0,99970	0,99750	0,99970	1,00000	0,99970	0,99810	0,99970	0,99710	0,99560
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
	0,97350	0,97490	0,96990	0,96700	0,97290	0,97270	0,97270	0,96860	0,97250	0,96560	0,96640
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1.00000
	0,99730	0,99730	0,99090	0,98730	0,99730	0,99450	0,99410	0,99180	0,99680	0,98680	0,98640
Classement (W|T|L) 5|17|0		0|16|6	0|16|6	0|16|6	0|12|10	0|16|6	0|17|5	0|16|6	0|15|7	0|16|6	0|13|9	0|14|8

Les valeurs en gras indiquent les meilleurs résultats.

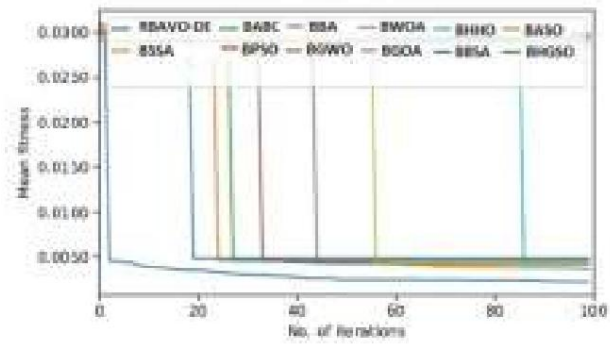
Annexe B. Graphiques de convergence des projets RBAVO-DE et Méthodes compétitives

Annexe B.1. Graphiques de convergence utilisant le modèle SVM

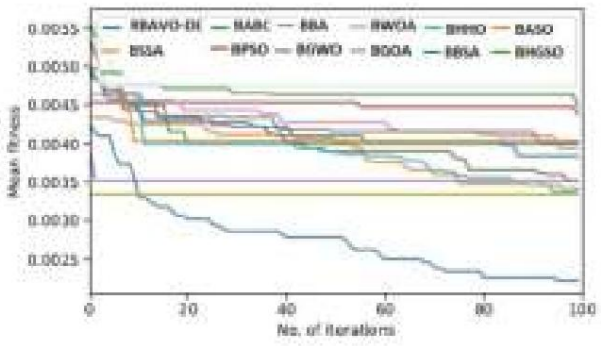


Graphique A1. Suite.

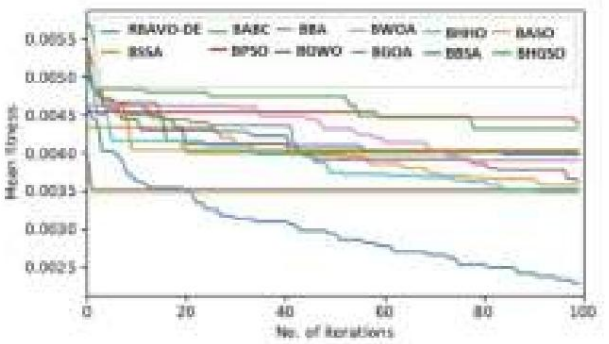
Machine Translated by Google



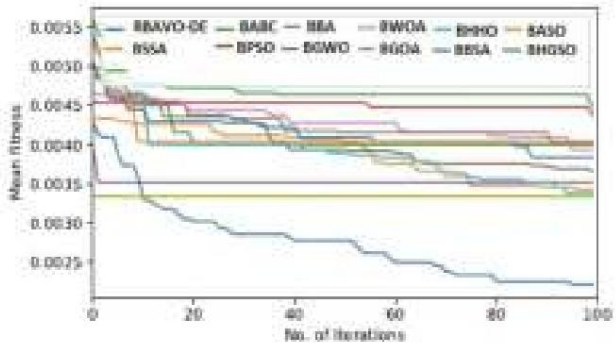
(e) ESCA



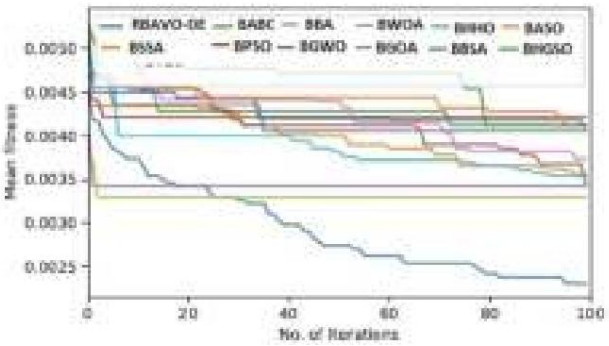
(f) GBM



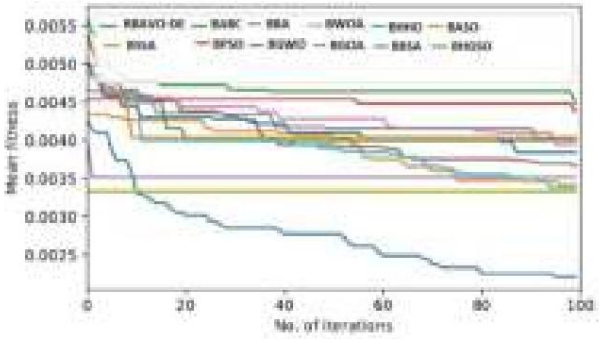
(g) HNSC



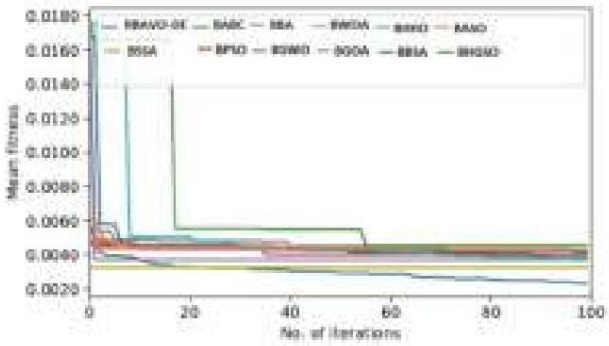
(h) KICH



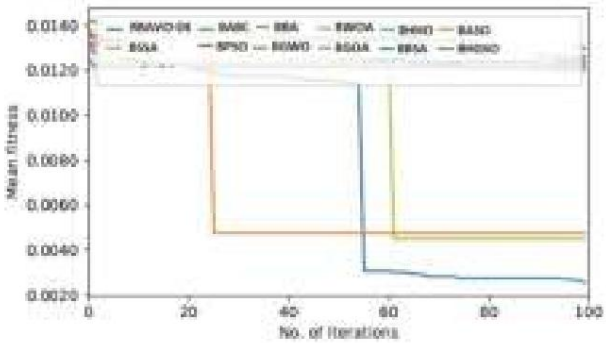
(i) KIRC



(j) KIRP



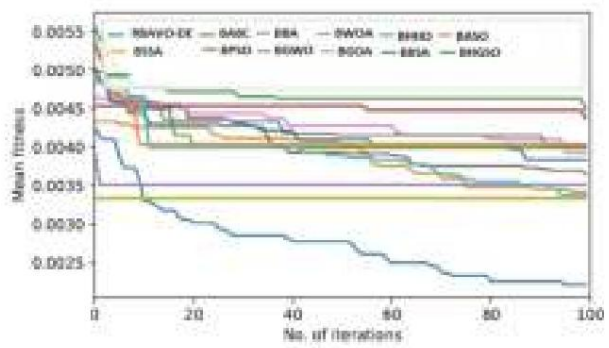
(k) LIHC



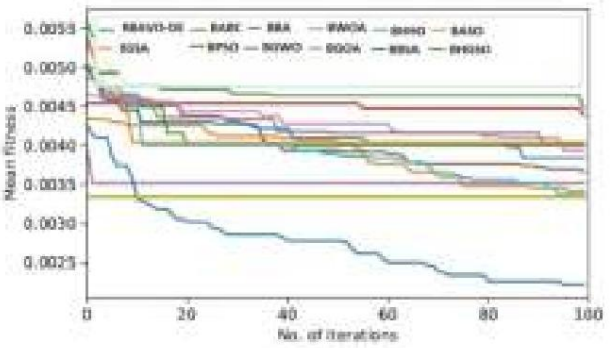
(l) LUAD

Graphique A1. Suite.

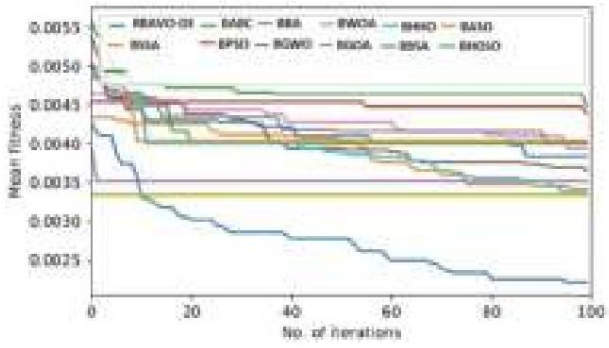
Machine Translated by Google



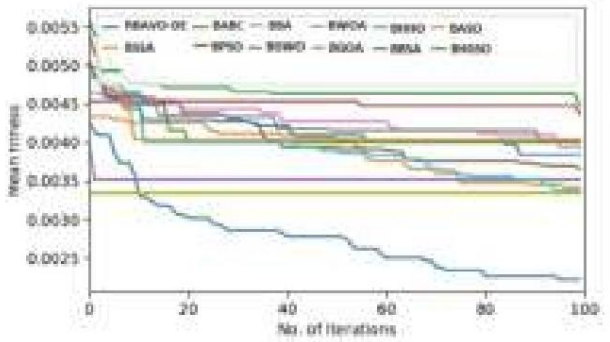
(m) LUSC



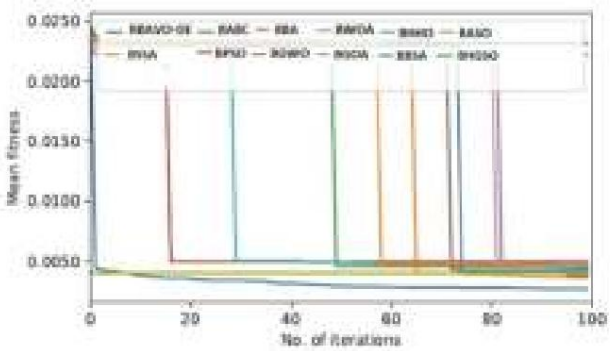
(n) PAAD



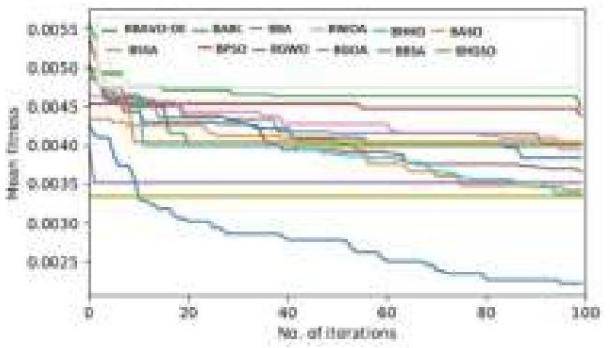
(o) PCPG



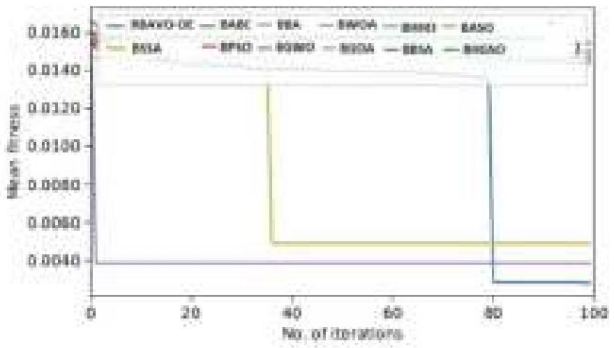
(p) READ



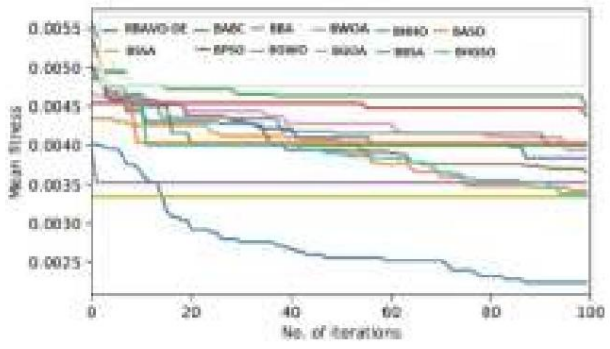
(q) SARC



(r) SKCM



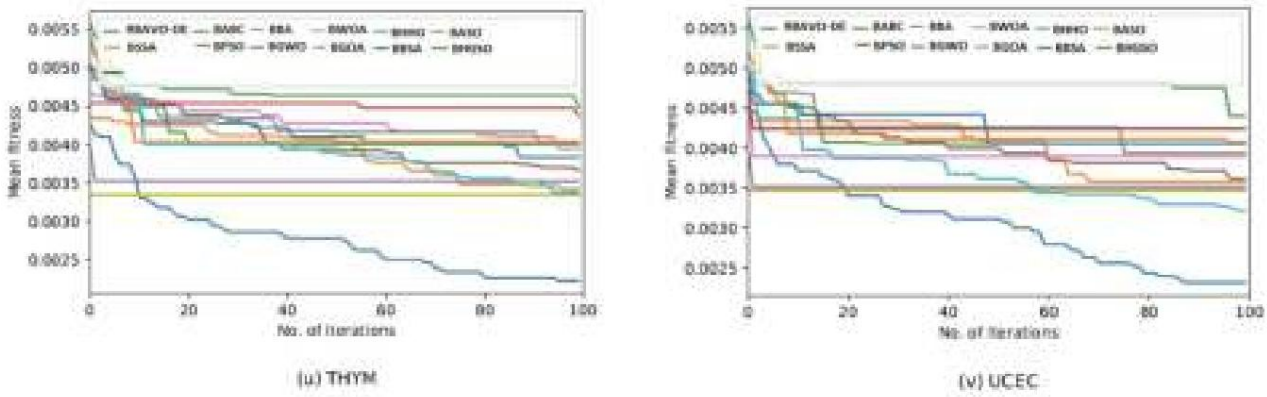
(s) STAD



(t) THCA

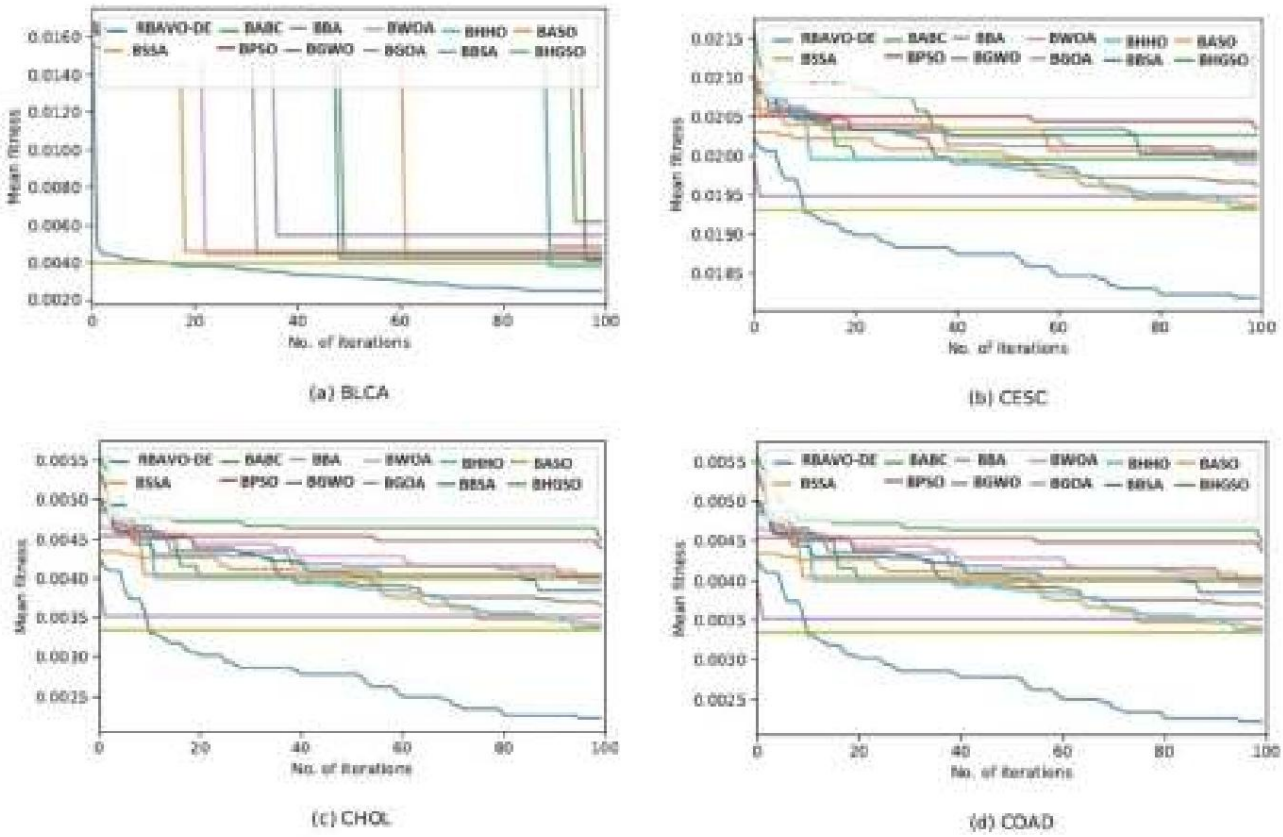
Graphique A1. Suite.

Machine Translated by Google



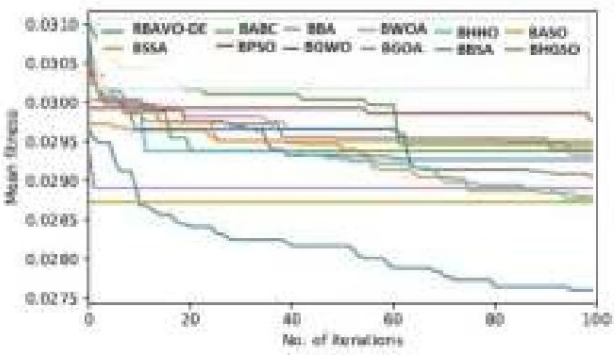
Graphique A1. Graphiques de convergence de l'algorithme RBAVO-DE proposé et des méthodes compétitives utilisant le modèle SVM sur les benchmarks globaux [57].

Annexe B.2. Graphiques de convergence utilisant le modèle k-NN

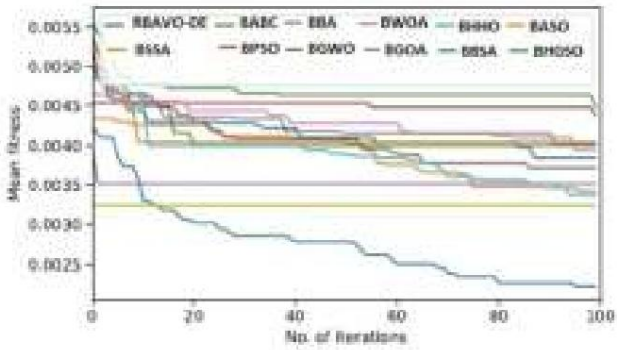


Graphique A2. Suite.

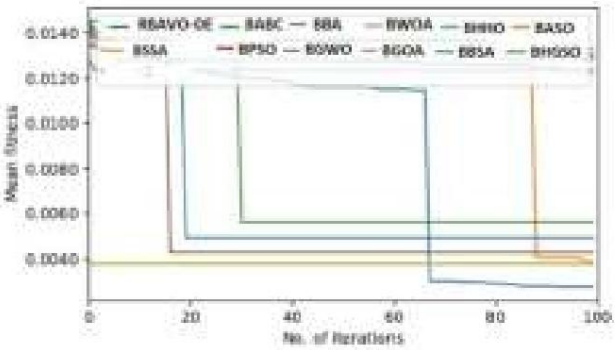
Machine Translated by Google



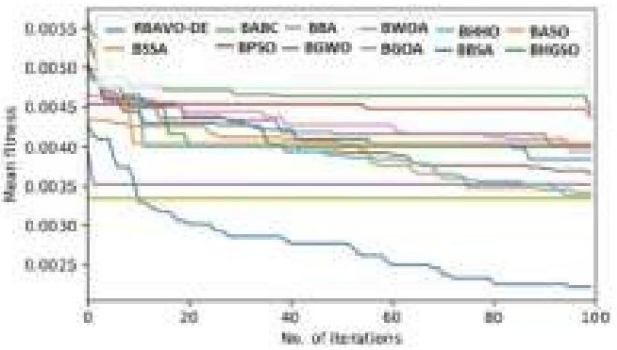
(e) ESCA



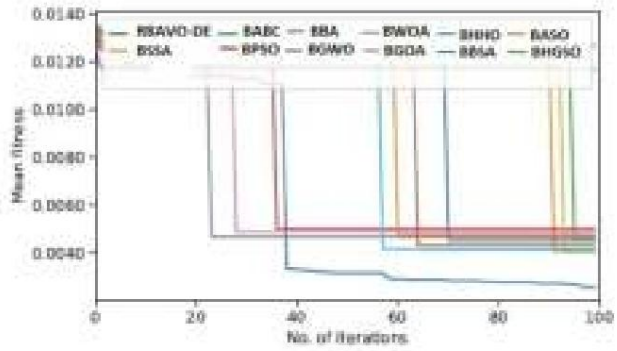
(f) GBM



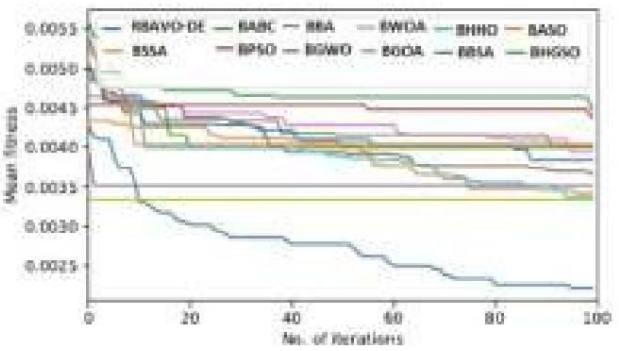
(g) HNSC



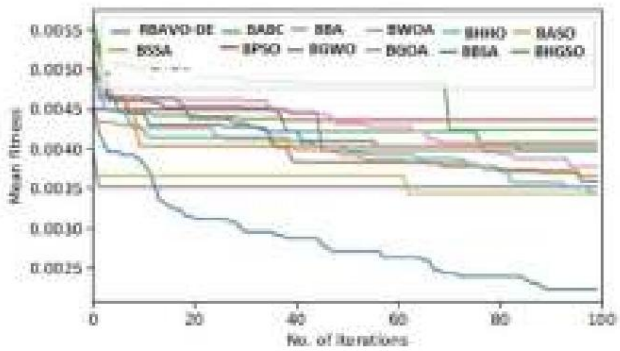
(h) KICH



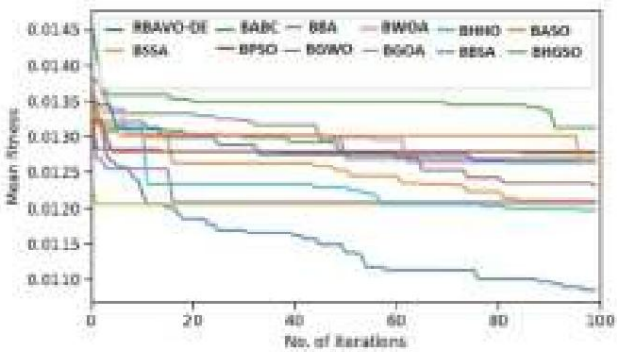
(i) KIRC



(j) KIRP



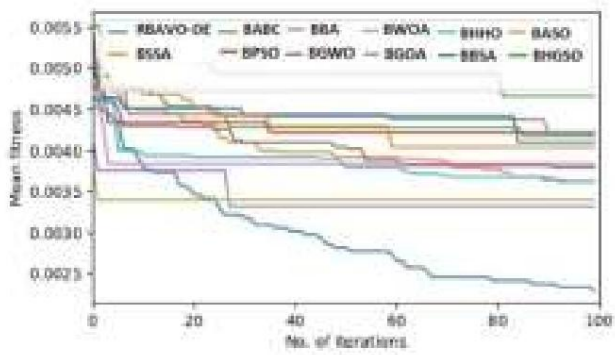
(k) LIHC



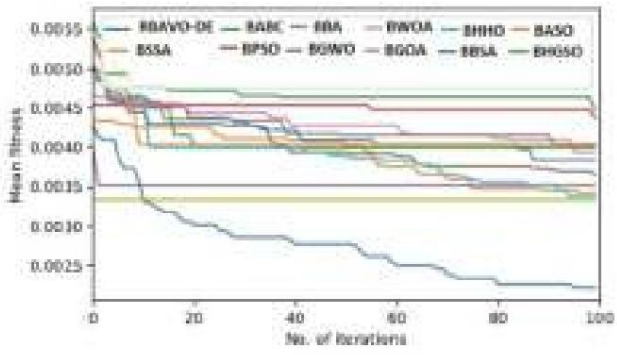
(l) LUAD

Graphique A2. Suite.

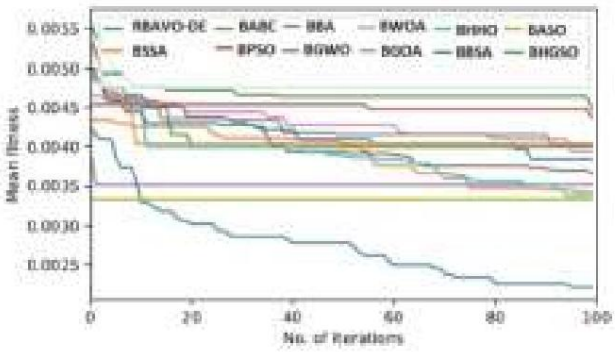
Machine Translated by Google



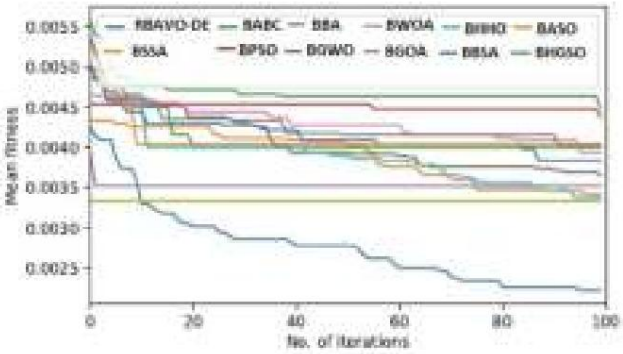
(m) LUSC



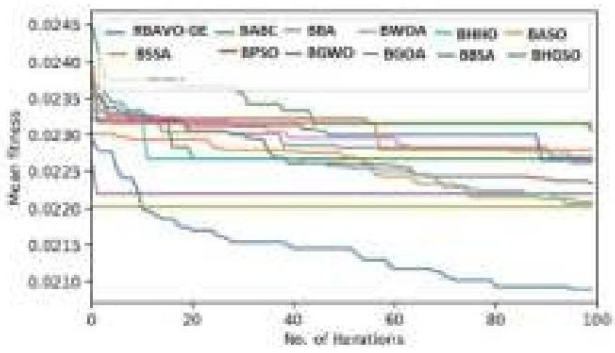
(n) PAAD



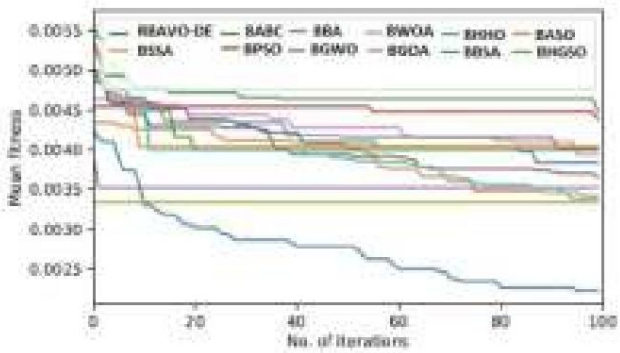
(o) PCPG



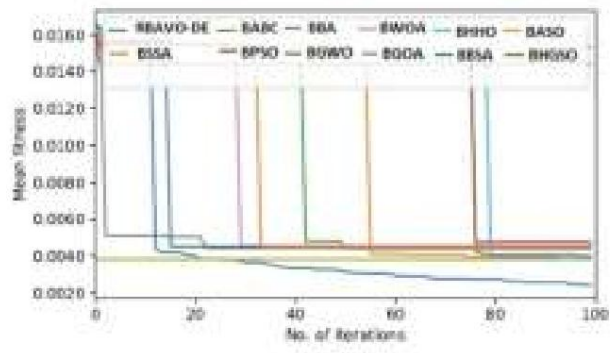
(p) READ



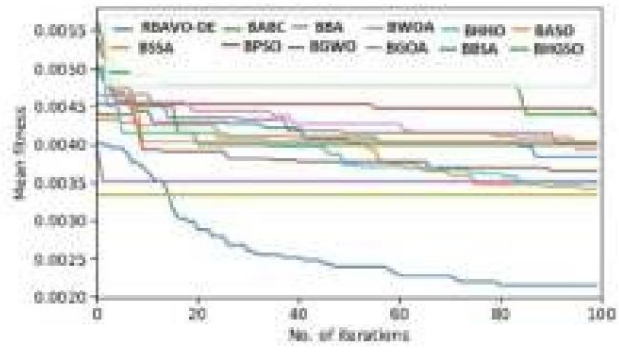
(q) SARC



(r) SKCM



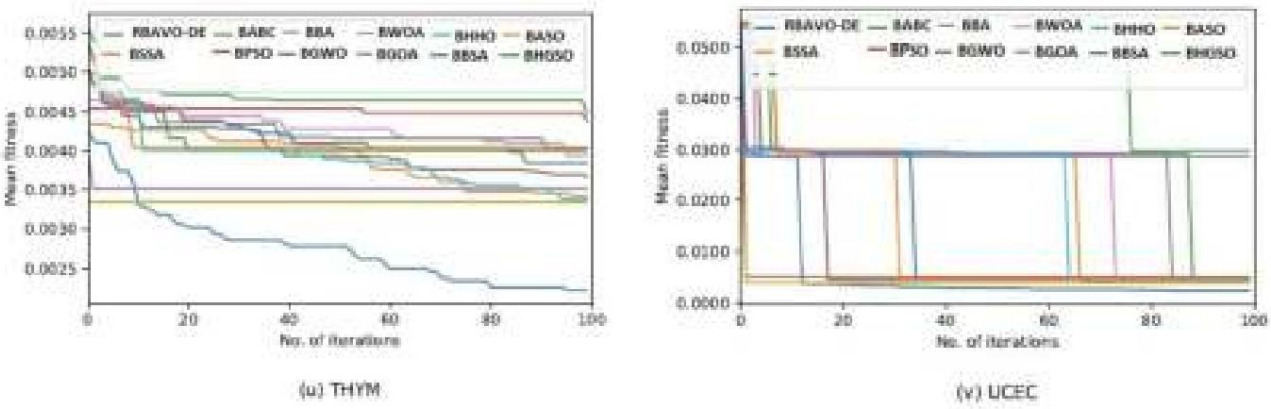
(s) STAD



(t) THCA

Graphique A2. Suite.

Machine Translated by Google



Graphique A2. Graphiques de convergence de l'algorithme RBAVO-DE proposé et des méthodes compétitives utilisant le modèle k-NN sur les benchmarks globaux (suite) [57].

Références

1. Estrada-Meza, C. ; Torres-Copado, A. ; Loreti González-Melgoza, L. ; Ruiz-Manriquez, LM; De Donato, M. ; Sharma, A. ; Pathak, S. ; Banerjee, A. ; Paul, S. Aperçus récents sur les microARN et la longue régulation médiée par l'ARN non codant des populations de cellules souches. 3 Biotechnologie 2022, 12, 270. [CrossRef] [Pub Med]
2. Kakati, T. ; Bhattacharyya, DK ; Kalita, JK; Norden-Krichmar, TM DEGnext : Classification de gènes exprimés différentiellement à partir de données de séquençage d'ARN à l'aide d'un réseau neuronal convolutif avec apprentissage par transfert. BMC Bioinformer. 2022, 23, 17. [Réf. croisée]
3. Zhao, S. ; Fung-Leung, WP ; Bittner, A. ; Ngo, K. ; Liu, X. Comparaison de l'ARN-Seq et de la puce à ADN dans le profilage du transcriptome des cellules T activées. PLoS ONE 2014, 9, e78644. [Référence croisée]
4. Chen, Z. ; Luo, Z. ; Zhang, D. ; Li, H. ; Liu, X. ; Zhu, K. ; Zhang, H. ; Wang, Z. ; Zhou, P. ; Ren, J. ; et coll. TIGER : un portail Web de ressources sur l'expression génique de l'immunothérapie tumorale. Génome. Protéome. Bioinformer. 2023, 21, 337-348. [Référence croisée] [Pub Med]
5. Nunez-Garcia, J. ; AbuOun, M. ; Étage, N. ; Brouwer, M. ; Delgado-Blas, J. ; Mo, SS ; Ellaby, N. ; Veldman, K. ; Haenni, M. ; Châtre, P. ; et coll. Harmonisation des méthodes de diagnostic et de surveillance basées sur le séquençage in silico de nouvelle génération. Sci. Rep.2022 , 12, 14372. [CrossRef]
6. Wang, Z. ; Gerstein, M. ; Snyder, M. RNA-Seq : Un outil révolutionnaire pour la transcriptomique. Nat. Révérend Genet. 2009, 10, 57-63. [Référence croisée]
7. Kim, WJ ; Choi, BR; Non, JJ ; Lee, YY; Kim, TJ ; Lee, JW; Kim, BG ; Choi, CH Comparaison de RNA-Seq et de microarray dans le prédiction de l'expression des protéines et prédiction de la survie. Devant. Genet. 2024, 15, 1342021. [CrossRef]
8. Wang, M. ; Chen, X. ; Dai, Y. ; Wu, D. ; Liu, F. ; Yang, Z. ; Chanson, B. ; Xie, L. ; Yang, L. ; Zhao, W. ; et coll. Étude de concordance d'un test de profilage génomique de nouvelle génération basé sur le séquençage de 520 gènes d'échantillons de tissus et de plasma. Mol. Diagnostic. Là. 2022, 26, 309-322. [Référence croisée]
9. Metzker, Technologies de séquençage ML : la prochaine génération. Nat. Révérend Genet. 2010, 11, 31-46. [Référence croisée] [Pub Med]
10. Pandey, D. ; Onkara Perumal, P. Une revue de la portée de l'apprentissage profond pour l'ARN-Seq de nouvelle génération. analyse des données. Fonction. Intégré. Génome. 2023, 23, 134. [Réf. croisée]
11. Liu, S. ; Yao, W. Prédiction du cancer du poumon à l'aide de l'expression génique et de l'apprentissage profond avec sélection de gènes de divergence KL. BMC Bioinformer. 2022, 23, 175. [Réf. croisée] [Pub Med]
12. Houssein, EH; Oliva, D. ; Celik, E. ; Emam, MM; Ghoniem, RM Algorithme d'optimisation de la sterne fuligineuse amélioré pour l'optimisation globale et la sélection des fonctionnalités. Système expert. Appl. 2023, 213, 119015. [Réf. croisée]
13. Joshi, AA; Aziz, RM Une approche basée sur la recherche de coucou en deux phases pour la sélection de gènes et la classification par apprentissage profond des maladies cancéreuses à l'aide de données d'expression génique avec une nouvelle fonction de fitness. Multimed. Outils Appl. 2024, 83, 71721-71752. [Référence croisée]
14. Ramaswamy, R. ; Kandhasamy, P. ; Palaniswamy, S. Sélection de fonctionnalités pour les données d'expression des gènes de la maladie d'Alzheimer à l'aide d'un binaire modifié optimisation des essaims de particules. IETE J. Rés. 2023, 69, 9-20. [Référence croisée]
15. Cui, X. ; Li, Y. ; Fan, J. ; Wang, T. Un nouvel algorithme de sélection de fonctionnalités de filtre basé sur le relief. Appl. Intell. 2022, 52, 5063-5081. [Référence croisée]
16. Alhenawi, E. ; Al-Sayyed, R. ; Hudaib, A. ; Mirjalili, S. Méthodes de sélection de caractéristiques sur les données de puces à ADN d'expression génique pour la classification du cancer : une revue systématique. Calculer. Biol. Méd. 2022, 140, 105051. [Réf. croisée] [Pub Med]
17. Parlak, B. ; Uysal, AK Une nouvelle méthode de sélection de fonctionnalités de filtre pour la classification de texte : le sélecteur de fonctionnalités étendu. J. Inf. Sci. 2023, 49, 59-78. [Référence croisée]
18. Albulayhi, K. ; Abu Al-Haija, Q. ; Alsuhibany, SA ; Jillepalli, AA; Ashrafuzzaman, M. ; Sheldon, FT Détection d'intrusion IoT à l'aide de l'apprentissage automatique avec une nouvelle méthode de sélection de fonctionnalités très performante. Appl. Sci. 2022, 12, 5015. [CrossRef]
19. Fatima, A. ; Nazir, T. ; Nazir, AK ; Din, AMU Une méthode efficace de sélection de sous-ensembles de gènes de gain d'informations basée sur un wrapper incrémental (IG basée sur IWSSr) pour le discernement des tumeurs. Multimed. Outils Appl. 2024, 83, 64741-64766. [Référence croisée]
20. Kaur, S. ; Kumar, Y. ; Koul, A. ; Kumar Kamboj, S. Une revue systématique des techniques d'optimisation métaheuristiques pour la sélection de caractéristiques dans le diagnostic des maladies : problèmes et défis ouverts. Cambre. Calculer. Méthodes Ing. 2023, 30, 1863-1895. [Référence croisée]

Machine Translated by Google

21. Abd El-Mageed, AA ; Gad, AG; Sallam, KM; Munasinghe, K. ; Abohany, AA Amélioration de la réduction de la dimensionnalité basée sur un algorithme d'optimisation adaptative binaire basée sur le vent pour la classification supervisée. Calculer. Ind. Ing. 2022, 167, 107904. [Référence croisée]
22. Gad, AG; Sallam, KM; Chakrabortty, RK ; Ryan, MJ ; Abohany, AA Un algorithme de recherche binaire amélioré pour les fonctionnalités sélection dans la classification des données. Calcul neuronal. Appl. 2022, 34, 15705-15752. [Référence croisée]
23. Hussien, RM; Abohany, AA; Abd El-Mageed, AA ; Hosny, KM Algorithme d'optimisation binaire suricate amélioré pour une sélection efficace des fonctionnalités de classification d'apprentissage supervisé. Système basé sur les connaissances. 2024, 292, 111616. [Réf. croisée]
24. Abd El-Mageed, AA ; Abohany, AA; Elashry, A. Stratégie efficace de sélection de fonctionnalités pour une classification supervisée basée sur un Algorithme d'optimisation binaire Aquila amélioré. Calculer. Ind. Ing. 2023, 181, 109300. [Réf. croisée]
25. Yin, Y. ; Jang-Jaccard, J. ; Xu, W. ; Singh, A. ; Zhu, J. ; Sabrina, F. ; Kwak, J. IGRF-RFE : Une méthode de sélection de fonctionnalités hybrides pour Détection d'intrusion réseau basée sur MLP sur l'ensemble de données UNSW-NB15. J. Big Data 2023, 10, 15. [CrossRef]
26. Nakao, H. ; Imaoka, M. ; Hida, M. ; Imai, R. ; Nakamura, M. ; Matsumoto, K. ; Kita, K. Détermination des facteurs individuels associés avec hallux valgus en utilisant SVM-RFE. Musculo-squelettique BMC. Désordre. 2023, 24, 534. [Réf. croisée] [Pub Med]
27. Sarafrazi, S. ; Nezamabadi-Pour, H. Face à la classification des problèmes binaires avec un système hybride GSA-SVM. Mathématiques. Calculer. Modèle. 2013, 57, 270-278. [Référence croisée]
28. Cadenas, JM; Garrido, MC; MartíNez, R. Filtre-wrapper de sélection de sous-ensemble de fonctionnalités basé sur des données de faible qualité. Système expert. Appl. 2013, 40, 6241-6252. [Référence croisée]
29. Oh, EST ; Lee, JS ; Moon, BR Algorithmes génétiques hybrides pour la sélection de caractéristiques. IEEETrans. Modèle Anal. Mach. Intell. 2004, 26, 1424-1437.
30. Abdollahzadeh, B. ; Gharehchopogh, FS ; Mirjalili, Algorithme d'optimisation des vautours d'Afrique du Sud : un nouvel algorithme métaheuristique inspiré de la nature pour les problèmes d'optimisation globale. Calculer. Ind. Ing. 2021, 158, 107408. [Réf. croisée]
31. El-Shafeiy, E. ; Hassanien, AE; Sallam, KM; Abohany, A. Approche pour la formation du réseau neuronal quantique afin de prédire la gravité de COVID-19 chez les patients. Calculer. Maître. Continuer. 2020, 66, 1745-1755. [Référence croisée]
32. Yaqoob, A. ; Verma, NK ; Aziz, RM Optimisation de la sélection des gènes et de la classification du cancer avec un algorithme de recherche hybride sinus cosinus et coucou . J.Méd. Système. 2024, 48, 10. [Réf. croisée]
33. Joshi, AA; Aziz, RM Approche d'apprentissage profond pour la classification des tumeurs cérébrales utilisant l'optimisation métaheuristique avec le gène données d'expression. Int. J. Système d'imagerie. Technologie. 2023, 34, e23007. [Référence croisée]
34. Mahto, R. ; Ahmed, SU ; Rahman, Ru ; Aziz, RM; Roy, P. ; Mallik, S. ; Li, A. ; Shah, MA Un cadre de classification du cancer nouveau et innovant grâce à une utilisation consécutive de la sélection de caractéristiques hybrides. BMC Bioinformer. 2023, 24, 479. [Réf. croisée] [Pub Med]
35. Neggaz, N. ; Neggaz, I. ; Abd Elaziz, M. ; Hussien, AG; Abulaigh, L. ; Damaševi˘cius, R. ; Hu, G. Boosting l'optimiseur de recherche de nourriture des raies manta par les opérateurs de trigonométrie : une étude de cas sur un ensemble de données médicales. Calcul neuronal. Appl. 2024, 36, 9405-9436. [Référence croisée]
36. Lyu, B. ; Haque, A. Classification des types de tumeurs basée sur l'apprentissage profond à l'aide de données d'expression génique. Dans Actes de la Conférence internationale ACM 2018 sur la bioinformatique, la biologie computationnelle et l'informatique de la santé, Washington, DC, États-Unis, 29 août-1er septembre 2018 ; pp. 89-96.
37. Khalifa, NEM ; Taha, MHN; Ali, DE; Slowik, A. ; Hassanien, AE Technique d'intelligence artificielle pour l'expression génique à partir de données ARN-Seq tumorales : une nouvelle approche d'apprentissage en profondeur optimisée. Accès IEEE 2020, 8, 22874-22883. [Référence croisée]
38. Dillies, MA; Rau, A. ; Aubert, J. ; Hennequet-Antier, C. ; Jeanmougin, M. ; Serviteur, N. ; Keime, C. ; Marot, G. ; Castel, D. ; Estelle, J. ; et coll. Une évaluation complète des méthodes de normalisation pour l'analyse des données de séquençage d'ARN à haut débit Illumina. Briefings Bioinformer. 2013, 14, 671-683. [Référence croisée]
39. Xiao, Y. ; Wu, J. ; Lin, Z. ; Zhao, X. Une méthode d'ensemble multimodèle basée sur l'apprentissage profond pour la prédiction du cancer. Calculer. Méthodes Programmes Biomed. 2018, 153, 1–9. [Référence croisée]
40. Liu, M. ; Xu, L. ; Yi, J. ; Huang, J. Une méthode de sélection de gènes caractéristiques basée sur ReliefF et PSO. Dans les actes de la 10e Conférence internationale 2018 sur les technologies de mesure et l'automatisation mécatronique (ICMTMA), Changsha, Chine, 7 et 8 janvier 2018 ; IEEE : New York, NY, États-Unis, 2018 ; pp. 298-301.
41. Kononenko, I. Estimation des attributs : analyse et extensions de RELIEF. Dans les actes de la Conférence européenne sur les machines Learning, Catane, Italie, 6-8 avril 1994 ; Springer : Berlin/Heidelberg, Allemagne, 1994 ; 171-182.
42. Faris, H. ; Mafarja, MM; Heidari, AA; Aljarah, moi; Ala'M, AZ ; Mirjalili, S. ; Fujita, H. Un algorithme d'essaim de salpes binaire efficace avec schéma croisé pour les problèmes de sélection de fonctionnalités. Système basé sur les connaissances. 2018, 154, 43-67. [Référence croisée]
43. Abdel-Basset, M. ; Ding, W. ; El-Shahat, D. Un algorithme d'optimisation hybride Harris Hawks avec recuit simulé pour la sélection de fonctionnalités. Artif. Intell. Rév.2020 , 54, 593-637. [Référence croisée]
44. Storn, R. ; Price, K. Evolution différentielle : une heuristique simple et efficace pour l'optimisation globale sur des espaces continus. J. Glob. Optimale. 1997, 11, 341-359. [Référence croisée]
45. Karaboga, D. ; Basturk, B. Un algorithme puissant et efficace pour l'optimisation des fonctions numériques : algorithme de colonie d'abeilles artificielles (ABC). J. Glob. Optimale. 2007, 39, 459-471. [Référence croisée]
46. Mirjalili, S. ; Gandomi, AH; Mirjalili, SZ; Saremi, S. ; Faris, H. ; Mirjalili, SM Salp Swarm Algorithm : un optimiseur bio-inspiré pour les problèmes de conception technique. Av. Ing. Logiciel. 2017, 114, 163-191. [Référence croisée]
47. Khanesar, MA; Technehlab, M. ; Shoorehdeli, MA Une nouvelle optimisation d'essaim de particules binaires. Dans Actes de la Conférence méditerranéenne 2007 sur le contrôle et l'automatisation, Athènes, Grèce, 27-29 juin 2007 ; IEEE : New York, NY, États-Unis, 2007 ; p. 1 à 6.

Machine Translated by Google

48. Mirjalili, S. ; Mirjalili, SM; Yang, XS Algorithme de chauve-souris binaire. Calcul neuronal. Appl. 2014, 25, 663-681. [Référence croisée]
49. Emary, E. ; Zawbaa, HM; Hassanien, AE Approches d'optimisation binaires du loup gris pour la sélection des fonctionnalités. Neuroinformatique 2016, 172, 371-381. [Référence croisée]
50. Hichem, H. ; Elkamel, M. ; Rafik, M. ; Mesaaoud, MT; Ouahiba, C. Un nouvel algorithme d'optimisation binaire de sauterelle pour le problème de sélection de fonctionnalités. J. King Saud Univ. Calculer. Inf. Sci. 2022, 34, 316-328. [Référence croisée]
51. Hussien, AG; Hassanien, AE; Houssein, EH; Bhattacharyya, S. ; Amin, M. Algorithme d'optimisation de baleine binaire en forme de S pour la sélection de fonctionnalités. Dans Tendances récentes en matière de traitement du signal et de l'image : ISSIP 2017 ; Springer : Berlin/Heidelberg, Allemagne, 2019 ; pp. 79-87.
52. Zhao, W. ; Wang, L. ; Zhang, Z. Optimisation de la recherche Atom et son application pour résoudre un problème d'estimation de paramètres hydrogéologiques. Système basé sur les connaissances. 2019, 163, 283-304. [Référence croisée]
53. Meng, XB ; Gao, XZ ; Lu, L. ; Liu, Y. ; Zhang, H. Un nouvel algorithme d'optimisation bio-inspiré : Bird Swarm Algorithm. J. Exp. Théorique. Artif. Intell. 2016, 28, 673-687. [Référence croisée]
54. Hashim, FA; Houssein, EH; Mabrouk, MS; Al-Atabany, W. ; Mirjalili, S. Henry optimisation de la solubilité des gaz : une nouvelle physique - algorithme basé sur. Futur Gén. Calculer. Système. 2019, 101, 646-667. [Référence croisée]
55. Heidari, AA; Mirjalili, S. ; Faris, H. ; Aljarah, moi; Mafarja, M. ; Chen, H. Harris prône l'optimisation : algorithme et applications. Futur Gén. Calculer. Système. 2019, 97, 849-872. [Référence croisée]
56. Ensemble de données d'expression génique RNA-Seq de niveau 3 normalisé. Disponible en ligne : https://gdac.broadinstitute.org/ (consulté le 20 décembre 2023).
57. El-Mageed, AAA ; Elkhouli, AE; Abohany, AA; Gafar, M. Sélection de gènes via un algorithme amélioré d'optimisation de la réaction nucléaire pour la classification du cancer dans des données de grande dimension. J. Big Data 2024, 11, 46. [CrossRef]
58. Mafarja, M. ; Mirjalili, S. Approches d'optimisation Whale pour la sélection des fonctionnalités du wrapper. Appl. Informatique douce. 2018, 62, 441-453. [Référence croisée]
59. Thaher, T. ; Heidari, AA; Mafarja, M. ; Dong, JS; Mirjalili, S. Optimiseur binaire Harris Hawks pour la sélection de caractéristiques de grande dimension et de faible taille d'échantillon. Dans les techniques évolutives d'apprentissage automatique ; Springer : Berlin/Heidelberg, Allemagne, 2020 ; pp. 251-272.
60. Derrac, J. ; García, S. ; Molina, D. ; Herrera, F. Un tutoriel pratique sur l'utilisation de tests statistiques non paramétriques comme méthodologie de comparaison des algorithmes d'intelligence évolutive et en essaim. Essaim Evol. Calculer. 2011, 1, 3-18. [Référence croisée]
61. Code Python pour la sélection de gènes via l'optimisation binaire des vautours africains en relief intégrée à l'évolution différentielle. Disponible en ligne : https://github.com/D-Amr-Atef/Gene_Selection_RBAVO-DE.git (consulté le 18 mai 2024).

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Revoir

Une revue de l'apprentissage par renforcement dans la planification de la production : Une perspective inférentielle

Vladimir Modrak [1,*], Ranjitharamasamy Sudhakarapandian [2], Arunmozhi Balamurugan [2] et Zuzana Soltysova [1]

[1] Faculté des technologies de fabrication, Université technique de Kosice, 080 01 Prešov, Slovaquie ; zuzana.soltysova@tuke.sk
[2] École de génie mécanique, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, Inde ; sudhakarapandian.r@vit.ac.in (RS); arunmozhi.b@vit.ac.in (AB)
* Correspondance : vladimir.modrak@tuke.sk

Résumé : Dans cette étude, une revue systématique de l'ordonnancement de la production basée sur des techniques d'apprentissage par renforcement (RL) utilisant notamment l'analyse bibliométrique a été réalisée. L'objectif de ce travail est, entre autres, de souligner l'intérêt croissant pour ce domaine et de souligner l'influence du RL en tant que type d'apprentissage automatique sur l'ordonnancement de la production. Pour y parvenir, l'article explore la planification de la production à l'aide de RL en étudiant les métadonnées descriptives des publications pertinentes contenues dans les bases de données Scopus, ScienceDirect et Google Scholar. L'étude se concentre sur un large éventail de publications couvrant les années 1996 à 2024. Les résultats de cette étude peuvent servir de nouvelles perspectives pour les futurs efforts de recherche dans le domaine de la planification de la production à l'aide des techniques RL.

Mots-clés : analyse bibliométrique ; planification de la production ; apprentissage par renforcement

check for updates

Citation : Modrak, V. ; Sudhakarapandian, R. ; Balamurugan, A. ; Soltysova, Z. Une revue sur l'apprentissage par renforcement dans la planification de la production : une perspective inférentielle. Algorithmes 2024, 17, 343. https://doi.org/10.3390/a17080343

Rédacteurs académiques : Alexandre Dolgui, David Lemoine, María I. Restrepo et Frank Werner

Reçu : 6 juin 2024
Révisé : 28 juillet 2024
Accepté : 2 août 2024
Publié : 7 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

La planification de la production est considérée comme l'un des éléments les plus critiques de la gestion manufacturière pour aligner les activités de production sur les objectifs commerciaux, pour assurer un flux fluide des ressources de marchandises et pour soutenir la capacité de l'entreprise à rester compétitive sur le marché. Les algorithmes de planification jouent un rôle important dans l'amélioration de l'efficience et de l'efficacité de la production et font donc depuis longtemps l'objet de recherches approfondies dans divers domaines interdisciplinaires, tels que l'ingénierie industrielle, l'automatisation et les sciences de gestion [1]. Les tâches de planification de la production peuvent être résolues en utilisant trois principaux types de procédures étape par étape telles que les algorithmes exacts, les algorithmes heuristiques et les algorithmes méta-heuristiques [2,3]. Bien qu'un algorithme exact puisse théoriquement garantir la solution optimale, la complexité NP des problèmes majeurs les rend impossibles à résoudre de manière efficace et efficiente [4]. Les heuristiques utilisent un ensemble de règles pour créer des solutions de planification rapidement et efficacement sans tenir compte de l'optimisation globale. De plus, la création de règles repose fortement sur une compréhension approfondie des par Alors que les méta-heuristiques peuvent produire de bonnes solutions de planification dans un temps de calcul raisonnable, la manière dont les opérateurs de recherche les créent dépend considérablement de la situation spécifique [9-13]. De plus, le processus de recherche itératif pose des défis en termes de consommation de temps et d'applicabilité dans des scénarios en temps réel lorsqu'il s'agit de problèmes à grande échelle . Les approches de planification basées sur l'apprentissage par renforcement (RL) se sont révélées être un outil utile à cet égard. L'apprentissage par renforcement est un sous-domaine du domaine plus large de l'apprentissage automatique. RL est considérée comme l'une des approches les plus prospectives pour une planification coopérative robuste, qui permet aux responsables de production d'interagir avec un environnement de fabrication complexe, d'apprendre de leurs expériences antérieures et de sélectionner les décisions optimales. Cela implique le processus par lequel un agent sélectionne et exécute de manière autonome des actions pour accomplir une tâche donnée. L'agent apprend par l'expérience et vis

Machine Translated by Google

récompenses qu'il reçoit dans certains scénarios. L'objectif principal de RL est d'optimiser la récompense cumulée obtenue par un agent grâce à l'évaluation et à la sélection d'actions dans un environnement dynamique [14-20]. Le développement le plus récent de la technologie de l'intelligence artificielle a permis une application réussie de la RL dans des problèmes de prise de décision séquentielle avec des objectifs multiples utilisables pour la planification et le contrôle des robots [21,22]. La recherche sur la planification de la production à l'aide de RL depuis 1998 évolue à mesure que les techniques d'optimisation progressent par rapport aux métaheuristiques. RL améliore considérablement l' efficacité informatique pour résoudre les problèmes de planification. De nombreux articles de recherche de RL sur la planification de la production ont été entrepris depuis sa création en 1996, établissant une base substantielle et précieuse (voir quelques-uns d'entre eux, par exemple [23-27]), mais seuls deux articles de synthèse récents [28,29] ont porté sur attention à ce sujet. Cependant, les deux articles publiés en 2021 se concentraient sur des perspectives différentes des algorithmes de planification basés sur RL par rapport à celles présentées dans cet article. En d'autres termes, le contenu de l'article proposé se situe dans une relation disjonctive et complémentaire aux deux études mentionné

La question de recherche de cet article comporte de multiples facettes, comprenant de multiples caractéristiques de ce domaine qui nécessitent des réponses distinctes. Les principales caractéristiques abordées dans cet article sont les suivantes : Quels sont les principaux domaines de recherche émergents dans le domaine ? Quels sujets connexes sont abordés dans la planification de la production basée sur RL ? Quels sont les domaines de mise en œuvre typiques au sein de RL appliqués à la planification de la production ?

Sa nouveauté réside dans la fourniture d'analyses et d'évaluations supplémentaires concernant le RL dans la planification de la production. Il englobe, par exemple, la tendance des citations de RL sur la planification de la production, les auteurs les plus influents dans ce domaine, les sources les plus pertinentes dans le domaine, la comparaison des types déterministes de méthodes de planification et des types incertains de méthodes de planification.

2. Matériels et méthodes

L'approche bibliométrique utilisée ici, comme principale méthode de recherche, présente un instrument quantitatif pour suivre et représenter le progrès scientifique en examinant et en visualisant l'information scientifique. L'acceptation croissante des méthodes bibliométriques dans plusieurs disciplines académiques indique que leur utilisation apporte les effets attendus [30–33].
La présente enquête sur la planification de la production basée sur RL est menée en utilisant la procédure comprenant cinq phases cohérentes, comme illustré dans la figure 1.

De plus, cette recherche suit une approche inférentielle, où l'échantillon de population est exploré pour déterminer ses caractéristiques [34,35]. De plus, le concept inférentiel de représentation scientifique proposé par Suárez [36] a été appliqué ici pour formuler les résultats de la recherche. Son essence est d'utiliser un raisonnement alternatif pour atteindre des résultats qui diffèrent de la vision isomorphe de la représentation scientifique dans le sens où la connaissance empirique joue un rôle important dans le raisonnement inductif [37,38].

Machine Translated by Google

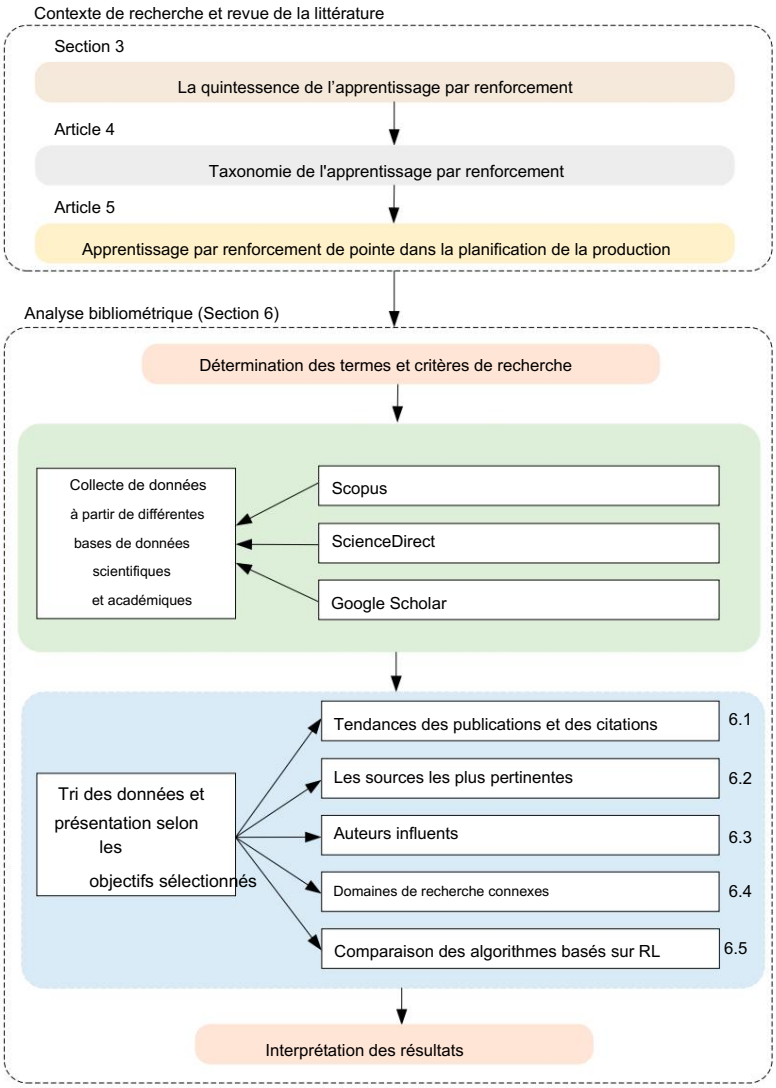


Figure 1. Cadre méthodologique de recherche.

3. La quintessence de l'apprentissage par renforcement

L'intelligence artificielle (IA) est devenue un aspect intégral et important de notre vie quotidienne à l'approche de la fin de cette décennie. L'intelligence artificielle est présente depuis sept décennies, mais a pris un essor considérable au cours des trois dernières décennies, conduisant à des recherches approfondies dans ce domaine [39]. L'IA fait référence à diverses techniques permettant aux ordinateurs d'acquérir des connaissances et de faire des choix en analysant des données. Parmi ces techniques, l'apprentissage automatique (ML) a réalisé des progrès significatifs au cours des deux dernières décennies, passant d'une simple curiosité en laboratoire à une technologie pratique largement utilisée dans les applications commerciales. En IA, le ML est devenu l'approche privilégiée pour créer des logiciels fonctionnels dans des domaines tels que la vision par ordinateur, le traitement du langage naturel, la reconnaissance vocale, le contrôle des robots et diverses autres applications [40].

De nombreux développeurs d'IA reconnaissent désormais que, dans plusieurs cas, il est plus pratique de former un système en démontrant un comportement d'entrée-sortie idéal plutôt que de le programmer manuellement pour anticiper la réponse souhaitée pour chaque entrée potentielle. Le ML a eu un impact significatif sur divers domaines de l'informatique et sur les secteurs qui traitent des problèmes liés aux données. Cela comprend les services aux consommateurs, le diagnostic des défauts dans des systèmes complexes et la gestion des chaînes logistiques. Les approches ML ont eu des effets divers sur plusieurs sciences empiriques, telles que la biologie, la cosmologie et les sciences sociales. Ces approches ont été utilisées de manière créative pour examiner de vastes données expérimentales [41].

Machine Translated by Google

Trois paradigmes majeurs du ML sont l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement [15]. L'apprentissage supervisé implique le processus de formation d'un modèle à l'aide d'un ensemble de données qui a été étiqueté, ce qui signifie que chaque entrée est associée à sa sortie précise correspondante. L'objectif est d'acquérir des connaissances sur une transformation d' entrées données en sorties correspondantes, qui peuvent ensuite être utilisées pour faire des prévisions précises sur des données nouvelles et non observées. L'apprentissage non supervisé implique l'identification de modèles cachés ou de structures inhérentes dans des données d'entrée non étiquetées. Le modèle vise à acquérir des connaissances sur la distribution ou le regroupement fondamental des données sans instructions explicites sur le résultat souhaité [28]. L'objectif du RL est d'apprendre à un agent à faire de nombreux choix en s'engageant dans un environnement et en recevant des récompenses ou des punitions en guise de feedback . L'agent acquiert la capacité d'optimiser ses activités pour maximiser les récompenses totales qu'il reçoit sur une période [42].

En raison de sa méthodologie dynamique et adaptative, RL est très efficace pour résoudre les problèmes dans le domaine de la recherche sur les systèmes de fabrication et la robotique [22,28]. Le RL se distingue de l'apprentissage supervisé, où un modèle est formé sur un ensemble de données prédéterminé, et de l'apprentissage non supervisé, où le modèle découvre des modèles cachés dans les données. Dans l'apprentissage par renforcement, un agent est formé à prendre une série de décisions en s'engageant dans un environnement. L'objectif de l'agent est d'acquérir une stratégie, parfois appelée politique, qui optimise les récompenses totales obtenues au fil du temps [43].

RL est un processus de prise de décision influencé par la psychologie comportementale. Cela implique qu'un agent apprenne à atteindre un objectif en agissant et en recevant des commentaires de l'environnement sous forme de récompenses ou de pénalités. Cette méthode d'essais et d'erreurs s'apparente à la façon dont les humains et les animaux apprennent de leur environnement. Les composants critiques de RL sont l'agent, l'environnement, les actions, les états et les récompenses. L'agent observe l'environnement, prend des mesures et reçoit des récompenses, qui sont utilisées pour guider les actions futures [14]. RL est particulièrement efficace dans les tâches aux résultats incertains, telles que jouer à des jeux, contrôler des robots et gérer des ressources. Dans ces situations, l'agent doit équilibrer l'exploration (essayer de nouvelles actions) et l'exploitation (utiliser des actions établies qui conduisent à des récompenses significatives) pour développer efficacement une politique optimale [42].

Le cadre mathématique du RL est généralement représenté à l'aide des processus de décision markoviens (MDP), qui offrent une approche systématique de la modélisation des problèmes de prise de décision , incluant à la fois des résultats aléatoires et contrôlés par un agent. Résoudre un MDP implique de déterminer une politique qui prescrit l'action optimale à entreprendre à chaque étape, dans le but de maximiser les récompenses cumulées prévues dans le futur. Les progrès récents en matière d'apprentissage profond ont considérablement amélioré les capacités du RL [15,16]. La combinaison de réseaux neuronaux profonds et d'algorithmes RL a abouti à l'émergence de l'apprentissage par renforcement profond (DRL), qui a démontré des résultats exceptionnels dans des situations complexes. Parmi les exemples marquants figurent AlphaGo, développé par Google, qui a triomphé des champions humains dans le jeu de Go, ainsi que diverses implémentations dans la conduite autonome, où les algorithmes DRL acquièrent la capacité de naviguer avec succès dans un environnement complexe et en constante évolution . L'apprentissage par renforcement reste un domaine dynamique et en évolution rapide dans l'apprentissage automatique, où les recherches en cours repoussent constamment les limites de ce que les agents artificiels peuvent accomplir. En approfondissant les complexités du RL, nous découvrons de nouvelles opportunités pour les systèmes intelligents possédant la capacité d'acquérir des connaissances et d'ajuster leur comportement en temps réel. RL ouvre la voie à un avenir où les ordinateurs pourront coopérer de manière transparente avec les gens et fonctionner de manière autonome d

4. Taxonomie des algorithmes d'apprentissage par renforcement

Sutton et Barto [42] soulignent le rôle crucial des trois composantes essentielles dans le processus RL : • Une politique détermine les actions à entreprendre dans chaque état environnemental. • Un signal de récompense classe ces actions comme bénéfiques ou préjudiciables en fonction du résultat immédiat de la transition entre les états.

Machine Translated by Google

• Une fonction de valeur évalue l'efficacité à long terme des actions en considérant non seulement la récompense immédiate d'un État mais également les récompenses futures anticipées.

L'objectif premier de l'agent est d'optimiser la récompense cumulée. Par conséquent, le signal de récompense constitue le fondement de la modification de la politique et de la fonction de valeur. Dans les systèmes RL spécifiques, il existe un composant supplémentaire appelé modèle d'environnement, qui est facultatif. Il reproduit le comportement du climat ou permet des hypothèses plus complètes sur le comportement de l'environnement. Les modèles sont utilisés pour la planification stratégique, qui implique de prendre des décisions basées sur l'anticipation de scénarios futurs avant qu'ils ne se produisent réellement [24]. Les méthodes basées sur des modèles sont utilisées pour gérer les problèmes d'apprentissage par renforcement en utilisant des modèles et une planification, contrairement aux méthodes plus simples sans modèle. Ils présentent un comportement d'apprentissage apparent par essais et erreurs et sont considérés comme l'inverse de la planification [26,27]. Les processus RL peuvent être représentés et analysés à l'aide d'un cadre mathématique appelé MDP. Le modèle mathématique stochastique peut être formalisé à l'aide d'un 5-tuple (s, a, p, r, γ), dans lequel le premier symbole « s » représente la collection finie de tous les états environnementaux possibles, tandis que st est l'état à un moment donné. heure précise 't'. Lettre « a » représente l'ensemble de toutes les actions possibles, tandis que l'action entreprise au moment « t » est marquée comme étant à . Le symbole « p » fait référence à la matrice des probabilités de transition, qui définit la probabilité conditionnelle de transition vers un nouvel état « s », avec une récompense « r », étant donné l'état actuel « s », et une action « a » (pour tous les états et actions) [44]. La figure 2 montre le RL du cycle de planification de la production.

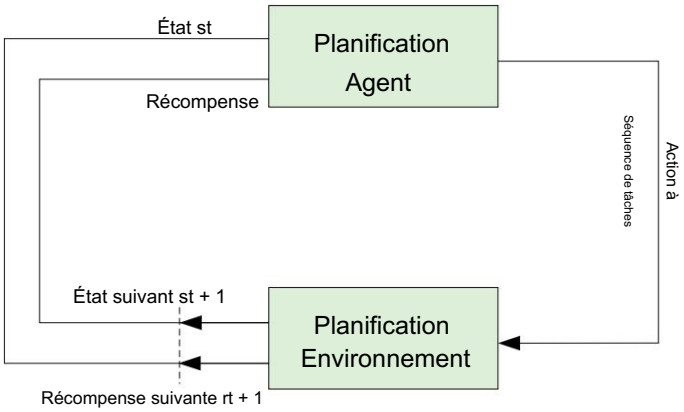


Figure 2. Cycle de planification de la production RL.

Les méthodologies d'apprentissage des algorithmes ou des agents RL peuvent être classées en deux catégories distinctes, basées sur un modèle et sans modèle. Le RL basé sur un modèle est souvent appelé apprentissage indirect. L'agent utilise un modèle prédictif pour apprendre la politique de contrôle de l' environnement à travers un nombre limité d'interactions. L'agent applique ensuite ce modèle aux épisodes suivants pour obtenir des récompenses [15]. Le RL sans modèle, ou apprentissage direct, fait référence à l'apprentissage dans lequel un agent apprend à prendre des décisions sans construire explicitement un modèle de l'environnement. L'agent acquiert des connaissances sur la politique de contrôle grâce à l'apprentissage expérientiel de l'environnement, en utilisant des méthodes d'essais et d'erreurs pour optimiser les récompenses sans s'appuyer sur un modèle préexistant. Cette approche met en valeur l'adaptabilité des algorithmes RL, leur permettant d'apprendre et d'évoluer dans des environnements dynar

Comme il existe de nombreux algorithmes RL différents, il est judicieux de comprendre la différence entre eux. A cet effet, des systèmes de classification ont été établis pour les catégoriser selon différents critères. Par exemple, ALMahamid et Grolinger [46] ont proposé de catégoriser les algorithmes RL en fonction du type d'environnement. Les algorithmes RL peuvent également être classés du point de vue de la politique : apprentissage sur politique ou hors politique [47,48]. Les RL sont principalement classés en fonction de leurs approches d'apprentissage [15,45]. À cet égard, une classification mise à jour des algorithmes RL du point de vue des approches d'apprentissage utilisées est fournie dans la figure 3.

Machine Translated by Google

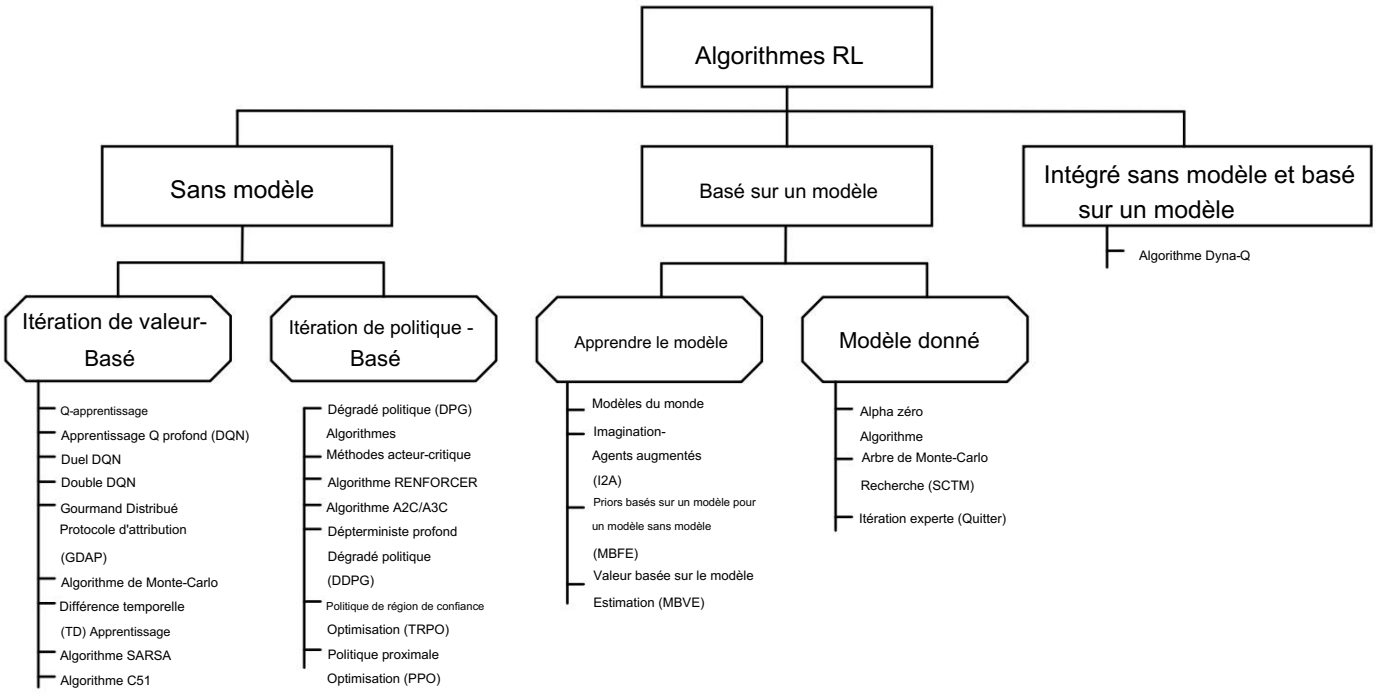


Figure 3. Classification des algorithmes RL.

5. État de l'art de l'apprentissage par renforcement dans la planification de la production

La planification déterministe englobe l'organisation et la planification de tâches selon des paramètres fixes et des contraintes connues. Son objectif est d'optimiser l'efficacité de la production et de minimiser délais d'exécution. Un travail complet [49] explore les principes fondamentaux et les techniques avancées de la planification déterministe, fournissant un aperçu de divers algorithmes.
et leurs applications. Les travaux de Pinedo [50] examinent en profondeur l'ordonnancement déterministe théorie, algorithmes et systèmes pratiques, soulignant leur importance dans la fabrication et les opérations de service. La planification Flowshop implique le séquençage des opérations sur plusieurs machines dans un ordre fixe afin de minimiser la durée de fabrication ou le temps de réalisation total. Allahverdi [51] discute de divers problèmes de planification de flowshop, y compris les configurations et examine les algorithmes et les approches pour y remédier. Panwalkar et Smith [52] fournissent une enquête fondamentale de recherches classiques et contemporaines sur la planification des flowshops, couvrant à la fois les méthodes heuristiques.

Les techniques d'apprentissage par renforcement (RL) ont gagné en popularité, affirmant l'intérêt pour les modèles basés sur des agents. Les études précédentes se sont principalement concentrées sur l'utilisation de RL pour résoudre les problèmes de planification des ateliers. Le secteur manufacturier est confronté à des défis tels tels que la satisfaction du client, la dégradation du système, la durabilité, les stocks et l'efficacité, ayant un impact sur la durabilité et la rentabilité de l'usine. Industrie 4.0 et offre de fabrication intelligente des solutions pour des opérations optimisées et des produits de haute qualité. Paraschos et col. [53] intégrer la RL à la production Lean Green pour créer un environnement de production durable, réduire l'impact environnemental grâce à une consommation de matériaux et un cycle de vie minimisés extension via des pratiques de production tirée, de maintenance prédictive et d'économie circulaire. Une analyse expérimentale rigoureuse valide son efficacité pour améliorer la durabilité et réutilisation des matériaux.

Récemment, des progrès significatifs ont été réalisés dans l'utilisation de RL pour résoudre plusieurs problèmes d'optimisation combinatoire, notamment la planification de la production, le problème de routage des véhicules et Problème de voyageur de commerce [54-59]. Dans RL, une tâche de planification de production peut être considérée comme un environnement dans lequel un agent opère, développant une politique grâce à une formation hors ligne par interagir avec cet environnement. Cette approche offre une nouvelle façon d'aborder la planification défis, nécessitant des contraintes strictes en temps réel, telles que la planification dynamique des ateliers de travail problèmes [22,60–64].

Machine Translated by Google

Dans les problèmes de planification de la production, des algorithmes RL basés sur des valeurs sont couramment utilisés, notamment le Q-learning, l'algorithme de différence temporelle TD(λ) , SARSA, ARL, le Q-learning informé, le dual Q-learning, le Q-learning approximatif, la descente de gradient TD(λ) , partage des revenus, apprentissage Q-III, RL relationnel, SMART détendu et apprentissage TD(λ). Dans l'apprentissage par renforcement profond (DRL), de nombreuses approches basées sur la valeur sont utilisées, telles que le DQN (Deep Q-Learning Networks), le DRL faiblement couplé, le DQN multiclasse et l' algorithme Q-network [43, 44, 65-73] .

Qu et al. [25] ont appliqué le Q-learning par approximation multi-agents, démontrant son efficacité à travers des expériences numériques dans des environnements statiques et dynamiques et divers scénarios de flow shop. L'objectif était de créer et de mettre en œuvre un calendrier de fabrication efficace prenant en compte des interactions réalistes entre les compétences de la main-d'œuvre et les machines adaptatives.
Luo [74] utilise la méthode DRL pour aborder la planification dynamique et flexible des ateliers, en se concentrant sur des scénarios visant à réduire les retards. Luo et coll. [75] ont d'abord utilisé « l'optimisation hiérarchique de la politique proximale multi-agents » (HMAPPO) pour le problème d'atelier de travail flexible multi-objectifs partiel sans attente (MOFJSP) en constante évolution avec de nouvelles insertions de tâches et des pannes de machines.

La détérioration des machines pendant la production augmente les dépenses opérationnelles et provoque des perturbations du flux de travail, nécessitant une maintenance corrective coûteuse. La maintenance préventive prolonge la durée de vie des machines mais entraîne des temps d'arrêt et des coûts. L'étude [76] répond à ces défis en optimisant les calendriers de production et de maintenance dans les systèmes multi-machines, en introduisant l'apprentissage par renforcement amélioré des connaissances pour améliorer l'efficacité du RL pour guider les décisions de production et favoriser la collaboration entre les machines. Des évaluations comparatives dans des scénarios déterministes et stochastiques mettent en évidence la capacité de l'algorithme à maximiser les récompenses commerciales et à prévenir les échecs.

Selon Wang et al. [54], le processus de planification de la production implique la fabrication de divers articles en utilisant un modèle de production hybride intégrant le modèle multi-agent Deep Reinforcement Learning (MADRL). Popper et coll. [60] ont proposé MADRL pour optimiser les usines de production flexibles en tenant compte de facteurs tels que l'efficacité et les objectifs environnementaux.
Du et coll. [77] ont utilisé la méthode Deep Q-Network (DQN) pour résoudre le problème de planification flexible de l'atelier de tâches (FJSP) au milieu de l'évolution des taux de traitement, du temps de configuration, du temps d'inactivité et du transport des tâches. Cette approche intègre des indicateurs d'état et des actions pour améliorer l' efficacité de la composante DQN. De plus, il comprend un composant d'analyse exploratoire des données (EDA) axé sur les problèmes pour améliorer l'exploration des données.

Li et coll. [78] ont utilisé une méthode d'apprentissage par renforcement profond (DRL) pour résoudre le problème discret de l'atelier flexible avec réutilisation inter-outils (DFJSP-ITR), abordant le problème d'optimisation multi-objectif consistant à minimiser la durée de fabrication combinée et la consommation totale d'énergie. La solution proposée comprend quelques caractéristiques d'état génériques, un espace d'action basé sur la programmation génétique et une fonction de récompense. Zhou et coll. [79] ont proposé d'utiliser des stratégies de planification en ligne basées sur RL avec des fonctions de récompense composites pour améliorer l'efficacité et la robustesse des systèmes industriels. Les travaux [18] ont utilisé un algorithme DRL avancé pour optimiser la planification de la production dans des ateliers complexes, mettant en évidence des avantages tels qu'une adaptabilité améliorée, une visibilité globale et une optimisation. Certains auteurs se sont concentrés sur le développement de divers algorithmes DRL capables de formuler des stratégies complexes de planification de la production. Par exemple, Luo et al. [26] ont créé un cadre de reprogrammation en ligne pour le problème de planification dynamique et multi-objectifs des ateliers flexibles, permettant de minimiser les retards totaux ou de maximiser le taux d'utilisation de la machine. Zhou et coll. [27] ont utilisé la technique du Q-learning profond pour aborder la planification dynamique dans la fabrication intelligente. Une autre méthode de planification dynamique utilisant le RL profond est proposée dans [80], employant l'optimisation de la politique proximale pour déterminer la politique de planification idéale. Wang et Usher [81] ont examiné l'utilisation de l'algorithme de renforcement Q-learning pour la planification de la production basée sur les agents. L' intégration de RL avec la planification de la production représente une avancée majeure dans l'optimisation des processus de fabrication. À mesure que la recherche et la technologie continuent d'évoluer, elles promettent une efficacité et une adaptabilité encore plus grandes dans l'industrie. Plonger dans les données bibliométriques peut

Machine Translated by Google

fournir des informations précieuses sur les tendances de la recherche, les travaux influents et les principaux contributeurs dans ce domaine en développement rapide.

L'Industrie 4.0 et la Smart Industry sont à la pointe de la fabrication et de la production contemporaines. L'Industrie 4.0 intègre des technologies avancées telles que l'Internet des objets (IoT), l'intelligence artificielle (IA), le Big Data et les systèmes cyber-physiques pour créer des environnements de production hautement automatisés et en réseau. Le concept d'industrie intelligente met l'accent sur l'analyse des données en temps réel, la maintenance prédictive et les processus de fabrication adaptables pour améliorer l'efficacité, la flexibilité et la production [27]. La mise en œuvre d'algorithmes RL est cruciale dans ces paradigmes, facilitant l'apprentissage automatique et la prédiction future grâce aux interactions dans l'environnement de production. Cela améliore la planification, l'allocation des ressources et le contrôle qualité, permettant aux entreprises d'atteindre une automatisation et une précision élevées dans la prise de décision , conduisant à une fabrication moderne plus résiliente et plus réactive [54]. Le tableau 1 illustre la contribution des algorithmes RL à la planification de la production

Tableau 1. Apports de l'apprentissage par renforcement sur la planification de la production.

Type de planification	Récompense	Contribution
Catégorie : Environnement		
Atelier de travaux complexes [18]	Pénalités	Algorithme DRL avancé optimisant la planification de la production, mettant en évidence l'adaptabilité, la visibilité et l'optimisation.
Boutique de flux [25]	Commande terminée ou pénalité de retard	Calendrier de fabrication efficace tenant compte des compétences de la main-d'œuvre et des machines adaptatives.
Atelier de travail flexible [74]	Retard/taux d'utilisation des machines ; coût de l'énergie; produits en retard	Reprogrammation en ligne minimisant les retards ou maximisant l'utilisation de la machine. Méthode MARL optimisant les objectifs d'efficacité et d'environnement.
Atelier de travail [54]	Pénalités	Technique DQN pour minimiser le temps d'exécution des tâches dans la planification dynamique. MADRL optimise les performances au niveau du système en tenant compte des interactions entre les pièces.
Catégorie : Méthode		
Atelier de travaux complexes [18]	Pénalités	Algorithme DRL pour optimiser la planification, améliorer l'adaptabilité, la visibilité et l'optimisation.
Boutique de flux [25]	Commande terminée ou pénalité de retard	Horaire efficace tenant compte des compétences de la main-d'œuvre et des interactions avec les machines.
Atelier de travail flexible [24]	Retard/taux d'utilisation des machines ; coût de l'énergie; produits en retard	Apprentissage par renforcement minimisant les délais d'exécution dans la planification dynamique des tâches.
Catégorie : Récompense		
Atelier de travaux complexes [62]	Pénalités	Planification des tâches avec plusieurs agents dans l'apprentissage par préemption des ressources. Optimisation des usines de production pour plus d'efficacité et de durabilité.
Boutique fluide [60]	Commande terminée ou pénalité de retard ; coût de l'énergie; produits en retard Retard/ taux d'utilisation des machines ; coût de l'énergie; produits en retard	Aborder l'algorithme de renforcement dans la planification, visant la prise de décision dans Flexible Job Shop.
Atelier de travail flexible [19]		

6. Analyse bibliométrique des études sur le RL dans la planification de la production

L'analyse bibliométrique du RL dans le contexte de la planification de la production se concentre ici sur ses fondements théoriques et ses implications pratiques, entre les années 1996 et 2024. Les deux premières phases impliquent la collecte de données bibliographiques, collectées via la base de données Scopus. La recherche a été limitée à cette base de données en raison de son importance en tant que référentiel complet de littérature scientifique et de son utilisation fréquente dans les évaluations universitaires. Les critères d'inclusion aux fins de l'analyse présentés dans les sections 6.1 à 6.4 englobaient les publications contenant les termes « apprentissage par renforcement », « production » et « planification ». En outre, les termes « apprentissage par renforcement » et « ordonnancement déterministe » ou « apprentissage par renforcement » et « ordonnancement incertain » ont été utilisés pour l'analyse effectuée dans la section 6.5. Au cours des deux phases suivantes de tri et de présentation des do

Machine Translated by Google

Les logiciels Microsoft Excel et VOSviewer 1.6.20 ont été utilisés pour extraire les informations nécessaires, telles que les résultats scientifiques annuels, les sources les plus pertinentes, l'auteur le plus cité et la cooccurrence des mots clés.

Le processus d'études sélectionnées pour la méthodologie de recherche mentionnée est synthétisé dans Figure 4, selon les directives PRISMA.

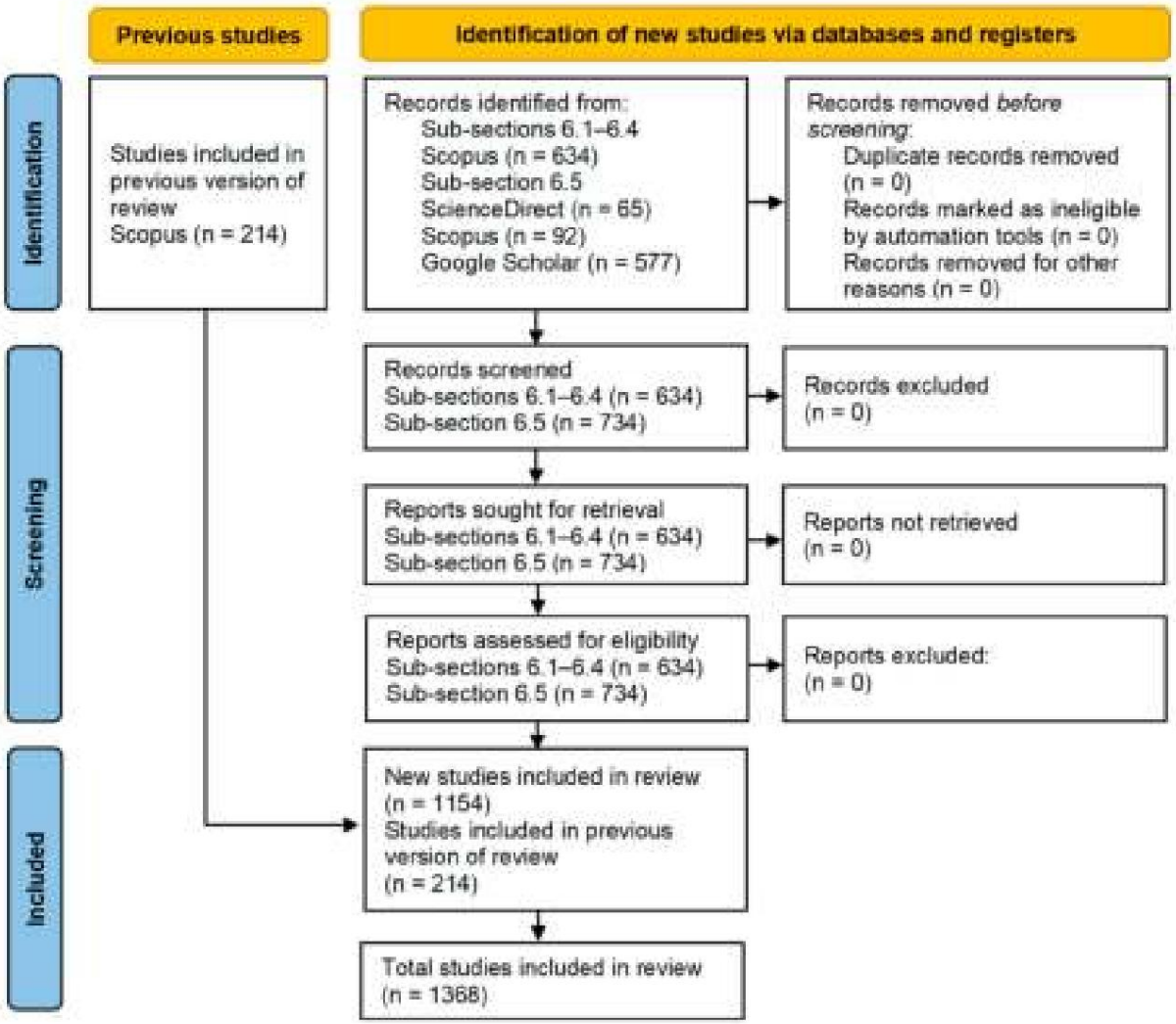


Figure 4. Diagramme PRISMA 2020 pour les examens systématiques mis à jour qui comprenaient des recherches dans les bases de données.

En utilisant les mots-clés et les critères de recherche spécifiés ci-dessus (pour les sections 6.1 à 6.4 n = 634 publications et pour la section 6.5 n = 734 publications), 1 368 articles co-écrits ont été trouvés jusqu'au 27 mai 2024, date à laquelle tous les types de documents ont été inclus dans la revue. analyse (nouvelles études n = 1 154 et études incluses dans la version précédente de la revue n = 214).

L'analyse des sections 6.1 à 6.4 a inclus 634 articles co-écrits recherchés dans « Titre de l'article, résumé et mots-clés ». Les principaux domaines de recherche qui reçoivent une attention particulière dans les articles comprennent l'informatique, l'ingénierie, les mathématiques, les sciences de la décision , les affaires, la gestion et le comptage, l'énergie, le génie chimique, la science des matériaux et d'autres sujets. Les 634 publications de notre échantillon ont été classées en 19 thèmes de recherche distincts. Les huit principaux domaines de recherche, ainsi que la répartition de leurs articles, sont présentés dans le tableau 2.

Machine Translated by Google

Tableau 2. Disciplines de recherche associées.

Domaines de recherche	Nombre de publications	Pourcentage
L'informatique	429	32,7%
Ingénierie	402	30,6%
Mathématiques	159	12%
Sciences de la décision	85	6,5%
Affaires, gestion et comptage	49	3,7%
Énergie	44	3,4%
Génie Chimique	34	2,6%
Science des matériaux	25	2%
Autres	85	6,5%

6.1. Tendances des publications et des citations

Le nombre d'articles publiés et leurs citations fournissent généralement des informations suffisamment fiables informations permettant d'anticiper le développement ultérieur du domaine de recherche examiné. Considérant le nombre de publications et de citations dans le domaine de la planification de la production par l'apprentissage algorithmes conservant environ 29 ans de données, le graphique d'analyse des tendances a été dérivé. Pour à cet effet, les mêmes termes de recherche et mots-clés ont été appliqués qu'en cas d'identification des grandes catégories de recherche, mais les types de documents ont été étendus à tous les types. La raison du changement de type de document était de connaître les premières initiatives de recherche dans ce domaine. domaine. Les mêmes conditions de recherche ont été appliquées dans le reste de l'article pour inclure le un plus grand échantillon de publications aux fins de l'enquête. La répartition annuelle des publications (sur un total de 634 éléments) et leurs citations issues de la même base de données au cours la période de 1996 à 2024 est illustrée à la figure 5.

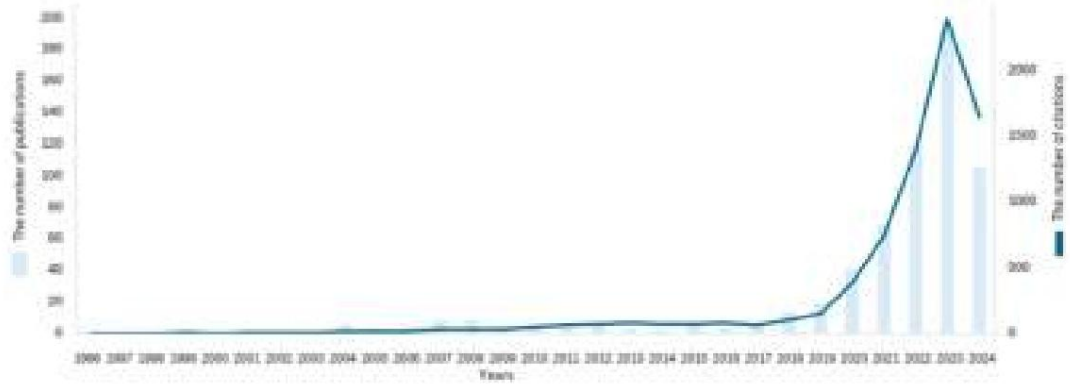


Figure 5. Tendances de publication et de citation de RL dans le domaine de la planification de la production de 1996 à 2024.

Un examen de la production scientifique annuelle entre 1996 et 2018 a largement démontré un faible nombre d'articles publiés chaque année, relativement stable. Tout au long de ces deux décennies, la nécessité du RL dans la planification de la production en conditions réelles apparemment n'est pas apparu. Au cours de la période 2019-2023, il y a eu une augmentation notable du nombre de publications enregistrées dans la base de données la plus reconnue pour son contenu évalué par les pairs. Ce phénomène peut être principalement attribué aux progrès de l'intelligence artificielle. Il convient de souligner que si cette tendance exponentielle à l'augmentation du nombre de les publications se poursuivent comme on peut le voir en 2024, alors on peut anticiper qu'au cours de la Au cours de la prochaine décennie, l'importance du RL dans la planification de la fabrication augmentera considérablement.

6.2. Sources les plus pertinentes

Une identification des publications les plus pertinentes à partir d'un ensemble de données initial présente approche courante dans la recherche bibliométrique puisque ces sources publient généralement des recherches influentes qui suscitent un large intérêt. En règle générale, les revues les plus productives ont la plus grande influence sur le développement de la science dans un domaine particulier puisqu'ils

Machine Translated by Google

publier plus d'articles et générer plus de citations [77-79]. Quant à la pertinence de la littérature dans le domaine exploré, les dix revues ayant publié le plus d'articles sont identifiées ici. En outre, les dix revues scientifiques les plus citées sont mentionnées dans cette sous-section. Les figures 6 et 7 classent les revues selon ces deux critères pour montrer qu'elles représentent des documents qui présentent la plus grande pertinence pour RL dans la planification de la production.

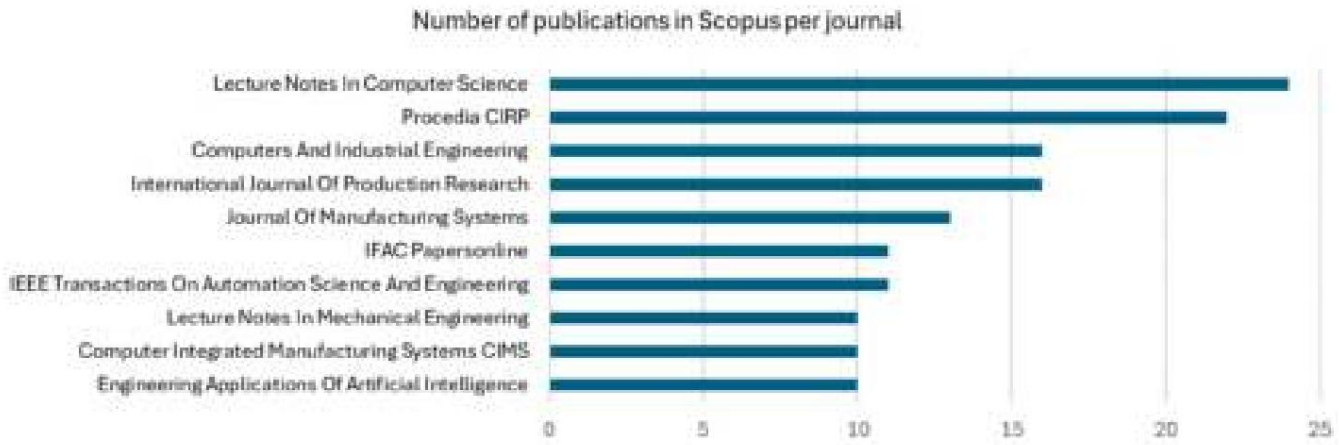


Figure 6. Dix revues les plus publiées pour RL dans le domaine de la planification de la production de 1996 à 2024.

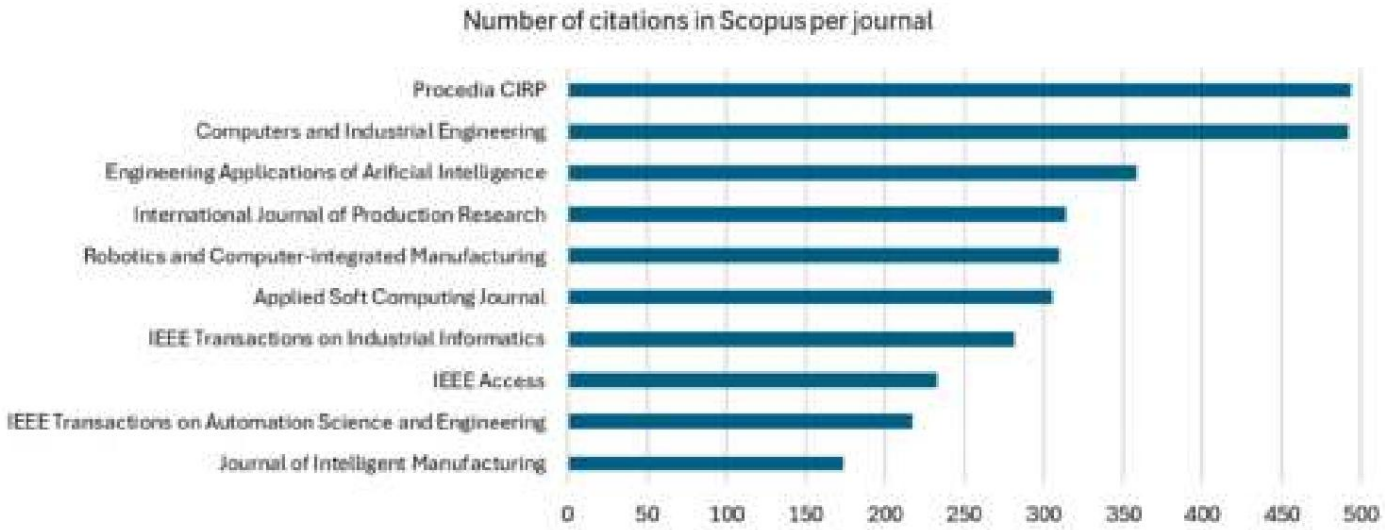


Figure 7. Dix revues les plus citées pour RL dans le domaine de la planification de la production de 1996 à 2024.

Les notes de cours en informatique, Procedia CIRP et Computers and Industrial Engineering ainsi que les douze autres revues sont considérées comme les publications scientifiques les plus pertinentes dans ce domaine. On peut également noter que les quinze revues répertoriées dans les figures 6 et 7, quels que soient les résultats des mesures utilisées, peuvent être empiriquement classées comme largement reconnues pour la diffusion de recherches avancées sur l'apprentissage par renforcement appliqué à la planification de la production. De plus, leur rigueur scientifique est également indiquée par le fait que sur les quinze revues identifiées, huit répondaient à des normes de qualité élevées car elles sont indexées pour les revues Current Content Connect avec un impact vérifiable sur l'orientation des pratiques et des comportements de recherche [80,81] .

6.3. Auteurs les plus cités

Depuis les années 1996, de nombreux auteurs ont apporté une contribution significative au développement de ce domaine. Dans cette sous-section, l'intention est de présenter certains des auteurs qui ont apporté d'importantes contributions intellectuelles à la recherche. L'analyse des auteurs les plus c

Machine Translated by Google

a été réalisée à partir des données de la base de données Scopus. De plus, une analyse de co-citation a été réalisée pour chaque source de publication (sur un total de 634 éléments) afin de révéler le réseau entre les études. A cet effet, le logiciel VOSviewer [82] a été utilisé. Pour obtenir des informations pertinentes et une représentation graphique claire de relations complexes, les filtres suivants ont été utilisés. Filtre 1 : Nombre maximum d'auteurs par document : 15 ; Filtre 2 : Nombre minimum de documents d'un auteur : 3 ; Filtre 3 : Nombre minimum de citations d'un auteur : 10. De plus, la méthode de comptage complet a été appliquée, ce qui signifie que les publications ayant des co-auteurs de plusieurs pays sont comptées comme une publication complète pour chacun de ces pays. Le réseau de co-citation de l'échantillon sélectionné de chercheurs utilisant ces paramètres est visualisé dans la figure 8. tt

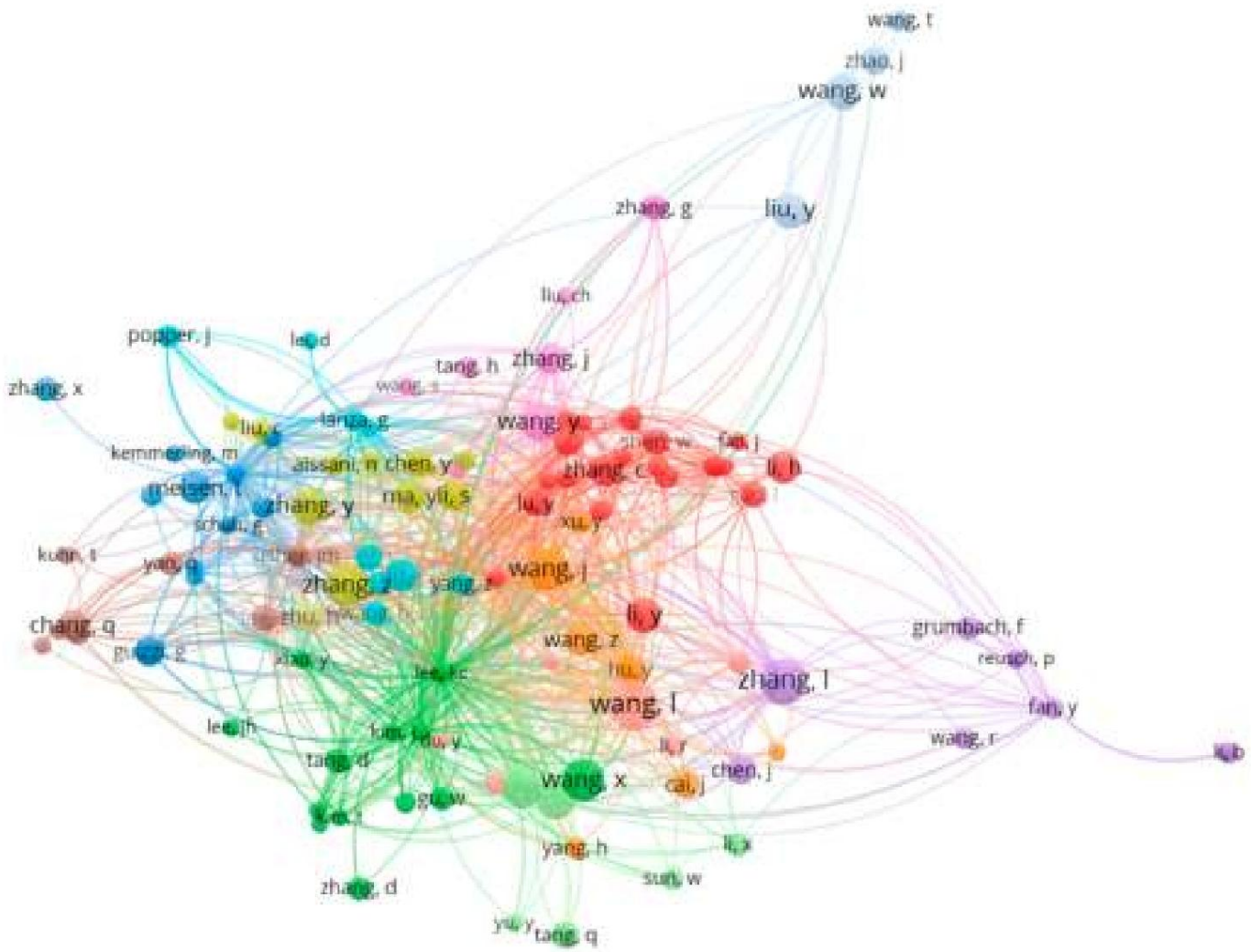


Figure 8. Réseau de co-citation des auteurs ayant contribué de manière significative au développement de ce domaine de 1996 à 2024.

Cette carte du réseau de co-citation montre, entre autres, les chercheurs qui ont reçu le plus grand nombre de citations au cours des 29 dernières années. Sur les 1376 auteurs cités, 138 répondent aux critères mentionnés ci-dessus. Chaque chercheur de chaque publication incluse est représenté par un nœud dans ce réseau. La taille de chaque nœud indique la fréquence de citation des travaux scientifiques du sujet. Une frontière est tracée entre deux nœuds si les deux savants ont été cités par un document commun. Pour classer les chercheurs influents dans un domaine donné en fonction de leurs taux de citation, les dix auteurs les plus cités ont été sélectionnés. Ces dix auteurs sont répertoriés dans le tableau 3.

Machine Translated by Google

Tableau 3. Auteurs les plus influents.

Nom	Nombre de Citations	Pays	Nom	Nombre de Citations	Pays
Waschenck, B.	386	Allemagne	Wang, YC	265	Taïwan
Wang, L.	375	Chine	Wang, H.	224	Chine
Wang, J.	354	Chine	Lin, CC	216	Taïwan
Bauernhansl, T.	342	Allemagne	Li, S.	209	Chine
Huissier, JM	265	USA	Beldjilali, B.	207	Algérie

6.4. Identification des domaines de recherche connexes

L'objectif de cette partie de l'article est d'identifier les domaines de recherche dans lesquels la problématique de la planification de la production basée sur RL est intéressante. A cet effet, l'analyse des mots clés L'application a été utilisée pour aider à séparer les thèmes de recherche importants qui ont a suscité un grand intérêt de la part des chercheurs de la part de chercheurs moins importants. Pour obtenir des informations pertinentes et catégories représentatives sans compter les catégories moins significatives, le paramètre suivant a été utilisé : nombre minimum d'occurrences de mots-clés : 14. Sur la base de cette restriction, 88 les mots-clés atteignent le seuil sur un total de 3 767 mots-clés. Ces principaux mots-clés qui ont été produits automatiquement à partir des titres des journaux selon un calendrier de production basé sur sur RL ainsi que leur apparition sont illustrés à la figure 9.

Comme on peut le constater, les sujets peuvent être divisés en cinq groupes sur la base d'un algorithme informatique, tandis que chaque cluster a une couleur différente comme le montre la figure 9. La cooccurrence des mots-clés La carte met en évidence ces clusters où plus la couleur est foncée, plus la valeur de densité est élevée. Il permet d'identifier des sujets de recherche pertinents et leurs relations mutuelles. Basé sur le obtenu des résultats bibliométriques extraits de VOSviewer, dix sujets connexes très proches du domaine de recherche exploré ont été identifiés comme le montre le tableau 4.

Les résultats du tableau 4 montrent leur cohérence avec la réalité pratique et le fonctionnement objectifs de recherche. Par exemple, ces dernières années, il y a eu un intérêt manifestement accru dans l'utilisation de l'apprentissage par renforcement pour l'optimisation des tâches de planification des tâches en temps réel [76–86]. Ce fait peut être corrélé à la tendance continue de personnalisation de masse dans la production de biens de consommation [87,88]. Comme on le sait, la personnalisation de masse est caractéristique de répondre aux exigences des utilisateurs en évolution dynamique dans le temps, tandis que les produits personnalisés ont besoin à réaliser dans des délais différents. En conséquence, une planification efficace des tâches en temps réel les algorithmes basés sur le DRL deviennent indispensables. La prochaine méthode importante classée parmi les 10 principaux mots-clés de cooccurrence se trouve le contrôle de la production. Il utilise un contrôle différent techniques pour atteindre les objectifs de production concernant les calendriers de production et la qualité (voir, par exemple, [89–92]). Le prochain mot clé co-occurrence important dans le tableau 3 : les systèmes multi-agents. En général, intégrer des systèmes multi-agents dans l'apprentissage par renforcement pour la production la planification offre de nombreux avantages en termes de flexibilité, d'évolutivité et d'adaptabilité [93-95]. En permettant une prise de décision décentralisée et un apprentissage continu, ces les systèmes peuvent gérer efficacement les complexités et la dynamique des environnements de production modernes, conduisant à des solutions de planification plus efficaces et plus résilientes [96]. Parmi les Avec des mots-clés de cooccurrence, il est possible de souligner également la « fabrication intelligente » qui représente le domaine de mise en œuvre de la planification de la production basée sur RL. Même si la fabrication intelligente est également devenue un mot à la mode, qui a aussi ses inconvénients, la conception s'impose progressivement comme le nouveau paradigme de fabrication. Sur le d'autre part, la complexité des infrastructures de réseaux de fabrication intelligente devient plus élevée et plus, et l'incertitude d'un tel environnement de fabrication devient un problème sérieux problème [25]. Ces faits conduisent à la nécessité d'appliquer des solutions avancées de planification dynamique qui incluent également la planification de la production à l'aide de RL. Ce paragraphe simultanément réponses à la principale question de recherche formulée dans la section 1.

De plus, l'application du RL dans la planification de la production sera ici analysée à travers des moyens bibliométriques du point de vue de différents problèmes d'ordonnancement.

Machine Translated by Google

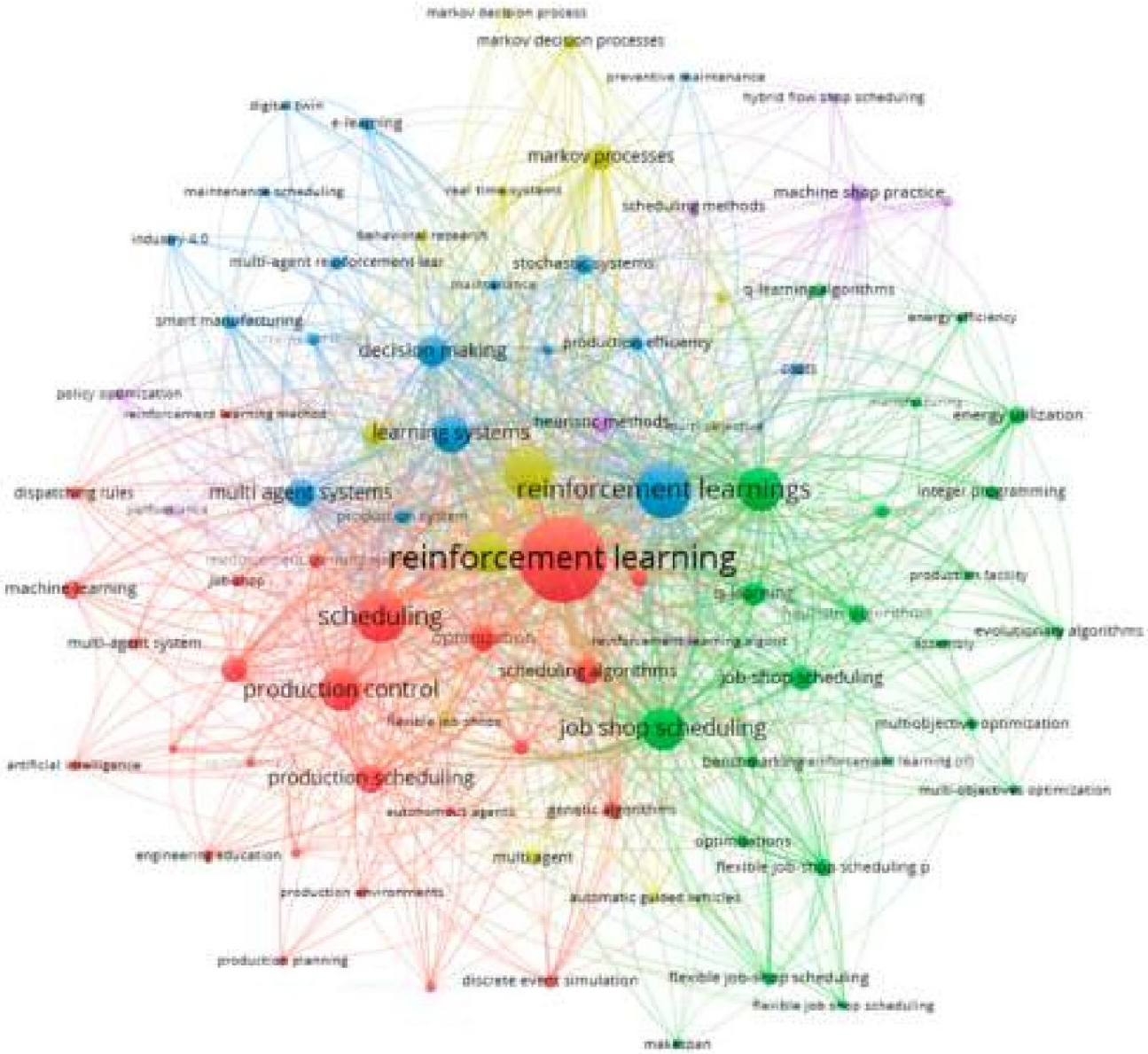


Figure 9. Mots-clés de cooccurrences pour RL dans le domaine de la planification de la production de 1996 à 2024.

Tableau 4. La plupart des méthodologies et domaines de mise en œuvre associés à la planification de la production basée sur sur RL.

Domaines de recherche connexes	Nombre de cooccurrences	Termes/Sujets connexes	Nombre de cooccurrences
Apprentissage par renforcement	526	Contrôle de production	161
Planification	220	Apprentissage par renforcement profond	148
Apprentissage profond	215 et	Systèmes d'apprentissage ffi	134
Algorithmes d'apprentissage	suiv . 178	Prise de décision	99
Planification de l'atelier de travail	168	Systèmes multi-agents	86

6.5. Comparaison des applications de RL dans différents problèmes de planification

En général, les méthodes de planification sont classées en fonction du moment où la décision est prise. pris en compte dans les dynamiques et les statiques. La planification dynamique est liée aux systèmes en temps réel qui nécessitent de répondre aux exigences changeantes de la demande, tandis que la planification statique est hors ligne et se concentre sur un horizon temporel à court terme en fixant un calendrier fixe pour l'achèvement du processus. Il a été découvert par Wang et al. [28] que la planification de la production basée sur les approches RL

Machine Translated by Google

sont largement adoptés pour résoudre les problèmes de planification dynamique. Les problèmes de planification peuvent également être classé selon la nature de l'environnement de planification comme déterministe, lors du traitement les paramètres sont connus et invariables, et non déterministes, lorsque les paramètres d'entrée sont incertains [97]. Les problèmes de planification de la production selon ce deuxième critère peuvent être classés comme le montre la figure 10.

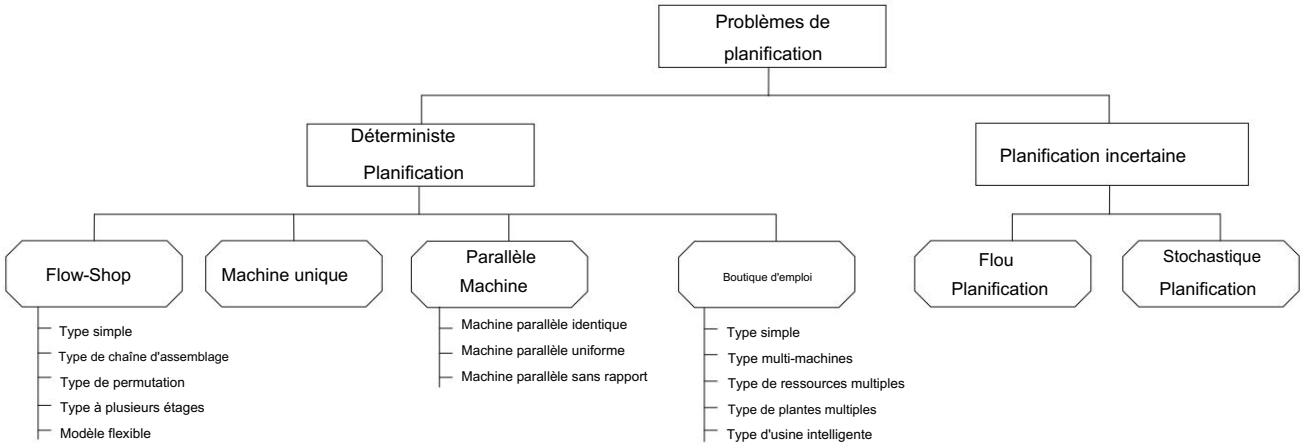


Figure 10. Classification de base des problèmes de planification.

Comprendre les différences entre des problèmes de planification spécifiques est crucial pour sélectionner tion des techniques d'optimisation et d'ordonnancement appropriées dans le but d'améliorer efficacité et productivité dans diverses industries. La principale différence entre la planification Les problèmes ffi peuvent être observés, par exemple, dans la flexibilité et la complexité du routage. Par exemple, job-shop les problèmes ont une flexibilité et une complexité de routage élevées, tandis que pour les flow-shops et les singles Les problèmes de machine sont typiques d'une faible flexibilité et les problèmes de machine parallèle sont modérés. flexibilité et complexité du routage [98,99]. Leurs différences résident également dans la nature de l'application domaines auxquels ils sont destinés. Par exemple, pour une fabrication sur mesure, c'est la planification typique d'un atelier de travail [100,101], pour les chaînes d'assemblage, est principalement considérée comme un atelier de flux problèmes de planification [102,103], et la planification ouverte des ateliers de travail est fréquemment applicable dans soins de santé [104,105].

La classification est en outre utilisée pour explorer une fréquence d'application de RL dans le catégories identifiées de problèmes de planification. Premièrement, déterministe versus incertain la planification est comparée selon cette vue. À cette fin, les termes de recherche ont été défini en combinant les mots-clés suivants : « Apprentissage par renforcement » avec « Ordonnancement déterministe » ou « Ordonnancement incertain », respectivement. Données avec les critères d'inclusion ont été rassemblés en recherchant : (i) Base de données ScienceDirect – Tous les champs, Tous années et Tous les types de documents ; (ii) Base de données Scopus – Tous les champs, Toutes les années, Tous les documents types, et (iii) la base de données Google Scholar : toutes les années et tous les types de documents. Ensemble 734 publications ont été analysées. Web of Science n'a pas été utilisé dans cette procédure en raison de la faible fréquence des travaux scientifiques sur les sujets. Les données ainsi obtenues sont indiqué dans le tableau 5.

Tableau 5. Comparaison des méthodes d'ordonnancement déterministes et incertaines basées sur RL.

Termes de recherche utilisés	Science directe Base de données	Scopus Base de données	Google Scholar Base de données
« Ordonnancement déterministe » et « apprentissage par renforcement »	53	68	474
« Planification incertaine » et « apprentissage par renforcement »	12	24	103

Le tableau 5 montre clairement que l'ordonnancement déterministe basé sur RL est dans une position dominante. position contre une planification incertaine à l'aide de RL. Par conséquent, pour identifier les plus prometteurs

Machine Translated by Google

Les domaines de l'ordonnancement de la production dans le contexte des approches RL, d'ordonnancement déterministe et d'ordonnancement incertain ont été examinés ci-après. Les mots clés utilisés à cet effet étaient :

(a) Planification ET « apprentissage par renforcement » ET problèmes de planification déterministes spécifiques (« job shop », « flow shop », « open shop », « single machine » et « machines parallèles »).

(b) « Apprentissage par renforcement » ET problèmes d'ordonnancement incertains spécifiques (« ordonnancement flou » et « ordonnancement stochastique »).

Les données ont été extraites de la base de données ScienceDirect à l'aide des filtres « Articles de recherche » et « Toutes les années » ; Base de données Scopus utilisant les filtres « Titres d'articles », « Résumés », « Mots-clés », Toutes les années et Tous les types de documents, et base de données Google Scholar : Toutes les années et Tous les types de documents. Les résultats obtenus sont présentés graphiquement et comparés en pourcentage les uns aux autres dans la figure 11.

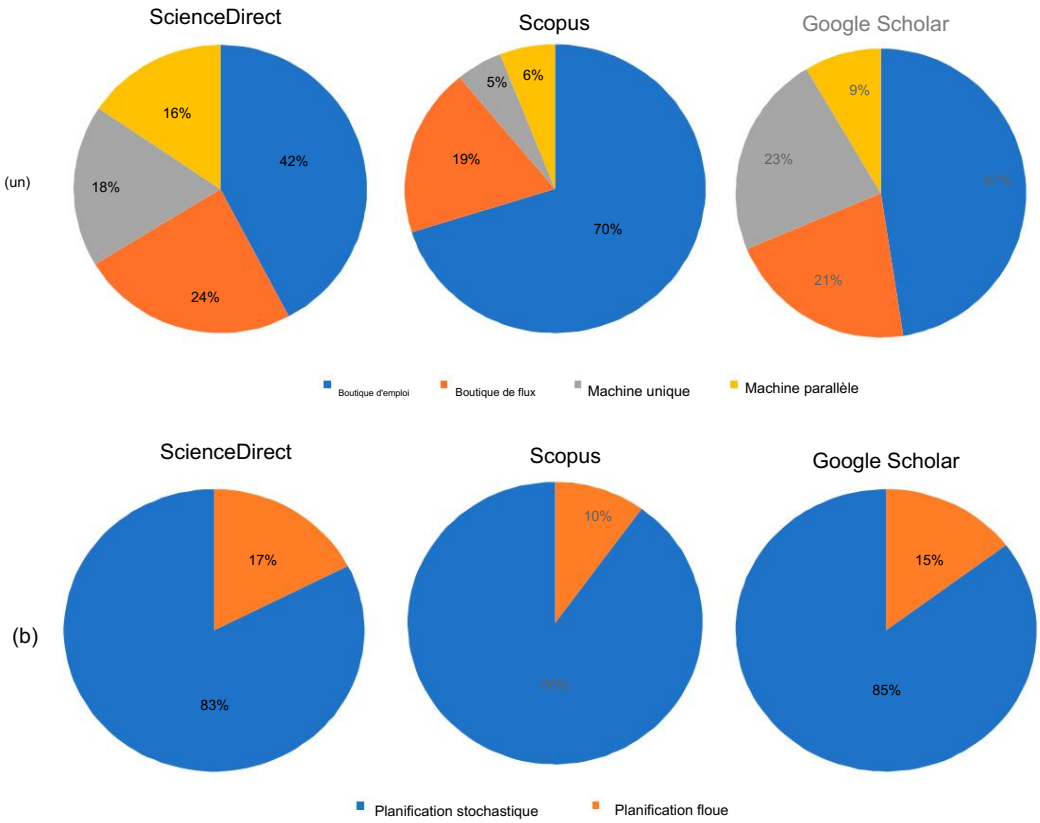


Figure 11. Comparaison de (a) les méthodes de planification déterministes et (b) les méthodes de planification incertaines à partir de leurs occurrences dans la littérature.

On peut voir sur la figure 11a que les algorithmes de planification basés sur RL sont principalement utilisés pour résoudre les problèmes de planification des ateliers de travail [106], tandis que l'environnement de production de l'atelier de flux est le deuxième type de problèmes de planification le plus important, où RL est appliqué. De toute évidence, les algorithmes basés sur RL ont trouvé des applications dans d'autres types de problèmes de planification. Comme les plus remarquables d'entre eux, on peut citer la planification de l'efficacité énergétique, la planification multi-objectifs et la planification distributive. Le dernier problème d'ordonnancement mentionné est exploité en particulier dans les systèmes de fabrication intelligents [107]. En analysant les résultats de la figure 11b, on observe que RL est principalement utilisé dans les méthodes de planification stochastique, par rapport à son application dans les méthodes de planification floue. De plus, l'efficacité de la planification stochastique utilisant RL a également été démontrée par des études comparatives [29, 108, 109]. Bien que l'application de RL dans la planification floue soit prometteuse, elle reste moins courante en raison des complexités supplémentaires impliquées dans l'intégration de la logique floue.

Machine Translated by Google

avec les méthodes RL [110,111]. Cependant, les résultats de cette approche quantitative de comparaison de l'application de RL dans différents problèmes d'ordonnancement fournissent des informations managériales utiles. un aperçu de la pratique de la planification. Néanmoins, des explorations et des recherches plus approfondies sur les aspects qualitatifs de la planification basée sur RL, en particulier dans le contexte de la fabrication intelligente, est requis.

7. Comparaison des recherches présentées et des articles de revue précédents

Comme cela a déjà été mentionné en introduction, ce document de synthèse poursuit le travail du deux articles précédents [28,29] en analysant quelques nouveaux aspects pertinents de la planification de la production méthodes utilisant RL et fournit des informations mises à jour sur les développements récents dans le champ. Le but de cette section est de fournir une vision claire des différences et similitudes des trois œuvres en termes de sujets. Le tableau 6 suivant contient une comparaison sommaire des zones d'investigation incluses dans les travaux mentionnés.

Tableau 6. Comparaison des articles de synthèse sur RL dans la planification de la production.

L'objet de l'enquête	Les articles de synthèse comparés		
	Wang et al., 2021 [28]	Kayhan et Yildiz, 2021 [29]	Recherche proposée
La tendance des citations pour RL en production programmation de 1996 à 2024			
Les publications les plus pertinentes dans le domaine			
Les auteurs les plus influents dans le domaine			
La comparaison entre déterministe et incertain méthodes de planification basées sur RL			
La comparaison du stochastique et du flou méthodes de planification basées sur RL			
La tendance de publication pour RL en production planification			
Les sources les plus pertinentes ont publié des recherches sur RL en planification de production			
La comparaison des plus fréquents méthodes de planification déterministes basées sur RL			
La comparaison des évaluations fondées sur la valeur et méthodes de planification basées sur des politiques basées sur RL			
La comparaison du statique et du dynamique méthodes de planification basées sur RL			
La catégorisation de la littérature sur RL dans ordonnancement de la production selon plusieurs aspects			
La tendance de publication pour RL en production planification selon fabrication environnement			
La comparaison des différents types de RL algorithmes pour la planification de la production			
La comparaison des valeurs stochastiques et déterministes méthodes de planification basées sur RL			

De ce tableau, il ressort que les articles comparés sont pour la plupart complémentaires à l'un l'autre.

Machine Translated by Google

8. Conclusions

Le corpus de recherche existant sur la planification de la production se compose principalement d'études menées dans les domaines de l'informatique, de l'ingénierie, des mathématiques, des sciences de la décision, des affaires, de la gestion et du comptage, de l'énergie, du génie chimique et de la science des matériaux. Ces disciplines scientifiques, notamment « l'ingénierie » et « l'informatique », ont exercé une influence significative sur le développement de la planification de la production basée sur RL.

L'examen des résultats bibliométriques indique fréquemment qu'une augmentation de la quantité d'articles publiés est associée à la reconnaissance de tendances progressistes dans le domaine. C'est également pour cette raison que l'analyse bibliométrique devient de plus en plus bénéfique dans divers domaines universitaires, car elle rend la cartographie de l'information scientifique et l'analyse du développement de la recherche objectives et reproductibles. L'utilisation de cette méthode permet d'identifier les réseaux de collaboration scientifique, d'établir des liens entre de nouveaux thèmes d'étude et des courants de recherche, ainsi que de montrer les liens entre les citations, les co-citations et la productivité publiée dans le domaine.

La principale contribution de cet article peut être résumée en deux étapes : (i) Cette revue apporte des informations supplémentaires sur le RL dans la planification de la production en fournissant les nouvelles fonctionnalités suivantes : • La tendance des citations des RL dans le domaine de la planification de la production de 1996 à 2024 (voir Figure 5) a été réalisée et montre que la tendance croissante à partir de 2019 jusqu'à aujourd'hui se poursuit rapidement avec une augmentation notable de leur nombre. • L'analyse des sources les plus pertinentes du point de vue du nombre de leurs citations (voir Figure 7) a été réalisée pour identifier les sources les plus impactées où les dernières connaissances dans ce domaine sont disponibles. • Les auteurs les plus influents dans ce domaine ont été identifiés (voir tableau 3) pour la détermination de l'état de pointe en matière d'ordonnancement de production basé sur RL. • La comparaison quantitative des méthodes d'ordonnancement déterministes et incertaines basées sur RL a été réalisée (voir tableau 5). Il a montré que les méthodes d'ordonnancement déterministes utilisant RL occupent une position dominante par rapport aux techniques d'ordonnancement incertaines basées sur RL. • La comparaison quantitative des méthodes d'ordonnancement stochastique et floue a été présentée dans la figure 11b. À partir de la comparaison, il a été observé que RL est principalement utilisé dans les méthodes d'ordonnancement stochastique, par rapport à son application dans les méthodes d'ordonnancement floues.

(ii) La contribution de ce document peut également être vue dans les activités suivantes et les résultats déclarations:

- La tendance de publication du RL dans le domaine de la planification de la production a été mise à jour par période de cartographie de 1996 à 2024 (voir Figure 5). Il a été démontré que la croissance exponentielle des publications à partir de 2019 [28] se poursuit rapidement avec une augmentation notable de leur nombre.
- La liste mise à jour des sources les plus pertinentes publiées dans les recherches sur le RL dans la planification de la production a été compilée dans la figure 6. Il a été constaté que cinq des dix sources identifiées jusqu'en 2021 appartiennent aux dix principales sources identifiées jusqu'en 2024. Cela signifie qu'en raison de l'intérêt croissant pour ce domaine de recherche, le nombre de revues traitant de ce sujet augmente simultanément. • L'actualisation de la comparaison quantitative des méthodes d'ordonnancement déterministe les plus fréquentes basées sur RL a été fournie dans la figure 11a. Dans ce contexte, la tendance identifiée précédemment a été confirmée selon laquelle les algorithmes de planification basés sur RL sont principalement utilisés pour résoudre les problèmes de planification en atelier, tandis que la production en atelier de flux présente le deuxième type de problèmes de planification en importance. • La catégorisation mise à jour des algorithmes RL nouvellement développés a été élaborée et présentée dans la figure 3, ce qui peut fournir un meilleur support de mise en œuvre pour la prise de décision dans des problèmes du monde réel.

Machine Translated by Google

• La classification mise à jour des méthodes d'ordonnancement est fournie dans la figure 10. En principe, elle peut être utilisée pour faciliter la sélection des techniques d'ordonnancement pour résoudre des types spécifiques de problèmes.

En plus des résultats mentionnés ci-dessus, il serait nécessaire de se concentrer sur d'autres défis à considérer à l'avenir, tels que la planification des processus de fabrication avec un soutien intégré au partage des connaissances, la demande croissante d'améliorations de « l'intelligence » omniprésente dans la fabrication. processus, y compris la conception et la mise en œuvre d'algorithmes intelligents, et la nécessité d'outils de planification robustes pour des systèmes de fabrication collaboratifs agiles.

Contributions de l'auteur : Conceptualisation, VM ; méthodologie, RS et AB ; validation, ZS; analyse formelle , VM et RS ; enquête, VM et AB ; rédaction : préparation de l'ébauche originale, AB et VM ; rédaction : révision et édition, RS et VM ; visualisation, ZS ; administration du projet, VM Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche a été financée par le projet SME 5.0 grâce au financement reçu du programme de recherche et d'innovation Horizon de l'Union européenne dans le cadre de la convention de subvention Marie Skłodowska-Curie n° 101086487 et du projet KEGA (Agence de subventions culturelles et éducatives) n° 1. 044TUKE-4/2023 accordé par le ministère de l'Éducation de la République slovaque.

Déclaration du comité d'examen institutionnel : sans objet.

Déclaration de consentement éclairé : sans objet.

Déclaration de disponibilité des données : les données sont contenues dans l'article.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts. Les bailleurs de fonds n'ont joué aucun rôle dans la conception de l'étude ; dans la collecte, l'analyse ou l'interprétation des données ; dans la rédaction du manuscrit ; ou dans la décision de publier les résultats.

Références

1. Pinedo, M. Planification et ordonnancement dans la fabrication et les services ; Springer : New York, NY, États-Unis, 2005. 2.
Beheshti, Z. ; Shamsuddin, SMH Une revue des algorithmes méta-heuristiques basés sur la population. Int. J.Adv. Informatique douce. Appl. 2013, 5, 1–35.
3. Xhafa, F. ; Abraham, A. (Eds.) Métaheuristiques pour la planification dans les applications industrielles et manufacturières ; Springer : Berlin/ Heidelberg, Allemagne, 2008 ; Tome 128.
4. Abdel-Kader, Optimisation des essaims de particules RF pour la planification d'instructions contraintes. VLSI Des. 2008, 2008, 930610. [Réf. croisée]
5. Balamurugan, A. ; Ranjitharamasamy, SP Une heuristique modifiée pour l'optimisation de la taille des lots avec un temps combiné dans un Système de fabrication personnalisé en masse. Int. J. Ind. Eng. Application théorique. Entraînez-vous. 2023, 30, 1090-1115.
6. Ghassemi Tari, F. ; Olfat, L. Règles heuristiques pour les problèmes de retard dans les flow shop avec des dates d'échéance intermédiaires. Int. J.Adv. Fab. Technologie. 2014, 71, 381-393. [Référence croisée]
7. Modrak, V. ; Pandian, algorithme de planification d'atelier RS Flow pour minimiser le temps d'exécution pour le problème m-machines à n tâches. Eh bien. Vjesn. 2010, 17, 273-278.
8. Thérasu, M. ; Rameshkumar, K. ; Rousseau, J. ; Anbuudayasankar, SP Développement et analyse de règles de décision prioritaires à l'aide de l'approche MCDM pour une planification flexible des ateliers de travail : une étude de simulation. Simul. Modèle. Entraînez-vous. Théorie 2022, 114, 102416. [CrossRef]
9. Pandian, S. ; Modrak, V. Possibilités, obstacles et défis de l'algorithme génétique dans la fabrication de la formation cellulaire. Av. Logiste. Système. 2009, 3, 63-70.
10. Abdulredha, Minnesota ; Bara'a, AA; Jabir, AJ Modèles d'optimisation heuristiques et méta-heuristiques pour la planification des tâches dans le brouillard nuageux systèmes : une revue. Irakien J. Electr. Électron. Ing. 2020, 16, 103-112. [Référence croisée]
11. Modrak, V. ; Pandian, RS ; Semanco, P. Calibrage des paramètres GA pour les problèmes d'optimisation de la conception de configuration à l'aide de la conception de expériences. Appl. Sci. 2021, 11, 6940. [CrossRef]
12. Keshanchi, B. ; Souri, A. ; Navimipour, NJ Un algorithme génétique amélioré pour la planification des tâches dans les environnements cloud à l'aide des files d'attente prioritaires : vérification formelle, simulation et tests statistiques. J. Syst. Logiciel. 2017, 124, 1-21. [Référence croisée]
13. Jans, R. ; Degraeve, Z. Méta-heuristiques pour le dimensionnement dynamique des lots : examen et comparaison des approches de solutions. EUR. J. Opér. Rés. 2007, 177, 1855-1875. [Référence croisée]
14. Han, BA; Yang, JJ Une solution basée sur l'apprentissage par renforcement profond pour les problèmes de planification flexible des ateliers. Int. J. Simul. Modèle. 2021, 20, 375-386. [Référence croisée]
15. Shyalika, C. ; Silva, T. ; Karunananda, A. Apprentissage par renforcement dans la planification dynamique des tâches : une revue. SN Calcul. Sci. 2020, 1, 306. [Référence croisée]
16. Wang, X. ; Zhang, L. ; Ren, L. ; Xie, K. ; Wang, K. ; Oui, F. ; Chen, Z. Bref examen de l'application de l'apprentissage par renforcement aux problèmes de planification des ateliers . J. Syst. Simul. 2022, 33, 2782-2791.

Machine Translated by Google

17. Dima, IC ; Gabrara, J. ; Modrak, V. ; Piotr, P. ; Popescu, C. Utiliser les systèmes experts dans la gestion opérationnelle de la production. Dans Actes de la 11e Conférence internationale WSEAS sur les mathématiques et les ordinateurs dans les affaires et l'économie (MCBE'10), Iasi, Roumanie, 13-15 juin 2010 ; WSEAS Press : Stevens Point, WI, États-Unis, 2010.
18. Waschneck, B. ; Reichstaller, A. ; Belzner, L. ; Altenmüller, T. ; Bauernhansl, T. ; Knapp, A. ; Kyek, A. Optimisation du global planification de la production avec apprentissage par renforcement profond. Procedia CIRP 2018, 72, 1264-1269. [Référence croisée]
19. Yan, J. ; Liu, Z. ; Zhang, T. ; Zhang, Y. Méthode de prise de décision autonome du processus de transport pour un problème de planification d'atelier flexible basé sur l'apprentissage par renforcement. Dans Actes de la Conférence internationale 2021 sur l'apprentissage automatique et l'ingénierie des systèmes intelligents, MLISE, Chongqing, Chine, 9-11 juillet 2021 ; Institute of Electrical and Electronics Engineers Inc. : Piscataway, NJ, États-Unis ; pp. 234-238.
20. Modrak, V. ; Pandian, RS Recherche en gestion des opérations et systèmes de fabrication cellulaire ; IGI Global : Hershey, Pennsylvanie, États-Unis, 2010.
21. Huang, Z. ; Liu, Q. ; Zhu, F. Apprentissage par renforcement hiérarchique avec planification adaptative pour le contrôle des robots. Ing. Appl. Artif. Intell. 2023, 126, 107130. [Réf. croisée]
22. Arviv, K. ; Stern, H. ; Edan, Y. Apprentissage collaboratif par renforcement pour un problème de planification d'atelier de transfert de tâches à deux robots. Int. J.Prod. Rés. 2016, 54, 1196-1209. [Référence croisée]
23. Wen, X. ; Zhang, X. ; Xing, H. ; Oui, G. ; Li, H. ; Zhang, Y. ; Wang, H. Un algorithme génétique amélioré basé sur l'apprentissage par renforcement pour un problème de planification de l'assemblage d'avions. Calculer. Ind. Ing. 2024, 193, 110263. [Réf. croisée]
24. Aydin, MOI ; Öztemel, E. Planification dynamique d'atelier à l'aide d'agents d'apprentissage par renforcement. Robot. Auton. Système. 2000, 33, 169-178. [Référence croisée]
25. Qu, S. ; Wang, J. ; Govil, S. ; Leckie, JO Planification adaptative optimisée d'un système de processus de fabrication avec une main-d'œuvre multi-compétences et plusieurs types de machines : une approche d'apprentissage par renforcement multi-agents basée sur l'ontologie. Procedia CIRP 2016, 57, 55-60. [Référence croisée]
26. Luo, S. ; Zhang, L. ; Fan, Y. Planification multi-objectifs dynamique pour un atelier de travail flexible par apprentissage par renforcement en profondeur. Calculer. Indiana Ing. 2021, 159, 107489. [Réf. croisée]
27. Zhou, L. ; Zhang, L. ; Horn, BKP Planification dynamique basée sur l'apprentissage par renforcement profond dans la fabrication intelligente. Procedia CIRP 2020, 93, 383-388. [Référence croisée]
28. Wang, L. ; Poêle, Z. ; Wang, J. Un examen de l'optimisation intelligente basée sur l'apprentissage par renforcement pour la planification de la fabrication. Système complexe. Modèle. Simul. 2021, 1, 257-270. [Référence croisée]
29. Kayhan, BM; Yildiz, G. Applications d'apprentissage par renforcement aux problèmes de planification automatique : une revue complète de la littérature. J. Intell. Fab. 2021, 34, 905-929. [Référence croisée]
30. Broadus, RN Vers une définition de la « bibliométrie ». Scientométrie 1987, 12, 373-379. [Référence croisée]
31. Arunmozhi, B. ; Sudhakarapandian, R. ; Sultan Batcha, Y. ; Rajay Vedaraj, IS Une analyse inférentielle de l'acier inoxydable en additif fabrication à l'aide d'indicateurs bibliométriques. Mater aujourd'hui Proc. 2023, sous presse. [Référence croisée]
32. Randhawa, K. ; Wilden, R. ; Hohberger, J. Une revue bibliométrique de l'innovation ouverte : établir un programme de recherche. J.Prod. Innover. Gérer. 2016, 33, 750-772. [Référence croisée]
33. van Raan, A. Méthodes bibliométriques avancées comme noyau quantitatif d'exercices d'évaluation et de prospective basés sur l'examen par les pairs. Scientométrie 1996, 36, 397-420. [Référence croisée]
34. Brandom, RB Articuler les raisons : une introduction à l'inférentialisme ; Harvard University Press : Cambridge, MA, États-Unis, 2001.
35. Kothari, CR Méthodologie de recherche : méthodes et techniques ; New Age International : New Delhi, Inde, 2004.
36. Suárez, M. Une conception inférentielle de la représentation scientifique. Philos. Sci. 2004, 71, 767-779. [Référence croisée]
37. Contessa, G. Représentation scientifique, interprétation et raisonnement de substitution. Philos. Sci. 2007, 74, 48-68. [Référence croisée]
38. Govier, T. Problèmes d'analyse et d'évaluation des arguments ; Université de Windsor : Windsor, Ontario, Canada, 2018 ; Tome 6.
39. Munusamy, R. ; Mukherjee, A. ; Vasudevan, K. ; Venkateswaran, B. Conception et simulation d'un cerveau d'intelligence artificielle (IA) pour un système de navigation de véhicule 2D. INCAS Taureau. 2022, 14, 53-64. [Référence croisée]
40. Dunjko, V. ; Briegel, HJ Apprentissage automatique et intelligence artificielle dans le domaine quantique : un examen des progrès récents. Rép. Prog. Phys. 2018, 81, 074001.
41. Horvitz, E. ; Mulligan, D. Données, confidentialité et bien commun. Sciences 2015, 349, 253-255. [Référence croisée]
42. Sutton, RS; Barto, AG Apprentissage par renforcement : une introduction ; MIT : New York, NY, États-Unis, 2018.
43. Kuhnle, A. ; Mai, MC ; Schäfer, L. ; Lanza, G. Apprentissage par renforcement explicable dans le contrôle de la production du système de fabrication en atelier . Int. J.Prod. Rés. 2022, 60, 5812-5834. [Référence croisée]
44. Estéso, A. ; Peidro, D. ; Mula, J. ; Díaz-Madroñero, M. Apprentissage par renforcement appliqué à la planification et au contrôle de la production. Int. J. Prod. Rés. 2023, 61, 5772-5789. [Référence croisée]
45. Khan, MAM ; Khan, MRJ; Tooshil, A. ; Sikder, N. ; Mahmoud, député ; Kouzani, Arizona ; Nahid, AA Une revue systématique de la robotique basée sur l'apprentissage par renforcement au cours de la dernière décennie. Accès IEEE 2020, 8, 176598-176623. [Référence croisée]
46. AlMahamid, F. ; Grolinger, K. Algorithmes d'apprentissage par renforcement : un aperçu et une classification. Dans Actes de la Conférence canadienne de l'IEEE 2021 sur le génie électrique et informatique (CCECE), Kingston, ON, Canada, 12-17 septembre 2021 ; IEEE : Piscataway, New Jersey, États-Unis ; p. 1 à 7.
47. Akalin, N. ; Loutfi, A. Approches d'apprentissage par renforcement en robotique sociale. Capteurs 2021, 21, 1292. [CrossRef]
48. Zhang, H. ; Yu, T. Taxonomie des algorithmes d'apprentissage par renforcement. Dans Apprentissage par renforcement profond : principes fondamentaux, recherche et Candidatures ; Springer : Singapour, 2020 ; pp. 125-133.

Machine Translated by Google

49. Boulanger, KR ; Trietsch, D. Principes de séquençage et de planification ; John Wiley & Sons : Hoboken, NJ, États-Unis, 2009.
50. Pinedo, ML Scheduling : théorie, algorithmes et systèmes ; Springer : Berlin/Heidelberg, Allemagne, 2016.
51. Allahverdi, A. ; Ng, CT ; Cheng, TCE ; Kovalyov, MY Une enquête sur les problèmes de planification liés aux délais ou aux coûts de préparation. EUR. J. Opér. Rés. 2008, 187, 985-1032. [Référence croisée]
52. Panwalkar, SS ; Smith, ML Enquête sur la recherche sur la planification des ateliers de flux. Opéra. Rés. 1977, 25, 45-84. [Référence croisée]
53. Paraschos, PD; Koulinas, GK; Koulouriotis, DE Renforcement de l'optimisation basée sur l'apprentissage pour une production durable et Lean dans le contexte de l'industrie 4.0. Algorithmes 2024, 17, 98. [CrossRef]
54. Wang, S. ; Li, J. ; Luo, Y. Planification intelligente pour une production flexible et hybride avec un apprentissage par renforcement approfondi multi-agents. Dans les actes de la 2e conférence internationale de l'IEEE 2021 sur les technologies de l'information, le Big Data et l'intelligence artificielle, ICIBA, Chongqing, Chine, 17-19 décembre 2021 ; Institute of Electrical and Electronics Engineers Inc. : Piscataway, NJ, États-Unis ; pp. 288-294.
55. Tang, J. ; Haddad, Y. ; Salonitis, K. Planification du système de fabrication reconfigurable : une approche d'apprentissage par renforcement profond. Procedia CIRP 2022, 107, 1198-1203. [Référence croisée]
56. Shahrabi, J. ; Adibi, MA; Mahootchi, M. Une approche d'apprentissage par renforcement pour l'estimation des paramètres dans la planification dynamique des ateliers . Ingénieur informatique. 2017, 110, 75-82. [Référence croisée]
57. Yang, J. ; Toi, X. ; Wu, G. ; Hassan, MM; Almogren, A. ; Guna, J. Application de l'apprentissage par renforcement dans la tâche du cluster UAV planification. Futur Gén. Calculer. Système. 2019, 95, 140-148. [Référence croisée]
58. Yuan, X. ; Poêle, Y. ; Yang, J. ; Wang, W. ; Huang, Z. Étude sur l'application de l'apprentissage par renforcement dans l'optimisation du fonctionnement du système CVC. En simulation de bâtiment ; Presses universitaires Tsinghua : Pékin, Chine, 2021 ; Volume 14, p. 75-87.
59. Kourinov, I. ; Orzechowski, G. ; Hamäläinen, P. ; Mikkola, A. Pelle automatisée basée sur l'apprentissage par renforcement et multicorps dynamique du système. Accès IEEE 2020, 8, 213998-214006. [Référence croisée]
60. Popper, J. ; Motsch, W. ; David, A. ; Petzsche, T. ; Ruskowski, M. Utilisation de l'apprentissage par renforcement profond multi-agents pour une planification flexible des ateliers de travail dans une perspective durable. Dans Actes de la Conférence internationale sur l' ingénierie électrique, informatique, des communications et mécatronique 2021, ICECCME, Maurice, Maurice, 7-8 octobre 2021 ; Institute of Electrical and Electronics Engineers Inc. : Piscataway, NJ, États-Unis.
61. Xiong, H. ; Fan, H. ; Jiang, G. ; Li, G. Une étude basée sur la simulation des règles de répartition dans un problème de planification d'atelier dynamique avec libération par lots et contraintes de préséance technique étendues. EUR. J. Opér. Rés. 2017, 257, 13-24. [Référence croisée]
62. Palacio, JC ; Jiménez, YM; Schietgat, L. ; Van Doninck, B. ; Nowé, A. Un algorithme Q-Learning pour une planification flexible des ateliers dans un scénario de fabrication réel. Procedia CIRP 2022, 106, 227-232. [Référence croisée]
63. Chang, J. ; Yu, D. ; Hu, Y. ; Lui, W. ; Yu, H. Apprentissage par renforcement approfondi pour une planification dynamique et flexible des ateliers de travail avec un caractère aléatoire Arrivée du travail. Processus 2022, 10, 760. [CrossRef]
64. Liu, R. ; Piplani, R. ; Toro, C. Apprentissage par renforcement approfondi pour la planification dynamique d'un atelier de travail flexible. Int. J.Prod. Rés. 2022, 60, 4049-4069. [Référence croisée]
65. Samsonov, V. ; Kemmerling, M. ; Paegert, M. ; Lutticke, D. ; Sauermann, F. ; Gützlaff, A. ; Schuh, G. ; Meisen, T. Contrôle de fabrication dans des environnements d'atelier avec apprentissage par renforcement. Dans Actes de la 13e Conférence internationale sur les agents et l'intelligence artificielle (ICAART 2021), en ligne, 4-6 février 2021 ; SciTePress : Setúbal, Portugal ; pp. 589-597.
66. Cunha, B. ; Madureira, AM; Fonseca, B. ; Coelho, D. Apprentissage par renforcement profond en tant que solution de planification d'atelier : une revue de la littérature. Dans Actes de la 18e Conférence internationale sur les systèmes intelligents hybrides (HIS 2018), Porto, Portugal, 13-15 décembre 2018 ; Madureira, AM, Abraham, A., Gandhi, N., Varela, ML, éd. ; Springer International Publishing : Cham, Suisse, 2018.
67. Wang, X. ; Zhang, L. ; Lin, T. ; Zhao, C. ; Wang, K. ; Chen, Z. Résoudre les problèmes de planification des tâches dans un environnement de préemption des ressources avec apprentissage par renforcement multi-agents. Robot. Calculer. Intégré. Fab. 2022, 77, 102324. [Réf. croisée]
68. Oh, SH; Cho, YI ; Woo, JH Apprentissage par renforcement distributionnel avec les apprenants indépendants pour un problème de planification d'atelier flexible avec une grande variabilité. J. Informatique. Des. Ing. 2022, 9, 1157-1174. [Référence croisée]
69. Zhang, Y. ; Zhu, H. ; Tang, D. ; Zhou, T. ; Gui, Y. Planification dynamique des ateliers de travail basée sur un apprentissage par renforcement profond pour les systèmes de fabrication multi-agents. Robot. Calculer. Intégré. Fab. 2022, 78, 102412. [Réf. croisée]
70. Liang, Y. ; Soleil, Z. ; Chanson, T. ; Chou, Q. ; Fan, W. ; Fan, J. ; Rui, Y. ; Zhou, Q. ; Bai, J. ; Yang, C. ; et coll. Lenovo programme un ordinateur portable Fabrication utilisant l'apprentissage par renforcement profond. Interfaces 2022, 52, 56-68. [Référence croisée]
71. Chen, Y. ; Guo, W. ; Liu, J. ; Shen, S. ; Lin, J. ; Cui, D. Une approche de contrôle du refroidissement à plusieurs points de consigne pour les centres de données refroidis par air utilisant le algorithme de réseau Q profond. Mesures. Contrôle 2024, 57, 782-793. [Référence croisée]
72. Théate, T. ; Ernst, D. Une application de l'apprentissage par renforcement profond au trading algorithmique. Système expert. Appl. 2021, 173, 114632. [Référence croisée]
73. Sanaye, S. ; Sarrafi, A. Une nouvelle méthode de gestion de l'énergie basée sur l'algorithme Deep Q Network pour un faible coût d'exploitation d'un système hybride intégré. Rapport énergétique 2021, 7, 2647-2663. [Référence croisée]
74. Luo, S. Planification dynamique pour un atelier d'emploi flexible avec de nouvelles insertions d'emploi par apprentissage par renforcement profond. Appl. Informatique douce. J. 2020, 91, 106208. [Réf. croisée]
75. Luo, S. ; Zhang, L. ; Fan, Y. Planification en temps réel pour un atelier de travail flexible, multiobjectif, dynamique, partiel, sans attente, par Deep Apprentissage par renforcement. IEEETrans. Automatique. Sci. Ing. 2022, 19, 3020-3038. [Référence croisée]

Machine Translated by Google

76. Hu, J. ; Wang, H. ; Tang, H.-K.; Kanazawa, T. ; Gupta, C. ; Farahat, A. Apprentissage par renforcement amélioré par les connaissances pour la planification intégrée de la production et de la maintenance multi-machines. Calculer. Ind. Ing. 2023, 185, 109631. [Réf. croisée]
77. Du, Y. ; Li, JQ; Chen, XL ; Duan, PY ; Pan, QK Apprentissage par renforcement basé sur les connaissances et algorithme d'estimation de la distribution pour un problème de planification flexible des ateliers. IEEETrans. Émerger. Haut. Calculer. Intell. 2023, 7, 1036-1050. [Référence croisée]
78. Li, Y. ; Gu, W. ; Yuan, M. ; Tang, Y. Planification dynamique basée sur les données en temps réel pour un atelier flexible avec un transport insuffisant ressources utilisant un réseau hybride Deep Q. Robot. Calculer. Intégré. Fab. 2022, 74, 102283. [Réf. croisée]
79. Zhou, T. ; Zhu, H. ; Tang, D. ; Liu, C. ; Cai, Q. ; Shi, W. ; Gui, Y. Apprentissage par renforcement pour l'optimisation en ligne de la planification des ateliers dans une usine de fabrication intelligente. Av. Mécanique. Ing. 2022, 14, 16878132221086120. [CrossRef]
80. Wang, L. ; Hu, X. ; Wang, Y. ; Xu, S. ; Maman, S. ; Yang, K. ; Wang, W. Planification dynamique des ateliers dans la fabrication intelligente utilisant des apprentissage par renforcement. Calculer. Réseau. 2021, 190, 107969. [Réf. croisée]
81. Wang, Y. ; Usher, JM Application de l'apprentissage par renforcement pour la planification de la production basée sur les agents. Ing. Appl. Artif. Intell. 2005, 18, 73-82. [Référence croisée]
82. Cancino, Californie ; Merigó, JM; Coronado, FC Une analyse bibliométrique des principales universités en recherche sur l'innovation. J.Innov. Connaître. 2017, 2, 106-124. [Référence croisée]
83. Varin, C. ; Cattelan, M. ; Firth, D. Modélisation statistique de l'échange de citations entre revues de statistique. JR Stat. Soc. Ser. Une statistique. Soc. 2016, 179, 1-63. [Référence croisée] [Pub Med]
84. Moral-Muñoz, JA; Herrera-Viedma, E. ; Santisteban-Espejo, A. ; Cobo, MJ Outils logiciels pour effectuer des analyses bibliométriques dans science : une revue à jour. Prof. De La Inf./Inf. Prof.2020 , 29, e290103. [Référence croisée]
85. Curry, S. Allons au-delà de la rhétorique : il est temps de changer la façon dont nous jugeons la recherche. Nature 2018, 554, 147-148. [Référence croisée] [Pub Med]
86. Al-Hoorie, A. ; Vitta, JP Les sept péchés de la recherche L2 : une revue de la qualité statistique de 30 revues et de leurs CiteScore, SJR, SNIP, Facteurs d'impact du RJC. Lang. Enseigner. Rés. 2019, 23, 727-744. [Référence croisée]
87. Waltman, L. ; Van Eck, New Jersey ; Noyons, ECM Une approche unifiée de la cartographie et du regroupement des réseaux bibliométriques. J. Informetr. 2010, 4, 629-635. [Référence croisée]
88. Cheng, F. ; Huang, Y. ; Tanpure, B. ; Sawalani, P. ; Cheng, L. ; Liu, C. Planification des tâches tenant compte des coûts pour les instances cloud à l'aide de Deep apprentissage par renforcement. Groupe. Calculer. 2022, 25, 619-631. [Référence croisée]
89. Thaipisutikul, T. ; Chen, Y.-C.; Hui, L. ; Chen, SC ; Mongkolwat, P. ; Shih, TK La question de l'apprentissage par renforcement profond vers des applications pratiques de l'IA. Dans Actes de la 12e Conférence internationale sur l'informatique Ubi-Media, Bali, Indonésie, 5-8 août 2019 ; p. 24-29.
90. Yan, J. ; Huang, Y. ; Gupta, A. ; Gupta, A. ; Liu, C. ; Li, J. ; Cheng, L. Systèmes économes en énergie pour la planification des tâches en temps réel dans les données cloud centres : une approche d'apprentissage par renforcement profond. Calculer. Electr. Ing. 2022, 99, 107688. [Réf. croisée]
91. Piller, FT Personnalisation de masse : Réflexions sur l'état du concept. Int. J. Flex. Fab. Système. 2004, 16, 313-334. [Référence croisée]
92. Suzic, N. ; Forza, C. ; Trentin, A. ; Aniši´c, Z. Lignes directrices de mise en œuvre pour la personnalisation de masse : caractéristiques actuelles et suggestions d'amélioration. Prod. Plan. Contrôle 2018, 29, 856-871. [Référence croisée]
93. Altenmüller, T. ; Stüker, T. ; Waschneck, B. ; Kuhnle, A. ; Lanza, G. Apprentissage par renforcement pour un contrôle de production intelligent et autonome d'ateliers complexes sous contraintes de temps. Prod. Ing. 2020, 14, 319-328. [Référence croisée]
94. Zhao, Y. ; Zhang, H. Application de l'apprentissage automatique et de la planification de règles dans un système de contrôle de production en atelier. Int. J. Simul. Modèle 2021, 20, 410-421. [Référence croisée]
95. Kuhnle, A. ; Kaiser, JP; Theiss, F. ; Stricker, N. ; Lanza, G. Conception d'un système de contrôle de production adaptatif utilisant l'apprentissage par renforcement. J. Intell. Fab. 2021, 32, 855-876. [Référence croisée]
96. Panzer, M. ; Cintreuse, B. ; Gronau, N. Apprentissage par renforcement profond dans la planification et le contrôle de la production : une revue systématique de la littérature. Dans Actes de la Conférence sur les systèmes de production et la logistique, en ligne, 10-11 août 2021.
97. Wojakowski, P. ; Warzołek, D. La classification des problèmes de planification sous incertitude de production. Rés. Logiste. Prod. 2014, 4, 245-256.
98. Blackstone, JH; Phillips, DT; Hogg, GL Une enquête de pointe sur les règles de répartition pour les opérations des ateliers de fabrication. Int. J.Prod. Rés. 1982, 20, 27-45. [Référence croisée]
99. Blazewicz, J. ; Ecker, KH ; Pesch, E. ; Schmidt, G. ; Weglarz, J. Manuel sur la planification : de la théorie aux applications ; Springer : Berlin/Heidelberg, Allemagne, 2007.
100. Ivanov, D. ; Dolgui, A. ; Sokolov, B. Une approche dynamique de la planification d'atelier en plusieurs étapes dans un système d'assemblage flexible basé sur l'industrie 4.0. Dans Avancées des systèmes de gestion de la production. La voie vers une fabrication intelligente, collaborative et durable : Conférence internationale IFIP WG 5.7, APMS 2017, Hambourg, Allemagne, 3-7 septembre 2017, Actes, Partie I ; Springer International Publishing : Cham, Suisse ; pp. 475-482.
101. Modrak, V. (Ed.) Fabrication personnalisée de masse : concepts théoriques et approches pratiques ; CRC Press : Boca Raton, FL, États-Unis, 2017.
102. Komaki, directeur général ; Cheikh, S. ; Malakooti, B. Problèmes de planification des ateliers Flow avec les opérations d'assemblage : un examen et de nouvelles tendances. Int. J.Prod. Rés. 2019, 57, 2926-2955. [Référence croisée]
103. Yang, Y. ; Li, X. Un algorithme heuristique constructif basé sur la connaissance pour le problème de planification de l'atelier de flux bloquant l'assemblage distribué. Système expert. Appl. 2022, 202, 117269. [Réf. croisée]

Machine Translated by Google

104. Nasiri, MM; Yazdanparast, R. ; Jolai, F. Une approche d'optimisation de simulation pour la planification en temps réel dans un environnement d'atelier ouvert en utilisant une règle de répartition composite. Int. J. Informatique. Intégré. Fab. 2017, 30, 1239-1252. [Référence croisée]
105. Abdelmaguid, TF Planification d'atelier ouvert multiprocesseur dynamique bi-objectif pour la maintenance et les diagnostics de santé. Expert Système. Appl. 2021, 186, 115777. [Réf. croisée]
106. Tremblet, D. ; Thévenin, S. ; Dolgui, A. Estimation Makespan dans un environnement de planification d'atelier flexible utilisant l'apprentissage automatique. Int. J.Prod. Rés. 2024, 62, 3654-3670. [Référence croisée]
107. Fu, Y. ; Hou, Y. ; Wang, Z. ; Wu, X. ; Gao, K. ; Wang, L. Problèmes de planification distribuée dans les systèmes de fabrication intelligents. Tsinghua Sci. Technologie. 2021, 26, 625-645. [Référence croisée]
108. Zhang, Y. ; Zhang, X. ; Ding, Y. Une approche basée sur l'apprentissage par renforcement pour la planification stochastique des ateliers. IEEETrans. Automatique. Sci. Ing. 2020, 17, 72-83.
109. Rinciog, A. ; Meyer, A. Vers la standardisation des approches d'apprentissage par renforcement pour la planification stochastique de la production. arXiv 2021, arXiv :2104.08196.
110. Zeng, B. ; Yang, Y. Une approche hybride d'apprentissage par renforcement et de logique floue pour la planification des ateliers. J. Intell. Fab. 2017, 28, 1189-1201.
111. Zhang, G. ; Huang, Y. Apprentissage par renforcement flou pour la planification dynamique multi-objectifs d'un système de fabrication flexible. J. Intell. Fab. 2005, 16, 293-304.

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Démélange basé sur l'inversion dans le domaine médian commun Utilisation du radon haute résolution dans le domaine temporel

Kai Zhuang [1], Daniel Trad 1,* et Amr Ibrahim [2]

1 Département de la Terre, de l'Énergie et de l'Environnement, Faculté des sciences, Université de Calgary, Calgary, AB T2N 1N4, Canada ;
2 kzhuang@ucalgary.ca Département de physique, Faculté des sciences, Université Beni-Suef, Beni-Suef 62521, Égypte ;
amr.ibrahim@ucalgary.ca * Correspondance : daniel.trad@ucalgary.ca

Résumé : Nous implémentons une méthode de démélange basée sur l'inversion dans les collectes de points médians communs (CMP) comme alternative aux méthodes de domaine standard de collecte de récepteurs communs (CRG). Le principal avantage de la démélange dans le domaine CMP est que les réflexions des couches plongeantes sont centrées autour des décalages nuls. En conséquence, les regroupements CMP présentent une structure plus simple par rapport aux regroupements de récepteurs communs (CRG), où ces réflexions sont décalées au sommet. Par conséquent, nous pouvons utiliser un opérateur Radon hyperbolique à décalage nul pour traiter les collectes CMP. Cet opérateur est une alternative informatiquement plus efficace au Radon hyperbolique à déplacement d'apex requis pour le traitement des collectes CRG. Les transformations clairsemées , telles que la transformée de Radon, peuvent empiler les réflexions et produire des modèles clairsemés capables de séparer les sources mélangées. Nous utilisons l'opérateur Radon pour développer un cadre de démélange basé sur l'inversion qui intègre une contrainte de modèle clairsemée. L'inclusion d'une contrainte de parcimonie dans le processus d'inversion améliore la focalisation de la transformation et améliore la récupération des données. La démélange basée sur l'inversion nous permet de prendre en compte toutes les données observées en incorporant l'opérateur de mélange dans la fonction de coût. Nos exemples de données synthétiques et de données de terrain démontrent que la démélange basée sur l'inversion dans le domaine CMP peut séparer efficacement les sources mélangées.

Mots-clés : démélange ; radon clairsemé; inversion

check for updates

Citation : Zhuang, K. ; Trad, D. ; Ibrahim, A. Basé sur l'inversion Démélange dans le domaine médian commun à l'aide du radon haute résolution du domaine temporel. Algorithmes 2024, 17, 344. https://doi.org/10.3390/a17080344

Rédacteur académique : Frank Werner

Reçu : 16 mai 2024
Révisé : 25 juin 2024
Accepté : 5 août 2024
Publié : 7 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

La recherche d'une acquisition efficace de données sismiques est un défi bien connu dans l'exploration géophysique. L'acquisition simultanée, également connue sous le nom d'acquisition mixte, constitue une solution, offrant une méthode qui réduit considérablement le temps passé sur le terrain en déclenchant simultanément plusieurs sources. Cette technique, initialement proposée par Garottu [1] pour l'acquisition de vibroseis et développée plus tard par Beasley et al. [2], Beasley [3] pour les environnements marins, permettent d'interférer des sources au moment de l'acquisition qui sont ensuite séparées lors du traitement. Malgré leurs avantages en matière de gain de temps, les modèles et algorithmes existants permettant de décomposer ces données présentent souvent des limites.

Cette étude est motivée par la nécessité de remédier à ces limitations et d'améliorer la précision et l'efficacité de la fusion des données sismiques. Les méthodes actuelles, notamment le tir aléatoire et le codage de phase, ont ouvert la voie à des approches d'acquisition de données. Cependant, le processus de démélange reste un goulot d'étranglement important, avec deux écoles de pensée prédominantes : l'approche basée sur le débruitage [4], qui se concentre sur l' atténuation du bruit erratique et aléatoire, et l'approche basée sur l'inversion [5], qui cherche à reconstruire le bruit. signal en inversant le processus de mélange. Cet article se concentrera principalement sur le traitement des données issues de l' approche de tir aléatoire.

L'introduction de petits retards aléatoires dans les temps d'allumage de la source, comme suggéré par Berkhout [6], est une étape critique dans le processus de démélange. Le pseudo-démélange consiste à remapper les plans mélangés à un temps de tir nul et à dupliquer les images en fonction du nombre de sources simultanées. Cette manipulation due aux délais aléatoires rend

Machine Translated by Google

l'interférence des sources voisines sous forme de bruit aléatoire dans certains des domaines de données communs, ne laissant que le signal cible à récupérer en tant qu'événements cohérents. Les données pseudo-démélangées montrent alors des structures incohérentes au sein des domaines de collecte de point médian commun (CMP), de collecte de récepteur commun (CRG) et de collecte de décalage commun (COF), facilitant ainsi le processus de démélange. Il existe diverses méthodes pour exploiter l'incohérence du bruit de mélange, comme les travaux de séparation de Lin et Herrmann [7] et des techniques ultérieures telles que le débruitage Stolt Radon [8], le filtrage SSA [9], le débruitage à l'aide d'un filtre médian variable dans l'espace. [10], le préconditionnement avec un filtre courbelet discret [11] et le préconditionnement avec seuillage robuste dans le domaine FK [12], chacun avec ses propres avantages et inconvénients, souvent confrontés à des problèmes pour distinguer efficacement le signal du bruit (interférence de source) pour les approches de débruitage ou des opérateurs de transformation très coûteux et complexes pour les approches basées sur l'inversion.

Cet article présente une approche de démélange basée sur l'inversion utilisant le domaine CMP, qui promet une solution à grande vitesse utilisant un opérateur de transformation de Radon. L'utilisation de la transformée hyperbolique de Radon pour la démélange remplit de multiples fonctions et contribue de manière significative à la récupération du signal sismique. Compte tenu de la nature intrinsèque de l'acquisition sismique, les réponses acoustiques émanant de l'intérieur de la Terre se manifestent par des arrivées hyperboliques à la surface. La transformée hyperbolique de Radon, en agrégeant des données basées sur des trajectoires hyperboliques, est parfaitement adaptée à la cartographie des données sismiques. Opter pour le domaine CMP plutôt que pour la méthode de domaine CRG traditionnelle permet d'utiliser une version plus rationalisée de l'opérateur de transformation Radon, ce qui entraîne des temps d'exécution

2. Théorie

2.1. Opérateur de

mélange L'opérateur de mélange nous permet de faire la transition entre le domaine mélangé, où les sources se chevauchent, appelé super plan, vers le domaine non mélangé, où les sources appartiennent à leurs enregistrements de tir respectifs avec des temps de tir ajustés à zéro pour chaque tir. Cela peut être représenté par l'opérateur de mélange Γ, qui contient les informations de chevauchement et de temps de retard . Nous pouvons alors représenter notre cube de données non mélangées par D et nos données mélangées par b avec la relation suivante [6,13] :

$$b = \Gamma D. \quad (1)$$

En appliquant l'adjoint de l'opérateur de mélange aux données mélangées, nous obtenons les données pseudo-démélangées, dans l'espace non mélangé, chaque plan occupant sa propre trame de données, par opposition à plusieurs plans par trame de données. Cela peut être exprimé par ce qui suit [6,13] :

$$\tilde{D} = \Gamma^H b . \quad (2)$$

Ici, D˜ représente les données pseudo-démélangées dans des images de prise de vue standard. Le rôle de la matrice de mélange est de fusionner les plans avec des temps de retard aléatoires, en superposant efficacement les données d'enregistrement des plans individuels. Cela signifie également que le pseudo-démélange agit pour redistribuer les données dans leurs trames respectives. Cela forme une représentation surcomplète des données avec des informations redondantes stockées dans chaque panneau source pseudo-mixte. Il est important de noter que l'opérateur adjoint n'élimine pas les interférences de source des données ; il le copie simplement.

Le succès de la fusion des données à l'aide de la technique de tir aléatoire dépend de l' inclusion de temps de retard aléatoires. Ces retards sont essentiels pour convertir les interférences de la source en ce qui ressemble à du bruit aléatoire lors du remappage des données via l' opérateur de mélange adjoint. Il est essentiel que ces temps de retard soient suffisamment aléatoires au point qu'aucun modèle n'apparaisse dans les domaines de données remappés, car des modèles cohérents dans les temps de retard réduisent l'efficacité de notre contrainte de cohérence. En outre, les délais doivent être suffisamment longs pour garantir la séparabilité des événements, en respectant la longueur de délai minimale déterminée par une fonction de la fréquence la plus basse et du pli de mélange, comme détaillé dans Jiang

Machine Translated by Google

et Abma [14]. Parallèlement, ils doivent être suffisamment courts pour que nous puissions bénéficier de temps d'acquisition réduits.

Les retards aléatoires introduits dans le calendrier de mélange, et par conséquent dans l' opérateur de mélange, sont cruciaux pour simplifier le démélange. Lorsque l'opérateur de mélange adjoint (pseudo-démélange) est appliqué à l'ensemble de données mélangées, les retards aléatoires deviennent cruciaux.
Le mélange peut être conceptualisé comme une opération sur des données d'acquisition standard qui introduit des interférences de source.

L'influence de l'addition sur les temps de retard aléatoires peut être observée lorsque les données sont triées dans l'un des autres domaines géométriques après pseudo-démélange. Après l' application de l'opérateur de mélange adjoint, des événements cohérents peuvent être observés en plus des événements incohérents observés avant l'application de la transformation ; ce sont ces événements cohérents que nous cherchons à extraire, car ils représentent la partie des données dont les temps de retard sont correctement corrigés, avec les données spontanées dues aux tirs interférents qui doivent être supprimées.
Grâce à l'utilisation d'une méthode d'inversion et d'une contrainte de cohérence telle que la transformée de Radon haute résolution, nous pouvons supprimer efficacement le signal interférent et démélanger nos données.

2.2. Transformation de

Radon Pour résoudre le problème sous-déterminé observé dans le problème d'inversion, nous utilisons la transformée hyperbolique de Radon (HRT) pour contraindre davantage le problème. Cet opérateur composite, composé à la fois de l'opérateur de mélange et de l'opérateur de transformation de Radon, reformule désormais l'inversion en un problème peu contraint. Le HRT est une variante de la transformée de Radon qui additionne les fonctions de base hyperboliques ; en géophysique, cette équation est définie en fonction des temps d'arrivée, du décalage source-récepteur et de la vitesse d'intervalle, comme le montrent Thorson et Claerbout [15] :

$$m(p, \tau) = \int_{h1}^{h2} d(h, t = \tau \sqrt{}^{2} + p\, 2h2\,)dh. \quad (3)$$

Ici, m(p, τ) représente les données de l'espace Radon, p désigne la lenteur, τ est le temps de trajet aller-retour , h1 et h2 sont respectivement les limites de décalage inférieure et supérieure, et d est l'espace de données à transformer. . La lenteur p est définie comme l'inverse de la vitesse 1/v.
La cartographie basée sur les arrivées hyperboliques permet une représentation précise et clairsemée des données sismiques, car les signaux sismiques apparaissent comme des événements hyperboliques sur les levés sismiques standards. La transformée de Radon a été utilisée dans diverses applications, notamment l'analyse de vitesse [15], la suppression multiple [16], l'interpolation dans le domaine fréquentiel ou temporel [17] et la démélange via le débruitage [8].

Bien que le HRT soit bien adapté aux données sismiques, il présente plusieurs limites. Principalement, la transformation nécessite que les arrivées hyperboliques soient symétriques et centrées dans l'espace des données, ce qui n'est pas nécessairement le cas dans les données du monde réel. En réalité, le pendage des structures souterraines provoque un déplacement des arrivées hyperboliques vers le haut, ce qui entraîne des arrivées non centrées. La symétrie est remise en question par deux facteurs : l'anisotropie et les conditions d'enquête. Les réseaux de capteurs de surface ne sont généralement pas placés sur une surface parfaitement plane, ce qui entraîne des écarts par rapport aux arrivées hyperboliques. Ce problème peut être facilement résolu grâce à une correction statique des traces, qui fait partie du flux de traitement standard. Pour le problème de l'anisotropie, une équation modifiée peut être utilisée pour tenir compte de l'étirement de la courbe dans des cas simples. L'abandon des localisations centrées présente un problème quelque peu complexe avec de multiples solutions potentielles. La première solution implique la mise en œuvre d'une transformée de Radon hyperbolique à déplacem
Dans cette approche, plusieurs panneaux Radon sont générés pour une seule image de données, chacun correspondant à une série d'emplacements de sommets décalés attendus. Bien que cette solution soit robuste et corresponde bien aux données sismiques, elle augmente considérablement les temps d'exécution car il n'existe aucun raccourci pour générer des panneaux Radon et les temps de traitement sont multipliés par le nombre de sommets analysés. La deuxième solution au problème de centrage consiste à transformer les données dans le domaine CMP, où les sommets sont centrés même pour les réflexions plongeantes. Cependant, ceci

Machine Translated by Google

Cette solution présente un inconvénient majeur : le domaine CMP a généralement un échantillonnage moins bon que les autres domaines de données, ce qui fait de l'alias un problème important.

2.3. Démélange basé sur l'inversion

Compte tenu de notre opérateur de transformation déjà défini, l'opérateur de mappage Γ qui mappe l'espace de données b à l'espace modèle D dans l'équation (1), nous pouvons présenter le problème comme un problème d'inversion qui inverse l'opération [19]. En inversant cette opération, nous devons formuler une fonction objectif qui minimise l'inadéquation entre les données et les données modélisées, en d'autres termes, minimisant les informations du modèle qui ne correspondent pas aux données. Cette fonction objectif est souvent présentée comme suit [17] :

$$\begin{aligned} &\text{minimiser } ||D||2\ 2 \\ &\text{Sous réserve de } ||(b - \Gamma D)||2\ 2\ . \end{aligned} \tag{4}$$

La fonction objectif d'inversion des données combinées étant écrite sous la forme

$$J = ||b - \Gamma D||, \tag{5}$$

L'inversion directe de l'opérateur de mélange est mal posée ; ainsi, aucune solution unique ne peut être évaluée en inversant simplement l'opérateur de mélange Γ. Ce caractère mal posé est dû au fait que l'opérateur de mélange ne sert qu'à collecter et déplacer les données, tandis que l'opérateur adjoint ne sert qu'à les diffuser. Pour contourner ce problème, nous avons reformulé le schéma d'inversion en ajoutant l'opérateur Radon comme contrainte de cohérence.

$$J = ||b - \Gamma Lm|| + \mu||m||, \tag{6}$$

où L est la transformée de Radon. En formulant le problème de cette façon, nous passons d' un problème mal posé à un problème peu contraint.

2.4. Régularisation clairsemée

Nous utilisons la régularisation clairsemée pour notre schéma d'inversion comme moyen de résoudre la norme ℓ1 dans l'espace modèle, comme solution plus robuste à la transformée de Radon [17] et, par conséquent, au schéma de démélange. Pour résoudre la solution clairsemée dans cet article, nous utilisons la méthode d'inversion des moindres carrés itérativement repondérée (IRLS) [20]. IRLS est un moyen populaire de résoudre des solutions inverses clairsemées et/ou robustes en raison de sa simplicité. En effet, IRLS reformule le calcul de la solution de la moindre valeur absolue comme une répétition d' itérations des moindres carrés en utilisant des poids dans les espaces de données et de modèle. Cette approche est plus efficace que la résolution directe de la solution de la plus petite valeur absolue, car nous pouvons utiliser des solveurs de moindres carrés plus rapides comme le gradient conjugué pour réduire les temps de calcul. L'utilisation d'IRLS permet également des inversions de normes mixtes ; pour cet article, nous utilisons une norme ℓ2 dans l'espace des données et une norme ℓ1 dans l'espace modèle.

Comme de nombreux chercheurs l'ont exploré, l'application d'une contrainte clairsemée à la transformée de Radon offre de nombreux avantages et peu d'inconvénients pour la cartographie des données sismiques [17]. Ces améliorations résolvent des problèmes tels que les données manquantes dues au crénelage et aux informations d'ouverture limitées lors de la cartographie et de la reconstruction. Les avantages s'étendent également à l'utilisation de la transformée Radon en démélange. Le principal avantage de l'application d'une contrainte clairsemée dans l'espace modèle pour notre méthode est que, en tant qu'intégrale de chemin, le bruit erratique et aléatoire issu des temps de retard aléatoires est mal mappé dans l'espace de Radon. Cette mauvaise cartographie entraîne des frottis de faible amplitude dans l'espace de Radon, contrairement aux points focaux de forte amplitude observés lors d'arrivées hyperboliques cohérentes. Ceci est complété par le fait que dans les images de prise de vue auxquelles appartiennent les données, les signaux s'additionnent et se concentrent correctement. Par conséquent, le solveur convergera vers la solution la plus clairsemée en réduisant la pondération du bruit mélangé et en augmentant la pondération des points focaux cohérents. Il s'agit également de la solution la plus démixée, car les données sont remappées dans les images où elles semblent les plus cohérentes, et le bruit de mélange s'explique entièrement par le mélange des plans voisins, détendant ainsi efficac

Machine Translated by Google

L'IRLS existe comme un moyen de résoudre indirectement la solution clairsemée d'un problème d'inversion en le présentant comme plusieurs itérations d'un problème des moindres carrés, dans ce cas, le gradient conjugué (CG), les itérations externes mettant à jour les données et les poids du modèle pour obtenir le CG. algorithme résolu, conduisant la solution à la norme ℓ1. L'application de la norme ℓ1 dans l'espace de données permet de réduire la sensibilité aux valeurs aberrantes qui autrement auraient un impact négatif sur l'inversion, souvent appelée inversion robuste. L'application de la norme ℓ1 dans l'espace modèle résout la solution la plus clairsemée, principalement le modèle qui explique les données avec le plus petit nombre de points de données ; c'est ce qu'on appelle la solution clairsemée. L'utilisation de l'IRLS nous permet de mélanger et de faire correspondre les normes que nous souhaitons utiliser via le calcul des poids et, ainsi, de trouver une solution de normes mixtes.

En ajoutant des poids de modèle et de données, nous pouvons ensuite dicter à l'algorithme de converger vers la solution de la moindre valeur absolue (ℓ1) ou des moindres carrés (ℓ2) en reformulant l' équation des moindres carrés d'origine grâce à la minimisation de l'équation modifiée :

$$\begin{aligned} &\text{Réduire } ||Wmm||2\ 2 \\ &\text{Sous réserve de } ||Wr(b - \Gamma Lm)||2\ 2\ . \end{aligned} \tag{7}$$

où Wr est l'opérateur de pondération des données et Wm est l'opérateur de pondération du modèle. Ces poids permettent à l'utilisateur d'ajuster la préférence soit au résidu, soit au modèle, où Wm peut être modifié pour obtenir un certain ajustement à un modèle clairsemé ℓ1 ou lisse ℓ2 , et Wr peut être modifié pour préférer les événements non spontanés ou le « poids » " tous les événements de manière égale dans les données d'entrée. Les opérateurs sont généralement des matrices diagonales. Par exemple, Wr peut être construit comme suit :

$$\begin{aligned} &\text{diag}(Wr) = |r|\ (p-2)/2 \\ &\text{où } r = \Gamma Lm - b. \end{aligned} \tag{8}$$

où p vaut 1 ou 2 pour mesurer la taille du modèle en utilisant la norme de données ℓ1 ou ℓ2 .
Wm est normalement construit comme

$$\text{diag}(Wm) = |m|\ (2-p)/2. \tag{9}$$

où p vaut 1 ou 2 pour une inversion basée sur la norme du modèle ℓ1 ou ℓ2 . En raison des pondérations des données créées par division, un facteur d'amortissement est choisi pour éviter la division par zéro. Le facteur d'amortissement est déterminé sous forme de centile des données. Le principal avantage de l'approche d'inversion pour le démélange est que tout le signal et le bruit de mélange peuvent être expliqués dans d'autres domaines où ils semblent cohérents, par opposition à l' approche de débruitage, qui cherche à supprimer le bruit de mélange image par image.

Nous pouvons également utiliser l'inversion clairsemée pour démontrer l'approche de débruitage en matière de dé-mélange en utilisant la transformée de Radon pour débruiter les regroupements de récepteurs individuels. L' approche de débruitage applique d'abord un pseudo-démélange aux données mélangées pour renvoyer un ensemble de données dans des images d'acquisition standard, comme illustré sur la figure 1b, qui permet ensuite la séparation des prises de vue par débruitage dans les autres domaines. Le processus général de débruitage utilisant Radon clairsemé/robuste applique d'abord l'opérateur de mélange adjoint aux données mélangées, b :

$$D\tilde{} = b\Gamma^{H}, \tag{10}$$

Ensuite, les données pseudo-démélangées sont utilisées dans l'inversion avec la fonction objectif suivante pour débruiter à l'aide de la transformée de Radon :

$$||Wr(D\tilde{} - Rm)||_p + \mu||Wmm||_q{}_q, \tag{11}$$

où D˜ est les données pseudo-mélangées, R est l'opérateur de transformation de Radon, m est le modèle, μ est le paramètre de compromis et p et q sont respectivement les données et les normes du modèle.
Une représentation graphique du débruitage du Radon peut être vue sur la figure 2, en utilisant la norme

Machine Translated by Google

Radon hyperbolique pour réduire la pondération des événements spontanés dans l'enregistrement du tir dans le but de récupérer le signal cohérent.

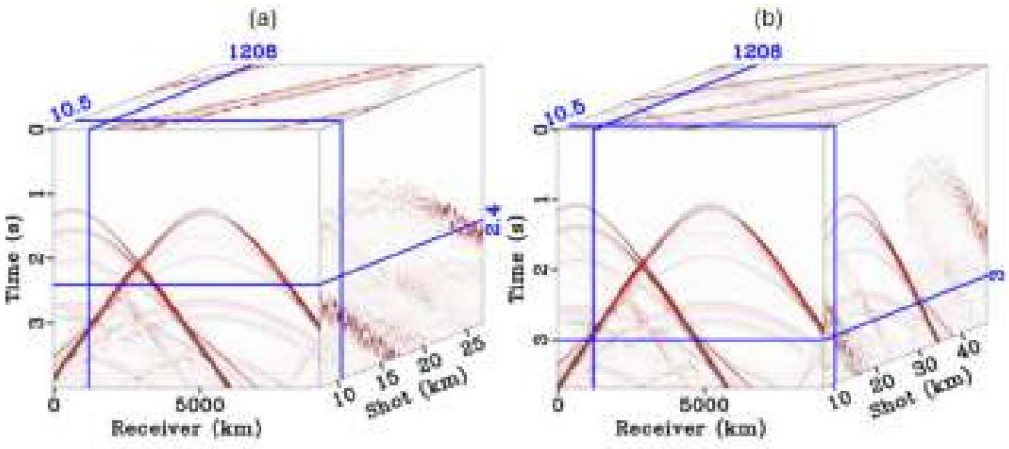


Figure 1. Application de l'opérateur adjoint : (a,b) est le résultat ; on peut voir dans d'autres domaines géométriques que notre cible est cohérente alors que les signaux interférents restent incohérents.

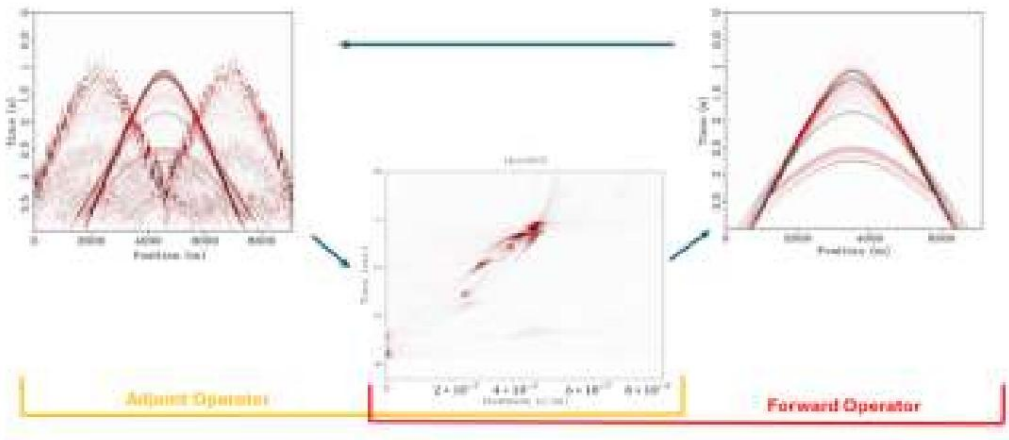


Figure 2. Approche de débruitage pour le démélange : L'approche de débruitage pour le démélange utilise du Radon clairsemé/robuste pour supprimer le bruit de mélange de chaque trame de récepteur pseudo-démélangée.

Étant donné que le bruit de mélange d'un enregistrement de plan est un signal provenant d'un autre, notre bruit de mélange sporadique a les mêmes amplitudes que celui du plan lui-même. Ce problème à un pli de mélange élevé se manifeste par le bruit de haute amplitude qui submerge complètement les sections d'événements de faible amplitude dans le signal, ce qui rend les méthodes de récupération traditionnelles basées sur le filtrage presque impossibles. Cependant, l'utilisation de l'inversion nous permet d'expliquer les événements interférents de forte amplitude et de remapper efficacement uniquement l'amplitude interférente, nous permettant ainsi de récupérer les signaux de plus faible amplitude. Le schéma de l' approche basée sur l'inversion est visible dans la figure 3. Cependant, cette approche n'est pas sans inconvénients , principalement en termes de besoins en mémoire, car l'ensemble des données de mélange ainsi que le modèle correspondant doivent être conservés en mémoire. Malgré les attentes initiales, cette méthode ne présente pas de différence de durée d'exécution significative par rapport à une approche similaire de débruitage basée sur l'inversion, puisqu'une inversion doit être effectuée dans les deux cas. Par rapport à une approche de débruitage Radon, l'approche d'inversion utilise une fonction objectif différente car elle cherche à incorporer à la fois l'opérateur de mé Nous pouvons formuler la fonction objectif en utilisant les poids déjà évoqués ainsi que l'équation (6) :

$$J = ||Wr(b - \Gamma Lm)||_p + \mu||Wmm||_q . \quad (12)$$

Machine Translated by Google

Ici, l'entrée n'est pas les données pseudo-démélangées, mais les données mélangées, et l'opérateur contient à la fois les opérateurs de mélange et Radon.

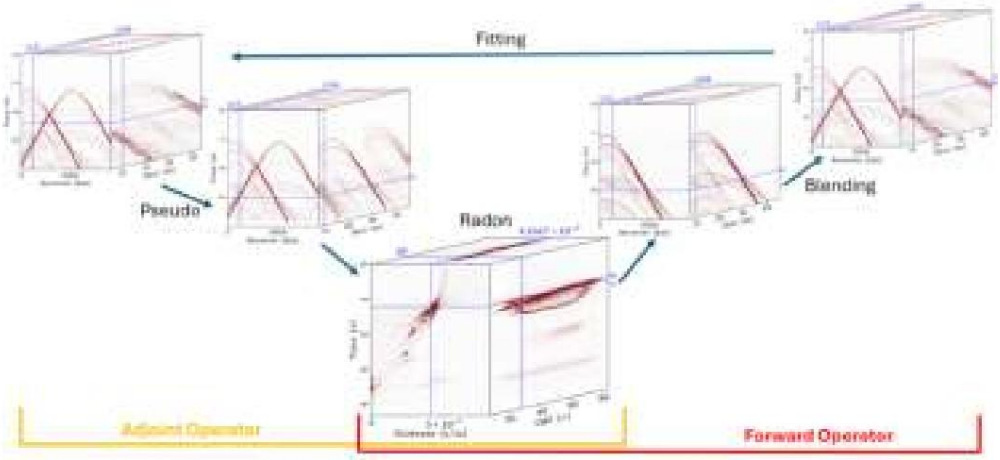


Figure 3. Schéma du démélange basé sur l'inversion : combinaison de l'opérateur de mélange avec le Radon La transformation nous permet de remapper les interférences de mélange sporadiques en signal cohérent dans son emplacement approprié, expliquant efficacement, plutôt que supprimant, les interférences.

En implémentant le démélange en tant qu'opération basée sur l'inversion en utilisant le mélange et Radon ensemble, le système agit comme un dictionnaire trop complet, où les les informations sont représentées et redondantes dans les trames pseudo-démixées. Cela aide à convergence car les informations redondantes s'expliquent mutuellement grâce au remappage des données vers leurs images de prise de vue respectives.

3. Considérations relatives au calcul haute performance (HPC)

La transformée de Radon hyperbolique clairsemée est utilisée dans notre algorithme de démélange pour reformater l'opérateur de mélange mal posé en un problème d'inversion qui utilise une régularisation clairsemée. Cet opérateur Radon nous permet de comparer les performances d'un système hautement parallélisable opération utilisant une série de méthodes de traitement parallèles. Pour coder correctement le THS en fonctionner efficacement dans openMP, nous envoyons des boucles de décalage externes à des threads séparés pour calculer le transformer en opérateur adjoint ; pour l'opérateur forward, nous envoyons ensuite chaque boucle de lenteur externe à la place. Le pseudo-code de la transformée de Radon peut être écrit comme le montre Algorithme 1.

Algorithme 1 Pseudo-code du radon

```
1 : fonction RADON
2 : #PRAGMA OMP PARALLEL FOR pour q = lenteur do                    insert pour parallélisation de boucle
3 :
4 :         pour h = décalage do
5 :             déplacement = h 2 q
6 :             pour ça = 0 à ne pas faire
7 :                 temps = (il dt) / (2 + déménagement)
8 :                 modèle[iqNt + it]+ = données[ihNt + INT(time/dt)]
9 :             fin pour
10 :        fin pour
11 : fin pour
12 : fin de la fonction
```

Pour notre implémentation dans CUDA, nous avons mis en place une configuration de thread bidimensionnelle avec une dimension pour chacune des dimensions temps et lenteur dans le code standard, puis calculé la transformation indépendamment pour chaque cellule de sortie pour chaque thread. Pour mettre en œuvre le HRT dans openMPI, nous envoyons les données relatives à chaque plan de la transformation pour séparer

Machine Translated by Google

nœuds, puis calculez la transformation indépendamment sur chaque nœud en utilisant les méthodes CUDA ou openMP indiquées ci-dessus. Nous récupérons ensuite les données chez le maître à la fin du calcul. Le pseudo-code CUDA est visible dans l'algorithme 2.

Algorithme 2 CUDA Radon Pseudo Code 1 :

```
__global__ 2 :
fonction CUDA_ RADON int it =
3 :     blockIdx.x * blockDim.x + threadIdx.x int iq = blockIdx.y *
4 :     blockDim.y + threadIdx.y if (iq >= nq || it >= nt) return int
5 :     iqNt = iq     nt double timemax = dt
6 :     nt pour h=offset do
7 :     int ihNt = ih     nt double moveout
8 :     = h 2     q double
9 :         time = (it     dt)
10 :        modèle[iqNt + it] + = data[ihNt
11 :        + (INT)(time/dt)] 13 : fin pour     2 + déménagement
12 :        14 : fin de la fonction
```

Étant donné qu'un nombre important d'applications parallèles utilisent des données multidimensionnelles, l' API CUDA regroupe les blocs de threads en plusieurs dimensions, jusqu'à trois. Ceci est bénéfique pour la mise en œuvre de Radon car il existe généralement trois dimensions de données pour la transformation, une pour le décalage, le temps et la vitesse. Cependant en affectant chacune de nos variables à une dimension, nous pouvons éliminer leurs boucles respectives, mais ce n'est pas la manière la plus efficace d' implémenter l'algorithme. Étant donné que le modèle est une accumulation sur la dimension de décalage, nous serions confrontés à une course aux données si la dimension de décalage h était exécutée en parallèle. Si le décalage était défini comme troisième dimension de thread dans CUDA, un opérateur atomique serait requis ; cela a été initialement testé et a produit des résultats nettement plus lents que l'algorithme fourni ci-dessus. En parallélisant uniquement les deux dimensions de la grille, nous évitons d'avoir recours à des opérateurs d'accumulation atomique, qui peuvent provoquer des ralentissements importants, chaque thread effectuant un travail pour chaque cellule de la grille dans ce que l'on peut appeler une optimisation alignée sur la sortie.

Cet algorithme est intéressant dans la mesure où la majeure partie peut être optimisée en parallèle, mais il existe de nombreux défis pour une bonne parallélisation. Le principal défi de l'optimisation et de la parallélisation de l'algorithme est dû aux échanges de mémoire importants requis pour chaque itération. Ces problèmes de mémoire surviennent en raison du besoin constant de changer de domaine dans l'opérateur de mélange, ce qui entraîne un accès non séquentiel à grande échelle aux données. L'autre défi auquel nous sommes confrontés est que cet algorithme ne peut pas être correctement parallélisé sur les systèmes de mémoire distribuée (serveurs) utilisant openMPI, car le calcul du gradient à chaque itération doit être calculé à l'aide de l'opérateur mixte qui nécessite le recours à toutes les données. Étant donné que la mémoire distribuée dans les clusters nécessite de diviser l'ensemble de données entre les nœuds au lieu de tout conserver dans la RAM du nœud maître, le processus de recours impose un défi très difficile pour une implémentation openMPI. En raison de cette exigence, chaque nœud d'un cluster devra renvoyer des données après chaque calcul pour calculer l'étape suivante, puis le maître devra renvoyer les résultats à chaque nœud. Pour notre implémentation dans CUDA, au lieu d'utiliser plusieurs nœuds, nous utilisons une seule machine hautes performances, où les algorithmes de descente de Radon et des moindres carrés sont réécrits à l'aide d

4. Résultats

Comme nos données de test sont principalement des données synthétiques et marines, nous supposons que l'espace de données est exempt de valeurs aberrantes ; ainsi, nous avons utilisé le ℓ2 dans l'espace de données, pour une inversion de norme mixte ℓ2-ℓ1 . Pour nos tests, nous avons utilisé trois ensembles de données, deux créés synthétiquement par différences finies et un ensemble de données marines réelles du golfe du Mexique qui a été mélangé numériquement. Pour les ensembles de données synthétiques, une modélisation par différences finies a été exécutée avec 300 tirs et 300 récepteurs régulièrement espacés à la surface. Le premier ensemble de données et le plus simpliste

Machine Translated by Google

est l'ensemble de données du modèle en coin illustré à la figure 4. Le modèle en coin se compose de deux couches, un coin triangulaire et un ensemble de points durs pour tester la reconstruction des diffractions de faible amplitude . Les données du coin ont été mélangées à deux prises de vue superposées avec des temps de retard randomisés entre 0 et 400 échantillons.

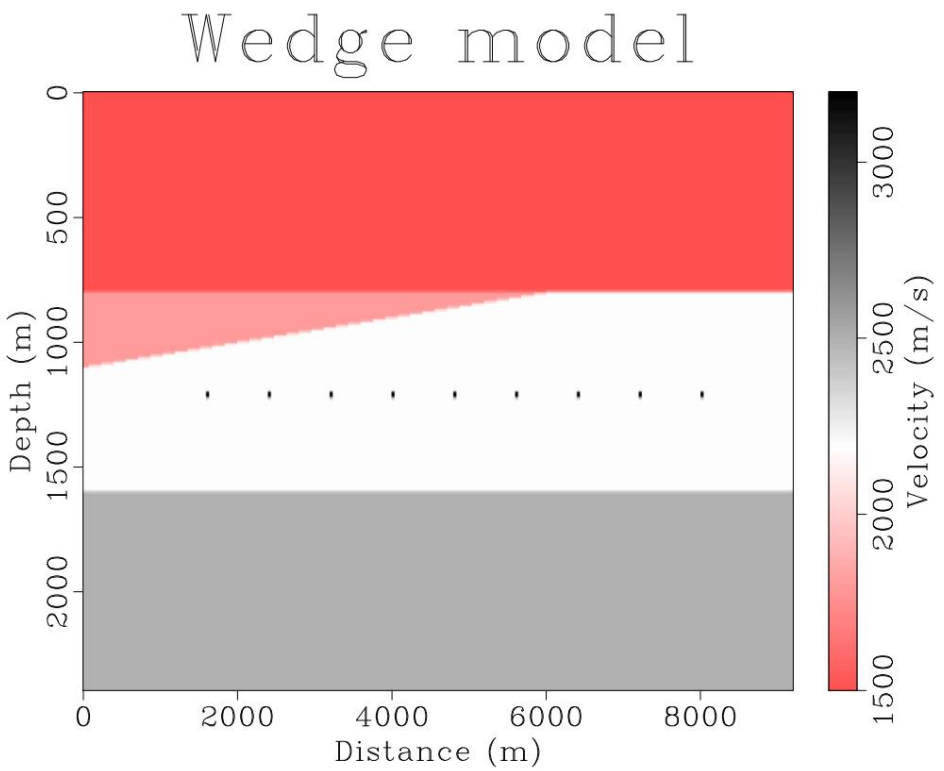


Figure 4. Modèle de vitesse 1 : modèle de coin composé de deux points de diffraction de couche 9 et d'un coin.

La démélange est visible sur la figure 5, où le bruit de mélange observé dans les données pseudo-démélangées sur la figure 5b est supprimé par inversion sur la figure 5c. Par la suite, les résultats ont été transformés dans le domaine du tir, comme le montre la figure 6.

Malgré quelques interférences résiduelles dans le tir mélangé, le processus de démélange a effectivement éliminé le tir secondaire interférent et a bien conservé le tir cible, y compris les diffractions. Ceux-ci peuvent être observés près du haut des figures 6b et c.

Une comparaison de notre méthode d'inversion et d'une méthode de débruitage est présentée à la figure 7. La méthode de débruitage utilise également le HRT clairsemé comme moteur de débruitage pour supprimer le bruit de mélange aléatoire dans les regroupements pseudo-démélangés. La principale distinction par rapport à la méthode d'inversion est que l'approche de débruitage n'implique pas l'opérateur de mélange. Dans le domaine CMP, les amplitudes élevées du bruit de mélange rendent difficile l'élimination directe par les méthodes générales de réduction du bruit. En effet, différencier le bruit des données devient de plus en plus difficile car le bruit de mélange a la même amplitude que le signal que nous souhaitons récupérer. Sans les informations supplémentaires fournies par l'opérateur de mélange, cette similarité rend difficile la distinction entre les parties des données qui sont du bruit et celles qui sont des signaux cohérents. Comme le montre la figure 7, le résultat du débruitage a réussi à démêler les données, mais au prix des données de plus faible amplitude des arrivées de diffraction, qui ont été supprimées avec le bruit de mélange.

Machine Translated by Google

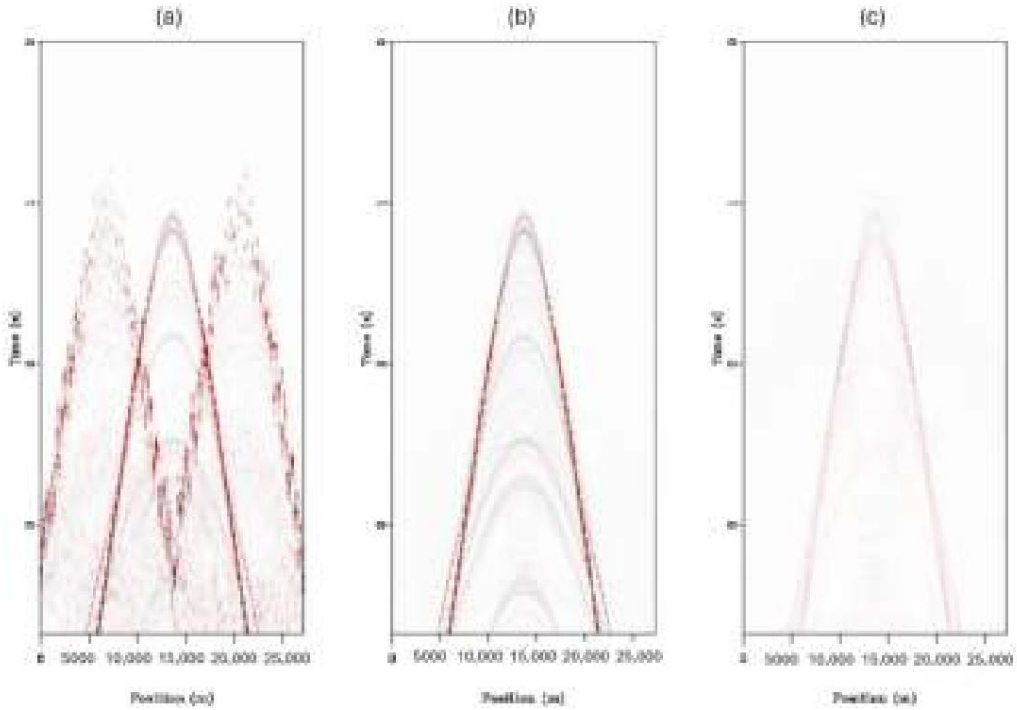


Figure 5. Résultats du test du modèle Wedge (CMP) : test de démélange 1 dans le domaine CMP avec (a) les données pseudo-démélangées, (b) les données démélangées et (c) l'erreur de reconstruction affichée, respectivement.

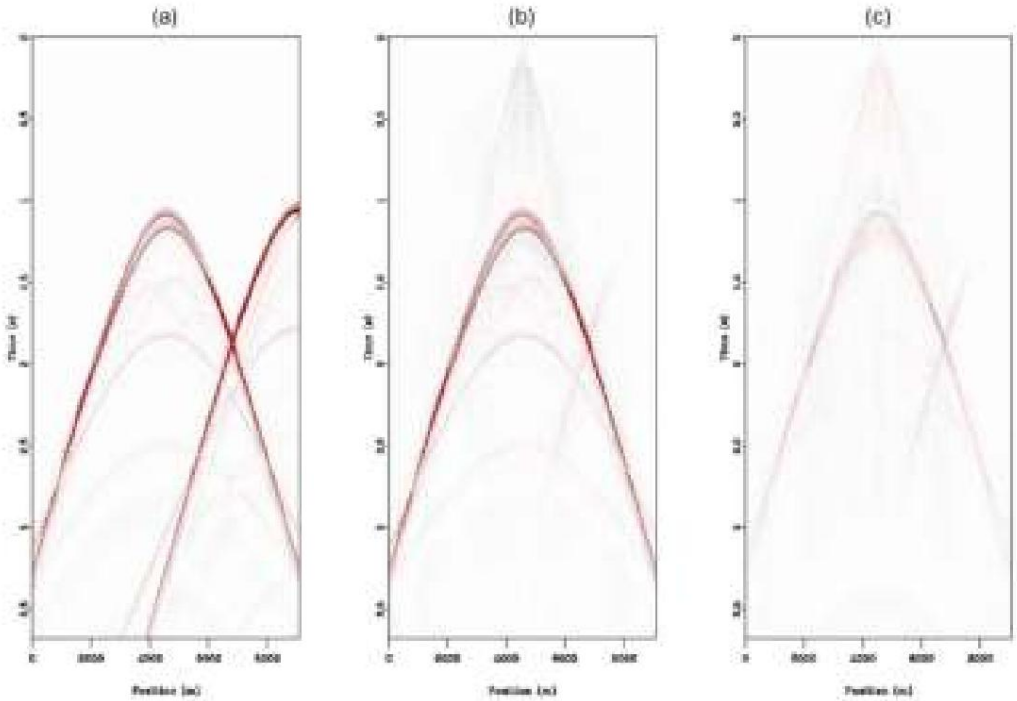


Figure 6. Résultats du test du modèle Wedge (domaine de tir) : Test de démélange 1 dans le domaine de tir avec (a) le données pseudo-mélangées, (b) les données démixées et (c) l'erreur de reconstruction affichée, respectivement.

Machine Translated by Google

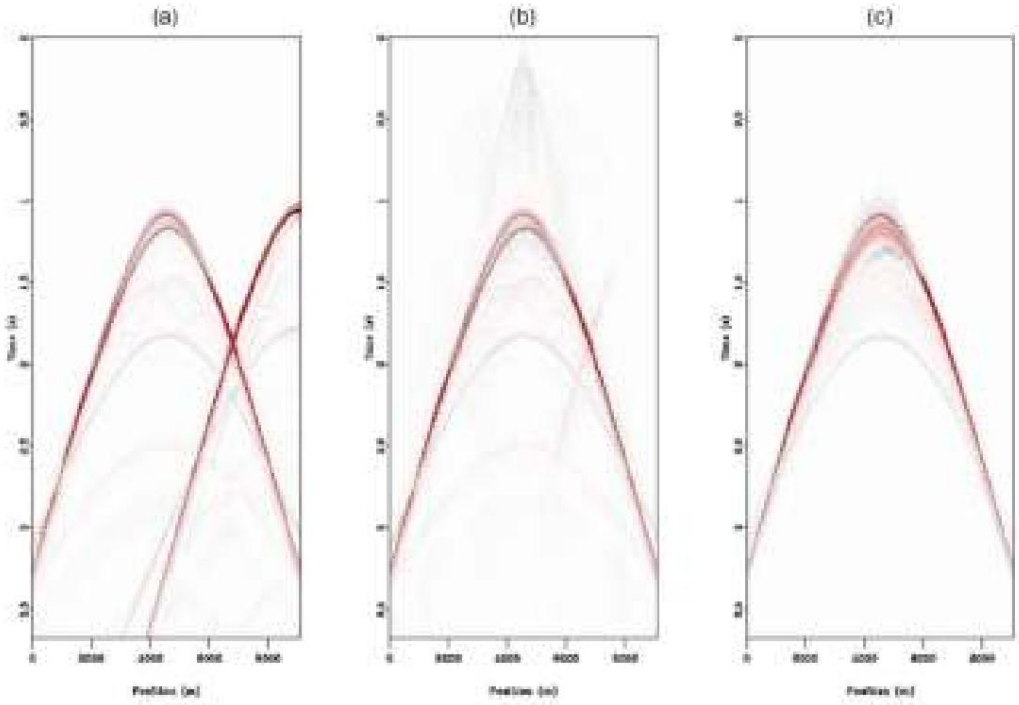


Figure 7. Comparaison des résultats du modèle de coin entre les approches d'inversion et de débruitage : approches d'inversion et de débruitage dans le domaine du tir, avec (a) mélangé, (b) inversion débruitée et (c) débruitage débrouillé.

Le test synthétique numéro 2 a été généré par différences finies en utilisant le modèle standard de Marmousi [21]. Comme le test 1 a montré que notre moteur de fusion réussissait efficacement à décomposer nos données, nous avons décidé de faire du test numéro 2 davantage un test de résistance, dans lequel une méthode de fusion non basée sur l'inversion aurait du mal à démêler efficacement. Nous avons considérablement augmenté la difficulté en réduisant la plage de nos temps de retard aléatoires et en augmentant également le nombre de tirs superposés de deux à cinq, des conditions que nous ne nous attendons pas à voir dans le monde réel actuellement. La figure 8a montre les données combinées dans le domaine CMP ; il est facile de visualiser à quel point il serait difficile pour un débruiteur normal sans l'opérateur de mélange de distinguer les données des interférences dans cette image.

En analysant les résultats du deuxième test, nous pouvons voir sur la figure 8 que notre récupération des réflexions de faible amplitude est efficace même lorsqu'elle est complètement submergée par les interférences de haute amplitude provenant d'autres prises de vue. En examinant l'erreur de reconstruction pour le domaine de tir sur la figure 9, nous pouvons voir que les principales différences sont les amplitudes réduites des événements de plus forte amplitude et l'introduction d'un bruit de fond de très faible amplitude ; nous pensons que l'erreur d'amplitude peut provenir d'une trop grande parcimonie dans l'espace du modèle, ce qui provoque généralement ces problèmes dans le Radon haute résolution, ou d'une séparation incomplète des plans de mélange et de la diffusion des amplitudes plus élevées sur les données sous forme de bruit de fond.

Notre test final était un ensemble de données du monde réel extrait du golfe du Mexique, connu sous le nom d'ensemble de données BP Tiber WATS [8,22], qui a ensuite été mélangé numériquement. Le schéma de mélange a été configuré différemment des autres, car il s'agissait également d'un test d'un autre programme d'acquisition appelé écoute continue. En écoute continue, nous avons effectivement un seul supershot, car les données sont enregistrées jusqu'à ce que tous les coups soient tirés et que les coups puissent se chevaucher. L'ensemble de données réduit comprenait 90 tirs et 90 récepteurs, avec un chevauchement plan par plan de 70 %, randomisé entre 0 et 200 échantillons. Les résultats de la fusion de l'ensemble de données du golfe du Mexique peuvent être vus dans le domaine CMP sur la figure 10 et dans le domaine des tirs sur la figure 11. En analysant les résultats, nous pouvons voir que l' algorithme de fusion est exceptionnellement efficace pour la séparation des tirs sur le terrain réel. ensemble de données mondial, avec une bonne récupération de tous les événements observés dans les données. La principale erreur que l'on peut observer est que da

Machine Translated by Google

domaine, les plans ne sont pas parfaitement mélangés et une certaine amplitude du plan arrière est encore visible dans le plan récupéré. Cela peut être observé vers 5,2 s sur la figure 11c.

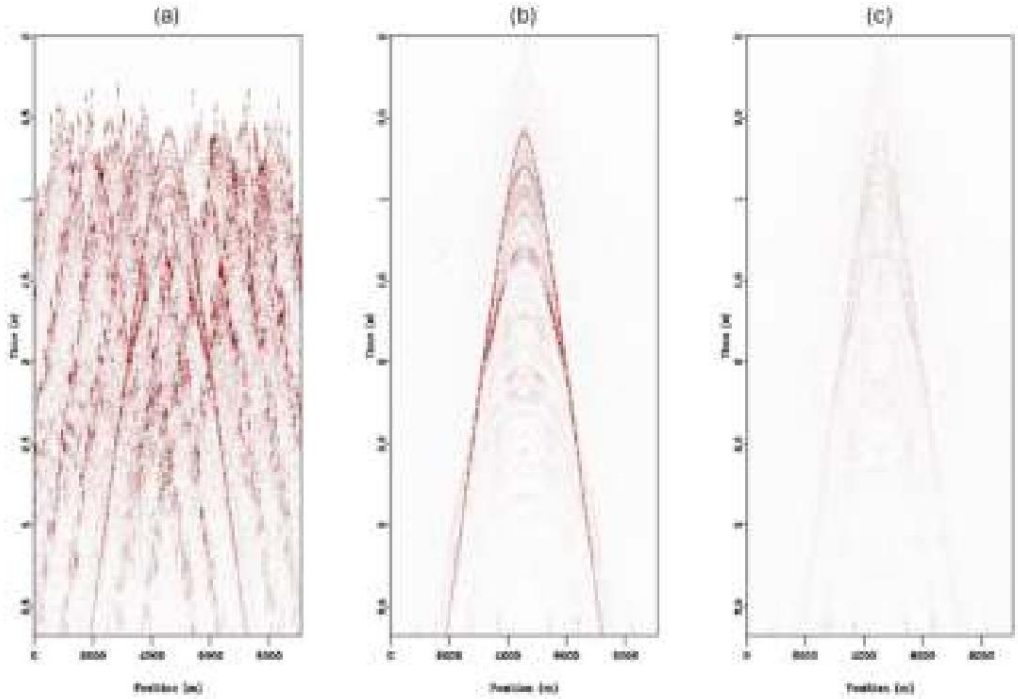


Figure 8. Résultats du test du modèle Marmousi (CMP) : test de démélange 2 avec 5 tirs, avec (a) mélange, (b) démélange et (c) erreur de reconstruction, respectivement.

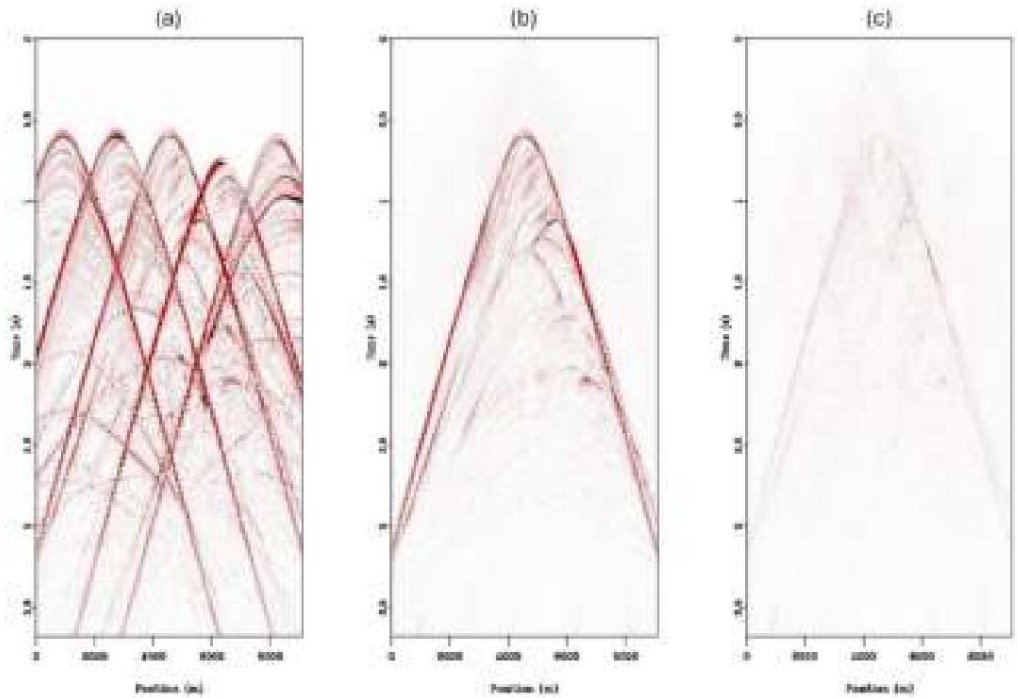


Figure 9. Résultats du test du modèle Marmousi (domaine Shot) : test de démélange 2 avec 5 tirs, avec (a) mélange, (b) démélange et (c) erreur de reconstruction, respectivement.

Machine Translated by Google

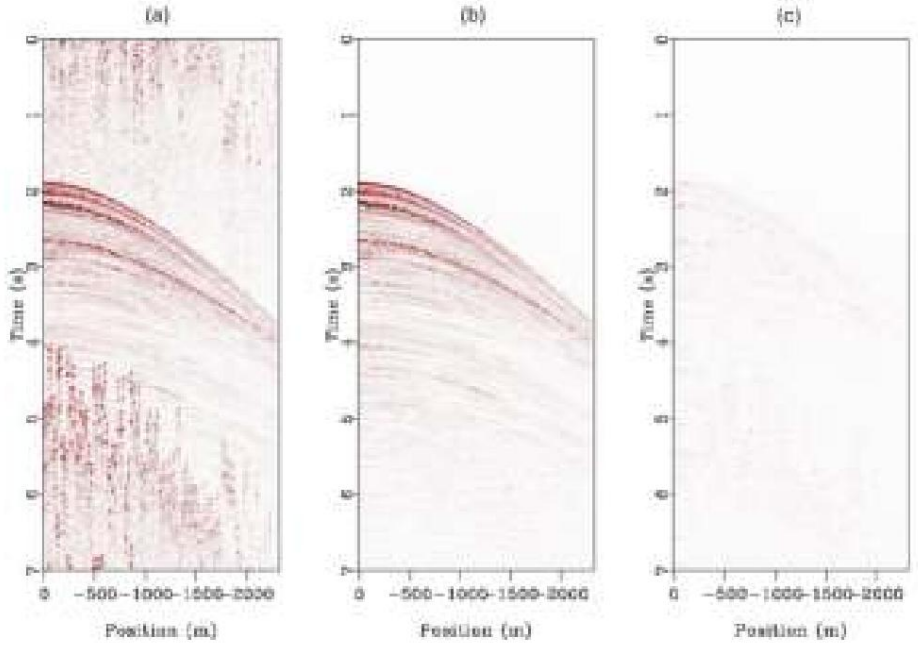


Figure 10. Données du Golfe du Mexique dans le domaine CMP : Démélange à l'aide d'un ensemble de données marines du Golfe du Mexique avec des résultats dans le domaine CMP, avec des données mélangées en (a), démixées en (b) et l'erreur de reconstruction en (c).

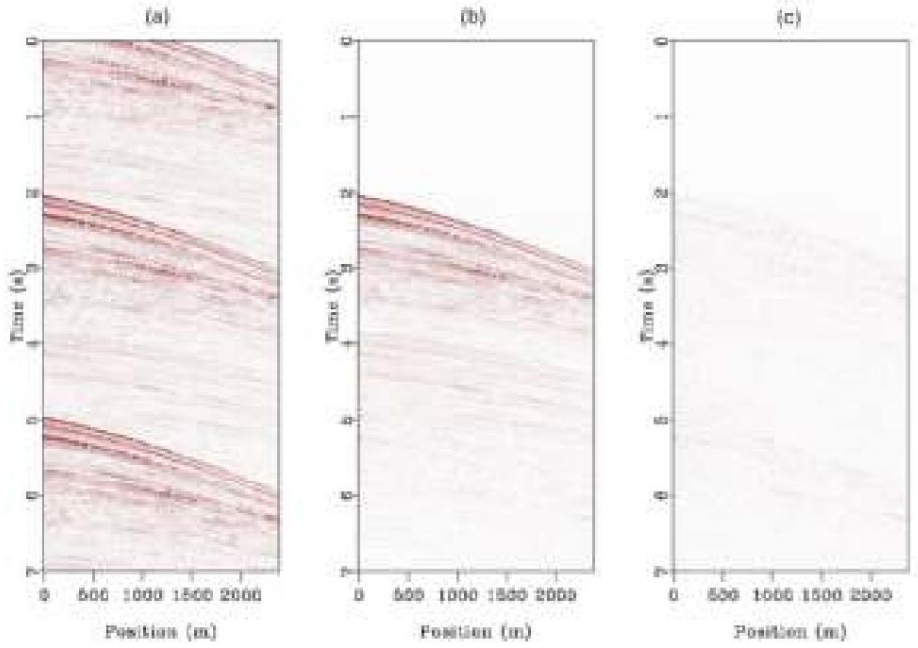


Figure 11. Données du Golfe du Mexique : Démélange à l'aide d'un ensemble de données marines du Golfe du Mexique avec une interférence de temps de prise de vue de 70 %, avec des données mélangées en (a), démixées en (b) et une erreur de reconstruction en (c).

Les limites de cette approche résident dans les hypothèses inhérentes à l'utilisation de la transformée hyperbolique de Radon. L'hypothèse principale de cette approche est la trajectoire hyperbolique de toutes les arrivées sismiques, qui devraient être représentées avec précision dans l'espace de Radon. Cependant, cette présomption devient problématique lorsque l'on considère les caractéristiques uniques des données sismiques terrestres. La statique inhérente aux données terrestres peut induire des déplacements verticaux des échantillons, s'écartant de leur alignement hyperbolique. De telles divergences conduisent à une cartographie inefficace dans le domaine de Radon et compromettent potentiellement l'intégrité du processus de démél

Machine Translated by Google

Nous avons exécuté à la fois l'algorithme de démélange et l'algorithme de transformation de Radon sur une gamme de différents matériels pour montrer les performances des algorithmes en utilisant les différentes méthodes de parallélisation dans différentes conditions : un ordinateur portable, deux postes de travail et notre interne cluster de serveurs. L'ordinateur portable utilise un processeur Intel i7 8750H à 6 cœurs avec un Nvidia RTX 2060 GPU MAX-Q, la première station de travail est équipée d'un AMD Threadripper 3960× 24-CPU principal, avec un Nvidia RTX 3080Ti, et la deuxième station de travail dispose d'un Intel i7 9800× Processeur à 8 cœurs équipé d'un super GPU Nvidia RTX 2060. Ces résultats peuvent être vus dans le tableau 1. Le temps d'exécution entre la transformée de Radon et l'algorithme de démélange ne peuvent pas être directement comparés car ils effectuaient des tâches différentes ; la différence relative dans le temps d'exécution entre les différents matériels montre l'effet du tri sur le différentes mises en œuvre.

Tableau 1. Tableau des durées d'exécution moyennes pour la fusion du radon et des moindres carrés sur différentes machines.

Processeur (API)	Temps de transformation du radon (s)	Temps (s) de démélange LS
i7 6800k 6 cœurs (openMP) i5	4462	2590
8600 6 cœurs (openMP) i7	3773	1748
8750H 6 cœurs (openMP) i7 9800	4044	2203
× 8 cœurs (openMP)	2715	1390
TR 3960 × 24 cœurs (openMP)	1123	581
RTX 3060 (CUDA)	147	174
RTX 3080Ti (CUDA)	108	142
RTX 2060 Max-Q (CUDA)	184	206
RTX 2060 Super (CUDA)	157	198
TR 3960 × (openMPI)	420	—

On peut voir qu'il y a un avantage informatique significatif à passer de Traitement CPU (openMP) vers traitement basé sur GPU (CUDA). Les résultats openMPI montrent une accélération significative par rapport aux résultats openMP et combler l'écart avec le CUDA résultats; cette accélération peut principalement être attribuée à la nature plus grossière de l'openMPI. L'implémentation openMPI partage moins de ressources entre les threads ; donc globalement moins la gestion de la mémoire doit être effectuée par le système, au détriment de l'espace RAM global à mesure que davantage de copies de données sont effectuées. Dans les tests standard de transformation de Radon, le test basé sur CUDA le code a réduit le temps de calcul de plus d'un ordre de grandeur par rapport à openMP version; après l'introduction de l'algorithme des moindres carrés, l'avantage a diminué de une grande quantité. Grâce à openMPI distribuant tout le travail, y compris l'optimisation LS entre les nœuds, il peut exécuter l'optimisation LS de manière parallèle, sans que ni l'un ni l'autre openMP ni le code CUDA ne le peuvent, ce qui entraîne un temps de calcul globalement plus rapide par rapport à openMP et approches de la mise en œuvre de CUDA. Cependant, le démélange OpenMPI ne peut pas être correctement implémenté car le nœud hôte doit collecter toutes les données de l'autre des nœuds pour effectuer le calcul LS à chaque itération ; cela entraîne des transferts de mémoire importants sur le réseau, ce qui entraîne une latence et un taux de transfert importants. Les frais généraux associé à l'implémentation d'openMPI est si important que nous n'avons pas remarqué d'avantage dans notre prototypage du programme par rapport aux environnements d'exécution même à thread unique. Il Il faut cependant noter que c'est openMPI qui a utilisé la plus grande quantité de RAM dans cet exercice, environ 122 Go, par rapport aux implémentations openMP et CUDA qui utilisaient 8 Go de RAM et 8 Go + 8 Go de VRAM (en miroir), respectivement. Le principal compromis entre un Une approche basée sur CUDA/GPU et une approche OpenMP seraient celles de la vitesse par rapport à portabilité; l'approche CUDA est nettement plus rapide mais nécessite un GPU Nvidia, tandis que l'approche OpenMP fonctionnera sur tout appareil prenant en charge la parallélisation OpenMP dans le compilateur, qui est n'importe quel processeur moderne existant sur le marché.

5. Conclusions

Nous avons proposé une méthode de démélange qui utilise une inversion linéaire d'un composite opérateur de transformation dans le domaine CMP. Cette inversion, lorsqu'elle est utilisée avec le mélange

Machine Translated by Google

opérateur, nous permet de prédire et de remapper, plutôt que d'éliminer, les informations de tir interférentes. Cela facilite à son tour la récupération précise des événements de faible amplitude, même dans les données fortement mélangées. L'implémentation d'une contrainte parcimonie sur l'algorithme d'inversion nous permet d'arriver à une solution unique qui améliore la cohérence du signal en arrivées hyperboliques. De plus, l'utilisation de transformations hyperboliques de Radon dans le domaine CMP nous fournit un opérateur efficace et bien adapté à nos données, favorisant la cohérence de notre récupération et évitant les problèmes d'alias. Par conséquent, nous sommes convaincus que notre approche d'inversion est une stratégie efficace pour démêler des données provenant de sources simultanées tout en protégeant les événements de faible amplitude. Nous avons également exploré des méthodes supplémentaires pour accélérer ce framework via des applications sur le CPU et le GPU, chacune avec son propre ensemble d'avantages et d'inconvénients.

Contributions des auteurs : Conceptualisation, KZ et DT ; méthodologie, KZ et DT ; logiciel, KZ; validation, KZ, DT et AI ; analyse formelle, KZ ; enquête, KZ; conservation des données, KZ et AI ; rédaction – préparation du projet original, KZ ; rédaction : révision et édition, IA et DT ; supervision, DT Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Nous remercions les sponsors du Consortium de recherche en sismologie d'exploration par ondes élastiques (CREWES) pour leur soutien continu. Ce travail a été financé par les sponsors industriels du CREWES et du CRSNG (Conseil de recherches en sciences naturelles et en génie du Canada) grâce à la subvention CRDPJ 543578-19.

Déclaration de disponibilité des données : Aucune nouvelle donnée n'a été créée ou analysée dans cette étude. Le partage de données ne s'applique pas à cet article.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Abréviations

Les abréviations suivantes sont utilisées dans ce manuscrit :

COF	Décalage commun
CMP	Point médian commun
CRG	Collecte du récepteur commun
HRT	Transformation hyperbolique du radon
ASHRT	Transformation hyperbolique de Radon décalée par Apex
CG	Dégradé conjugué
IRLS	Moindres carrés itératifs repondérés
CUDA	Architecture de périphérique unifiée de calcul (API de programmation Nvidia)
API	Interface de programmation d'applications

Références

1. Garottu, R. Enregistrement simultané de plusieurs lignes sismiques vibroseis. Dans Résumés élargis du programme technique SEG 1983 ; Society of Exploration Geophysicists : Houston, Texas, États-Unis, 1983 ; pp. 308-310. [Référence croisée]
2. Beasley, CJ ; Chambres, RE ; Jiang, Z. Un nouveau regard sur les sources simultanées. Dans Résumés élargis du programme technique SEG 1998 ; Society of Exploration Geophysicists : Houston, Texas, États-Unis, 1998 ; pp. 133-135. [Référence croisée]
3. Beasley, CJ Un nouveau regard sur les sources marines simultanées. Plomb. Bord 2008, 27, 914-917. [Référence croisée]
4. Mahdad, A. ; Doulgeris, P. ; Blacquière, G. Séparation des données mixtes par estimation itérative et soustraction du mélange bruit d'interférence. Géophysique 2011, 76, Q9–Q17. [Référence croisée]
5. Abma, R. ; Howe, D. ; Favoriser, M. ; Ahmed, moi; Tanis, M. ; Zhang, Q. ; Arogunmati, A. ; Alexander, G. Acquisition et traitement simultanés indépendants de sources. Géophysique 2015, 80, WD37–WD44. [Référence croisée]
6. Berkhout, AJ Changer la mentalité en matière d'acquisition de données sismiques. Plomb. Bord 2008, 27, 924-938. [Référence croisée]
7. Lin, T. ; Herrmann, F. Conception d'acquisitions simultanées avec détection de compression. Dans Actes de la 71e conférence et exposition EAGE intégrant SPE EUROPEC, Amsterdam, Pays-Bas, 8-11 juin 2009 ; p. cp-127-00269. [Référence croisée]
8. Ibrahim, A. Séparation des sources sismiques simultanées à l'aide d'une inversion robuste du radon et d'opérateurs de migration. doctorat Thèse, Université de l'Alberta, Edmonton, AB, Canada, 2015.
9. Liu, B. ; Sacchi, M. Interpolation de norme pondérée minimale des enregistrements sismiques. Géophysique 2004, 69, 1560-1568. [Référence croisée]
10. Chen, Y. Démélange à l'aide d'un filtre médian variable dans l'espace. Explorer. Géophysique. 2015, 46, 332-341. [Référence croisée]

Machine Translated by Google

11. Mahmoudian, F. ; Curtis, J. ; De Meersman, K. ; Kennedy, R. ; Staples, G. Démélange de données sources simultanées. Disponible en ligne : https://geoconvention.com/wp-content/uploads/abstracts/2023/91390-deblending-of-simultaneous-source-.pdf (consulté le 25 juin 2024).

12. Stanton, A. ; Wilkinson, K. Démélange robuste de données sismiques sources simultanées. arXiv 2018, arXiv : 1812.06040.
13. Berkhout, AJ; Blacquière, G. ; Verschuur, DJ Le concept de double blending : combiner des prises de vue incohérentes et incohérentes détection. Géophysique 2009, 74, A59-A62. [Référence croisée]
14. Jiang, Z. ; Abma, R. Une analyse sur l'imagerie simultanée de données sources simultanées. Dans le programme technique SEG élargi Résumés 2010 ; Society of Exploration Geophysicists : Houston, Texas, États-Unis, 2010 ; pp. 3115-3119. [Référence croisée]
15. Thorson, JR ; Claerbout, JF Inversion stochastique à pile de vitesse et à pile inclinée. Géophysique 1985, 50, 2727-2741. [Référence croisée]
16. Foster, DJ ; Mosher, CC Suppression des réflexions multiples à l'aide de la transformée de Radon. Géophysique 1992, 57, 386-395. [Référence croisée]
17. Trad, D. ; Ulrych, T. ; Sacchi, M. Dernières vues de la transformation clairsemée de Radon. Géophysique 2003, 68, 386-399. [Référence croisée]
18. Trad, D. ; Hargreaves, N. ; verWest, B. ; Wombell, R. Atténuation multiple à l'aide d'une transformation de radon par déplacement d'apex. Dans SEG étendu Résumés ; Society of Exploration Geophysicists : Houston, Texas, États-Unis, 2004. [CrossRef]
19. Claerbout, JF Analyse des sondages terrestres : traitement versus inversion ; Blackwell Scientific Publications : Londres, Royaume-Uni, 1992.
20. Balances, JA ; Gersztenkorn, A. ; Treitel, S. Solution Fast Ip de grands systèmes linéaires clairsemés : application à la tomographie du temps de trajet sismique. J. Informatique. Phys. 1988, 75, 314-333. [Référence croisée]
21. Brougois, A. ; Bourget, M. ; Lailly, P. ; Poulet, M. ; Ricarte, P. ; Versteeg, R. Marmousi, modèle et données. Dans les actes de l'EAEG atelier-aspects pratiques de l'inversion des données sismiques, Copenhague, Danemark, 28 mai-1er juin 1990.
22. Naranjo, JC; Ekstrand, EJ; Etgen, J. ; Hargrove, KL ; Brewton, JG; García, O. ; Astvatsaturov, G. ; Hite, D. ; Howieson, B. ; Bai, B. ; et coll. Conception d'enquête et mise en œuvre d'une stratégie d'acquisition sismique à plusieurs streamers remorqués à grand azimut au Tiber Discovery ; eaux profondes du golfe du Mexique, États-Unis. Dans Résumés élargis du programme technique SEG 2011 ; Society of Exploration Geophysicists : Houston, Texas, États-Unis, 2011 ; 132-136. [Référence croisée]

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Algorithme de construction de table de priorité pour les CFG, qu'ils soient ou non des OPG

Leonardo Lizcano, Eduardo Angulo et José Márquez *

Département d'informatique et d'ingénierie, Universidad del Norte, Barranquilla 081007, Colombie ;
ldlizcano@uninorte.edu.co (LL); edangulo@uninorte.edu.co (EA)
* Correspondance : jmarquez@uninorte.edu.co

Résumé : Les grammaires à priorité d'opérateur (OPG) sont des grammaires hors contexte (CFG) qui se caractérisent par l'absence de deux symboles non terminaux adjacents dans le corps de chaque production (côté droit). Les langages à priorité d'opérateur (OPL) sont déterministes et sans contexte. Trois relations de préséance possibles entre paires de symboles terminaux sont établies pour ces langages. De nombreux CFG ne sont pas des OPG car la préséance des opérateurs ne peut leur être appliquée car ils ne respectent pas la règle de base. Pour résoudre ce problème, nous avons procédé à une redéfinition approfondie des ensembles de terminaux gauche et droit qui constituent la base du calcul des relations de préséance, et nous avons défini un nouvel ensemble de terminaux le plus à gauche. Les algorithmes permettant de les calculer sont également décrits en détail. La contribution la plus significative de notre travail est que nous établissons des relations de préséance entre les terminaux en surmontant la règle de base de ne pas avoir deux non-terminaux consécutifs en utilisant un algorithme qui permet de construire la table de préséance des opérateurs pour un CFG, qu'il s'agisse ou non d'un OPG. L' article montre la complexité des algorithmes proposés et les exceptions possibles aux règles proposées. Nous présentons des exemples en utilisant un OPG et deux non-OPG pour illustrer le fonctionnement des algorithmes proposés. Avec ceux-ci, la table de priorité des opérateurs est construite et l'analyse ascendante est effectuée correctement.

Mots-clés : langage de priorité des opérateurs ; grammaire de priorité des opérateurs ; tableau de priorité ; relations de préséance ; algorithme

check for updates

Citation : Lizcano, L. ; Angulo, E. ; Márquez, J. Algorithme de construction de table de priorité pour les CFG, qu'ils soient ou non des OPG. Algorithmes 2024, 17, 345. https://doi.org/10.3390/a17080345

Rédacteur académique : Dimitris Fotakis

Reçu : 24 juin 2024
Révisé : 29 juillet 2024
Accepté : 30 juillet 2024
Publié : 7 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Cet article se penche sur la famille des langages à priorité d'opérateur (OPL), créée par R. Floyd [1], qui est utilisée pour une analyse ascendante efficace [2,3]. L'inspiration de l'auteur de la structure des expressions arithmétiques, où la priorité est attribuée aux opérateurs multiplicatifs (et /) sur les opérateurs additifs (+ et −), est un aperçu clé. Grâce au processus d'analyse ascendante, le côté gauche (symbole non terminal) remplace complètement la production identifiée du côté droit (corps) dans une grammaire sans contexte (CFG), ne laissant aucune place à l'ambiguïté. La priorité des opérateurs a été étendue aux grammaires composées de terminaux sur le côté droit, même si des non-terminaux les séparaient. Une CFG est généralement considérée comme une grammaire de priorité d'opérateur (OPG) si elle ne contient pas de productions avec des symboles non terminaux adjacents sur son côté droit.

Les OPL ont grandement aidé à déduire des grammaires sans contexte. Un exemple est décrit dans [4–6], où une méthode est présentée pour déduire des grammaires sans préséance d'opérateur à partir d'échantillons entre parenthèses. La méthode est basée sur la construction d'arbres d'analyse et l'utilisation de fonctions qui produisent les grammaires. Les règles de priorité sont utilisées pour identifier les parenthèses entourant les éléments terminaux d'une chaîne dans l'échantillon, indiquant l'ordre dans lequel la chaîne doit être évaluée.

La recherche sur les langages formels pilotés par les entrées (IDL) [7–9] (renommés plus tard Visiblement Pushdown Languages (VPL) [10]) a conclu [11] que les propriétés de fermeture OP impliquent des propriétés de fermeture ID et que les langages ID sont un type spécifique. des langages OP caractérisés par des relations OP limitées [12].

Machine Translated by Google

Depuis les progrès du calcul parallèle, les propriétés d'analyse locale dérivées d'OPL ont été utilisés pour générer des analyseurs parallèles rapides [13].

La principale contribution de cet article est sa nouvelle approche pour construire des relations de préséance entre les symboles de terminaux dans les non-OPG. Cette approche permet la génération de tables de priorité des opérateurs et une analyse ascendante sur une chaîne dérivée du non-OPG avec le support de la table obtenue. La nouveauté du travail est un aspect important qui le distingue des recherches antérieures dans le domaine. • Les ensembles de terminaux Gauche et Droit proposés dans les travaux précédents [1,11,14] pour construire des relations de préséance entre les terminaux d'un OPG ont été redéfinis. • Un nouvel ensemble de terminaux appelé Leftmost a été défini pour prendre en charge la résolution des relations de priorité lorsqu'il y a deux ou plusieurs terminaux adjacents sur le côté droit des productions. Cet ensemble est important car il gère systématiquement les relations de préséance dans des scénarios aussi complexes, élargissant ainsi le processus d'analyse ascendante aux non-OPG.

- Un algorithme a été établi pour construire la table de préséance avec le support des trois ensembles de terminaux proposés.
- Enfin, l'application des algorithmes proposés permet d'obtenir à partir d'un non- OPG la table de relations et de préséance des opérateurs auparavant limitée aux seuls OPG. De plus, un analyseur ascendant de priorité des opérateurs pourrait être utilisé pour analyser toute chaîne générée par un non-OPG, une technique qui ne pourrait être appliquée qu'aux OPG.

Comme démontré dans ce travail, la solution proposée pour construire la table de priorité des opérateurs n'est pas seulement un concept théorique mais un outil pratique qui peut être appliqué aux OPG sans aucune restriction. À la fin des travaux, la fonctionnalité est démontrée en l'appliquant à un OPG et à deux non-OPG et en effectuant l'analyse ascendante respective . Cette démonstration pratique devrait inspirer confiance dans l'efficacité de la solution proposée et garantir son applicabilité pratique.

L'article est organisé comme suit : La section 2 fournit les définitions nécessaires pour comprendre ce travail. La section 3 présente les travaux antérieurs réalisés dans le domaine du CFG. La section 4 présente notre approche, détaillant le problème que nous cherchons à résoudre et les redéfinitions des concepts des travaux précédents utilisés pour développer nos algorithmes. La section 5 présente des exemples d'application de l'algorithme proposé à un OPG et à deux non-OPG, démontrant l'application de l'analyse ascendante avec les tables de priorité pour tester leur utilisation. La section 6 décrit plusieurs exceptions où les règles décrites dans les algorithmes proposés ne s'appliquent pas. Enfin, la section 7 présente les conclusions de notre travail.

2. Définitions de base

2.1. Grammaires et langues sans contexte

Au sein de la hiérarchie de Chomsky [15], les grammaires sans contexte jouent un rôle important dans les applications des langages de programmation et de compilation en abordant la structure syntaxique des langages de programmation.

Une grammaire sans contexte (CFG) G génère un langage L(G), communément appelé langage sans contexte (CFL), qui est composé des chaînes générées par la CFG. Le CFG est défini comme un 4-tuple G(S) =< T, N, S, P >, où : • T représente l'ensemble des symboles terminaux qui créent les chaînes dans la CFL générée par le CFG. • N représente l'ensemble des symboles non terminaux ou des variables syntaxiques qui déterminent les chaînes menant à la génération CFL. • S est le symbole non terminal initial à partir duquel toutes les chaînes définies par le GFC sont généré.

- P est l'ensemble des productions ou règles qui regroupent les terminaux et les non-terminaux pour générer les chaînes. Une production prend la forme A −→ α, où A représente un non-terminal et α est une combinaison de terminaux et de non-terminaux. Le −→ se prononce « produire ». Ainsi, la production se lit « A produire alpha ».

Machine Translated by Google

La notation suivante sera utilisée : les lettres minuscules du début de l' alphabet représentent les symboles terminaux (a, b, c, . . . T) ; les lettres majuscules du début de l'alphabet représentent des symboles non terminaux (A, B, C, . . . N) ; les lettres minuscules tardives dans l'alphabet représentent des chaînes terminales (u, v, w, x, y, z T +) ; Les lettres grecques minuscules représentent généralement des chaînes terminales et/ou non-terminales (α, β, γ, . . . (T N)), et la chaîne vide sera représentée par ε (cette lettre grecque sera l'exception à la convention ci-dessus).

2.2. Grammaires de priorité des opérateurs (règle de base)

Un CFG est considéré comme une grammaire d'opérateurs si aucune des productions de P n'a de symboles non terminaux adjacents sur son côté droit. Autrement dit, pour une production A −→ α dans G, α ne prend pas la forme uBCv.

2.3. Dérivations

Une chaîne η (T N) peut être dérivée en k étapes à partir d'un non-terminal A N s'il existe une séquence de chaînes de la forme η1, η2, . . . , ηk . En utilisant le symbole = , signifiant dérivation en une étape, la séquence de chaînes est organisée sous la forme : η1 = η2 = . . . = ηk .

Dans cette séquence, η1 correspond à A et ηk à η. Chaque intermédiaire ηi dans la séquence a la forme δBγ, où B N et il existe une production B −→ β, telle qu'en la substituant du côté droit, on obtient ηi+1(= δβγ) . La dérivation en une ou plusieurs étapes est notée par l'opérateur += . En général, il est indiqué que chaque chaîne de symboles terminaux w L(G) peut être établie comme S += w ou en développant la dérivation en une ou plusieurs étapes : S = η1 = η2 = . . . = ηk = w. Pour une grammaire G, une forme phrase η (T N) = η, ce qui signifie que η peut être obtenu avec zéro ou plusieurs dérivations de S. chaîne que S Si deux est un

grammaires* génèrent des langues identiques, elles sont considérées comme équivalentes.

2.4. Types de dérivations

D'après [16], S = η1 = η2 = . . . = ηk = w est une dérivation la plus à gauche de la chaîne w T lorsque chaque ηi , 2 ≤ i ≤ k− 1, a la forme xiAiβi avec xi T Ai N et βi (T N) . De plus, Ai −→ αi *, appartient à P, et alors, dans chaque ηi , le symbole Ai est remplacé par αi , ce qui donne la forme phrase ηi+1 . Chaque ηi est appelée la forme phrase de gauche.

Si chaque ηi , 2 ≤ i ≤ k − 1, prend la forme βiAixi avec xi T βi (T N) forme *, Ai N, et phrase droite. , alors une dérivation la plus à droite serait formée. Chaque ηi est appelé le

2.5. Analyser les arbres

D'après la référence [16], un arbre de dérivation pour un CFG G(S) =< T, N, S, P > est un arbre étiqueté et ordonné dans lequel chaque nœud reçoit une étiquette correspondant à un symbole de l'ensemble N T {ε}. Si un nœud non-feuille est étiqueté A et que ses descendants immédiats sont étiquetés X1, X2, . . . , Xn, puis A −→ X1X2 . . . Xn est une production dans P. Un arbre ordonné étiqueté D est un arbre de dérivation pour un CFG G(A) =< T, N, A, P > si

1. La racine de D est notée A.
2. Si D1, . . . , Dk sont les sous-arbres des descendants directs de la racine et la racine de Di est notée Xi , puis A −→ X1 . . . Xk est dans P.
3. Di est un arbre de dérivation pour G(Xi) =< T, N, Xi , P > si Xi est dans N ou Di est un seul nœud étiqueté Xi si Xi est dans T ou Di est un seul nœud étiqueté ε.

Les arbres d'analyse peuvent être construits à partir de n'importe quel type de dérivation.

Machine Translated by Google

2.6. Analyse ascendante

L'analyse syntaxique ascendante [2], également connue sous le nom d'analyse par shift-reduce, tente de construire un arbre d'analyse pour une chaîne d'entrée qui commence dans les feuilles (en bas de l'arbre) et se déplace vers la racine (sommet de l'arbre). Ce processus est considéré comme la réduction de la chaîne w en le symbole initial S d'un CFG. À chaque étape de réduction de l'analyse, une sous-chaîne particulière du la forme phrase qui correspond au côté droit d'une production est remplacée par le non terminal symbole du côté gauche de cette production. Si, à chaque étape, la sous-chaîne est choisie correctement, une dérivation la plus à droite est tracée inversement.

2.7. Poignée

Une poignée [2,17] d'une forme phrase droite γ est une production A −→ β et une position de γ où la chaîne β pourrait être trouvée et remplacée par A pour produire la forme précédente de phrase à droite dans une dérivation la plus à droite de γ. Généralement, si S $= ^{*} \alpha Aw = \alpha\beta w$, alors A −→ β si la position à côté de α est une poignée de αβw. Si la grammaire est ambiguë, plus que une poignée sera obtenue. Si la grammaire est sans ambiguïté, alors toute phrase droite le formulaire a exactement une poignée. Habituellement, la chaîne β est considérée comme un handle de αβw lorsque le les conditions pour y parvenir sont claires.

2.8. Implémentation d'une analyse ascendante

L'analyse ascendante par décalage et réduction utilise une pile pour stocker les symboles de grammaire et un tampon pour gérer la chaîne d'entrée à analyser. Le symbole $ est également utilisé pour délimiter le bas de la pile et le côté droit du tampon d'entrée. Le format du modèle de reconnaissance commence par la pile à $ et la chaîne w$ dans l'entrée, comme indiqué dans le tableau 1.

Tableau 1. Tableau d'analyse ascendant.

Empiler	Saisir	Action
$	w$	
...	...	Décaler/Réduire
...	...	Décaler/Réduire
$S	$	Accepter

L'analyse ascendante décale zéro ou plusieurs symboles d'entrée vers la pile jusqu'à ce qu'une poignée soit en haut. L'analyse réduit ensuite la poignée sur le côté gauche de la production appropriée pour obtenir la forme phrase correspondant à l'étape précédente de la dérivation la plus à droite. Lors de la réduction, l'analyse reconnaît que l'extrémité droite de la poignée se trouve en haut de la empiler. Par conséquent, l'analyse prend des mesures pour trouver son extrémité gauche dans la pile et déterminer la non terminal qui le remplacera selon le côté droit de certaines productions où il correspond à la poignée. Ces étapes sont répétées jusqu'à ce qu'une erreur soit détectée ou jusqu'à ce que la pile contient $S et l'entrée est avec le symbole $, auquel point l'analyse se termine et le La chaîne d'entrée est considérée comme valide pour le CFG.

3. Travaux antérieurs

D'après [1,11,14], pour un CFG G, le terminal gauche définit LG(A) et le terminal droit les ensembles RG(A) sont définis comme suit :

$$LG(UNE) = une|UNE = ^{*} \gamma a\alpha, \gamma \in N \cup \{\varepsilon\}.$$

$$RG(UNE) = une|UNE = ^{*} \alpha a\gamma, \gamma \in N \cup \{\varepsilon\}.$$

Les relations de préséance résultent de l'établissement par l'arbre d'analyse de relations binaires entre des terminaux consécutifs ou ceux qui deviennent séquentiels après une approche ascendante. processus vers le symbole non terminal S.

Machine Translated by Google

Les relations de préséance sont établies comme suit : a .= b, si et seulement

1. si A −→ αaγbβ P, γ N {ε}.
2. a b, si et seulement si A −→ αaDβ P, b LG(D). 3. a b, si et seulement si A −→ αDbβ P, a RG(D).

Selon [2], les relations de préséance sont utilisées pour délimiter un handle sous une forme phrase droite. Autrement dit, sous la forme phrase αβw, la sous-chaîne β est un handle s'il existe une production A −→ β dans le CFG ; par conséquent, il peut être délimité par les relations de préséance, obtenant α β w. La relation .= entre les symboles utilisés lorsque la poignée a plus d'un symbole terminal consécutif. Autrement dit, si β = a1a2 . . . an, alors la relation ai .= i = ai+1 , 1, . . . , n − 1 doit être établi. Supposons qu'une forme de phrase droite soit A0a1A1a2 . . . anAn, chaque Ai est un non-terminal. Dans ce cas, les relations peuvent être établies comme ai Aiai+1 , avec un maximum d'un non-terminal (principe de grammaire des opérateurs) entre les terminaux. Les non-terminaux pourraient être éliminés sous la bonne forme de phrase, ne laissant que les relations entre terminaux.

En bref, chaque fois qu'une poignée est obtenue, elle peut être enfermée entre les symboles et . De plus, deux symboles de terminaux consécutifs peuvent se trouver à l'intérieur d'une poignée (non-terminaux au milieu), et la relation doit être établie .=. Enfin, les symboles non terminaux peuvent être supprimés de la forme de phrase droite lorsqu'un handle est trouvé et verrouillé entre les relations de préséance.

Selon [2], les concepts abordés dans la section 2.8 et les relations de préséance, l'algorithme d'analyse de préséance 1 est construit comme suit :

Algorithme 1 Algorithme d'analyse de préséance Exiger :

```
Ensemble de relations de préséance. procédure
    PrecedenceParsing(w) stack ← [$] end ← f
        alse pos ← 0
        do a ← stack[top]
        b ← w$

        [pos] si a = $ et b = $
            alors

                print ("Action : Accepter") fin ←
                vrai
            autre
                si a b ou a .= b alors pos
                    ++
                    push(stack, b)
                    print("Action: Shift") sinon si a
                  b
                    alors
                        faire
                            e ← pull(stack) while
                        not (stack[top] e) print(« Action :
                        Réduire ») else print(« Action :

                    Erreur ») end ← true end if end if
                        end if while not
                    end end
                procédure
```

Machine Translated by Google

4. Approche

Ce travail propose une redéfinition des ensembles de terminaux Gauche et Droit , correspondant respectivement à LG(A) et RG(A) . De plus, la définition du nouvel ensemble le plus à gauche est présentée. Les nouvelles règles pour déterminer les relations de préséance sont démontrées même si le CFG ne respecte pas la règle fondamentale de la grammaire des opérateurs, ce qui signifie que le CFG peut avoir plus d'un symbole non terminal adjacent sur le côté droit de ses productions.

Nous présentons les algorithmes de construction des trois ensembles proposés : Left, Leftmost et Right. De plus, nous démontrons l'algorithme qui intègre les règles de préséance nouvellement établies.

4.1. Définitions

Il existe les symboles suivants :

• Une chaîne de bornes x T +. •
Une chaîne de non-terminaux β N . •
Une forme phrase γ (T N) Les ·

ensembles suivants sont définis :

• La gauche, notée L(A), est l'ensemble des symboles terminaux qui peuvent apparaître sur le côté gauche de toute forme phrase comme produit d'une dérivation la plus à droite de A.

$$L(UNE) = \{une \quad T \mid A \overset{*}{=} \beta a\gamma\}.$$

• Le plus à gauche, noté Lm(A), est l'ensemble des symboles terminaux qui peuvent apparaître à la fin du côté gauche dans toute dérivation de A, en l'absence d'autres symboles grammaticaux les plus à gauche (ou qui dérivent vers ε en leur absence).

$$Lm(UNE) = \{une \quad T \mid UNE = \overset{*}{} \ une\gamma\}.$$

• La droite, notée R(A), est l'ensemble des symboles terminaux qui peuvent apparaître à l'extrémité droite des dérivations les plus à droite de A ou apparaître plus à droite dans toute production de A.

$$R(UNE) = \{une \quad T \mid A \overset{*}{=} \gamma a \quad A \longrightarrow \gamma a\beta\}.$$

D'après ce qui précède, les règles de préséance suivantes sont définies : b

- Si C −→ γ1aβBγ2 , L(B), alors a b.
- C −→ γ1AβBγ2 , = a ε, a R(A) a b b. Lm(B) β Si *
- Si C −→ γ1Aβbγ2 , a R(A) β = ε, * alors a b.
- Si C −→ γ1aβbγ2, alors a .= b.

4.2. Analyse du problème

4.2.1. Grammaire des opérateurs absolus

Étant donné la série de dérivations la plus à droite de la forme :

$$S \mathrel{+}= \gamma 1Ax1 = \gamma 1\omega 1x1, \text{ où } A \longrightarrow \omega 1 \text{ et } \omega 1 = \gamma 2B\gamma 3. \quad (1)$$

Il est à noter que ω1 est un handle selon la définition précédente de la section 2.7.
La forme phrase à droite de (1) peut être réécrite comme suit :

$$\gamma 1\omega 1x1 = \gamma 1\gamma 2B\gamma 3x1 \mathrel{+}= \gamma 1\gamma 2Bx2x1 = \gamma 1\gamma 2\omega 2x2x1, \text{ où } B \longrightarrow \omega 2. \quad (2)$$

De la dernière forme phrase de (2), il s'ensuit que ω2 est un handle selon la définition de la section 2.7.

Machine Translated by Google

Pour cette raison, l'analyse se concentrera sur les productions nées de A, qui ont au moins un non-terminal B à droite, comme le montre (1). Les productions dans A seront parcourues à travers la grammaire pour trouver les relations de préséance du non-terminal B situé à droite.

Dans le contexte de ω1, devant B −→ ω2, et ω2 est un handle, les formes suivantes peut être obtenu à partir de sa définition en (1) : (a)

ω1 = γ4aB, il s'ensuit que a L(B). Ici γ2 = γ4a et γ3 = ε. (b) ω1 = Bbγ5, il s'ensuit que R(B) b. Ici γ2 = ε et γ3 = bγ5. (c) ω1 = γ4aBbγ5, il s'ensuit que a L(B) et R(B) b. Ici γ2 = γ4a et γ3 = bγ5.

Pour le symbole de début non terminal, S, une production factice initiale est définie E −→ S et la règle (c) est appliquée.

Pour une production A −→ γ1aXbγ2, X N {ε}, il s'ensuit que a .= b.

4.2.2. Grammaire générale avec deux ou plusieurs symboles non terminaux consécutifs Redéfinition de gauche, de droite et définition de l'extrême gauche.

Gauche

(B) : étant donné la dérivation la plus à droite :

$$\gamma 1aBx1 = \gamma 1a\omega 1x1. \quad (3)$$

On observe que ω1 est une poignée dans une analyse ascendante (dérivation inverse la plus à droite) puisque B = ω1 ; par conséquent, à partir de (3), on peut établir que γ1a ω1 x1. De plus, ω1 peut avoir les formes suivantes : (a) ω1 = β1bγ2. (b) ω1 = β1Cγ3.

En continuant avec la dérivation la plus à droite dans (3) et en prenant chacune des options de ω1, on obtient que :

(a) γ1aBx1 = γ1aβ1bγ2x1 puisque ω1 est une poignée, par conséquent γ1a β1bγ2 x1 ; là-donc, a b où B −→ β1bγ2 et b L(B). (b) γ1aBx1 = $\gamma 1a\beta 1C\gamma 3x1 = \gamma 1a\beta 1Cx2x1 = \gamma 1a\beta 1\beta 2\overset{*}{D}x3x2x1 = \gamma 1a\beta 1\beta 2\overset{*}{\beta}3b\gamma 4\ x3x2x1$. On observe que C += β2Dx3 et D −→ β3bγ4 ; donc, β3bγ4 est une poignée, et par conséquent γ1aβ1β2 β3bγ4 x3x2x1 et a b.

En bref, γ1aBx1 += γ1aβ1β2β3bγ4x3x2x1 ; en prenant β = β1β2β3 et γ5 = γ4x3x2, on obtient la dérivation sommaire γ1aBx1 += γ1aβbγ5x1, où B += βbγ5 et b L(B).

Droite (B) :

étant donné la dérivation la plus à droite :

$$\gamma 1Bax1 = \gamma 1\omega 1ax1. \quad (4)$$

On observe que ω1 est un handle dans une analyse ascendante (dérivation inverse la plus à droite), puisque B = ω1 ; par conséquent, à partir de (4), on peut établir que γ1 ω1 ax1. De plus, ω1 peut avoir les formes suivantes : (a) ω1 = γ2bβ. (b) ω1 = γ3Cβ1, $\overset{*}{=}$ ε.

β1 En continuant avec la dérivation la plus à droite dans (4) et en prenant chacune des options de ω on obtient que :

(a) γ1Bax1 = γ1γ2bβax1 puisque ω1 est une poignée, par conséquent γ1 γ2bβ ax1 ; donc, b a où B −→ γ2bβ et b R(B). (b) γ1Bax1 = $\gamma 1\gamma 3C\beta 1ax1 = \gamma 1\gamma 3\gamma 4Dax1 = \gamma 1\gamma 3\gamma 4\gamma 5b\beta 2ax1$. On note $\overset{*}{\text{que}}$ C += γ4D et D −→ γ5bβ2 ; donc, γ5bβ2 est un handle, et par conséquent,

Machine Translated by Google

γ1γ3γ4 γ5bβ2 ax1 et b a.

En bref, γ1Bax1 += γ1γ3γ4γ5bβ2ax1 ; en prenant γ6 = γ3γ4γ5, on obtient la dérivation sommaire γ1Bax1 += γ1γ6bβ2ax1, où B += γ6bβ2 et b R(B).

Le plus à gauche (B) :

Étant donné la production avec deux symboles non terminaux à droite (corps) :

$$S - \rightarrow AB \tag{5}$$

$$A - \rightarrow \gamma 1b1\beta 1 \tag{6}$$

La dérivation la plus à droite de S dans (5) est :

$$S = \underset{\text{(ii)}}{AB +=} Aa\gamma += \underset{\text{(iv)}}{Aax1 =} \gamma 1b1\beta 1ax1 = \gamma 2Cax1 += \underset{\text{(vi)}}{\gamma 2\gamma 3\gamma 4b2\beta 2ax1.} \underset{\text{(vii)}}{\text{(i) (iii) (v)}}$$

En résumé:

• Dans (v), la relation γ1b1β1 ax1 peut s'établir puisque la partie droite du la production (6) est une poignée. Par conséquent, b1 a et b1 R(A).

• Dans (vi), on sait que β1 = C0C1 . . . CnC, où chaque Ci N et C est le non- terminal, premier dérivé de la droite. On note que γ2 = γ1b1C0 . . . CN. • Dans (vii), C += γ3γ4b2β2 étant γ4b2β2 le résultat de la dernière dérivation la plus à droite effectuée et, à son tour, une poignée ; par conséquent, la relation γ2γ3 γ4b2β2 ax1 est établie, puisque γ4b2β2 est une poignée, donc b2 a. • D'après (iv), Aax1 += γ2γ3γ4b2β2ax1. En établissant γ5 = γ2γ3γ4, alors A += γ5b2β2, donc b2 R(A). • D'après (ii), B += aγ est établi, à partir duquel nous définissons le nouvel ensemble le plus à gauche de B comme Lm(B) = {a T | B += uneγ}.

4.2.3. Une manière intuitive d'observer les éléments de chaque ensemble de vision de gauche (·) :

étant donné le CFG : A −→ CDE

E −→ eh D

−→ d C

−→ cF F

−→ f La dérivation la plus à droite de la chaîne cf deh comporte la manière suivante de créer la gauche f t(A) ou L(A).

$$A = \underset{e}{CDE =} \underset{d}{CDeh =} \underset{c}{Cdeh =} cFdeh = cfdeh.$$

Les symboles terminaux e, d et c sont vus de gauche et enregistrés dans L(A). Le symbole e est le premier vu de la gauche et est enregistré dans L(A), et le symbole h est bloqué par e puisque les deux apparaissent lorsque E est substitué. Ensuite, D est remplacé par d, qui est observé et enregistré dans L(A). Lorsque cF remplace C, nous observons et enregistrons c dans L(A).

Enfin, lorsque la substitution de F par f est effectuée, elle n'est plus observable (bloquée par c) et ne peut être enregistrée. Par conséquent:

$$L(UNE) = \{e, d, c\}.$$

Vision du Droit(·):

Compte tenu du CFG :

A −→ abBC

B −→ eF

Machine Translated by Google

C −→ cD D
−→ d F
−→ f La

dérivation la plus à droite de la chaîne abe f cd comporte la manière suivante de créer le Right(A) ou R(A).

$$A = \underset{b}{abBC} = \underset{c}{abBcD} = \underset{d}{abBcd} = abeFcd = abefcd.$$

Les symboles terminaux b, c et d sont vus de droite et enregistrés dans R(A).
Le symbole b est le premier observé et enregistré dans R(A). Le symbole a est bloqué par b, puisque les deux apparaissent lorsque A est dérivé. C est alors remplacé par cD et c est observé et enregistré dans R(A). Lorsque D est substitué à d, nous observons et enregistrons d dans R(A).
Alors B est dérivé de eF mais e ne peut pas être enregistré puisque l'inclusion précédente de c et d dans R(A) le rend invisible ou l'empêche d'être observé depuis la droite. Enfin, en substituant F à f , il ne peut pas non plus être observé depuis la droite (bloqué par c et d) et ne peut pas être enregistré. Par conséquent:

$$R(UNE) = \{b, c, d\}.$$

Vision de l'extrême gauche(·):
Compte tenu du CFG :
S −→ dAB
A − → fa
B − → Cb
C− → c

La dérivation la plus à droite de la chaîne df acb comporte la manière suivante de créer le
Le plus à gauche (B) ou Lm (B), comme le montre la figure 1.

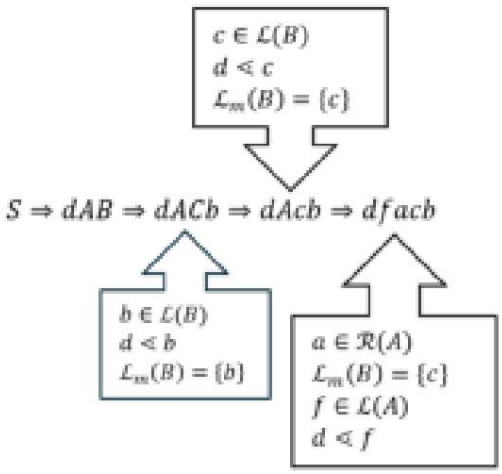


Figure 1. Dérivation de la chaîne df acb et de la séquence de création de l'ensemble le plus à gauche (B) .

On voit que L(B) = {b, c}. Lorsque B dérive Cb, nous observons que Lm(B) = {b} car aucun symbole terminal n'est vu jusqu'à ce point. Dans la dérivation suivante, C est remplacé par c ; par conséquent, c est situé plus à gauche de la sous-chaîne cb, laissant Lm(B) = {c}.

Machine Translated by Google

4.2.4. Résumé des relations entre les symboles grammaticaux des productions et des décors

Le tableau 2 montre les relations entre les symboles grammaticaux des productions et les ensembles de terminaux.

Tableau 2. Productions générales et relations de préséance obtenues.

Produire	Relation
A −→ aβb	une .= b
A − → Cβb, β $=^*$ ε	R(C)
A −→ aβC	ba L(C)
A − → BβC, β $=^*$ ε	R(B) Lm (C)

4.3. Algorithmes

Les algorithmes suivants ont été développés sur la base des définitions présentées à la section 4.1.

L'algorithme 2 reçoit un non-terminal A dont les productions sont connues ; pour chacun d' entre eux, on vérifie s'ils correspondent à la forme βBγ ; si tel est le cas, l'algorithme 2 est à nouveau exécuté , envoyant le non-terminal B comme paramètre (à condition que B = A), et sa réponse est incluse dans l'ensemble L(A).

Par la suite, on vérifie si la production correspond à la forme βaγ, et si tel est le cas, le terminal a est ajouté à l'ensemble L(A).

Une fois la vérification de toutes les productions de A terminée, l'ensemble L(A) est renvoyé comme sortie de l'algorithme 2.

Algorithme 2 Algorithme de gauche d'un non-terminal A

```
fonction Gauche(A)
    L(A) ←
    pour chaque production de A
        do si A −→ βBγ alors
            L(A) Gauche(B) // L(A) contient une extrémité

        Gauche(B) si si A −→
            βaγ alors a L(A) // a est dans
        L(A) fin
    si fin pour
    retourner L(A)
fonction de fin
```

L'algorithme 3 reçoit un non-terminal A dont les productions sont connues ; pour chacun d'entre eux, on vérifie s'ils correspondent à la forme βBγ et β = ε ; si tel est le cas, l'algorithme 3 est à nouveau exécuté, envoyant le non-terminal B comme paramètre (à condition que B = A), et sa réponse est incluse dans l'ensemble Lm(A).

Par la suite, on vérifie si la production correspond à la forme βaγ et β. Si tel est le cas, le $=^*$ ε. terminal a est ajouté à l'ensemble Lm(A).

Une fois la vérification de toutes les productions de A terminée, l'ensemble Lm(A) est renvoyé comme résultat de l'algorithme 3.

Machine Translated by Google

Algorithme 3 Algorithme le plus à gauche d'un non-terminal A

```
fonction la plus à gauche (A)
    Lm(A) ←
    pour chaque production de A do si
        A −→ βBγ et β = ε alors *
            Lm(A) Fin la plus à gauche
        (B) si si
        A −→ βaγ et β = ε alors a * Lm(A)
            fin si fin pour
        retour
    fonction
    de fin Lm(A)
```

L'algorithme 4 reçoit un non-terminal A dont les productions sont connues ; pour chacun d'entre eux, on vérifie s'ils correspondent à la forme γBβ et β = ε ; si tel est le cas, l'algorithme 4 est à nouveau exécuté, envoyant le non-terminal B comme paramètre (à condition que B = A) et sa réponse soit incluse dans l'ensemble R(A).

Par la suite, on vérifie si la production correspond à la forme γaβ, et si tel est le cas, le terminal a est ajouté à l'ensemble R(A).

Une fois que toutes les productions de A ont été vérifiées, l'ensemble R(A) est renvoyé comme sortie de l'algorithme 4.

Algorithme 4 Algorithme de droite d'un non-terminal A

```
fonction Droite(A)
    R(A) ←
    pour chaque production de A faire
        si A −→ γBβ et β = ε alors *
            R(A) Extrémité
        droite(B) si
        si A −→ γaβ alors a
            R(A) fin si
        fin pour
    retour
    fonction fin R(A)
```

L'algorithme 5 reçoit une grammaire G(S) =< T, N, S, P > et les algorithmes 2–4. Initialement, des relations de préséance sont ajoutées entre le symbole délimitant $ et les ensembles Gauche et Droite du symbole non terminal de départ S, selon les règles (A.1) et (A.2), respectivement.

L'algorithme 5 fonctionne de manière itérative. Il parcourt toutes les productions de l' ensemble P, et pour chaque itération, il initialise deux variables temporaires, u et l, respectivement à null et à une liste vide. Il effectue ensuite une série de processus pour chaque paire de symboles grammaticaux contigus X et Y dans une production.

1. L'algorithme vérifie si XY a la forme ab ; si c'est le cas, il ajoute la priorité correspondante relation entre a et b, suivant la règle (B).
2. L'algorithme vérifie si XY a la forme aB ; si tel est le cas, il attribue à u la valeur de a et ajoute les relations de préséance correspondantes entre a et les terminaux dans Left(B), en suivant la règle (C).
3. L'algorithme vérifie si XY a la forme Ab ; si tel est le cas, il ajoute les relations de préséance correspondantes entre b et les terminaux dans Right(A), en suivant la règle (D.1). Ensuite, vérifiez si u = null, et si c'est le cas, ajoutez la relation de préséance correspondante entre u et b, en suivant la règle (D.2), et réinitialisez u à null. Enfin, vérifie si la liste l contient des éléments, et si c'est le cas, ajoute les relations de priorité correspondantes entre b et les terminaux dans Right(B), en suivant la règle (D.3), et réinitialise la liste l à vide.

Machine Translated by Google

4. L'algorithme vérifie si XY a la forme AB ; si tel est le cas, il ajoute les relations de préséance correspondantes entre les terminaux de droite (A) et les terminaux de gauche (B), conformément à la règle (E.1). Ensuite, il vérifie si u = null, et si c'est le cas, ajoute la relation de préséance correspondante entre u et les terminaux dans Left(B), en suivant la règle (E.2). Enfin , il vérifie si la liste l contient des éléments et si c'est le cas, ajoute les relations de préséance correspondantes entre les terminaux de Right(C), avec C l et les terminaux de Leftmost(B), en suivant la règle (E.3), et réinitialise la liste l si nécessaire.

Algorithme 5 Algorithme pour obtenir la table de priorité des opérateurs d'une grammaire G

```
Exiger : G(S) =< T, N, S, P > procédure
    Table(G, Gauche(·), Droite(·), L'extrême gauche(·))
        calculer Gauche(A), Droite(A), Gauche(A), A N ajouter $ a,
          a Gauche(S) ajouter a                    (A.1)
        $, a Droite(S) pour chaque                 (A.2)
        A −→ γ dans P est-ce que tu ←
            null
            l ← [ ] // liste vide pour
            chaque XY contigu dans γ faire si XY = ab
                alors
                    ajouter a .= b                 (B)
                finir si
                si XY = aB alors
                    tu ← un
                    ajouter a b, b Fin             (C)

                gauche(B) si si XY =
                    Ab alors ajouter a b, a        (D.1)
                    Droite(A) si u = null
                        alors ajouter              (D.2)
                        u .= bu ←
                    fin nulle
                    si si l = [ ] alors
                        pour B l
                            ajouter a b, a extrémité   (D.3)
                        droite(B)
                        pour l
                    ← [ ] fin
                si fin si
                si XY = AB alors
                    ajoutez a b, a Droite(A) b Le plus à gauche(B)   (E.1)
                    si u = null alors
                        ajoutez u b, b Gauche(B)       (E.2)
                    finir si
                    si l = [ ] alors
                        pour C l faire
                            ajouter c b, c Droite(C) b Extrémité la plus à   (E.3)
                        gauche(B) pour
                    finir si
                    si B = *ε alors
                        ajouter A à l
                    sinon
                        l ← [ ]
                    fin si fin
                si fin
            pour fin
        pour fin
    procédure
```

Machine Translated by Google

4.4. Analyse de complexité

Les variables suivantes seront établies pour le calcul de la complexité de les quatre algorithmes proposés :

- t : Cardinalité de l'ensemble des symboles terminaux.
- p : Cardinalité de l'ensemble des productions.
- g : Nombre de symboles grammaticaux dans les productions.
- n : Cardinalité de l'ensemble des symboles non terminaux.

4.4.1. Complexité de l'algorithme 2 (à gauche)

La fonction Gauche vérifie dans quelles productions le symbole A apparaît sur le côté gauche. Dans le premier bloc if, un appel récursif de Left est observé, ce qui implique que tous les symboles non terminaux de G pourraient être parcourus à la fin de l'exécution dans la recherche de Left(A).
Par conséquent, toutes les productions les plus défavorables seraient parcourues, ce qui donnerait un temps de calcul de O(p). Le deuxième bloc if a un temps de calcul de O(1), ce qui ne contribue pas au temps de calcul global O(p).

4.4.2. Complexité de l'algorithme 3 (le plus à gauche)

La fonction Leftmost a une structure similaire à la précédente ; par conséquent, le le temps d'exécution du calcul est O(p).

4.4.3. Complexité de l'algorithme 4 (à droite)

La fonction Right a une structure similaire à Left ; par conséquent, le temps d'exécution du calcul est O(p).

4.4.4. Complexité de l'algorithme 5 (tableau)

Selon le premier cycle for, le nombre de fois où il sera exécuté sera de p. Immédiatement après le précédent, le cycle for interne est exécuté dans le pire des cas g fois, en supposant que le nombre total de symboles grammaticaux apparaisse dans chaque production.
Par conséquent, le coût de calcul des deux cycles est O(p × g).

Les coûts de calcul pour la construction de Left(A), Right(A) et Leftmost(A), A N, sont déjà calculés. Lors du calcul de la complexité du tableau, la longueur de chaque ensemble généré sera prise en compte, qui dans le pire des cas est t.

L'analyse des quatre blocs if à l'intérieur du deuxième for-cycle est la suivante :

- Dans le premier bloc if, après vérification, l'instruction exécutée est exécutée en constante temps et est noté O(1).
- Dans le deuxième bloc if, après vérification, la première instruction est exécutée en un temps constant O(1), et la seconde est exécutée en un temps de calcul O(t) ; par conséquent, le temps de calcul du bloc if est O(t).
- Dans le troisième bloc if, après vérification, trois sous-blocs sont distingués ; le premier correspond au parcours de Right(A), exécuté en un temps de calcul O(t). Après avoir vérifié u = null, le deuxième sous-bloc présente deux instructions exécutées en un temps constant O(1). Le troisième sous-bloc, après vérification de l = [ ], présente un coût de calcul O(n × t), déterminé par le cycle for exécuté, dans le pire des cas, O(n) fois, correspondant au nombre de symboles non terminaux dans G et le voyage qu'il effectue depuis Right(B) avec un coût O(t). En résumé, le coût de calcul du troisième bloc if est O(n × t), correspondant au pire des cas.
- Dans le quatrième bloc if, après vérification, quatre sous-blocs sont distingués. La première correspond aux combinaisons formées en parcourant Right(A) et Leftmost(B) pour obtenir les relations de préséance a b. Puisque la taille maximale de chaque ensemble est t, le coût de calcul de ces combinaisons est O (t^2). Après avoir vérifié u = null, le Le deuxième sous-bloc présente une instruction où Left(B) est parcouru ; par conséquent, le coût de calcul est O(t). Dans le troisième sous-bloc -block, après vérification de l = [ ], il présente un cycle for qui exécute au plus n fois les combinaisons formées à partir de

Machine Translated by Google

en parcourant Right(C) et Leftmost(B) pour obtenir les relations de préséance c b. Ces combinaisons sont parcourues au maximum en temps de calcul O(t) car t est la taille des deux ensembles et, par conséquent, le coût de calcul du troisième sous-bloc est O(n × t). Le dernier sous-bloc aura un coût de O(1). En résumé, le coût de calcul du quatrième bloc est O(n × t. En résumé, le coût de calcul des quatre blocs if dans le deuxième cycle for), correspondant au pire des cas.

de l'algorithme 5 est O (n × t le), ceci étant le pire des cas ; par conséquent, le coût de calcul de corps principal de l'algorithme 5 formé par les deux cycles for imbriqués et les quatre blocs if est O (p × g × n × t des symboles terminaux) qui est supérieur au coût de calcul de l'un des trois ensembles Gauche (A), Droite (A), et Leftmost(A), A N, au démarrage de l' algorithme, et correspondant à O(p).

5. Exemples

5.1. Exemple 1

5.1.1. Grammaire

Étant donné l'OPG suivant : E
−→ E + T | TT −→
T F | FF −→ (E)
| id Les ensembles
sont construits étape par étape en suivant les algorithmes.

5.1.2. Gauche

Les ensembles de gauche sont calculés comme détaillé dans le tableau 3.

Tableau 3. Gauche Définit le calcul pour le CFG dans l'exemple 1.

Gauche(E)			
Pour E −→ E + T			
1. Évaluation du premier conditionnel de l'algorithme 2.	A = E β = ε B = E y = +T L(E) L(E)	2. Évaluation du deuxième conditionnel de l'algorithme 2.	A = E β = E a = + γ = T L(E) {+}
Pour E −→ T			
3. Évaluation du premier conditionnel de l'algorithme 2.	A = E β = ε B = T γ = ε L(E) = {+} L(T)		

Gauche(T)			
Pour T −→ T F			
1. Évaluation du premier conditionnel de l'algorithme 2.	UNE = T β = ε B = T γ = F L(T) L(T)	2. Évaluation du deuxième conditionnel de l'algorithme 2.	UNE = T β = T une = y = F L(T) { }
Pour T −→ F			
3. Évaluation du premier conditionnel de l'algorithme 2.	UNE = T β = ε B = F γ = ε L(T) = { } L(F)		

Machine Translated by Google

Tableau 3. Suite

Gauche(F)			
Pour F −→ (E)		Pour F −→ identifiant	
1. Évaluation du deuxième conditionnel de l'algorithme 2.	A = F β = ε une = (γ = E) L(F) {(}	2. Évaluation du deuxième conditionnel de l'algorithme 2.	A = F β = ε a = identifiant γ = ε L(F) = {(, identifiant}

Résumé des ensembles gauches de non-terminaux : • Gauche(E) = {+} L(T). • Gauche(T) = { } L(F). • Gauche(F) = {(, id}.

Remplacement des ensembles : • Gauche(E) = {+, ,(, id}. • Gauche(T) = { ,(, id}. • Gauche(F) = {(, id}.

5.1.3. Droite

Les ensembles de droites sont calculés comme détaillé dans le tableau 4.

Tableau 4. Calcul des ensembles de droits pour le CFG dans l'exemple 1.

Droite(E)			
Pour E −→ E + T			
1. Évaluation du premier conditionnel de l'algorithme 4.	A = E γ = E+ B = T β = ε R(E) R(T)	2. Évaluation du deuxième conditionnel de l'algorithme 4.	A = E γ = E a = + β = T R(E) R(T) {+}
Pour E −→ T			
3. Évaluation du premier conditionnel de l'algorithme 4.	A = E γ = ε B = T β = ε R(E) = R(T) {+}		

Droite(T)			
Pour T −→ T F			
1. Évaluation du premier conditionnel de l'algorithme 4.	UNE = T γ = T B = F β = ε R(T) R(F)	2. Évaluation du deuxième conditionnel de l'algorithme 4.	UNE = T y = T une = β = F R(T) R(F) { }
Pour T −→ F			
3. Évaluation du premier conditionnel de l'algorithme 4.	UNE = T γ = ε B = F β = ε R(T) = R(F) { }		

Machine Translated by Google

Tableau 4. Suite

Droite(F)			
Pour F −→ (E)		Pour F −→ identifiant	
1. Évaluer le deuxième conditionnel de l'algorithme 2.	A = F γ = (E a =) β = ε R(F) {)}	2. Évaluer la seconde conditionnel de l'algorithme 2.	A = F γ = ε a = identifiant β = ε R(F) = {), identifiant}

Résumé des ensembles de droits de non-terminaux :

- Droite(E) = R(T) {+}.
- Droite(T) = R(F) { }.
- Droite(F) = {), id}.

Remplacement des ensembles :

- Droite(E) = {), id, , +}.
- Droite(T) = {), id, }.
- Droite(F) = {), id}.

5.1.4. Le plus à gauche

Les éléments les plus à gauche sont calculés comme détaillé dans le tableau 5.

Tableau 5. Calcul des ensembles les plus à gauche pour le CFG dans l'exemple 1.

Le plus à gauche (E)			
Pour E −→ E + T		Pour E −→ T	
1. Évaluation du premier conditionnel de l'algorithme 3.	A = E β = ε B = E y = +T Lm(E) Lm(E)	2. Évaluation du premier conditionnel de l'algorithme 3.	A = E β = ε B = T γ = ε Lm(E) = Lm(T)

Le plus à gauche (T)			
Pour T −→ T F		Pour T −→ F	
1. Évaluation du premier conditionnel de l'algorithme 3.	UNE = T β = ε B = T γ = F Lm(T) Lm(T)	2. Évaluation du premier conditionnel de l'algorithme 3.	UNE = T β = ε B = F γ = ε Lm(T) = Lm(F)

Le plus à gauche (F)			
Pour F −→ (E)		Pour F −→ identifiant	
1. Évaluer le deuxième conditionnel de l'algorithme 3.	A = F β = ε une = (γ = E) Lm(F) {(}	2. Évaluer la seconde conditionnel de l'algorithme 3.	A = F β = ε a = identifiant γ = ε Lm(F) = {(, identifiant}

Résumé des ensembles de non-terminaux les plus à gauche :

- Le plus à gauche(E) = Lm(T).

Machine Translated by Google

• Le plus à gauche(T) = Lm(F).
• Le plus à gauche(F) = {(, id}.

Remplacement des ensembles :
• Le plus à gauche(E) = {(, id}.
• Le plus à gauche(T) = {(, id}.
• Le plus à gauche(F) = {(, id}.

5.1.5. Études de cas

Pour l'algorithme 5, l'application de certaines relations de préséance est présentée.

1. Pour le symbole délimiteur $:
   (a) $ L(E) −→ Règle (A.1).
   • $ +
   • $
   • $ (
   • $ identifiant
   (b) R(E) $ −→ Règle (A.2).
   •) $
   • identifiant $
   • $
   • + $
2. Pour E −→ E + T
   (a) X = E, Y = + −→ Règle (D.1).
   R(E) +
   •) +
   • identifiant +
   • +
   • + +
   u = nul, l = [ ]
   (b) X = +, Y = T −→ Règle (C).
   + L(T)
   • +
   • + (
   • + identifiant
   u = +, l = [ ]
3. Pour F −→ (E)
   (a) X = (, Y = E −→ Règle (C).
   (L(E)
   • (+
   • (
   • ((
   • (identifiant
   u = (, l = [ ]
   (b) X = E, Y =) −→ Règle (D.1).
   R(E))
   •))
   • identifiant)
   •)
   • +)
   u = null −→ u .= b −→ (.=) −→ Règle (D.2).
   u = nul, l = [ ]

Machine Translated by Google

4. Dans les cas suivants, la forme XY n'est pas remplie dans la partie droite de la production ; aucune règle ne peut donc être appliquée :

(un) E − → T

(b) T − → F

(c) F − → identifiant

Les relations de préséance restantes sont dérivées en utilisant la même méthode que les cas démontrés précédemment.

5.1.6. Tableau de priorité des opérateurs

A la fin de l'exécution de l'algorithme, le tableau de priorité des opérateurs 6 est obtenu.

Tableau 6. Tableau de priorité de l'exemple 1.

	+	*	(	)	identifiant	$
+						
*						
				.=		
( )						
identifiant $						

5.1.7. Analyse ascendante

Soit id + ((id + id) (id)) id une chaîne obtenue par dérivation la plus à droite, puis l'application de l'algorithme 1 aboutit à l'analyse ascendante détaillée dans le tableau 7.

Tableau 7. Analyse ascendante de l'exemple 1.

Empiler	ID	Action
$	d'entrée + ((id + id) (id)) id $ +((id +	Changement
$id	id) (id)) id$ +((id + id) (id))	Réduire
$	id$ ((id + id) (id)) id$ (id + id)	Changement
$+	(id)) id$ id + id) (id))	Changement
$ +	id$ +id) (id)) id$ +id)	Changement
($ +	(id)) id$ id) (id)) id$)	Changement
(($ +	(id)) id$) (id))	Réduire
((id $	id$) (id)) id$ (id))	Changement
+ (($ +	id$ (id)) id$ (id))	Changement
((+ $ +	id$ id)) id$))	Réduire
((+id $	id$)) id$	Réduire
+ ((+		Changement
$ + (($		Réduire
+ (()		Changement
$ +		Changement
($ +		Changement
($ +		Réduire
(($ + ((id $ + ((		Changement

Machine Translated by Google

Tableau 7. Suite

Empiler	Entrée)	Action
$ + (()	id$)	Réduire
$ +	id$)	Réduire
($	id$	Changement
+ ($ + ()	id$	Réduire
$+	id$	Changement
$ +	identifiant$	Changement
$ + id	$	Réduire
$ +	$	Réduire
$+	$	Réduire
$	$	Accepter

5.2. Exemple 2

5.2.1. Grammaire

Compte tenu des non-OPG suivants :

S −→ SD ; | D;

D −→ Tid(L)

T −→ T | int

L − → je | ε

je −→ T | T, je

Les ensembles sont construits étape par étape en suivant les algorithmes.

5.2.2. Gauche

Les ensembles de gauche sont calculés comme détaillé dans le tableau 8.

Tableau 8. Gauche Définit le calcul pour le CFG dans l'exemple 2.

Gauche(S)			
Pour S −→ SD ;			
1. Évaluation du premier conditionnel de l'algorithme 2.	UNE = S β = ε B = S y = D ; L(S) L(S)	2. Évaluation du premier conditionnel de l'algorithme 2.	UNE = S β = S B = D y = ; L(S) L(D)
3. Évaluer le deuxième conditionnel de l'algorithme 2.	UNE = S β = SD une =; γ = ε L(S) L(D) {; }		
Pour S −→ D ;			
4. Évaluation du premier conditionnel de l'algorithme 2.	UNE = S β = ε B = D y = ; L(S) L(D) {; } L(D)	5. Évaluation du deuxième conditionnel de l'algorithme 2.	UNE = S β = D une =; γ = ε L(S) = L(D) {; }

Machine Translated by Google

Tableau 8. Suite

Gauche(D)			
Pour D −→ Tid(L)			
1. Évaluation du premier conditionnel de l'algorithme 2.	A = D β = ε B = T γ = identifiant(L) L(D) L(T)	2. Évaluation du deuxième conditionnel de l'algorithme 2.	A = D β = T a = identifiant y = (L) L(D) = L(T) {id}

Gauche(T)			
Pour T −→ T			
1. Évaluation du premier conditionnel de l'algorithme 2.	UNE = T β = ε B = T γ = L(T) L(T)	2. Évaluation du deuxième conditionnel de l'algorithme 2.	UNE = T β = T une = γ = ε L(T) { }
Pour T −→ int			
3. Évaluation du deuxième conditionnel de l'algorithme 2.	UNE = T β = ε a = int γ = ε L(T) = { , int}		

Gauche (L)			
Pour L −→ I			
1. Évaluation du premier conditionnel de l'algorithme 2.	A = Lβ = ε B = je γ = ε L(L) = L(Je)		

Gauche(I)			
Pour I −→ T			
1. Évaluation du premier conditionnel de l'algorithme 2.	A = je β = ε B = T γ = ε L(I) L(T)		
Pour I −→ T, je			
2. Évaluation du premier conditionnel de l'algorithme 2.	A = je β = ε B = T γ =, je L(I) L(T)	3. Évaluation du deuxième conditionnel de l'algorithme 2.	A = je β = T a =, γ = I L(I) = L(T) {, }

Résumé des ensembles gauches de non-terminaux : • Gauche(S) = L(D) {; }. • Gauche(D) = L(T) {id}. • Gauche(T) = { , int}. • Gauche(L) = L(I). • Gauche(I) = L(T) {, }.

Machine Translated by Google

Remplacement des ensembles : • Left(S) = {; , identifiant, , int}. • Gauche(D) = {id, , int}. • Gauche(T) = { , int}. • Gauche(L) = {, , , int}. • Gauche(I) = {, , , int}.

5.2.3. Droite

Les ensembles de droites sont calculés comme détaillé dans le tableau 9.

Tableau 9. Calcul des ensembles de droits pour le CFG dans l'exemple 2.

Droits)			
Pour S −→ SD ;		Pour S −→ D ;	
1. Évaluation du deuxième conditionnel de l'algorithme 4.	UNE = S γ = SD une =; β = ε R(S) {; }	2. Évaluation du deuxième conditionnel de l'algorithme 4.	UNE = S y = D une =; β = ε R(S) = {; }

Droite (D)	
Pour D −→ Tid(L)	
1. Évaluation du deuxième conditionnel de l'algorithme 4.	A = D γ = Tid(L a =) β = ε R(D) = {)}

Droite(T)			
Pour T −→ T		Pour T −→ int	
1. Évaluation du deuxième conditionnel de l'algorithme 4.	UNE = T y = T une = β = ε R(T) { }	2. Évaluation du deuxième conditionnel de l'algorithme 4.	UNE = T γ = ε a = int β = ε R(T) = { , int}

Droite (L)	
Pour L −→ I	
1. Évaluation du premier conditionnel de l'algorithme 4.	A = L γ = ε B = je β = ε R(L) = R(Je)

Machine Translated by Google

Tableau 9. Suite

C'est vrai (je)			
Pour I −→ T			
1. Évaluation du premier conditionnel de l'algorithme 4.	A = je γ = ε B = T β = ε R(I) R(T)		
Pour I −→ T, je			
2. Évaluation du premier conditionnel de l'algorithme 4.	A = je y = T, B = je β = ε R(I) R(T) R(I)	3. Évaluation du deuxième conditionnel de l'algorithme 4.	A = je y = T une =, β = je R(I) = R(T) {, }

Résumé des ensembles Right de non-terminaux : • Right(S) = {; }. • Droite(D) = {)}. • Droite(T) = { , int}. • Droite(L) = R(I). • Droite(I) = R(T) {, }.

Remplacement des ensembles : • Right(S) = {; }. • Droite(D) = {)}. • Droite(T) = { , int}. • Droite(L) = {, , , int}. • Droite(I) = {, , , int}.

5.2.4. Le plus à gauche

Les ensembles les plus à gauche sont calculés comme détaillé dans le tableau 10.

Tableau 10. Calcul des ensembles les plus à gauche pour le CFG dans l'exemple 2.

Le plus à gauche(S)			
Pour S −→ SD ;		Pour S −→ D ;	
1. Évaluation du premier conditionnel de l'algorithme 3.	UNE = S β = ε B = S y = D ; Lm(S) Lm(S)	2. Évaluation du premier conditionnel de l'algorithme 3.	UNE = S β = ε B = D y = ; Lm(S) = Lm(D)

Le plus à gauche (D)	
Pour D −→ Tid(L)	
1. Évaluation du premier conditionnel de l'algorithme 3.	A = D β = ε B = T γ = identifiant(L) Lm(D) = Lm(T)

Machine Translated by Google

Tableau 10. Suite

Le plus à gauche (T)			
Pour T −→ T		Pour T −→ int	
1. Évaluation du premier conditionnel de l'algorithme 3.	UNE = T β = ε B = T γ = Lm(T) Lm(T)	2. Évaluer la seconde conditionnel de l'algorithme 3.	UNE = T β = ε a = entier γ = ε Lm(T) = {int}

Le plus à gauche (L)	
Pour L −→ I	
1. Évaluation du premier conditionnel de l'algorithme 3.	A = L β = ε B = je γ = ε Lm(L) = Lm(I)

Le plus à gauche (I)			
Pour I −→ T		Pour I −→ T, je	
1. Évaluation du premier conditionnel de l'algorithme 3.	A = je β = ε B = T γ = ε Lm(I) Lm(T)	2. Évaluation du premier conditionnel de l'algorithme 3.	A = je β = ε B = T γ =, je Lm(I) = Lm(T)

Résumé des ensembles de non-terminaux les plus à gauche :

- Le plus à gauche(S) = Lm(D).
- Le plus à gauche(D) = Lm(T).
- Le plus à gauche(T) = {int}.
- Le plus à gauche(L) = Lm(I).
- Le plus à gauche(I) = Lm(T).

Remplacement des ensembles :

- Le plus à gauche(S) = {int}.
- Le plus à gauche(D) = {int}.
- Le plus à gauche(T) = {int}.
- Le plus à gauche(L) = {int}.
- Le plus à gauche(I) = {int}.

5.2.5. Études de cas

Pour l'algorithme 5, l'application de certaines relations de préséance est présentée.

1. Pour le symbole délimiteur $:
   (a) $ L(S) −→ Règle (A.1).
   - $;
   - $ identifiant
   - $
   - $ entier

   (b) R(S) $ −→ Règle (A.2).
   - ; $

2. Pour S −→ SD ;

Machine Translated by Google

(a) X = S, Y = D −→ Règle (E.1).

R(S) Lm(D)

- ; entier

u = nul, l = [ ]

(b) X = D, Y = ; −→ Règle (D.1).

R(D) ;

-) ;

u = nul, l = [ ]

3. Pour D −→ Tid(L)

(a) X = id, Y = (−→ Règle (B).

- identifiant .= (

u = nul, l = [ ]

(b) X = (, Y = L −→ Règle (C).

(L(L)

- ,
- ((
- (entier

u = (, l = [ ]

(c) X = L, Y =) −→ Règle (D.1).

R(L))

- ,)
-)
- entier)

u = null −→ u .= b −→ (.=) −→ Règle (D.2).

u = nul, l = [ ]

4. Dans les cas suivants, la forme XY n'est pas remplie dans la partie droite de la production ; aucune règle ne peut donc être appliquée :

(a) T −→ entier

(b) L − → Je

(c) L − → ε

(d) je −→ T

Les relations de préséance restantes sont dérivées en utilisant la même méthode que les cas démontrés précédemment.

5.2.6. Tableau de priorité des opérateurs

A la fin de l'exécution de l'algorithme, le tableau de priorité des opérateurs 11 est obtenu.

Tableau 11. Tableau de priorité de l'exemple 2.

	;	identifiant	(	)	*	int	,	$
:								
			.=					
identifiant (				.=				
)								
*								
int								
,								
$								

Machine Translated by Google

5.2.7. Analyse ascendante

Soit int id(); int identifiant(int, int); être une chaîne obtenue par dérivation la plus à droite, puis en appliquant L'algorithme 1 aboutit à l'analyse ascendante détaillée dans le tableau 12.

Tableau 12. Analyse ascendante de l'exemple 2.

Empiler	Saisissez	Action
$	int id(); int identifiant(int, int); $	Changement
$int	identifiant(); int identifiant(int,	Réduire
$	int); $ identifiant(); int	Changement
$id	identifiant(int, int); $ (); int	Changement
	identifiant(int, int); $); int	Changement
	identifiant(int, int); $; int	Réduire
$id($id() $	identifiant(int, int); $; int	Changement
$;	identifiant(int, int); $ int	Réduire
$	identifiant(int, int); $ int	Changement
$int	identifiant(int, int);	Réduire
$	$ id(int, int); $	Changement
$id	id(int, int); $	Changement
	(int, int); $ int,	Changement
$id($id(int	int); $,	Réduire
	entier);	Changement
	$,	Changement
$id($id(,		Réduire
$id(,		Réduire
int		Changement
$id(, $id($id()	entier); $ int); $); $); $); $; $	Réduire
$	; $	Changement
$;	$	Réduire
$	$	Accepter

5.3. Exemple 3

5.3.1. Grammaire

Compte tenu des non-OPG suivants :

S − → ABC

UNE −→ uneUNE | un

B −→ bB | b | ε

C −→ CDc | c

ré −→ ré

Les ensembles sont construits étape par étape en suivant les algorithmes.

5.3.2. Gauche

Les ensembles de gauche sont calculés comme détaillé dans le tableau 13.

Machine Translated by Google

Tableau 13. Gauche Définit le calcul pour le CFG dans l'exemple 3.

Gauche(S)			
Pour S −→ ABC			
1. Évaluation du premier conditionnel de l'algorithme 2.	A = S β = ε B = UNE γ = BC L(S) L(UNE)	2. Évaluation du premier conditionnel de l'algorithme 2.	A = S β = A B = B y = C L(S) L(A) L(B)
3. Évaluation du premier conditionnel de l'algorithme 2.	UNE = S β = AB B = C γ = ε L(S) = L(A) L(B) L(C)		

Gauche(A)			
Pour A −→ aA			
1. Évaluation du deuxième conditionnel de l'algorithme 2.	UNE = UNE β = ε une = une y = A L(UNE) {une}	2. Évaluation du premier conditionnel de l'algorithme 2.	UNE = UNE β = une B = UNE γ = ε L(UNE) {une} L(UNE)
Pour A −→ a			
3. Évaluation du deuxième conditionnel de l'algorithme 2.	UNE = UNE β = ε une = une γ = ε L(UNE) = {une}		

Gauche(B)			
Pour B −→ bB			
1. Évaluation du deuxième conditionnel de l'algorithme 2.	A = B β = ε a = b γ = B L(B) {b}	2. Évaluation du premier conditionnel de l'algorithme 2.	A = B β = b B = B γ = ε L(B) {b} L(B)
Pour B −→ b			
3. Évaluation du deuxième conditionnel de l'algorithme 2.	A = B β = ε a = b γ = ε L(B) = {b}		

Machine Translated by Google

Tableau 13. Suite

Gauche(C)			
Pour C −→ CDc			
1. Évaluation du premier conditionnel de l'algorithme 2.	A = Cβ = ε B = C γ = Dc L(C) L(C)	2. Évaluation du premier conditionnel de l'algorithme 2.	A = Cβ = C B = D y = c L(C) L(D)
3. Évaluation du deuxième conditionnel de l'algorithme 2.	A = C β = CD une = c γ = ε L(C) L(D) {c}		
Pour C −→ c			
4. Évaluation du deuxième conditionnel de l'algorithme 2.	A = C β = ε une = c γ = ε L(C) = L(D) {c}		

Gauche(D)	
Pour D −→ d	
1. Évaluation du deuxième conditionnel de l'algorithme 2.	A = D β = ε a = d γ = ε L(D) = {d}

Résumé des ensembles gauches de non-terminaux : • Gauche(S) = L(A) L(B) L(C). • Gauche(A) = {a}. • Gauche(B) = {b}. • Gauche(C) = L(D) {c}. • Gauche(D) = {d}.

Remplacement des ensembles : • Gauche(S) = {a, b, d, c}. • Gauche(A) = {a}. • Gauche(B) = {b}. • Gauche(C) = {d, c}. • Gauche(D) = {d}.

5.3.3. Droite

Les ensembles de droites sont calculés comme détaillé dans le tableau 14.

Machine Translated by Google

Tableau 14. Calcul des ensembles de droits pour le CFG dans l'exemple 3.

Droits)			
Pour S −→ ABC			
1. Évaluation du premier conditionnel de l'algorithme 4.	A = S γ = AB B = C β = ε R(S) = R(C)		

Droite(A)			
Pour A −→ aA			
1. Évaluation du premier conditionnel de l'algorithme 4.	UNE = UNE γ = une B = UNE β = ε R(UNE) R(UNE)	2. Évaluation du deuxième conditionnel de l'algorithme 4.	UNE = UNE γ = ε une = une β = UNE R(UNE) {une}
Pour A −→ a			
3. Évaluation du deuxième conditionnel de l'algorithme 4.	UNE = UNE γ = ε une = une β = ε R(UNE) = {une}		

Droite (B)			
Pour B −→ bB			
1. Évaluation du premier conditionnel de l'algorithme 4.	A = B y = b B = B β = ε R(B) R(B)	2. Évaluation du deuxième conditionnel de l'algorithme 4.	A = B γ = ε a = b β = B R(B) {b}
Pour B −→ b			
3. Évaluation du deuxième conditionnel de l'algorithme 4.	A = B γ = ε a = b β = ε R(b) = {b}		

Droite(C)			
Pour C −→ CDc		Pour C −→ c	
1. Évaluation du deuxième conditionnel de l'algorithme 4.	A = C γ = CD une = c β = ε R(C) {c}	2. Évaluation du deuxième conditionnel de l'algorithme 4.	A = C γ = ε une = c β = ε R(C) = {c}

Machine Translated by Google

Tableau 14. Suite

Droite (D)	
Pour D −→ d	
1. Évaluation du deuxième conditionnel de l'algorithme 4.	A = D γ = ε a = d β = ε R(D) = {d}

Résumé des ensembles Right de non-terminaux : • Right(S) = R(C). • Droite(A) = {a}. • Droite(B) = {b}. • Droite(C) = {c}. • Droite(D) = {d}.

Remplacement des ensembles : • Right(S) = {c}. • Droite(A) = {a}. • Droite(B) = {b}. • Droite(C) = {c}. • Droite(D) = {d}.

5.3.4. Le plus à gauche

Les ensembles les plus à gauche sont calculés comme détaillé dans le tableau 15.

Tableau 15. Calcul des ensembles les plus à gauche pour le CFG dans l'exemple 3.

Le plus à gauche(S)	
Pour S −→ ABC	
1. Évaluation du premier conditionnel de l'algorithme 3.	A = S β = ε B = UNE γ = BC Lm(S) = Lm(A)

Le plus à gauche (A)			
Pour A −→ aA		Pour A −→ a	
1. Évaluation du deuxième conditionnel de l'algorithme 3.	UNE = UNE β = ε une = une γ = A Lm(A) {a}	2. Évaluation du deuxième conditionnel de l'algorithme 3.	UNE = UNE β = ε une = une γ = ε Lm(UNE) = {une}

Le plus à gauche (B)			
Pour B −→ bB		Pour B −→ b	
1. Évaluation du deuxième conditionnel de l'algorithme 3.	A = B β = ε a = b γ = B Lm(B) {b}	2. Évaluation du deuxième conditionnel de l'algorithme 3.	A = B β = ε a = b γ = ε Lm(B) = {b}

Machine Translated by Google

Tableau 15. Suite

Le plus à gauche (C)			
Pour C −→ CDc		Pour C −→ c	
1. Évaluation du premier conditionnel de l'algorithme 3.	A = Cβ = ε B = C γ = Dc Lm(C) Lm(C)	2. Évaluation du deuxième conditionnel de l'algorithme 3.	A = Cβ = ε une = c γ = ε Lm(C) = {c}

Le plus à gauche (D)	
Pour D −→ d	
1. Évaluation du deuxième conditionnel de l'algorithme 3.	A = D β = ε a = d γ = ε Lm(D) = {d}

Résumé des ensembles de non-terminaux les plus à gauche : • Le plus à gauche(S) = Lm(A). • Le plus à gauche(A) = {a}. • Le plus à gauche(B) = {b}. • Le plus à gauche(C) = {c}. • Le plus à gauche(D) = {d}.

Remplacement des ensembles : • Le plus à gauche(S) = {a}. • Le plus à gauche(A) = {a}. • Le plus à gauche(B) = {b}. • Le plus à gauche(C) = {c}. • Le plus à gauche(D) = {d}.

5.3.5. Études de cas

Pour l'algorithme 5, l'application de certaines relations de préséance est présentée.

1. Pour le symbole délimiteur $:

   (a) $ L(S) −→ Règle (A.1). $ a
   - $ b
   - $ d
   - $ c
   -

   (b) R(S) $ −→ Règle (A.2). c $
   -

2. Pour S −→ ABC

   (a) X = A, Y = B −→ Règle (E.1).
   R(A) Lm(B) a
   - b

   u = nul, l = [A] (b)
   X = B, Y = C −→ Règle (E.1).
   R(B) Lm(C) b
   - c

Machine Translated by Google

A l = [ ] −→ Règle (E.3).
R(A) Lm (C)
• une c

u = nul, l = [ ]

3. Pour A −→ aA

(a) X = a, Y = A −→ Règle (C).
une L(UNE)
• une une

u = a, l = [ ]

4. Pour B −→ bB

(a) X = b, Y = B −→ Règle (C).
b L(B)
• b b

u = b, l = [ ]

5. Pour C −→ CDc

(a) X = C, Y = D −→ Règle (E.1).
R(C) Lm(D)
• c ré

u = nul, l = [ ]

(b) X = D, Y = c −→ Règle (D.1).
R(D) c
• d c

u = nul, l = [ ]

6. Dans les cas suivants, la forme XY n'est pas remplie dans la partie droite de la production ; aucune règle ne peut donc être appliquée :

(une) UNE −→ une
(b) B − → b
(c) B − → ε
(d) C −→ c
(e) D −→ d

Les relations de préséance restantes sont dérivées en utilisant la même méthode que les cas démontrés précédemment.

5.3.6. Tableau de priorité des opérateurs

A la fin de l'exécution de l'algorithme, le tableau de priorité des opérateurs 16 est obtenu.

Tableau 16. Tableau de priorité de l'exemple 3.

	un	b	c	d	$
un					
b					
c					
d					
$					

5.3.7. Analyse ascendante

Soit aabbcdc une chaîne obtenue par dérivation la plus à droite, puis en appliquant l'algorithme 1 aboutit à l'analyse ascendante détaillée dans le tableau 17.

Machine Translated by Google

Tableau 17. Analyse ascendante de l'exemple 3.

Empiler	Saisir	Action
$	aabbcdc$	Changement
$un	abcdc$	Changement
$aa	bbcdc$	Réduire
$un	bbcdc$	Réduire
$	bbcdc$	Changement
$b	bcdc$	Changement

Tableau 17. Suite

Empiler	Saisir	Action
$bb	cdc$	Réduire
$b	cdc$	Réduire
$	cdc$	Changement
$c	$ CC	Réduire
$	$ CC	Changement
$d	$ CA	Réduire
$	$ CA	Changement
$c	$	Réduire
$	$	Accepter

5.4. Discussion

Dans les exemples, trois CFG sont considérés : un OPG et deux non-OPG. Dans le premier CFG, on constate qu'il respecte la règle de base, alors que les suivants le font pas puisqu'ils présentent des productions avec des symboles non terminaux consécutifs à droite côté. Quelle que soit la nature du CFG, les algorithmes proposés jouent un rôle crucial dans obtenir la table de priorité des opérateurs. Utilisation de l'algorithme 1 appliqué à une chaîne spécifique pour chaque CFG, une analyse ascendante est effectuée, vérifiant que les chaînes sont générées par leurs grammaires respectives. Ce processus souligne la fonctionnalité de la proposition pour effectuer une analyse ascendante basée sur la priorité des opérateurs, même lorsque les CFG ne le font pas remplir les exigences de base des OPG.

6. Exceptions

Si le CFG présente l'une des formes suivantes, les règles décrites dans les algorithmes de ce le papier ne s'appliquerait pas :

1. A −→ aBb

   B − → βa

   Les décors pour B dans la deuxième production sont :

   L(B) = {une}.

   R(B) = {une}.

   Les relations de préséance basées sur les sections de l'algorithme 5 présentent :

   • Pour la règle (C) : a a et u = a.

   • Pour la règle (D.1) : a b.

   • Pour la règle (D.2) : a .= b.

   L'application des règles (D.1) et (D.2) génère le conflit.

Machine Translated by Google

2. A −→ aBb B −→ aβ

Étant β = ε β = γa. * *

Selon les conditions établies, les ensembles pour B dans la deuxième production sont L(B) = {a}.

R(B) = {une}.

Les relations de préséance basées sur les sections de l'algorithme 5 présentent : • Pour la règle (C) : a a et u = a. • Pour la règle (D.1) : a b. • Pour la règle (D.2) : a .= b.

L'application des règles (D.1) et (D.2) génère le conflit.

Si le R(B) ne contient pas a, alors le conflit ne survient pas.

3. A −→ bBa B −→ aβ Les ensembles pour B dans la deuxième production sont : L(B) = {a}.

R(B) = {une}.

Les relations de préséance basées sur les sections de l'algorithme 5 montrent : • Pour la règle (C) : b a et u = b. • Pour la règle (D.1) : a a. • Pour la règle (D.2) : b .= a.

L'application des règles (C) et (D.2) génère le conflit.

4. A −→ bBa

B −→ βa

Étant β = ε * β = aγ. *

Selon les conditions établies, les ensembles pour B dans la deuxième production sont : L(B) = {a}.

R(B) = {une}.

Les relations de préséance basées sur les sections de l'algorithme 5 montrent : • Pour la règle (C) : b a et u = b. • Pour la règle (D.1) : a a. • Pour la règle (D.2) : b .= a.

L'application des règles (C) et (D.2) génère le conflit.

Si le L(B) ne contient pas a, alors le conflit ne survient pas.

Lorsque les productions CFG prennent la forme suivante, il n'y aura aucun conflit :

A −→ aBb

B −→ γ

où γ n'a ni a à l'extrémité du côté droit ni b à l'extrémité du côté gauche.

7. Conclusions

Des travaux antérieurs ont porté sur l'application des OPG pour résoudre des problèmes liés à l' analyse de diverses formes de langages ; cependant, peu d'accent a été mis sur la définition des ensembles Gauche et Droite ni sur la génération d'un algorithme pour trouver la table de priorité des opérateurs.

Dans ce travail, nous avons redéfini les ensembles Right et Left, précédemment exposés par d'autres auteurs, et introduit un nouvel ensemble appelé Leftmost. Ces ensembles constituent la base d'un algorithme qui construit la table de priorité des opérateurs. Chaque ensemble est construit à partir d'un algorithme qui facilite son obtention, ajoutant une nouvelle dimension à la recherche existante.

L'algorithme proposé pour construire la table de préséance, tout en cassant (à quelques exceptions près) la définition de base d'un OPG, ouvre de nouvelles possibilités. Il permet la construction d'une table de priorité des opérateurs à partir d'un CFG non-OPG, ouvrant la voie à un algorithme d'analyse ascendant pour reconnaître une chaîne générée par le CFG.

Machine Translated by Google

Trois exemples de CFG sont présentés : un OPG et deux non-OPG. Les algorithmes proposés sont systématiquement appliqués à ces exemples, étape par étape, jusqu'à l'obtention des tables de priorité des opérateurs, quelle que soit la nature du CFG. De plus, l'algorithme d'analyse ascendante est appliqué à des chaînes spécifiques pour chaque grammaire, permettant ainsi de vérifier qu'elles les ont générées.

Il est important de noter que même si l'algorithme et la définition des nouveaux ensembles de symboles terminaux constituent des avancées significatives, ils présentent néanmoins des limites. Il n'est pas toujours possible de résoudre tous les cas dans lesquels deux ou plusieurs symboles de borne adjacents se trouvent sur le côté droit. Il existe des exceptions et nous devons en être conscients.

Contributions des auteurs : Conceptualisation, LL et JM ; méthodologie, JM; logiciels, LL; validation, LL et JM ; analyse formelle, LL, EA et JM ; enquête, LL et JM; ressources, JM; conservation des données , EA ; rédaction : préparation de l'ébauche originale, EA et JM ; rédaction – révision et édition, EA et JM ; visualisation, EA; supervision, JM; administration de projet, JM; acquisition de financement, JM; Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche n'a reçu aucun financement externe et l'APC a été financée par l'Universidad del Norte.

Déclaration de disponibilité des données : le partage de données ne s'applique pas à cet article.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Références

1. Floyd, RW Analyse syntaxique et priorité des opérateurs. J.ACM (JACM) 1963, 10, 316-333. [Référence croisée] 2.
Alfred, V. ; Monique, S. ; Ravi, S. ; Jeffrey, DU Compilateurs Principes, Techniques ; Pearson Education : Londres, Royaume-Uni, 2007.
3. Grune, D. ; Jacobs, CJH Parsing Techniques (Monographies en informatique) ; Springer : Berlin, Heidelberg, 2006.
4. Crespi-Reghizzi, S. Un modèle efficace pour l'inférence grammaticale. Dans Traitement de l'information 71 ; Gilchrist, B., éd.; Elsevier : North-Holland, New York, 1972 ; pp. 524-529.
5. Crespi-Reghizzi, S. Réduction du dénombrement dans l'acquisition de la grammaire. Dans les actes de la 2e Conférence conjointe internationale sur l'intelligence artificielle, San Francisco, Californie, États-Unis, 1er-3 septembre 1971 ; pp. 546-552.
6. Crespi-Reghizzi, S. ; Melkanoff, MA; Lichten, L. L'utilisation de l'inférence grammaticale pour la conception de langages de programmation. Commun. ACM 1973, 16, 83-90. [Référence croisée]
7. von Braunmühl, B. ; Verbeek, R. Les langages pilotés par les entrées sont reconnus dans l'espace log n. Dans Thèmes de théorie du calcul ; Études mathématiques de Hollande du Nord ; Karplnski, M., van Leeuwen, J., éd. ; Elsevier : Hollande du Nord, Pays-Bas, 1985 ; Volume 102, pages 1 à 19.
8. Mehlhorn, K. Chaînes de montagnes Pebbling et son application à la reconnaissance DCFL. En Automates, Langages et Programmation ; de Bakker, J., van Leeuwen, J., éd.; Springer : Berlin/Heidelberg, Allemagne, 1980 ; pp. 422-435.
9. Mandrioli, D. ; Pradella, M. Généraliser les langages pilotés par les entrées : avantages théoriques et pratiques. Calculer. Sci. Rév. 2018, 27, 61-87. [Référence croisée]
10. Alur, R. ; Madhusudan, P. Ajout d'une structure d'imbrication aux mots. J.ACM 2009, 56, 1–43. [Référence croisée]
11. Crespi Reghizzi, S. ; Mandrioli, D. Priorité des opérateurs et propriété visiblement pushdown. J. Informatique. Système. Sci. 2012, 78, 1837-1867. [Référence croisée]
12. Crespi Reghizzi, S. ; Pradella, M. Au-delà des grammaires et des langages à priorité d'opérateur. J. Informatique. Système. Sci. 2020, 113, 18-41. [Référence croisée]
13. Barenghi, A. ; Crespi Reghizzi, S. ; Mandrioli, D. ; Panella, F. ; Pradella, M. L'analyse parallèle rendue pratique. Sci. Calculer. Programme. 2015, 112, 195-226. [Référence croisée]
14. Crespi-Reghizzi, S. ; Mandrioli, D. ; Martin, DF Propriétés algébriques des langages de priorité des opérateurs. Inf. Contrôle 1978, 37, 115-133. [Référence croisée]
15. Chomsky, N. Trois modèles pour la description du langage. IRE Trans. Inf. Théorie 1956, 2, 113-124. [Référence croisée]
16. Ah, AV ; Ullman, JD La théorie de l'analyse, de la traduction et de la compilation ; Prentice-Hall Englewood : Cliffs, NJ, États-Unis, 1973 ; Tome 1.
17. Louden Kenneth, C. Construction du compilateur : principes et pratique. Dans le cours Technologie ; PWS Publishing Co. : Montgomery, Illinois, États-Unis, 1997.

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Sélection des canaux EEG pour la réadaptation des patients victimes d'un AVC Optimiseur de BAT

Mohammed Azmi Al-Betar 1,*, Zaid Abdi Alkareem Alyasseri [2,3,*], Noor Kamal Al-Qazzaz [4], Sharif Naser Makhadmeh 1,5, Nabeel Salih Ali [2] et Christoph Guger [6]

1 Centre de recherche sur l'intelligence artificielle (AIRC), Collège d'ingénierie et de technologie de l'information, Université d'Ajman, Ajman PO Box 346, Émirats arabes unis ; sharif.makhadmeh@uop.edu.jo
2 Centre de recherche et de développement des technologies de l'information (ITRDC), Université de Kufa, Najaf 54001, Irak ; nabeel@uokufa.edu.iq
3 Collège d'ingénierie, Université de Warith Al-Anbiyaa, Karbala 56001, Irak
4 Département de génie biomédical, Collège d'ingénierie AL-Khwarizmi, Université de Bagdad, Bagdad 47146, Irak ; noorbme@kecbu.uobaghdad.edu.iq
5 Département de science des données et d'intelligence artificielle, Faculté des technologies de l'information, Université de Petra, Amman 1196, Jordan
6 G.Tec Medical Engineering GmbH, 4521 Schiedlberg, Autriche ; guger@gtec.at *
Correspondance : m.albetar@ajman.ac.ae (MAA-B.) ; zaid.alyasseri@uokufa.edu.iq (ZAAA)

Résumé : Les accidents vasculaires cérébraux sont une cause majeure de mortalité dans le monde, ils perturbent le flux sanguin cérébral, entraînant de graves lésions cérébrales. L'hémiplégie, une conséquence courante, entraîne une perte des tâches motrices d'un côté du corps. De nombreux survivants d'un AVC sont confrontés à des déficiences motrices à long terme et nécessitent une rééducation approfondie. Les électroencéphalogrammes (EEG) constituent une méthode non invasive pour surveiller l'activité cérébrale et ont été utilisés dans les interfaces cerveau-ordinateur (BCI) pour faciliter la rééducation. Les tâches d'imagerie motrice (IM), détectées par EEG, sont essentielles au développement de BCI qui aident les patients à retrouver leur objectif moteur. Cependant, l'interprétation des signaux EEG pour les tâches MI reste difficile en raison de leur complexité et de leur faible rapport signal/bruit. L'objectif principal de cette étude est de se concentrer sur l'optimisation de la sélection des canaux dans les BCI basés sur l'EEG spécifiquement pour la rééducation après un AVC. La détermination des canaux EEG les plus informatifs est cruciale pour capturer les signaux neuronaux liés aux déficiences motrices chez les patients victimes d'un AVC.

Dans cet article, une méthode d'optimisation basée sur un algorithme binaire de chauve-souris (BA) est proposée pour sélectionner les canaux les plus pertinents adaptés aux changements neurophysiologiques uniques chez les patients victimes d'un AVC. Cette approche est capable d'améliorer les performances du BCI en améliorant la précision de la classification et en réduisant la dimensionnalité des données. Nous utilisons des attributs temps-entropie-fréquence (TEF), traités par analyse automatisée des composants indépendants avec débruitage par transformée en ondelettes (AICA-WT), pour améliorer la clarté du signal. Les canaux et caractéristiques sélectionnés sont prouvés grâce à un classificateur k-plus proche voisin (KNN) utilisant des ensembles de données publics BCI, démontrant une classification améliorée des tâches MI et le potentiel d'obtenir de meilleurs résultats de rééducation.

Mots clés : EEG ; extraction de fonctionnalités ; sélection des chaînes ; les patients victimes d'un AVC ; Algorithme BAT

check for updates

Citation : Al-Betar, MA ; Alyasseri, ZAA; Al-Qazzaz, NK ; Makhadmeh, SN; Ali, Nouvelle-Écosse; Guger, C. Canal EEG Sélection pour les patients victimes d'un AVC Rééducation à l'aide de BAT Optimizer. Algorithmes 2024, 17, 346. https://doi.org/10.3390/a17080346

Rédactrice académique : Gabriella Trucco

Reçu : 21 mai 2024
Révisé : 14 juillet 2024
Accepté : 15 juillet 2024
Publié : 8 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

L'Organisation mondiale de la santé (OMS) classe l'accident vasculaire cérébral comme la deuxième cause de mortalité dans le monde en raison des dommages qu'il provoque au cerveau en raison de perturbations de la circulation cérébrale [1,2]. L'hémiplégie est l'une des formes les plus débilitantes de lésions du système nerveux . Elle décrit une paralysie partielle ou complète d'un côté du corps, notamment du bras, de la jambe, du pied et de la main. L'accident vasculaire cérébral ischémique, le type d'accident vasculaire cérébral le plus courant, provoque une interruption de la perfusion cérébrale, ce qui entraîne une perte rapide des fonctions cérébrales [3,4]. Soixante pour cent ou plus des survivants d'un AVC nécessitent une rééducation en raison d'une déficience permanente de la fonction motrice [5]. Cependant, les survivants d'un AVC souffrent d'un large éventail de handicaps, notamment des déficits visuels et cognitifs, qui ont un effet cumulatif dévastateur sur leur capacité

Machine Translated by Google

effectuer même les tâches les plus élémentaires. En conséquence, la recherche sur le traitement efficace et la réadaptation des victimes d'accident vasculaire cérébral est une priorité depuis de nombreuses années [6].

Un électroencéphalogramme (EEG) est une méthode non invasive qui estime l'activité électrique du cerveau avec une bonne résolution temporelle afin de définir les résultats d'une variété de tâches mentales. En traduisant les signaux EEG cérébraux en instructions de contrôle, les interfaces cerveau-ordinateur (BCI) aident les personnes ayant une déficience physique à interagir avec le monde sensoriel extérieur [7]. L'imagerie motrice (IM) est l'imagination mentale du mouvement du corps sans mouvement musculaire réel, et les activités rythmiques du cerveau qui l'accompagnent peuvent être enregistrées et utilisées comme signaux d'entrée des systèmes BCI. La désynchronisation/synchronisation liée à un événement (ERD/ERS) fait référence à la détection des changements de puissance rythmique dans la zone sensorimotrice dans les bandes de fréquences Rolandic mu (9-13 Hz) et bêta (13-30 Hz) , qui peuvent être utilisé pour distinguer les différents types de tâches MI [8]. Étant donné que l'IM ne nécessite aucun stimuli externe, les BCI basées sur l'IM ont une grande variété d'utilisations potentielles [9].

La volatilité et la complexité des signaux EEG posent cependant des difficultés pour le décodage MI-BCI conventionnel [7]. Des problèmes cruciaux et difficiles demeurent dans l'amélioration de la capacité de décodage EEG, l'extraction d'informations discriminantes à partir d'EEG à faible rapport signal/bruit et la réalisation d'une classification fiable de diverses tâches MI [8].

La sélection des canaux dans les investigations basées sur l'EEG est importante pour déterminer le sous-ensemble de canaux le plus informatif en termes de capture des signaux neuronaux appropriés pour l'application choisie (10). Néanmoins, ces considérations et leurs défis correspondants peuvent différer dans d'autres contextes d'application, et cela peut être constaté lorsque l'on compare la rééducation après un AVC avec le traitement des émotions et la détection de la démence à l'aide d'un ensemble de données basé sur l'EEG (11). Par exemple, la rééducation après un accident vasculaire cérébral se caractérise par le fait qu'en cas de défauts neurologiques résultant de l'accident vasculaire cérébral, certaines modifications se produisent dans l'activité du cerveau et dans le fonctionnement des réseaux neuronaux. Les données EEG peuvent également être caractérisées par des caractéristiques spécifiques qui les distinguent des sujets sains ou d'autres patients, telles que des changements dans l'organisation topologique des réseaux, une activité aberrante et l'apparition de nouvelles stratégies de régulation [12].

Par conséquent, l'approche visant à choisir les canaux EEG les plus importants pour la rééducation après un AVC devra peut-être être plus sélective et spécifique. L'idée est de découvrir lequel des canaux fournit le meilleur enregistrement des caractéristiques neuronales de la déficience motrice, cognitive ou comportementale particulière qu'un patient victime d'un AVC peut présenter, et ce ne sera pas le même que les canaux qui seraient idéaux pour d'autres applications [13]. Cependant, il existe des approches distinctes, telles que le traitement des émotions et la détection de la démence, qui peuvent utiliser différents mécanismes neurophysiologiques et/ou modèles spectrotemporels. Par exemple, dans le traitement des émotions, la bande de fréquences liée au traitement des émotions peut être répartie sur différentes zones corticales, et le but de la sélection des canaux pourrait être de capturer cette activité de fréquence distribuée. De même, dans le cas de la détection de la démence, l'altération de la structure et des fonctions cérébrales qui définissent les troubles cognitifs pourrait représenter une répartition spatiale différente de la charge qui nécessite des règles différentes pour le choix des canaux [14]. En se concentrant directement sur le problème de sélection des canaux dans le contexte de la rééducation après un AVC, l'étude proposée vise à apporter une solution efficace qui peut améliorer les performances d'un BCI et ainsi améliorer le processus de rééducation.

La sélection des canaux est au centre de cette étude. De nombreuses électrodes sont utilisées dans les traitements médicaux et diagnostiques. Pour les applications pratiques du BCI, de nombreuses électrodes sont nécessaires pour la précision de la classification ; cependant, placer de nombreuses électrodes sur le cuir chevelu prend du temps et est sujet à un surapprentissage du signal EEG. Pour résoudre ces problèmes, les chercheurs doivent identifier les électrodes qui ne peuvent pas être classées. La sélection des électrodes peut évaluer les connaissances neurologiques des individus en raison de réactions variables et du placement idéal des électrodes en fonction du sujet [15]. Le travail de classification dépendant de l'utilisateur nécessite une détermination automatique de la pertinence des électrodes. Ce défi peut être surmonté en considérant une large gamme d'électrodes et en utilisant de nombreuses méthodes pour identifier les canaux appropriés pour chaque patient. Une technique d'optimisation basée sur le BA des particules a été utilisée pour sélectionner le modèle le p

Machine Translated by Google

Canaux EEG pour la rééducation des patients victimes d'un AVC. Chaque canal de données EEG offre des fonctionnalités de classification uniques. La sélection de canal réduit le stockage et le traitement des données, accélère la formation du classificateur en utilisant des données plus simples et réduit la « malédiction de la dimensionnalité ».

De nombreuses études ont examiné la sélection des canaux chez des personnes en bonne santé [16,17], tandis que la présente étude se concentre sur la rééducation après un AVC, où les anomalies et limitations neurophysiologiques sous-jacentes des patients victimes d'un AVC peuvent exiger une stratégie de sélection de canaux personnalisée .

Le choix des canaux EEG pour le traitement des émotions, l'évaluation cognitive et le diagnostic clinique implique souvent la connaissance des mécanismes neuronaux et des régions cérébrales liées aux activités souhaitées. Les schémas courants d'activité cérébrale et la sélection optimale des canaux sont bien établis pour les personnes en bonne santé. Cependant, les accidents vasculaires cérébraux peuvent altérer la structure, la connectivité et le fonctionnement du cerveau. Ces altérations peuvent provoquer des schémas cérébraux anormaux et des processus compensatoires [18,19]. Ainsi, les meilleurs canaux EEG pour la rééducation après un AVC peuvent différer de ceux recommandés pour les personnes en bonne santé.

Pour capturer les corrélats neuronaux des anomalies motrices, cognitives et comportementales des patients victimes d'un AVC, la sélection des canaux doit être plus spécialisée et personnalisée en raison de la neuroplasticité induite par l'AVC. Cibler les patients victimes d'un AVC avec une solution personnalisée capable de reconnaître correctement les empreintes neuronales et les modèles d'activité cérébrale vise à relier les déficiences liées à l'AVC. Cette étape est cruciale pour développer des thérapies contre l'AVC basées sur la BCI qui améliorent les résultats de la rééducation (20).

La littérature sur la sélection des canaux chez les personnes en bonne santé peut être utile, mais cette étude reconnaît la nécessité de répondre aux problèmes et considérations uniques des survivants d'un AVC. Les scientifiques ont examiné les différences de sélection des canaux entre les patients victimes d'un AVC et les personnes en bonne santé afin de mieux comprendre les changements neurophysiologiques et les configurations BCI optimales pour la rééducation après un AVC.

Par conséquent, dans cette étude, une méthode d'optimisation basée sur un algorithme binaire de chauve-souris (BA) est utilisée pour identifier les canaux EEG les plus importants pour la rééducation après un AVC. Ainsi, cette méthode est adaptée aux besoins des survivants d'un AVC en optimisant la technologie BCI et l'efficacité de la rééducation.

Les données EEG brutes sont segmentées et filtrées à l'aide de filtres standard et de l' approche de débruitage automatisée et indépendante des composants avec transformation en ondelettes (AICA-WT) dans le but de résoudre les problèmes mentionnés ci-dessus. Ensuite, les attributs temps-entropie-fréquence (TEF) sont créés en combinant les caractéristiques efficaces des domaines temps, entropie et fréquence. En conséquence, la méthode d'optimisation basée sur BA est utilisée pour optimiser les caractéristiques du TEF afin de choisir les canaux efficaces qui améliorent la stationnarité et la résilience du système. Après avoir formé une technique de classificateur KNN, nous menons une série d'expériences pour évaluer notre cadre proposé en utilisant 25 sessions BCI basées sur MI avec des visites d'évaluation de suivi pour examiner les changements fonctionnels avant et après la neurorééducation EEG à partir d'ensembles de données publics du concours BCI. Les attributs hybrides temps- entropie-fréquence (TEF) utilisés dans le cadre AICA-WT-TEF-Chs pour la classification MI-BCI sont destinés à exploiter efficacement les informations sous-jacentes des domaines temporel, entropie et fréquentiel sur les performances de classification.

L'approche suggérée dans cet article apporte les contributions suivantes. Premièrement, il examine comment les variables temps-entropie-fréquence (TEF) affectent les performances de classification.
Deuxièmement, il utilise les canaux EEG efficaces générés par une méthode d'optimisation basée sur BA pour détecter les changements avant et après la rééducation. Troisièmement, il teste l'efficacité du cadre AICA-WT-TEF-Chs en utilisant une variété de modèles de classification et de techniques de pointe . En éliminant les dimensions superflues, une sélection efficace des canaux diminue la quantité de données qui doivent être stockées et traitées, accélère la formation du classificateur en utilisant des données plus simples et empêche le surajustement du classificateur en atténuant l'impact de la « malédiction de la dimensionnalité ».

2. Travaux connexes

Les BCI basés sur l'EEG en particulier ont une bonne résolution temporelle [21], mais l' activité des ondes enregistrées peut être corrompue de diverses manières en fonction des artefacts survenus. Le prétraitement des signaux EEG bruts est donc essentiel dans la classification des signaux basés sur l'IM. .

Machine Translated by Google

La plupart du temps, ces artefacts peuvent imiter ou superposer le comportement anormal du cerveau.
Des études ont étudié les effets des artefacts sur les signaux EEG, notamment les clignements des yeux, les mouvements oculaires, les artefacts cardiaques et l'activité musculaire, et le bruit des lignes électriques peut également croiser les fréquences d'un EEG [22,23]. En conséquence, évaluer l'efficacité des signaux EEG en présence de bruit de fond peut s'avérer difficile. Les filtres spatiaux peuvent produire un signal plus localisé pour chaque électrode dans cette situation, car les BCI basés sur l'EEG ont en particulier une mauvaise résolution spatiale.

Pour choisir les meilleures caractéristiques pour l'analyse EEG basée sur BCI-MI, l'extraction des caractéristiques est effectuée pour chaque canal enregistré dans plusieurs domaines. La technique la plus populaire consiste à utiliser des modèles spatiaux communs (CSP) pour obtenir des caractéristiques discriminantes à partir de signaux EEG de haute dimension, qui peuvent être utilisées pour identifier les caractéristiques spatiales liées à l'ERD produites par diverses tâches d'IM (24). Cependant, l'efficacité des CSP est facilement affectée par le bruit et est très sensible au choix des bandes de fréquences [16]. Plusieurs techniques d'extraction de caractéristiques ont été développées pour analyser les données EEG dans les domaines temps, fréquence et temps- fréquence afin de résoudre le problème de l'efficacité (25). De plus, des caractéristiques du bispectre de cohérence d'ondelettes (WC) ont été suggérées dans [26] pour distinguer les MI droit et gauche.

Les tailles des caractéristiques dérivées des signaux EEG sont généralement assez grandes pour chaque canal, et elles augmentent lorsque l'on passe au canal suivant. La classification de plusieurs caractéristiques nécessite des calculs et du temps supplémentaires. Pour résoudre ce problème, l' ensemble de fonctionnalités le plus efficace doit être obtenu en sélectionnant un sous-ensemble de canaux EEG plus étroitement liés au travail mental que d'autres. Les quatre méthodologies d'évaluation de filtrage, d'encapsulation, d'intégration et hybride peuvent être utilisées pour catégoriser les algorithmes de sélection de fonctionnalités [27]. Sur la base de critères statistiques spécifiques, la méthode de filtrage utilisée pour les activités EEG basées sur l'intention motrice est prise en compte. Par exemple, avant de classer l'activité intentionnelle motrice, He et al. [28] ont développé la méthode statistique du cap Bhattacharyya pour la sélection des canaux. Le meilleur sous-ensemble de candidats a été choisi à l'aide de cette méthode, qui utilisait également une stratégie de recherche CSP et séquentielle. Tam et coll. [29] ont suggéré une méthode statistique différente basée sur le rang CSP pour trier les coefficients de filtre avec des valeurs absolues, puis choisir les caractéristiques avec le plus grand coefficient successif. Bien que la technique de filtrage ait une faible précision, elle fonctionne rapidement et n'est pas affectée par le choix du sujet ou du classificateur.

D'un autre côté, le classificateur et le sujet sont très importants dans le processus de sélection du wrapper [30]. De cette manière, les sous-ensembles candidats sont évalués à l'aide de la précision de la classification et peuvent ainsi faire des prédictions plus précises que les techniques de filtrage. En conséquence, la stratégie wrapper est plus coûteuse en calcul que la technique de filtrage et est sujette au problème de surajustement [31], qui peut être évité en employant des mesures de validation croisée pour la prédiction. Pour choisir le meilleur canal, la majorité des stratégies wrapper utilisent la recherche séquentielle vers l'avant, la stratégie d'élimination vers l'arrière [32] et la méthode de recherche heuristique [33,34]. Sans mentionner de condition d'arrêt, les procédures de filtrage et d'emballage ont également été couplées à des fins de sélection. Cette stratégie hybride a été créée pour gérer d'énormes ensembles de données. Gaur et al., par exemple [35], ont proposé la meilleure technique de sélection de canal possible basée sur les CSP. Avant de classer tous les canaux en fonction de leurs scores et d'utiliser la précision de la classification pour évaluer la combinaison de canaux idéale choisie, ils ont d'abord appliqué la norme L1 d'un CSP pour regrouper les scores de contribution des canaux. Enfin, en utilisant les critères définis pour le processus d'apprentissage d'un classificateur particulier, l'approche intégrée a été utilisée pour choisir des canaux significatifs.

Pour choisir les canaux offrant les meilleurs résultats de classification, Shi et al. [36] ont utilisé des techniques de sélection de fonctionnalités , de suppression récursive de fonctionnalités et d'optimisation de norme zéro entraînées avec des machines à vecteurs de support. L'utilisation de ces techniques a permis de réduire significativement les canaux EEG [37].

L'étape cruciale dans la classification des données EEG consiste à utiliser des caractéristiques pertinentes qui affectent de manière significative la capacité du système à classer efficacement les données [38-40]. Par conséquent, pour obtenir les meilleures performances de classification, il est nécessaire d'utiliser des caractéristiques valides des signaux EEG [41].

Machine Translated by Google

De plus, les techniques d'optimisation globale et la classification de l'apprentissage automatique peuvent être combinées pour résoudre de tels problèmes dans le domaine BCI, y compris l'extraction et la sélection de fonctionnalités [42]. Plusieurs algorithmes ont suscité beaucoup d'intérêt pour leurs applications dans la sélection de caractéristiques, par exemple l'évolution différentielle (DE), le recuit simulé (SA), l'optimisation des essaims de particules (PSO), les colonies d'abeilles artificielles (ABC) et l'optimisation des colonies de fourmis (ACO). [43

Des recherches ont été menées sur la sélection de canaux pour différents paradigmes BCI, y compris l'élimination récursive des canaux pour les tâches d'imagerie motrice (44). Dans une étude menée par Rakotomamonjy et al. [45], les chercheurs ont utilisé une technique de sélection de canaux intégrée à la formation d'un classificateur SVM. L'espace des caractéristiques a été recherché à l'aide de GA dans les références [46, 47], et la fonction de fitness était la combinaison linéaire pondérée du nombre de caractéristiques et de la précision du SVM. Udhaya et coll. dans [48] ont suggéré une approche de sélection de caractéristiques d'un ensemble approximatif basée sur PSO pour déterminer le meilleur sous-ensemble de caractéristiques, et la précision d'un classificateur de quartier a été utilisée comme critère d'évaluation pour le sous-ensemble de caractéristiques. DE a été utilisé par Baig et al. [49], et PSO et GA ont été utilisés par Atyabi et al. [50] afin de trouver respectivement le meilleur sous-ensemble de fonctionnalités. Dans [51], le processus de sélection des caractéristiques pour l'identification des émotions basée sur l'EEG utilisait l'optimisation des colonies de fourmis, le recuit simulé, les algorithmes génétiques, l'optimisation des essaims de particules et l'évolution différentielle. Les performances de ces approches basées sur l'EA pour la sélection des fonctionnalités sont assez encourageantes.

Cependant, la taille de la zone de recherche est prédéterminée dans ces algorithmes ; par conséquent, tout au long de chaque itération, il est important de prendre une décision quant au choix ou non de chaque fonctionnalité. Dans ces circonstances, l'effort de calcul sera inutile sur certaines fonctionnalités superflues, inutiles ou de nature triviale. De plus, les algorithmes évolutionnaires (EA) convergent facilement vers un optimal local en raison du large espace de recherche ainsi que de l'interférence de fonctionnalités redondantes, non pertinentes et triviales. Dans cette étude, nous concevons des méthodes de sélection de caractéristiques basées sur l'EA pour l'imagerie motrice BCI grâce à la réduction de la dimensionnalité.

Les performances d'un classificateur diminueront considérablement à mesure que le nombre de fonctionnalités augmente en raison de la malédiction de la dimensionnalité, qui se produit lorsque le nombre de fonctionnalités dépasse un seuil spécifique. De plus, la procédure de formation prend plus de temps à mesure que la dimensionnalité des caractéristiques devient élevée [46].

Afin de localiser un grand nombre de chaînes existantes, une sélection de fonctionnalités est nécessaire avec une recherche approfondie pour extraire les chaînes associées [10,52]. Les chercheurs se sont concentrés sur les problèmes de sélection des fonctionnalités et ont suggéré des approches basées sur la recherche complète, la recherche gourmande et la recherche aléatoire [11]. Les coûts de calcul élevés et la convergence vers un optimal local restent des problèmes avec ces techniques de sélection [53].

3. Algorithme BAT Dans

cette section, l'algorithme d'optimisation des chauves-souris est présenté et illustré en termes de son inspiration, des microchauves-souris et de leur mode de vie dans la nature, ainsi que de la procédure d'optimisation générale et de la formulation mathématique de la méthode.

3.1. Inspiration de BAT

L'algorithme d'optimisation des chauves-souris a été proposé pour la première fois par Xin-She Yang en 2010 [54]. Xin -She Yang a proposé l'algorithme des chauves-souris basé sur les caractéristiques d'écholocation des micro-chauves-souris, utilisées pour chasser leurs proies, et leurs déplacements dans l'obscurité. Une microchauve-souris produit des impulsions sonores extrêmement fortes et entend leur écho provenant des objets et des proies qui l'entourent. Grâce à l'écho sonore, les chauves-souris peuvent définir le type de proie et la stratégie pour la chasser. Les chauves-souris produisent généralement 10 à 20 impulsions par seconde lorsqu'elles recherchent une proie. Une fois qu'ils sont proches de la proie, le nombre d'impulsions augmente jusqu'à 200 impulsions par seconde et l'impulsion sonore devient plus faible.

Dans la nature, les micro-chauves-souris ont un comportement de chasse collectif, où elles volent ensemble à une vitesse déterminée pour chasser simultanément des proies en utilisant la caractéristique d'écholocation, comme le montre la figure 1. La figure présente le mouvement des micro-chauves-souris en fonction de l'écholocalisation.

Machine Translated by Google

mécanisme de tion, dans lequel les microbats mettent à jour leurs emplacements et se déplacent vers des proies en fonction de l'écho et de l'emplacement d'autres microchauves-souris, déterminés en interagissant les uns avec les autres. De plus, la vitesse des microbats devient plus élevée et les impulsions diminuent lorsque les microbats sont proches de l'emplacement de la proie.

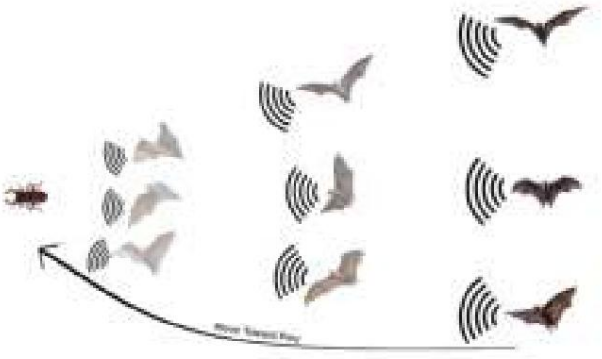


Figure 1. Mouvement des chauves-souris vers leurs proies.

Dans l'algorithme BAT, chaque microbat est considérée comme une solution dans la population, la position de la proie est la solution optimale pour un problème d'optimisation particulier, la localisation est présentée dans un vecteur x avec une plage comprise entre fmin et fmax, la vitesse de les microbats sont présentés dans un vecteur v et la longueur d'onde est λ avec une fréquence d'impulsion r [0, 1] et une plage d'intensité sonore A, où la plus grande est présentée comme Ao et la plus basse est Amin.

3.2. Procédure de

BAT Cette section présente la procédure générale d'optimisation de l'algorithme BAT. Cette procédure d'optimisation contient six étapes principales, présentées dans la figure 2 et discutées en détail ci-dessous :

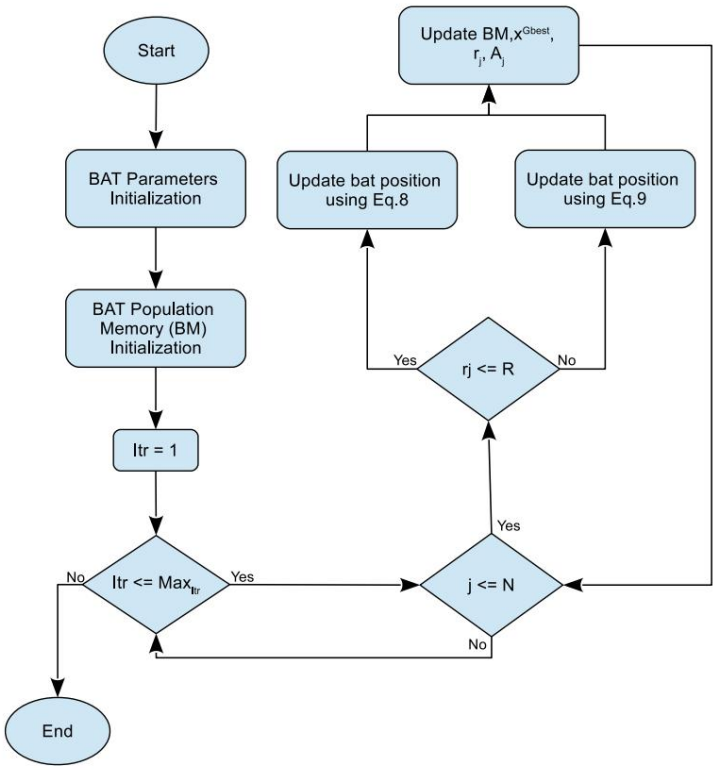


Figure 2. Organigramme de l'algorithme Bat.

• Étape 1 : initialisation des paramètres de l'algorithme BAT. Dans cette étape, les paramètres BAT sont initialisés avec des valeurs initiales. Ces paramètres sont le nombre de microbats

Machine Translated by Google

(solutions) (N), nombre maximum d'itérations (MaxItr), plage de fréquences max (fmax) et min (fmin) , vecteur vitesse (v), taux de sonie (A), fréquence du pouls (r), fréquence du pouls initiale (r plage (ε) dans la ^{0}), deux paramètres (λ et α) avec des valeurs constantes dans la plage [0, 1] et une bande passante plage [−1, 1].

• Étape 2 : initialisation de la mémoire de population BAT. Les solutions BAT sont générées de manière aléatoire au cours de cette étape à l'aide de l'équation (1), en tenant compte de l'algorithme BAT et des paramètres et contraintes particuliers du problème d'optimisation.

$$x_{ij} = lbi + (ubi - lbi) \times R, \quad (1)$$

où i = 1, 2, . . . , d, j = 1, 2, . . . , N et R est un nombre aléatoire compris entre 0 et 1.

Les solutions produites génèrent ensemble la population, comme le montre l'équation (2), et sont stockées dans la mémoire BAT (BM) par ordre croissant sur la base de leur fitness Gbest et stockées dans les solution avec les valeurs les plus adaptées est attribuée à x première position valeurs. La meilleure du BM.

$$MB = \begin{matrix} x_1^1 & x_2^1 & \cdots & x_d^1 \\ x_1^2 & x_2^2 & d2 \cdots & x_d^2 \\ \vdots & \vdots & \cdots & \vdots \\ x_1^N & x_2^N & \cdots & x_d^N \end{matrix}, \quad (2)$$

• Étape 3 : Intensification de la population de chauves-souris. Désormais, toutes les microchauves-souris x volent et changent de position en considérant une vitesse v définie par une fréquence f générée de manière aléatoire, comme le montrent les équations (3) et (4). En conséquence, les positions des chauves-souris sont mises à jour à l'aide de l'équation (5).

$$fi = fmin + (fmin - fmax) \times R, \quad (3)$$

$$v_{je}^{'j} = v_{je}^{jj} + (x_{je} - x\,Gbest) \times fi\,, \quad (4)$$

$$x_{je}^{'j} = x_{ii}\,, \quad j + v = {}^{'j} \quad (5)$$

Il est à noter que les microbats mettent à jour leurs emplacements pour se rapprocher des meilleurs mondiaux (x Gbest). Ainsi, la nouvelle position des microchauves-souris intensifie la position dans une direction Gbest vers x .

• Étape 4 : Diversification des populations de chauves-souris.

Dans cette étape, les positions des microbats sont mises à jour en fonction de paramètres aléatoires pour tenter de trouver de meilleures solutions globales. Une nouvelle solution est sélectionnée dans le BM sur la base d'une méthode de sélection et attribuée à x Gbest. Par conséquent, la solution actuelle est mis à jour en utilisant $x^{Meilleur}$ comme suit :

$$x_{je}^{'j} = x_{Gmeilleur} + \varepsilon \times A\,j\,, \quad (6)$$

où A ' est la moyenne du volume sonore. Les étapes 3 et 4 étant des étapes contradictoires, l' algorithme BAT en choisit une à la fois pour son exécution. Cette sélection est basée sur l' équation suivante :

$$x^{'j} = \begin{cases} x^{Meilleur} + \varepsilon \times A\,j\,' & \text{si } rj \le R \\ x_{j+v'j} & \text{sinon} \end{cases}, \quad (7)$$

• Étape 5 : mise à jour du BM. La position des microbats dans le BM sera mise à jour dans cette étape Gbest si le nouvel emplacement est plus adapté que l'ancien et R ≤ Aj . De plus, x

Machine Translated by Google

mis à jour si le nouveau BM contient une solution avec une meilleure valeur de fitness. Par la suite, les valeurs rj et Aj seront mises à jour conformément aux équations (8) et (9).

$$r_j = r_j^0 \left(1 - e^{(-\gamma \times itr)}\right), \tag{8}$$

$$A_j = \alpha \times A_j, \tag{9}$$

où itr désigne le numéro d'itération actuel.

- Étape 6 : Critère d'arrêt. Les étapes 3, 4 et 5 seront répétées jusqu'à ce que l'algorithme atteigne le critère d'arrêt.

4. Méthode proposée

Cette section présente la principale contribution de cet article. La méthode proposée comprend cinq phases, présentées dans la figure 3 et discutées en détail ci-dessous.

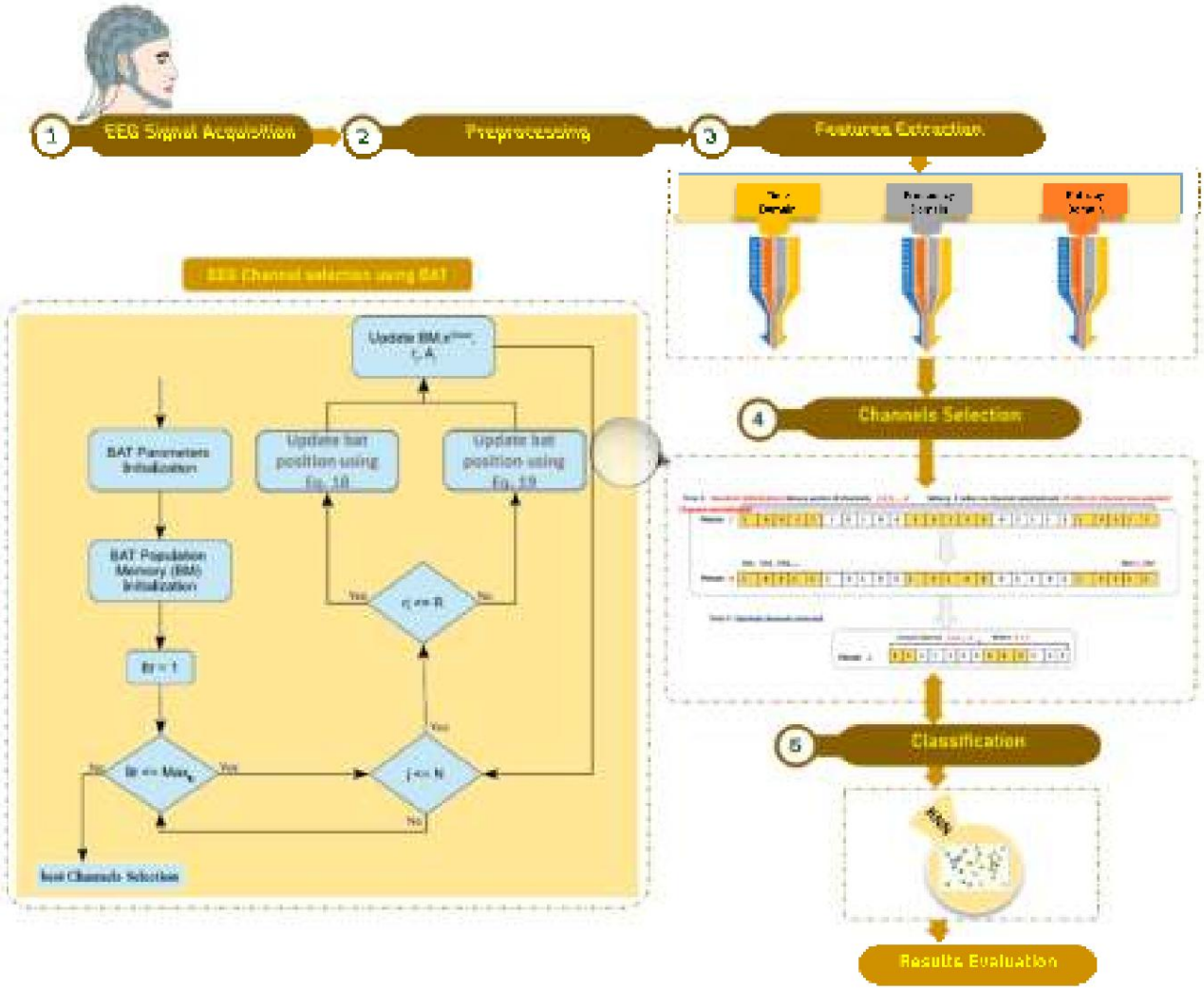


Figure 3. Une méthode proposée pour la sélection des canaux d'électroencéphalogramme.

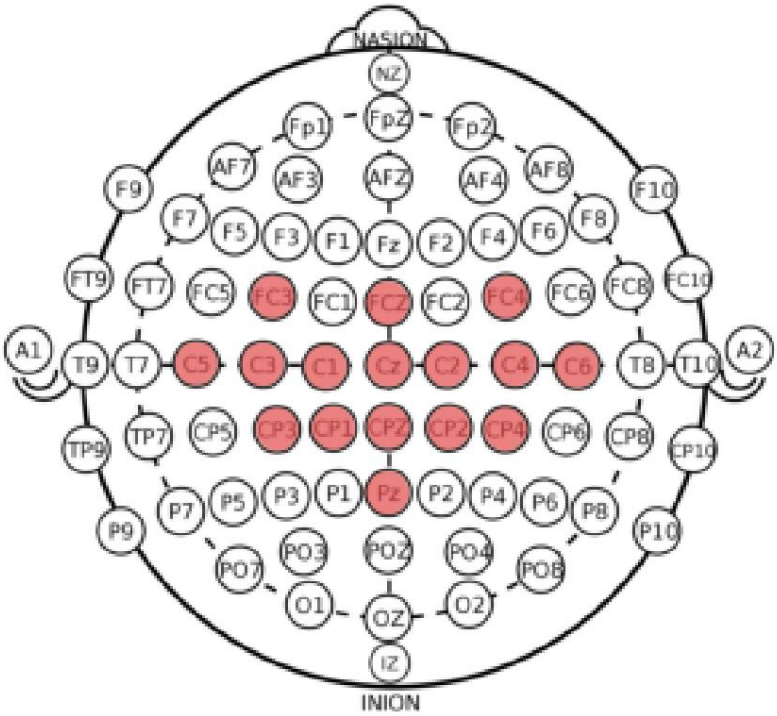
4.1. Phase I : Acquisition du signal EEG

Dans cette étude, la source de l'ensemble de données provient de g.tec Medical Engineering GmbH, un ensemble de données EEG standard ; les détails sur cet ensemble de données sont expliqués dans [8]. Les données EEG de patients post-AVC présentant une hémiparésie des membres supérieurs ont été étudiées. Cette étude a porté sur 8 patients post-AVC traités avec le système recoveryiX (g.tec Medical Engineering

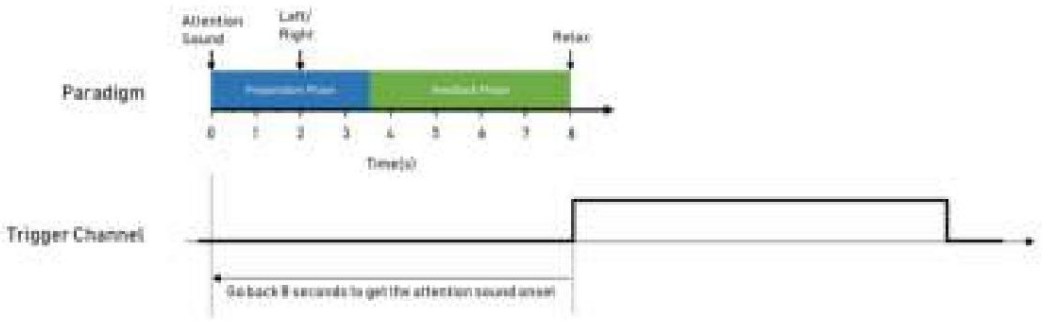
Machine Translated by Google

GmbH), avec un âge moyen de 22 ans (SD = 4,582). Chaque participant a reçu une formation MI basée sur BCI pendant trois mois, avec 2 séances de formation chaque semaine (pour un total de 25 séances de formation). L'équipe d'étude a mené et analysé deux évaluations (pré- et post-formation). L'évaluation pré-formation était programmée 30 à 35 jours avant l'intervention, et l' évaluation post-formation a été réalisée quelques jours après l'intervention (voir Figure 4). L' Ethikkommission des Landes Oberosterreich en Autriche (#D-42-17) a autorisé ce protocole d'étude et chaque patient a signé un formulaire de consentement éclairé avant la pré-éval

Enfin, cet ensemble de données est capturé avec une fréquence d'échantillonnage de 256 Hz.



(un)



(b)

Figure 4. (a) Distributions d'électrodes EEG basées sur le système 10-20 ; (b) diagramme schématique du protocole d'enregistrement EEG.

Selon les indications du système pour une tâche mentale MI, les patients devaient visualiser le mouvement dorsal du poignet. Avant et après la neurorééducation EEG, les patients ont participé à 25 séances de BCI basées sur l'IM avec des visites d'évaluation de suivi pour mesurer les changements fonctionnels. Chaque séance consistait en 240 répétitions MI avec les deux mains, réparties en trois séries de 80 essais. Chaque session a duré environ une heure, y compris le temps nécessaire à l'installation et au nettoyage. Les tâches BCI basées sur MI ont été illustrées avec des intervalles inter-essais randomisés dans un ordre pseudo-aléatoire.

Machine Translated by Google

Pour capturer tous les différents processus neuronaux qui ont lieu pendant la rééducation après un AVC – impliquant des composants cognitifs, affectifs et moteurs – des capuchons EEG actifs à seize canaux ont été utilisés. Les capuchons ont été fabriqués par la société g.tec Medical Engineering GmbH en Autriche.

Selon le système international 10-20, les électrodes EEG ont été positionnées comme suit : (FC5, FC1, FCz, FC2, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP1, CP2 et CP6).

Au sujet, une électrode de masse a été appliquée au niveau FPz et une électrode de référence a été fixée au lobe de l'oreille droite.

Ces canaux sont liés à différents processus cognitifs et sensorimoteurs qui sont d'une manière ou d'une autre altérés chez les patients après un accident vasculaire cérébral. En particulier, les canaux frontocentraux (FC5, FC1, FCz, FC2, FC6) sont associés à la régulation émotionnelle, à la mémoire de travail et à la zone motrice secondaire. De plus, il est important de noter que les canaux centraux (C5, C3, C1, Cz, C2, C4 et C6) sont liés à l'intégration sensori-motrice et à la représentation des mouvements corporels. De plus, les canaux centropariétaux (CP5, CP1, CP2, CP6) sont décrits comme étant impliqués dans le processus d'intégration des entrées extéroceptives et proprioceptives ainsi que dans la régulation des mouvements volontaires [22].

4.2. Phase II : Prétraitement Chaque

canal de l'ensemble de données EEG enregistré utilisait initialement deux filtres standards. Premièrement, un filtre passe-bande (BPF) avec des fréquences comprises entre (8 et 30 Hz) a été utilisé pour confiner la bande des données EEG enregistrées, et deuxièmement, un filtre coupe-bande Butterworth (BW) à (50 Hz) a été utilisé pour réduire le bruit. des interférences des lignes électriques. (CI) s(t) = [s1(t), . . . ,sn(t)], en utilisant l' algorithme FastICA proposé par [55]. L'ensemble s(t) de n composants inconnus qui ont été mélangés linéairement dans la matrice A et l'ensemble x(t) de n observations où x(t) = [x1(t), . . . , xn(t)] [56–58] représentent les EEG et sont liés à s(t), t, qui est le temps ; par conséquent, l'équation ICA est

$$x(t) = \text{Comme}(t) \quad (10)$$

Ensuite, trois mesures ont été utilisées pour évaluer les composants artefactuels, les CI : l'aplatissement (Kurt), l'asymétrie (Skw) et l'entropie de l'échantillon (SampEn). Si ces paramètres dépassaient le seuil pm1,2 pour chaque IC, le IC était marqué comme critique et débruité à l'aide de WT.

La valeur pratique du seuil a été déterminée par essais et erreurs et par des recherches antérieures [59-61]. La valeur seuil de ± 1, 2 n'est pas un inconvénient de la technique AICA – WT, car les CI artefactuels n'ont pas été rejetés mais débruités à l'aide de la technique WT. En conséquence, WT a débruité les CI marqués avant de renvoyer les composants améliorés à l' ensemble de données EEG d'origine [22,23].

4.3. Phase III : Extraction des caractéristiques EEG

Trois techniques différentes d'extraction de caractéristiques sont utilisées dans ce travail. Ces techniques incluent les caractéristiques du domaine temporel, les caractéristiques d'entropie et les caractéristiques du domaine fréquentiel telles qu'extraites dans [8].

4.4. Phase IV : sélection des canaux EEG à l'aide de BAT Optimizer

Cette phase constitue la principale contribution et comprend plusieurs étapes pour obtenir le sous-ensemble optimal de canaux EEG pouvant fournir le taux de précision le plus élevé. Les étapes suivantes représentent la manière dont nous adaptons l'optimiseur BAT au problème de sélection de canal EEG : • Étape 1 : initialisation des paramètres de l'algorithme BAT. Dans cette étape, les paramètres BAT sont initialisés avec des valeurs initiales. Ces paramètres sont le nombre de microbats (solutions) (N), le nombre maximum d'itérations (MaxItr), la plage de fréquences max (fmax) et min (fmin) , le vecteur vitesse (v), le taux de sonie (A), la fréquence du pouls (r), fréquence du pouls initiale (plage r (ε) dans la plage [−1, 1], comme indiqué dans le tableau 1.

^{0}), deux paramètres (λ et α) avec des valeurs constantes dans la plage [0, 1] et une bande passante

Machine Translated by Google

Tableau 1. Paramètres métaheuristiques.

Algorithme	Paramètres	Taille de la population	Fonctionne
CS	β = 1,5, Pa = 0,25 c1	20	30
OPS	= c2 = 2 A =	20	30
BATTE	0,5, r = 0,5 α = 0,5,	20	30
FFA	γ = 1, β0 = 0,2	20	30
GWO	-	20	30

• Étape 2 : initialisation de la mémoire de population BAT. Les solutions BAT sont générées aléatoirement au cours de cette étape à l'aide de l'équation (11), en tenant compte de l'algorithme BAT et de ses caractéristiques particulières. paramètres et contraintes du problème d'optimisation.

$$x_{j} = lbi + (ubi - lbi) \times R, \quad (11)$$

où i = 1, 2, . . . , d, j = 1, 2, . . . , N et R est un nombre aléatoire compris entre 0 et 1.

Les solutions produites génèrent ensemble la population, comme le montre l'équation (12), et sont stockées dans la mémoire BAT (BM) par ordre croissant en fonction de leur aptitude valeurs. La meilleure solution avec les valeurs les plus adaptées est attribuée à x^{Gbest} et stocké dans le première position du BM.

$$MB = \begin{matrix} x_1^1 & x_2^1 & \cdots & x_d^1 \\ x_1^2 & x_2^2 & \cdots & x_d^2 \\ \vdots & \vdots & \cdots & \vdots \\ x_1^N & x_2^N & \cdots & x_d^N \end{matrix}, \quad (12)$$

• Étape 3 : Intensification de la population de MTD. Désormais, toutes les microchauves - souris volent et changent de position. position considérant une vitesse v définie par une fréquence f générée aléatoirement, comme le montrent les équations (13) et (14). En conséquence, les positions des chauves-souris sont mises à jour en utilisant Équation (15).

$$fi = fmin + (fmin - fmax) \times R, \quad (13)$$

$$v_{j}^{'j} = v_{j}^{j} + (x_{j}^{j} - x_{j}^{G} \text{ meilleur}) \times fi, \quad (14)$$

$$x_{III,}^{'j} = x^{j} + v^{'j} \quad (15)$$

Il est à noter que les microbats mettent à jour leurs emplacements pour se rapprocher des meilleurs au monde. (x Gbest). Ainsi, la nouvelle position des microchauves-souris intensifie la position dans une direction vers $x_{Meilleur}$.

• Étape 4 : Diversification de la population BAT.
Dans cette étape, les positions des microbats sont mises à jour en fonction de paramètres aléatoires pour tenter de trouver de meilleures solutions globales. Une nouvelle solution est sélectionnée parmi le BM basé sur une méthode de sélection et attribué à x meilleur. Par conséquent, la solution actuelle est mis à jour en utilisant x best, comme suit :

$$x_{je}^{'j} = x_{\text{meilleur}} + \varepsilon \times A_{j.}^{'} \quad (16)$$

où $A^{'}$ est la moyenne du volume sonore. Les étapes 3 et 4 étant des étapes contradictoires, la MTD l'algorithme en choisit un à la fois pour l'exécution. Cette sélection est basée sur équation suivante :

$$x^{'j} = \begin{cases} x^{meilleur} + \varepsilon \times UNE_{j}^{'} & \text{si } rj \le R \\ xj + v^{'j} & \text{sinon} \end{cases}, \quad (17)$$

Machine Translated by Google

- Étape 5 : mise à jour du BM. La position des microbats dans le BM sera mise à jour dans cette étape Gbest si emplacement est plus adapté que l'ancien et R ≤ Aj . De plus, X est mis à jour si le nouveau le nouvel BM contient une solution avec une meilleure valeur de fitness. Par la suite, les valeurs rj et Aj seront mises à jour conformément aux équations (18) et (19).

$$rj = rj^{0}(1 - e^{(-\gamma \times itr)}), \quad (18)$$

$$Aj = \alpha \times Aj, \quad (19)$$

où itr désigne le numéro d'itération actuel.

- Étape 6 : critère d'arrêt. Les étapes 3, 4 et 5 seront répétées jusqu'à ce que l'algorithme atteigne le critère d'arrêt.

4.5. Phase V : Classification

Un classificateur KNN est utilisé dans ce travail. Dans cette méthode proposée, l'ensemble de données est divisé en 60 % de formation, 20 % de validation et 20 % de tests.

5. Expériences et résultats

Cette section explique en détail les performances de la méthode proposée (BAT) des canaux EEG pour la rééducation des patients victimes d'un AVC. L'approche proposée étant non déterministe, le taux de précision moyen sur 20 exécutions est calculé pour éviter des résultats biaisés.
Les expériences utilisent un PC Lenovo Ideapad 310, Chine. Processeur Intel Core i7® 2,59 Ghz, 16 Go de RAM et Windows 10 officiel. Les performances de la méthode proposée sont évaluées à l'aide de deux mesures, à savoir la précision et le nombre de canaux EEG sélectionnés, avec un éclairage sur la stabilité des algorithmes. en calculant la meilleure, la pire et la moyenne précision. De plus, un test statistique est adopté sur la base de la précision de la classification de l'ensemble de données EEG moteur/imagerie en utilisant la somme des rangs des algorithmes métaheuristiques.
La méthode proposée est testée en utilisant trois types de caractéristiques EEG : les caractéristiques du domaine temporel, les caractéristiques du domaine fréquentiel et les caractéristiques du domaine entropique. De plus, les résultats de performance de la méthode proposée sont évalués à l'aide d'un test statistique pour montrer la signification de la méthode proposée par rapport à d'autres algorithmes métaheuristiques.

Le tableau 1 montre les paramètres utilisés pour les algorithmes métaheuristiques sélectionnés qui sont utilisés dans ce travail.

5.1. Résultats du domaine temporel

Le tableau 2 présente les résultats de sept caractéristiques dans le domaine temporel : la dimension fractale de Higuchi (HFD) ; les paramètres Hjorth d'activité Hjorth (HjAc), de complexité Hjorth (Hj-Comp) et de mobilité Hjorth (HjMo) ; Hurst (Hur); aplatissement (Kurt); et l'asymétrie (Skw).
La précision et le nombre de canaux EEG sélectionnés sont deux mesures permettant d'évaluer les sept caractéristiques choisies pour l'ensemble de données EEG moteur/imagerie. Les mesures d'évaluation sont classées en trois catégories : meilleure, pire et moyenne précision et non. de canaux sélectionnés pour évaluer la stabilité des algorithmes.

Selon le tableau 2, les résultats de performances des algorithmes métaheuristiques PSO, GWO, FFA, CS et BAT avec plusieurs fonctionnalités ont permis d'obtenir une précision différente et remarquable concernant la meilleure, la pire et la moyenne précision et le nombre de canaux sélectionnés. Pour la fonction HFD, l'algorithme BAT a obtenu de meilleures précisions de 99, 97,5 et 92,8 pour la meilleure, moyenne et la pire précision et sélection de canal, respectivement. Pour la fonctionnalité HjAc, l' algorithme BAT a atteint une précision élevée, 97,7, pour la précision et non. de canaux sélectionnés dans le meilleur des cas, tandis que PSO a atteint un rapport précieux de 94,4 et 85,3 en termes de précision et de nombre de canaux sélectionnés dans les cas moyen et pire, respectivement. Pour la fonctionnalité HjComp, l'algorithme PSO a atteint une précision de 88,6 et 84,7 concernant les cas moyens et les pires, tandis que l'algorithme BAT a atteint la précision la plus élevée de 93,8 dans le meilleur des cas. Pour la fonctionnalité HjMo, BAT a obtenu 97,8, 95,4 et 88,6 pour le meilleur, la moyenne et le pire parmi d'autres algorithmes métaheuristiques. Pour la fonction Hur, PSO a obtenu des précisions élevées de 72,8, 65,2 et 56,6 pour le meilleur, la moyenne et le pire, respectivement. Pou

Machine Translated by Google

L'algorithme a obtenu des précisions précieuses de 73,2, 68,8 et 63,1 pour le meilleur, la moyenne et le pire précision et le nombre de canaux sélectionnés, respectivement.

De plus, pour Skw, BAT a atteint des précisions élevées de 79,2 et 72,5 pour la meilleure et la moyenne, respectivement, par rapport à PSO, qui a obtenu une précision de 58 dans le pire des cas d'exactitude. La figure 4 montre les résultats des caractéristiques du domaine temporel pour la convergence distribution de débit et de canal, en comparant les algorithmes BAT et PSO, qui ont obtenu la plus grande précision par rapport aux autres algorithmes métaheuristiques, c'est-à-dire GWO, FFA et CS. Sur la figure 4a, le taux de convergence montre les canaux et le nombre d'itérations pour les algorithmes BAT et PSO. De plus, la figure 4b présente les actions efficaces et émotionnelles canaux lorsque les canaux sont distribués dans les algorithmes BAT et PSO.

Tableau 2. Performances des algorithmes métaheuristiques avec plusieurs approches d'extraction de caractéristiques pour Ensemble de données EEG moteur/imagerie.

Fonctionnalité/algorithme			PSO	GWO	FFA	CS	BATTE
HFD	Meilleur	Précision	97,9	90,7	66,7	54,9	99
		Nombre de chaînes	8	3	1	1	6
	Pire	Précision	81,5	67,4	45.1	43.2	92,8
		Nombre de chaînes	12	6	3	2	13
	Signifier	Précision	94,4	76.2	56,7	50,6	97,5
		Nombre de chaînes	9	4	2	2	10
HjAc	Meilleur	Précision	97.1	88,4	68,8	55.1	97,7
		Nombre de chaînes	5	3	1	1	4
	Pire	Précision	85,3	70,4	41,7	42.1	77,9
		Nombre de chaînes	9	5	3	2	14
	Signifier	Précision	94,4	82.1	60,7	48,7	90,8
		Nombre de chaînes	8	4	3	2	8
HjComp	Meilleur	Précision	91.2	78,9	58,8	48,7	93,8
		Nombre de chaînes	7	3	2	1	5
	Pire	Précision	84,7	58.1	47.4	40,9	74,7
		Nombre de chaînes	15	6	3	2	13
	Signifier	Précision	88,6	69,9	53,9	44.2	88,2
		Nombre de chaînes	9	4	3	1	9
HjMo	Meilleur	Précision	96,5	89,3	67,9	53,5	97,9
		Nombre de chaînes	6	2	2	1	5
	Pire	Précision	86,4	51,8	53,9	43,8	88,6
		Nombre de chaînes	11	6	3	2	12
	Signifier	Précision	92,3	80,6	61.2	47,9	95,4
		Nombre de chaînes	9	4	3	1	9
Hur	Meilleur	Précision	72,8	57,9	47.2	44.1	70.1
		Nombre de chaînes	4	2	1	1	5
	Pire	Précision	56,6	44.1	41,5	41.3	53,9
		Nombre de chaînes	10	6	3	2	11
	Signifier	Précision	65.2	51,7	44,7	42,6	64,9
		Nombre de chaînes	8	4	3	2	9

Machine Translated by Google

Tableau 2. Suite

Fonctionnalité/algorithme			PSO	GWO	FFA	CS	BATTE
Kurt	Meilleur	Précision	71.2	61	48.4	43.4	73.2
		Nombre de chaînes	7	3	2	1	7
	Pire	Précision	56,5	48.2	43.3	40.3	63.1
		Nombre de chaînes	12	6	3	2	12
	Signifier	Précision	65,6	54,9	45.4	42	68,8
		Nombre de chaînes	9	5	3	2	10
Skw	Meilleur	Précision	78,7	63,7	50	44.4	79.2
		Nombre de chaînes	4	1	2	1	3
	Pire	Précision	58	41	43.2	40.4	48.2
		Nombre de chaînes	11	6	3	2	13
	Signifier	Précision	69,6	53.3	45,9	42.3	72,5
		Nombre de chaînes	7	4	2	2	9

5.2. Résultats du domaine d'entropie

Le tableau 3 montre les résultats de la précision des performances en utilisant sept caractéristiques d'entropie, à savoir, ConFuzEn, FuzEn, impe, MFEmu, RCMFEmu, SampEn et TsEn. Basé sur le précision et le nombre de canaux sélectionnés pour le meilleur, le pire et le cas moyen, ces entropie les fonctionnalités sont utilisées pour comparer les cinq algorithmes métaheuristiques, PSO, GWO, FFA, CS et BAT, respectivement. Par rapport aux sept autres caractéristiques d'entropie, la meilleure classification les performances ont atteint la plus haute précision en utilisant ConFuzEn avec l'algorithme PSO par rapport aux algorithmes GWO, FFA, CS et BAT. D'autres caractéristiques d'entropie produites plus élevées la précision des résultats avec l'algorithme BAT est supérieure à celle des algorithmes PSO, GWO, FFA et CS. Le L'algorithme BAT a enregistré des précisions de 95,5, 95,6, 88,1, 87,2, 91,4 et 84,6 en utilisant FuzEn, impe, MFEmu, RCMFEmu, SampEn et TsEn, respectivement. De plus, la figure 5 présente les résultats d'évaluation des performances entre les algorithmes BAT et PSO concernant le taux de convergence et distribution des canaux.

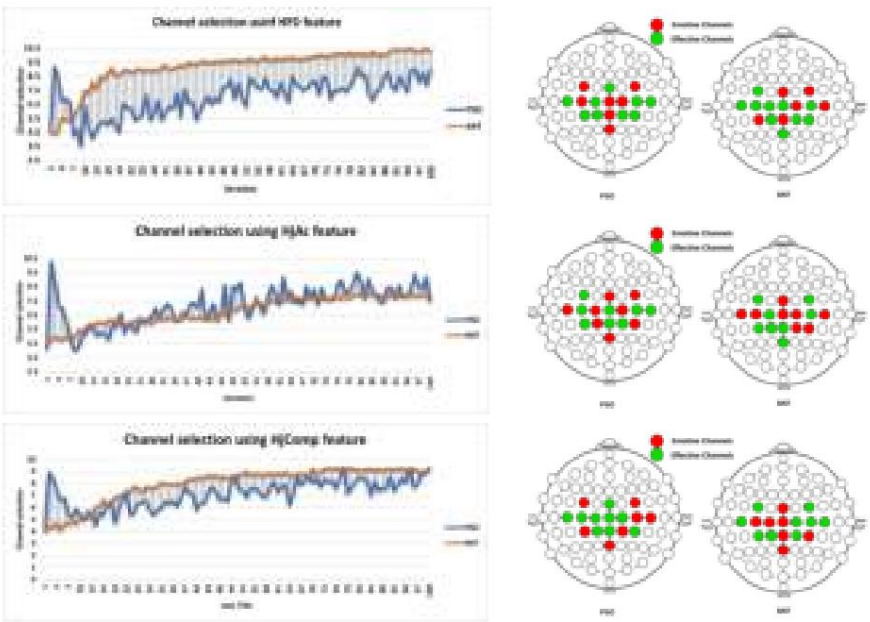


Figure 5. Suite.

Machine Translated by Google

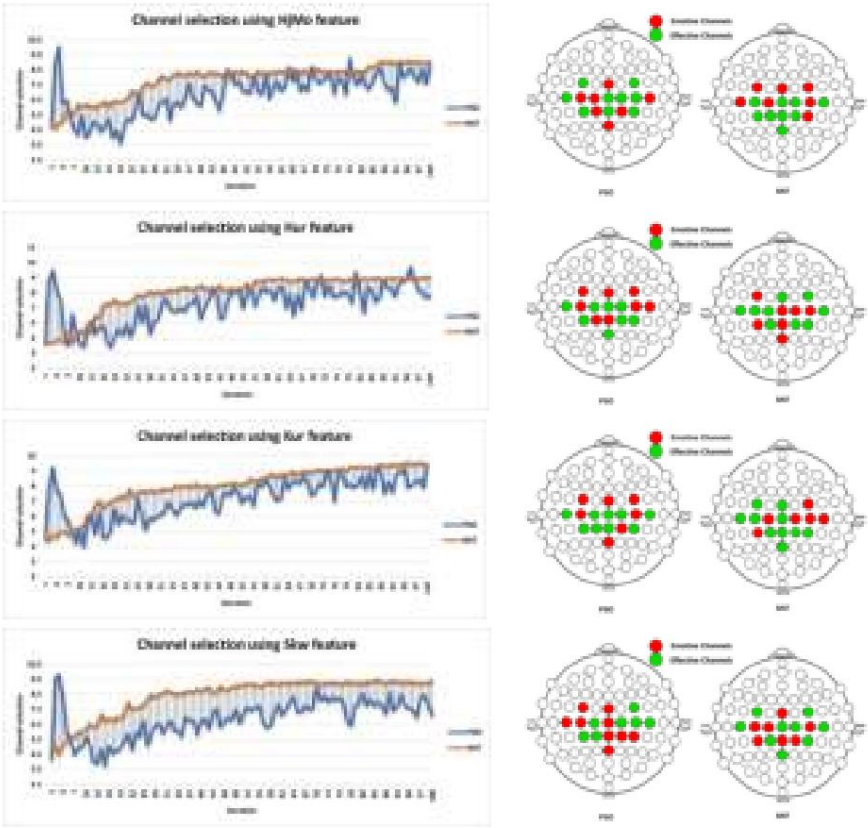


Figure 5. Taux de convergence et distribution des canaux.

Tableau 3. Performances des algorithmes métaheuristiques avec plusieurs approches d'extraction de caractéristiques pour Ensemble de données EEG moteur/imagerie.

Fonctionnalité/algorithme			PSO	GWO	FFA	CS	BATTE
ConvFuzEn	Meilleur	Précision	96,6	82,7	61,3	52	96,5
		Nombre de chaînes	3	3	1	1	7
	Pire	Précision	76,4	64,7	45,5	42,5	87.1
		Nombre de chaînes	11	6	3	2	14
	Signifier	Précision	90,2	73,6	54,9	45,8	93.2
		Nombre de chaînes	8	5	2	1	11
FuzEn	Meilleur	Précision	93,6	81,6	60,6	49.1	95,5
		Nombre de chaînes	2	4	2	1	8
	Pire	Précision	82,2	68,7	50,5	41,8	88,9
		Nombre de chaînes	10	6	3	2	12
	Signifier	Précision	88,6	74,4	55,7	45,5	92.1
		Nombre de chaînes	7	5	3	1	10
impen	Meilleur	Précision	94,7	78,6	59.3	49,8	95,6
		Nombre de chaînes	5	3	2	1	7
	Pire	Précision	75,8	56.1	47,9	44.3	84,9
		Nombre de chaînes	10	6	3	2	12
	Signifier	Précision	85,3	69,8	52.3	46,6	90,8
		Nombre de chaînes	8	5	2	2	10

Machine Translated by Google

Tableau 3. Suite

Fonctionnalité/algorithme			PSO	GWO	FFA	CS	BATTE
MFEmu	Meilleur	Précision	85,2	73,6	54.3	46,8	88.1
		Nombre de chaînes	7	3	2	1	7
	Pire	Précision	70,8	55,9	45.1	40,5	75,5
		Nombre de chaînes	12	6	3	2	12
	Signifier	Précision	80,7	66.1	49.4	44,8	84
		Nombre de chaînes	9	5	2	2	9
RCMFEmu pire	Meilleur	Précision	84,3	67,7	51,9	46,6	87,2
		Nombre de chaînes	4	2	2	1	8
		Précision	63,7	47	44,9	41.2	75.3
		Nombre de chaînes	12	6	3	2	13
	Signifier	Précision	77.3	56,5	47,7	43	80,3
		Nombre de chaînes	8	4	2	1	10
SampFr	Meilleur	Précision	87,9	71,5	52,9	51,9	91,4
		Nombre de chaînes	6	3	1	1	7
	Pire	Précision	75.4	53.2	42,6	40.4	75,9
		Nombre de chaînes	10	6	3	3	14
	Signifier	Précision	81,4	63,7	48.2	44.3	83,6
		Nombre de chaînes	8	5	2	2	10
TsEn	Meilleur	Précision	83,9	72,5	52.3	46,6	84,6
		Nombre de chaînes	7	3	1	1	4
	Pire	Précision	65,9	54,4	41.2	40.2	61,3
		Nombre de chaînes	12	7	3	2	13
	Signifier	Précision	78,7	64,5	47.2	42,8	79,8
		Nombre de chaînes	10	5	2	1	10

5.3. Résultats du domaine fréquentiel

Sur la base du tableau 4, les caractéristiques du domaine fréquentiel de moyenneF et medF ont obtenu des résultats élevés. des précisions de 96,7 et 83,4 pour les algorithmes PSO et BAT, respectivement. Dans le meilleur, dans le pire, et les cas moyens, MeanF a acquis une précision significative par rapport à la fonctionnalité medF sur le résultats de précision globaux pour les algorithmes PSO et BAT. La meilleure précision de medF était obtenu à 96,7 avec PSO, tandis que les pires et moyennes précisions obtenues étaient de 90,3 et 93,8 avec l'algorithme BAT, respectivement. De plus, l'algorithme BAT a obtenu une grande précision résultats de classification de 83,4 et 78,4 relatifs aux meilleures et aux pires précisions avec le caractéristique de medF, respectivement, tandis que PSO a gagné une précision de 69,5. Le taux de convergence et la distribution des canaux sont illustrées à la figure 6.



Figure 6. Suite.

Machine Translated by Google

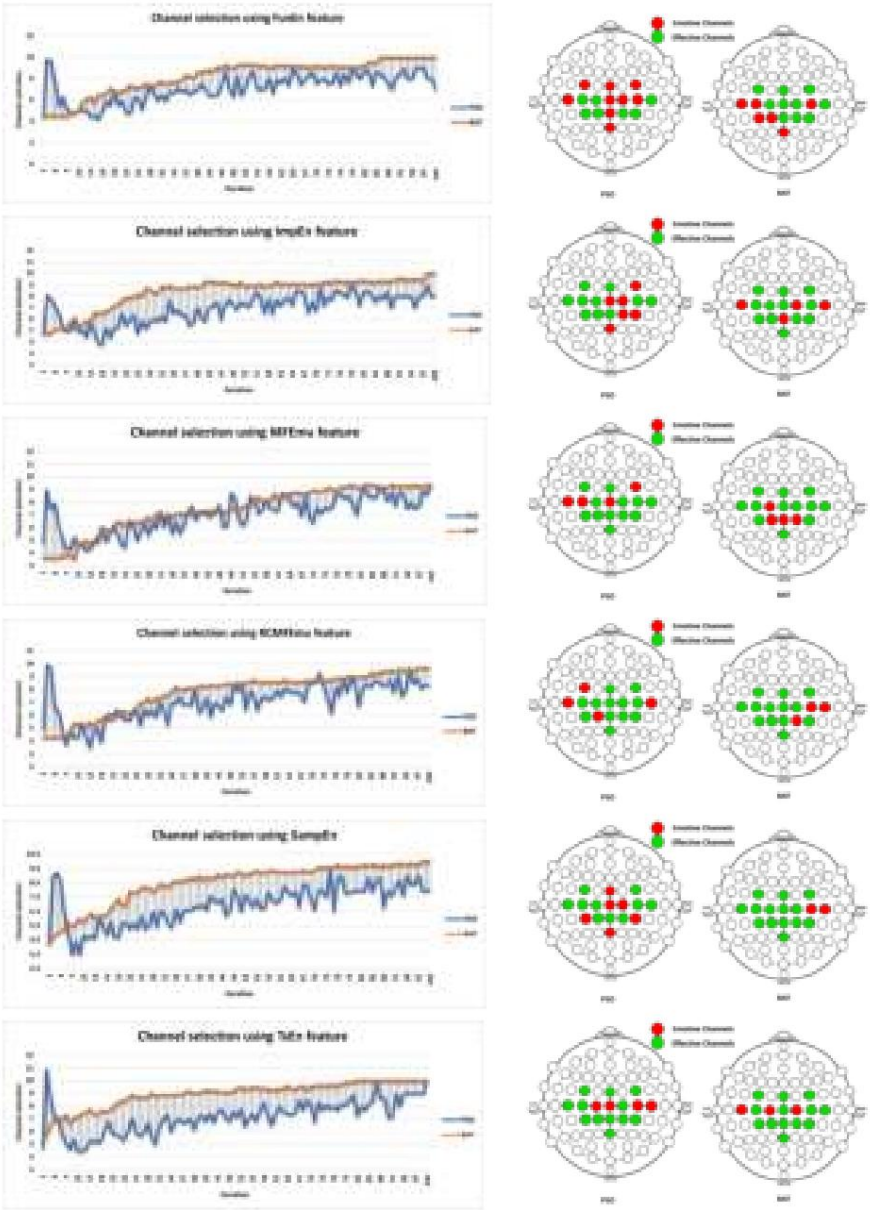


Figure 6. Taux de convergence et distribution des canaux.

Comme mentionné, le tableau 4 montre les performances de la méthode proposée (BAT) par rapport aux algorithmes PSO, GWO, FFA et CS utilisant les fonctionnalités du domaine fréquentiel. L'algorithme BAT a obtenu de meilleurs résultats que tous les autres algorithmes, sauf dans certains cas où l'algorithme PSO l'a légèrement surpassé, comme dans MeanF, où la meilleure précision était de 96,7 pour PSO et de 96,3 pour BAT. Cependant, les exécutions globales de l'algorithme BAT ont obtenu de meilleurs résultats avec des mesures de précision égales à 93,8, 91,2, 76,4, 54 et 50 pour les algorithmes BAT, PSO, GWO, FFA et CS, respectivement. Le taux de convergence et la répartition des canaux sont illustrés à la figure 7.



Figure 7. Suite.

Machine Translated by Google



Figure 7. Taux de convergence et distribution des canaux.

Tableau 4. Performances de l'algorithme BAT par rapport à plusieurs algorithmes métaheuristiques utilisant caractéristiques du domaine fréquentiel.

Fonctionnalité/algorithme			PSO	GWO	FFA	CS	BATTE
moyenneF	Meilleur	Précision	96,7	87,3	64.1	52,5	96,3
		Nombre de chaînes	5	3	2	1	7
	Pire	Précision	65,4	66	50,2	41,9	90,3
		Nombre de chaînes	10	6	3	2	12
	Signifier	Précision	91.2	76,4	54	50	93,8
		Nombre de chaînes	8	4	2	2	10
MedF	Meilleur	Précision	82,7	69,2	53,6	46,6	83,4
		Nombre de chaînes	6	3	2	1	6
	Pire	Précision	69,5	55,5	41	20	68,7
		Nombre de chaînes	10	6	3	2	14
	Signifier	Précision	76,3	60,6	46.1	36,5	78,4
		Nombre de chaînes	9	4	2	2	10

5.4. Résultats de l'analyse statistique

Avant de présenter nos conclusions pour l'exactitude et non. de chaînes sélectionnées pour En améliorant les performances de classification de l'analyse des caractéristiques de temps, d'entropie et de fréquence des signaux EEG d'AVC , nous aimerions examiner l'influence de la précision de la classification. dans l'ensemble de données EEG moteur/imagerie par l'adoption de la somme des rangs de la métaheuristique algorithmes. De cette façon, la fonctionnalité qui a amélioré la précision du classement des chaînes sélectionnées peuvent être identifiés.

Comme le montre le tableau 5, nous avons effectué la somme des rangs des PSO, GWO, FFA, CS, et les algorithmes métaheuristiques BAT en fonction du domaine de fonctionnalités individuel. Dans le domaine temporel, nous avons testé sept fonctionnalités. L'algorithme BAT a considérablement amélioré la rang de précision, notamment en HFD et Hjac, pour un rang de précision de 99 et 97,7, respectivement. De plus, d'autres fonctionnalités dans le domaine temporel, telles que HjComp, HjMo, Kurt et Skw, ont montré un impact influent sur les résultats de précision de la classification par rapport à la somme du rang de l'algorithme BAT à l'exception de la fonctionnalité Hur, qui a eu un effet important sur l'algorithme PSO. Cependant, lors de la phase de test statistique, l'algorithme BAT avait un faible résultats de sommation de rang parmi d'autres algorithmes métaheuristiques.

D'un autre côté, la somme des classements pour l'algorithme BAT a montré un impact significatif impact sur les résultats de précision de la classification par rapport aux sept caractéristiques de l'entropie domaine à l'exception de la fonctionnalité ConvFuzEn, qui a eu un effet remarquable sur le PSO algorithme avec une différence de 0,1 par rapport à BAT concernant les résultats de somme des rangs dans l'analyse statistique. Cependant, l'algorithme BAT avait de faibles résultats de sommation des classements. en phase de test statistique parmi les autres algorithmes métaheuristiques.

Cependant, dans la précision de la classification résulte des caractéristiques du domaine fréquentiel, MeanF a montré un effet d'amélioration important sur la précision avec l'algorithme PSO relatif au rang de précision de 96,7 par rapport à la fonction medF, qui avait une précision de 83,4

Machine Translated by Google

classer avec l'algorithme BAT. De plus, les algorithmes BAT et PSO avaient des résultats de sommation de rang faibles par rapport aux autres algorithmes métaheuristiques.

Tableau 5. La somme des classements des algorithmes métaheuristiques basés sur la précision de la classification sur un ensemble de données EEG moteur/imagerie.

Fonctionnalité/algorithme			PSO	GWO	FFA	CS	BATTE
TDF HjMo	HFD	Précision	97,9	90,7	66,7	54,9	99
		Rang	2	3	4	5	1
	HjAc	Précision	97.1	88,4	68,8	55.1	97,7
		Rang	2	3	4	5	1
	HjComp	Précision	91.2	78,9	58,8	48,7	93,8
		Rang	2	3	4	5	1
		Précision	96,5	89,3	67,9	53,5	97,9
		Rang	2	3	4	5	1
	Hur	Précision	72,8	57,9	47.2	44.1	70.1
		Rang	1	3	4	5	2
	Kur	Précision	71.2	61	48.4	43.4	73.2
		Rang	2	3	4	5	1
	Skw	Précision	78,7	63,7	50	44.4	79.2
		Rang	2	3	4	5	1
	Somme des rangs		13	21	28	35	8
FDF MedF	moyenneF	Précision	96,7	87,3	64.1	52,5	96,3
		Rang	1	3	4	5	2
		Précision	82,7	69,2	53,6	46,6	83,4
		Rang	2	3	4	5	1
	Somme des rangs		3	6	8	10	3
EDF MFEmu	ConvFuzEn	Précision	96,6	82,7	61,3	52	96,5
		Rang	1	3	4	5	2
	FuzEn	Précision	93,6	81,6	60,6	49.1	95,5
		Rang	2	3	4	5	1
	impen	Précision	94,7	78,6	59.3	49,8	95,6
		Rang	2	3	4	5	1
		Précision	85,2	73,6	54.3	46,8	88.1
		Rang	2	3	4	5	1
	RCMFEmu	Précision	84,3	67,7	51,9	46,6	87,2
		Rang	2	3	4	5	1
	SampFr	Précision	87,9	71,5	52,9	51,9	91,4
		Rang	2	3	4	5	1
	TsEn	Précision	83,9	72,5	52.3	46,6	84,6
		Rang	2	3	4	5	1
	Somme des rangs		13	21	28	35	8

La figure 8 montre les résultats de la vitesse des algorithmes métaheuristiques en secondes.

Machine Translated by Google

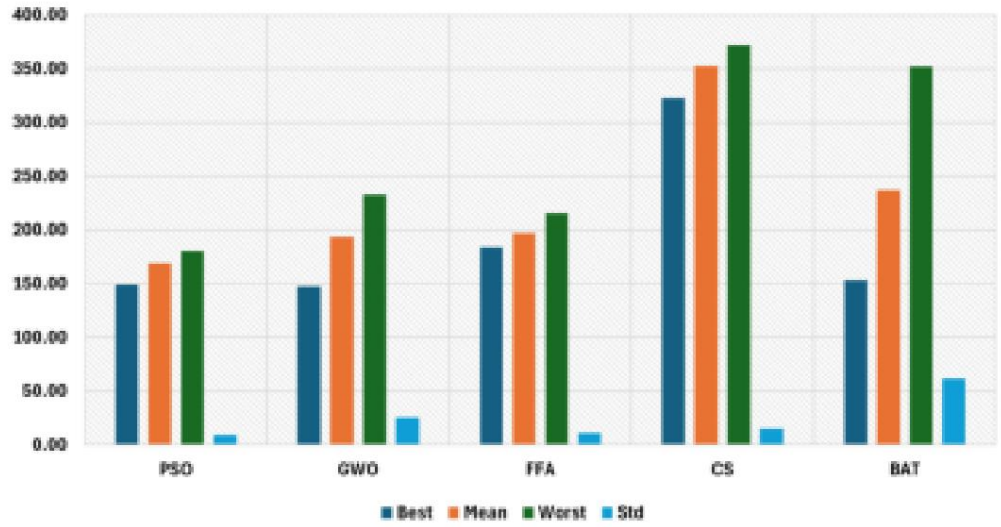


Figure 8. Vitesse des algorithmes métaheuristiques en secondes.

5.5. Discussion sur les résultats

Cette recherche vise à choisir les canaux EEG les plus pertinents qui produisent un taux de précision efficace pour la rééducation des patients victimes d'un AVC en utilisant une technique d'optimisation basée sur un algorithme de chauve-souris binaire pour optimiser les caractéristiques du TEF. Pour améliorer encore les performances de classification des signaux EEG d'AVC, trois types de caractéristiques (à savoir les caractéristiques du domaine temporel, entropique et fréquentiel) ont été utilisées comme base de référence pour la tâche de classification. Deux mesures ont été appliquées pour évaluer les performances de la méthode proposée, la précision et le non. des canaux EEG sélectionnés. En outre, nous avons adopté un test statistique utilisant la somme des classements de l'ensemble de données EEG moteur/imagerie pour examiner l'influence de la précision de la classification et la signification de la méthode suggérée par rapport aux algorithmes métaheuristiques de pointe tels que PSO, GWO, FFA et CS. Premièrement, l'exactitude et le non. des canaux EEG sélectionnés ont été évalués en calculant les valeurs de précision la meilleure, la pire et la moyenne pour capturer la stabilité de l'algorithme mis en œuvre. Dans le domaine temporel, sept fonctionnalités ont été testées, et la fonctionnalité HFD a eu l' influence la plus significative sur l'amélioration de la tâche de classification dans l'optimiseur BAT parmi d'autres algorithmes d'optimisation pour les meilleurs, moyens et pires cas de précision et le non. de canaux sélectionnés, respectivement, sur la base d'un classificateur kNN. De plus, les résultats BAT ont surpassé les valeurs de test pour le cas de meilleure précision par rapport à PSO, qui a obtenu les résultats les plus élevés par rapport aux algorithmes GWO, FFA et CS dans les fonctionnalités HjAc et HjComp. De plus, les résultats de l'optimiseur BAT ont enregistré des valeurs de précision élevées, moyennes et pires, ainsi que le no. des canaux sélectionnés, respectivement, avec les fonctionnalités HFD, HjMo et Kurt, respectivement, tandis que PSO avait l'influence la plus significative sur l'amélioration de la tâche de classification par rapport à BAT avec la fonctionnalité Hur basée sur la meilleure, la moyenne et la pire précision et la Non. des cl
Pour les meilleures valeurs, l'optimiseur BAT a surperformé dans les valeurs les plus précieuses et les plus élevées parmi PSO et les autres algorithmes métaheuristiques avec les six fonctionnalités HFD, HjAc, HjComp, HjMo, Kurt et Skw, respectivement (voir tableau 2). De plus, selon les résultats de la précision des performances utilisant sept fonctionnalités d'entropie basées sur le classificateur KNN, les meilleures performances de classification ont été obtenues avec la plus grande précision en utilisant ConFuzEn avec l'algorithme PSO par rapport aux algorithmes GWO, FFA, CS et BAT.

D'autres fonctionnalités d'entropie ont produit des résultats de précision plus élevés avec l'algorithme BAT qu'avec les algorithmes PSO, GWO, FFA et CS selon les valeurs de précision la meilleure, la moyenne et la pire et le nombre non. des chaînes sélectionnées, respectivement. L'algorithme BAT a enregistré des valeurs de précision optimales de 95,5, 95,6, 88,1, 87,2, 91,4 et 84,6 en utilisant respectivement les fonctionnalités FuzEn, impe, MFEmu, RCMFEmu, SampEn et TsEn, tandis que PSO a surpassé BAT avec des valeurs de résultat de précision pour le Fonctionnalité ConvFuzEn pour le classificateur KNN (voir tableau 3).

Machine Translated by Google

Les algorithmes BAT et PSO ont obtenu des précisions élevées de 96,7 et 83,4, respectivement, par rapport à d'autres algorithmes de la littérature pour les caractéristiques des domaines fréquentiels mediumF et medF . L'algorithme BAT a obtenu des résultats de classification de haute précision de 83,4 et 78,4 par rapport à la meilleure et à la pire précision avec la fonction medF, respectivement, tandis que PSO a obtenu une précision de 69,5 ; en outre, les algorithmes PSO et BAT ont gagné considérablement en précision par rapport à la fonction medF sur les résultats de précision globaux dans les cas de précision le meilleur, le pire et le moyen. Sur la base des résultats de classification basés sur les caractéristiques des trois types de domaines différents, les algorithmes BAT et PSO ont exercé des impacts significatifs sur les performances de classification par rapport aux résultats des algorithmes d'optimisation GWO, FFA et CS concernant la précision et le non-respect. . des chaînes sélectionnées, respectivement. Par conséquent, ils démontrent une amélioration significative de la précision moyenne par rapport au résultat de la classification sur la base des domaines de caractéristiques individuels, comme le montre la figure 5, ce qui explique la comparaison entre le taux de convergence et la distribution des canaux pour les algorithmes BAT et PSO.

Deuxièmement, nous avons examiné l'influence de la précision de la classification sur l' ensemble de données EEG moteur/imagerie en utilisant la somme adoptée des rangs d'algorithmes métaheuristiques. Ainsi, la fonctionnalité qui a amélioré la précision du classement des canaux sélectionnés peut être identifiée. L'algorithme BAT a considérablement amélioré le classement de précision, en particulier dans les fonctionnalités HFD et Hjac, avec un classement de précision de 99 et 97,7, respectivement, par rapport aux algorithmes PSO, GWO, FFA et CS selon la somme des résultats avec par rapport au domaine tempore De plus, des fonctionnalités telles que HjComp, HjMo, Kurt et Skw ont montré un impact influent sur les résultats de précision de la classification par rapport à la somme des classements de l'algorithme BAT, à l'exception de la fonctionnalité Hur, qui a eu un effet important sur l'algorithme PSO. Cependant, lors de la phase de test statistique, l'algorithme BAT avait des résultats de sommation de rang faibles parmi les autres algorithmes métaheuristiques. De plus, l'algorithme BAT a montré un impact significatif sur les résultats de précision de la classification pour la somme des classements relatifs aux sept caractéristiques du domaine d'entropie, à l'exception de la fonctionnalité ConvFuzEn, qui a eu un effet remarquable sur l'algorithme PSO avec une différence de 0,1 par rapport à BAT concernant la somme des résultats de classement dans l'analyse statistique. Cependant, lors de la phase de test statistique, l' algorithme BAT avait des résultats de sommation de rang faibles parmi les autres algorithmes métaheuristiques.
De plus, les algorithmes BAT et PSO avaient de faibles résultats de sommation de rang parmi les autres algorithmes métaheuristiques dans les caractéristiques du domaine fréquentiel. La fonctionnalité MeanF a montré un effet d'amélioration important sur la précision avec l'algorithme PSO par rapport au rang de précision de 96,7 par rapport à la fonctionnalité medF, qui avait un rang de précision de 83,4 avec l'algorithme BAT.

6. Conclusions et travaux futurs

Cet article propose une nouvelle méthode de sélection des canaux EEG basée sur un algorithme d'optimisation appelé algorithme inspiré de BAT. L'objectif principal de l'algorithme proposé est de sélectionner les canaux EEG les plus pertinents pouvant fournir un taux de précision plus élevé pour la rééducation des patients victimes d'un AVC.

La méthode proposée a été testée à l'aide d'un ensemble de données EEG standard collectées auprès de huit patients post-AVC présentant une hémiparésie des membres supérieurs. Les données EEG ont été capturées à l'aide de capuchons EEG avec 16 électrodes actives de g.Nautilus PRO, g.tec Medical Engineering GmbH, Autriche.

La méthode proposée utilisait plusieurs méthodes d'extraction de caractéristiques EEG telles que le domaine temporel, le domaine fréquentiel et le domaine entropique. De plus, la méthode proposée (BAT) a été comparée à quatre algorithmes métaheuristiques, à savoir l'optimisation par essaim de particules (PSO), l'optimiseur de loup gris (GWO), la recherche de coucou (CS) et l'algorithme de luciole (FFA). Les performances de la méthode proposée ont été évaluées en utilisant le taux de précision et le nombre de canaux sélectionnés. La méthode proposée a obtenu les meilleurs résultats parmi toutes les méthodes d'extraction de caractéristiques et a montré une amélioration significative en utilisant un test d'analyse statistique standard appelé test de sommation des classements.

Machine Translated by Google

Enfin, la méthode proposée a réussi à réduire le nombre de canaux EEG à moins de moitié tout en maintenant le taux de précision.

Concernant les travaux futurs, la méthode proposée souffre de quelques problèmes ; par exemple, la méthode proposée est incapable de générer de nouvelles solutions capables d' augmenter la précision (c'est-à-dire bloquées dans des minima locaux). Pour cela, la méthode proposée devra être améliorée pour surmonter ce problème en modifiant le mécanisme de l'algorithme BAT ou en hybridant l'algorithme BAT avec un autre algorithme métaheuristique pour pouvoir trouver la solution optimale. Concernant le deuxième problème, la version actuelle de l'algorithme BAT se concentre sur la recherche de la solution optimale offrant le taux de précision le plus élevé et ne prend pas en compte le nombre de canaux EEG dans la fonction objectif. Par conséquent, dans les travaux futurs , nous travaillerons sur la version BAT multi-objectif proposée en utilisant simultanément la plus grande précision et le plus petit nombre de canaux EEG.

Contributions des auteurs : Conceptualisation, MAA-B., ZAAA et NKA-Q. ; méthodologie, MAA-B., ZAAA et NKA-Q.; logiciel, MAA-B. et ZAAA ; validation, MAA-B. et SNM ; enquête, MAA-B. et ZAAA ; ressources, MAA-B., ZAAA et CG ; rédaction : version originale , MAA-B., ZAAA, NKA-Q., SNM, NSA et CG ; Visualisation, MAA-B., ZAAA, SNM, NSA et CG Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Ce travail a été soutenu par le doyen de la recherche et des études supérieures (DRG) de l'Université d'Ajman, Ajman, Émirats arabes unis (subvention n° 2022-IRG-ENIT-7).

Déclaration de disponibilité des données : les données sont contenues dans l'article.

Conflits d'intérêts : l'auteur Christoph Guger était employé par g.tec Medical Engineering GmbH. Les autres auteurs déclarent que la recherche a été menée en l'absence de toute relation commerciale ou financière pouvant être interprétée comme un conflit d'intérêts potentiel.

Références

1. Li, F. ; Fan, Y. ; Zhang, X. ; Wang, C. ; Hu, F. ; Jia, W. ; Hui, H. Méthode de fusion multi-fonctionnalité basée sur le signal EEG et son application dans la classification des accidents vasculaires cérébraux. J.Méd. Système. 2020, 44, 39. [Réf. croisée]
2. Khairunizam, W. ; Oui, CW ; Murugappan, M. ; Junoh, AK ; Razlan, ZM ; Shahriman, A. ; Mustafa, WAW ; Ibrahim, Z. ; Nurhafizah, S. Un cadre expérimental pour évaluer les émotions des patients victimes d'un AVC à l'aide de l'électroencéphalogramme (EEG). J. Phys. Conf. Ser. 2020, 1529, 052072. [Réf. croisée]
3. Carino-Escobar, RI ; Carrillo-Mora, P. ; Valdés-Cristerna, R. ; Rodriguez-Barragan, MA; Hernández-Arenas, C. ; Quinzaños-Fresnedo, J. ; Galice-Alvarado, MA ; Cantillo-Negrete, J. Analyse longitudinale des rythmes cérébraux des patients victimes d'un AVC lors d'une intervention avec une interface cerveau-ordinateur. Plast neuronal. 2019, 2019. [Réf. croisée] [Pub Med]
4. Grefkes, C. ; Fink, GR Approches basées sur la connectivité en matière d'AVC et de récupération fonctionnelle. Lancette Neurol. 2014, 13, 206-216. [Référence croisée]
5. Hachinski, V. ; Donnan, Géorgie ; Gorelick, PB; Hacke, W. ; Cramer, Caroline du Sud ; Kaste, M. ; Pêcheur, M. ; Brainin, M. ; Buchan, AM; Eh bien, hein ; et coll. Accident vasculaire cérébral : Travailler à un agenda mondial prioritaire. AVC 2010, 41, 1084-1099. [Référence croisée] [Pub Med]
6. Gao, W. ; Cui, Z. ; Yu, Y. ; Mao, J. ; Xu, J. ; Ji, L. ; Kan, X. ; Shen, X. ; Li, X. ; Zhu, S. ; et coll. Application d'un système d'interface cerveau-ordinateur avec rétroaction visuelle et motrice dans la rééducation fonctionnelle des membres et du cerveau après un AVC : rapport de cas. Science du cerveau. 2022, 12, 1083. [CrossRef]
7. Jin, J. ; Qu, T. ; Xu, R. ; Wang, X. ; Cichocki, A. Classification EEG de l'imagerie motrice basée sur l'optimisation clairsemée riemannienne et la fusion Dempster-Shafer de modèles multi-temps-fréquence. IEEETrans. Système neuronal. Rééducation. Ing. 2022, 31, 58-67. [Référence croisée]
8. Al-Qazzaz, NK ; Alyasseri, ZAA; Abdulkareem, KH ; Ali, Nouvelle-Écosse; Al-Mhiqani, MN ; Guger, C. Fusion de fonctionnalités EEG pour l'imagerie motrice : un nouveau cadre robuste pour la réadaptation des patients victimes d'un AVC. Calculer. Biol. Méd. 2021, 137, 104799. [Réf. croisée]
9. Al-Timemy, AH ; Bugmann, G. ; Outram, N. ; Escudero, J. Réduction des erreurs de classification pour le contrôle myoélectrique des mouvements de la main avec analyse des composants indépendants. Dans les actes de la 5e Conférence internationale sur les technologies de l'information, ICIT, Bali, Indonésie, 26-28 octobre 2011 ; Tome 2011.
10. Al-Qazzaz, NK ; Ali, SHBM ; Ahmad, SA; Escudero, J. Sélection optimale des canaux EEG pour l'identification de la démence vasculaire à l'aide d'un algorithme de recherche par gravitation binaire amélioré. Dans Actes de la Conférence internationale pour l'innovation en génie biomédical et en sciences de la vie, Penang, Malaisie, 10-13 décembre 2017 ; 125-130.
11. Al-Qazzaz, NK ; Sabir, MK; Ali, SHBM ; Ahmad, SA; Grammer, K. Optimisation multicanal avec des marqueurs hybrides d'entropie spectrale pour l'amélioration de l'identification de genre des EEG basés sur les émotions. Accès IEEE 2021, 9, 107059-107078. [Référence croisée]
12. Velasco, I. ; Sipols, A. ; De Blas, CS; Pasteur, L. ; Bayona, S. Classification des signaux EEG d'imagerie motrice avec une approche de séries chronologiques multivariées . Biomède. Ing. En ligne 2023, 22, 29. [CrossRef]

Machine Translated by Google

13. Jouzizadeh, M. Caractéristiques du réseau évalué par EEG et de la variabilité du signal chez les hommes et les femmes au cours d'une tâche visuospatiale. doctorat Thèse, Université d'Ottawa/Université d'Ottawa, Ottawa, ON, Canada, 2024.
14. Al-Qazzaz, NK ; Sabir, MK; Ali, S. ; Ahmad, SA; Grammer, K. Canaux EEG efficaces pour l'identification des émotions dans les régions du cerveau à l'aide d'un algorithme d'évolution différentielle. Dans Actes de la 41e conférence internationale annuelle 2019 de l'IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Allemagne, 23-27 juillet 2019 ; pages 4703 à 4706.
15. Dornhege, G. ; Millán, JdR; Hinterberger, T. ; McFarland, DJ ; Muller, KR Vers l'interface cerveau-ordinateur ; Presse du MIT : Cambridge, MA, États-Unis, 2007 ; Tome 63.
16. Kingphai, K. ; Moshfeghi, Y. Sur la sélection des canaux pour la classification de la charge de travail mentale basée sur l'EEG. Dans Actes de la Conférence internationale sur l'apprentissage automatique, l'optimisation et la science des données, Grasmere, Royaume-Uni, 22-26 septembre 2023 ; pp. 403-417.
17. Liu, T. ; Wu, Y. ; Ouais.; Cao, L. ; Cao, Y. Algorithme évolutif multi-objectifs clairsemé en deux étapes pour l'optimisation de la sélection de canaux dans les BCI. Devant. Hum. Neurosci. 2024, 18, 1400077. [Réf. croisée] [Pub Med]
18. Soleimani, B. ; Dallasta, moi; Das, P. ; Kulasingham, JP; Girgenti, S. ; Simon, JZ; Babadi, B. ; Marsh, EB Une connectivité fonctionnelle directionnelle altérée est à la base de la récupération cognitive post-AVC. Commune cérébrale. 2023, 5, fcad149. [Référence croisée] [Pub Med]
19. Séguin, C. ; Sporns, O. ; Zalesky, A. Communication en réseau cérébral : concepts, modèles et applications. Nat. Révérend Neurosci. 2023, 24, 557-574. [Référence croisée] [Pub Med]
20. Mang, J. ; Xu, Z. ; Qi, Y. ; Zhang, T. Favoriser le processus cognitif-moteur dans la boucle fermée de la récupération de la fonction motrice post-AVC médiée par le BCI : défis et approches. Devant. Neurorobotique 2023, 17, 1271967. [CrossRef] [Pub Med]
21. Nunez, PL; Srinivasan, R. ; Westdorp, AF; Wijesinghe, RS; Tucker, DM; Silberstein, RB; Cadusch, PJ Cohérence EEG : I : Statistiques, électrode de référence, conduction volumique, Laplaciens, imagerie corticale et interprétation à plusieurs échelles. Électroencéphalogramme. Clin. Neurophysiol. 1997, 103, 499-515. [Référence croisée] [Pub Med]
22. Al-Qazzaz, NK ; Hamid Bin Mohd Ali, S. ; Ahmad, SA; Islam, MS ; Escudero, J. Suppression automatique des artefacts dans l'EEG normal et des individus déments utilisant ICA – WT lors de tâches de mémoire de travail. Capteurs 2017, 17, 1326. [CrossRef] [Pub Med]
23. Al-Qazzaz, NK ; Ali, SHBM ; Ahmad, SA; Islam, MS ; Escudero, J. Discrimination des troubles cognitifs légers liés à un accident vasculaire cérébral et de la démence vasculaire à l'aide de l'analyse du signal EEG. Méd. Biol. Ing. Calculer. 2018, 56, 137-157. [Référence croisée] [Pub Med]
24. Zhang, Y. ; Nam, CS ; Zhou, G. ; Jin, J. ; Wang, X. ; Cichocki, A. Modèles spatiaux de groupes clairsemés contraints temporellement pour le moteur imagerie BCI. IEEETrans. Cybern. 2018, 49, 3322-3332. [Référence croisée] [Pub Med]
25. Chaudhary, S. ; Taran, S. ; Bajaj, V. ; Siuly, S. Une approche flexible basée sur la transformation analytique en ondelettes pour les tâches d'imagerie motrice classification dans les demandes BCI. Calculer. Méthodes Programmes Biomed. 2020, 187, 105325. [Réf. croisée]
26. Duan, L. ; Hongxin, Z. ; Khan, MS ; Fang, M. Reconnaissance des tâches d'imagerie motrice pour BCI utilisant CSP et chaotique PSO twin SVM. J. Université de Chine. Postes Télécommun. 2017, 24, 83-90. [Référence croisée]
27. Alotaiby, T. ; Abd El-Samie, FE ; Alshebeili, SA; Ahmad, I. Une revue des algorithmes de sélection de canaux pour le traitement du signal EEG. EURASIP J.Adv. Processus de signal. 2015, 2015, 66. [Réf. croisée]
28. Lui, L. ; Yu, Z. ; Gu, Z. ; Li, Y. Bhattacharyya, sélection de canaux basée sur les limites pour la classification des images motrices dans les signaux EEG. Dans Actes de la Conférence chinoise sur le contrôle et la décision de 2009, Guilin, Chine, 17-19 juin 2009 ; pages 2353 à 2356.
29. Tam, WK ; Ke, Z. ; Tong, KY Performance d'un modèle spatial commun sous un ensemble plus petit d'électrodes EEG dans l'interface cerveau-ordinateur chez des patients victimes d'un AVC chronique : une étude d'ensemble de données multi-sessions. Dans Actes de la conférence internationale annuelle 2011 de l'IEEE Engineering in Medicine and Biology Society, Boston, MA, États-Unis, 30 août-3 septembre 2011 ; pages 6344 à 6347.
30. Gómez-López, JC ; Escobar, JJ; Diaz, AF; Damas, M. ; Gil-Montoya, F. ; González, J. Stimuler la convergence d'un wrapper basé sur GA pour les problèmes de sélection de fonctionnalités sur des données de grande dimension. Dans Actes de la Genetic and Evolutionary Computation Conference Companion, Boston, MA, États-Unis, 9-13 juillet 2022 ; pp. 431-434.
31. Shen, C. ; Zhang, K. Algorithme d'optimisation Grey Wolf amélioré en deux étapes pour la sélection de caractéristiques lors de la classification de grande dimension. Intelligence complexe. Système. 2022, 8, 2769-2789. [Référence croisée]
32. Tiwari, A. ; Chaturvedi, A. Sélection automatique des canaux EEG pour la classification multiclasse de l'interface cerveau-ordinateur à l'aide algorithme de luciole amélioré multiobjectif. Multimed. Outils Appl. 2023, 82, 5405-5433. [Référence croisée]
33. Millán, JdR; Franzé, M. ; Mouriño, J. ; Cincotti, F. ; Babiloni, F. Caractéristiques EEG pertinentes pour la classification des spontanés tâches liées à la motricité. Biol. Cybern. 2002, 86, 89-95. [Référence croisée] [Pub Med]
34. Martínez-Cagigal, V. ; Santamaría-Vázquez, E. ; Hornero, R. Optimisation de la sélection des canaux de l'interface cerveau-ordinateur à l'aide de méta-heuristiques et d'algorithmes évolutifs. Appl. Informatique douce. 2022, 115, 108176. [Réf. croisée]
35. Gaur, P. ; Gupta, H. ; Chowdhury, A. ; McCreadie, K. ; Pachori, RB; Wang, H. Un modèle spatial commun à fenêtre coulissante pour améliorer la classification de l'imagerie motrice dans l'EEG-BCI. IEEETrans. Instrument. Mesures. 2021, 70, 4002709. [Réf. croisée]
36. Shi, B. ; Wang, Q. ; Yin, S. ; Yue, Z. ; Huai, Y. ; Wang, J. Un algorithme de recherche d'harmonie binaire comme méthode de sélection de canal pour le moteur BCI basé sur l'imagerie. Neuroinformatique 2021, 443, 12-25. [Référence croisée]
37. Ghaemi, A. ; Rashedi, E. ; Pourrahimi, AM; Kamandar, M. ; Rahdari, F. Sélection automatique des canaux dans les signaux EEG pour la classification des mouvements de la main gauche ou droite dans les interfaces cerveau-ordinateur à l'aide d'un algorithme de recherche par gravitation binaire amélioré. Biométhode. Processus de signal . Contrôle 2017, 33, 109-118. [Référence croisée]
38. Lotte, F. ; Congedo, M. ; Lécuyer, A. ; Lamarche, F. ; Arnaldi, B. Une revue des algorithmes de classification pour le cerveau-ordinateur basé sur l'EEG interfaces. J. Ing. neuronal. 2007, 4, R1. [Référence croisée]
39. Hassan, MM; Alam, MGR ; Uddin, MZ; Huda, S. ; Almogren, A. ; Fortino, G. Reconnaissance des émotions humaines à l'aide d'une architecture de réseau de croyances profondes. Inf. Fusion 2019, 51, 10-18. [Référence croisée]

Machine Translated by Google

40. Khosla, A. ; Khandnor, P. ; Chand, T. Une analyse comparative des méthodes de traitement du signal et de classification pour différents applications basées sur les signaux EEG. Biocyberne. Biomède. Ing. 2020, 40, 649-690. [Référence croisée]
41. Altan, G. ; Kutlu, Y. ; Allahverdi, N. Classification de l'activité cérébrale basée sur les réseaux de croyances profondes utilisant l'EEG à partir de potentiels corticaux lents en accident vasculaire cérébral. Int. J.Appl. Mathématiques. Électron. Calculer. 2016, 4, 205-210. [Référence croisée]
42. García-Laencina, PJ; Rodríguez-Bermudez, G. ; Roca-Dorda, J. Explorer la réduction de la dimensionnalité des caractéristiques EEG dans le moteur classification des tâches d'imagerie. Système expert. Appl. 2014, 41, 5285-5295. [Référence croisée]
43. Aler, R. ; Galván, IM Optimisation du nombre d'électrodes et de filtres spatiaux pour les interfaces cerveau-ordinateur au moyen d'un approche évolutive multi-objectifs. Système expert. Appl. 2015, 42, 6215-6223. [Référence croisée]
44. Lal, Tennessee ; Schröder, M. ; Hinterberger, T. ; Weston, J. ; Bogdan, M. ; Birbaumer, N. ; Scholkopf, B. Prise en charge de la sélection de canaux vectoriels dans BCI. IEEETrans. Biomède. Ing. 2004, 51, 1003-1010. [Référence croisée] [Pub Med]
45. Rakotomamonjy, A. ; Guigue, V. Concours BCI III : Ensemble de données II-ensemble de SVM pour l'orthographe BCI P300. IEEETrans. Biomède. Ing. 2008, 55, 1147-1154. [Référence croisée] [Pub Med]
46. Tan, P. ; Wang, X. ; Wang, Y. Réduction de la dimensionnalité dans la sélection de caractéristiques basée sur des algorithmes évolutifs pour l'imagerie motrice interface cerveau-ordinateur. Essaim Evol. Calculer. 2020, 52, 100597. [Réf. croisée]
47. Tsai, CF; Eberle, W. ; Chu, CY Algorithmes génétiques dans la sélection de fonctionnalités et d'instances. Système basé sur les connaissances. 2013, 39, 240-247. [Référence croisée]
48. Udhaya Kumar, S. ; Hannah Inbarani, H. Sélection de caractéristiques basée sur PSO et classification approximative basée sur des ensembles de quartiers pour BCI tâche d'imagerie motrice multiclasse. Calcul neuronal. Appl. 2017, 28, 3239-3258. [Référence croisée]
49. Baig, MZ; Aslam, N. ; Chum, HP ; Zhang, L. Algorithme d'évolution différentielle comme outil pour la sélection optimale de sous-ensembles de fonctionnalités dans imagerie motrice EEG. Système expert. Appl. 2017, 90, 184-195. [Référence croisée]
50. Atyabi, A. ; Luerssen, M. ; Fitzgibbon, S. ; Pouvoirs, sélection des fonctionnalités évolutives DM et réduction des électrodes pour la classification EEG. Dans Actes du Congrès IEEE 2012 sur le calcul évolutif, Brisbane, QLD, Australie, 10-15 juin 2012 ; p. 1 à 8.
51. Nakisa, B. ; Rastgoo, Minnesota ; Tjondronegoro, D. ; Chandran, V. Algorithmes de calcul évolutifs pour la sélection des caractéristiques de Reconnaissance des émotions basée sur l'EEG à l'aide de capteurs mobiles. Système expert. Appl. 2018, 93, 143-155. [Référence croisée]
52. Al-Qazzaz, NK ; Ali, SHM ; Ahmad, SA Algorithme de sélection de canaux basé sur l'évolution différentielle sur le signal EEG pour la détection précoce de la démence vasculaire chez les survivants d'un AVC. Dans Actes de la conférence IEEE-EMBS 2018 sur le génie et les sciences biomédicales (IECBES), Sarawak, Malaisie, 3-6 décembre 2018 ; pp. 239-244.
53. Alyasseri, ZAA; Khader, AT ; Al-Betar, MA ; João, PP ; Osama, AA Authentification de personne basée sur l'EEG utilisant un algorithme de pollinisation des fleurs multi-objectifs. Dans Proceedings of the Evolutionary Computation (CEC), Rio de Janeiro, Brésil, 8-13 juillet 2018 ; p. 15-22.
54. Yang, XS Un nouvel algorithme métaheuristique inspiré des chauves-souris. Dans Stratégies coopératives d'optimisation inspirées par la nature (NICSO 2010); Springer : Berlin/ Heidelberg, Allemagne, 2010 ; pp. 65-74.
55. Hyvarinen, A. Algorithmes à virgule fixe rapides et robustes pour l'analyse de composants indépendants. IEEETrans. Réseau neuronal. 1999, 10, 626-634. [Référence croisée] [Pub Med]
56. Escudero, J. ; Hornero, R. ; Abasolo, D. ; Fernández, A. Séparation aveugle des sources pour améliorer les caractéristiques spectrales et non linéaires des enregistrements magnétoencéphalogrammes. Application à la maladie d'Alzheimer. Méd. Ing. Phys. 2009, 31, 872-879. [Référence croisée] [Pub Med]
57. James, CJ; Hesse, CW Analyse indépendante des composants pour les signaux biomédicaux. Physiol. Mesures. 2005, 26, R15. [Référence croisée] [Pub Med]
58. Al-Qazzaz, NK ; Hamid Bin Mohd Ali, S. ; Ahmad, SA; Islam, MS ; Escudero, J. Sélection de fonctions d'ondelettes mères pour l'analyse du signal EEG multicanal lors d'une tâche de mémoire de travail. Capteurs 2015, 15, 29015-29035. [Référence croisée] [Pub Med]
59. Escudero, J. ; Hornero, R. ; Abasolo, D. ; Fernández, A. Évaluation quantitative de l'élimination des artefacts dans un magnétoencéphalogramme réel signaux avec séparation aveugle des sources. Anne. Biomède. Ing. 2011, 39, 2274-2286. [Référence croisée] [Pub Med]
60. Barbati, G. ; Porcaro, C. ; Zappasodi, F. ; Rossini, PM ; Tecchio, F. Optimisation d'une approche d'analyse de composants indépendante pour l'identification et l'élimination des artefacts dans les signaux magnétoencéphalographiques. Clin. Neurophysiol. 2004, 115, 1220-1232. [Référence croisée]
61. Mammon, N. ; La Foresta, F. ; Morabito, FC Rejet automatique des artefacts de l'EEG multicanal du cuir chevelu par ondelette ICA. IEEE Sens. J. 2011, 12, 533-542. [Référence croisée]

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Classification et régression des corrosions par sténopé sur les pipelines

Basé sur des signaux de fuite de flux magnétique utilisant la convolution Réseaux de neurones

Yufei Shen et Wenxing Zhou*

Département de génie civil et environnemental, Université Western Ontario, London, ON N6A 5B9, Canada * Correspondance : wzhou@eng.uwo.ca

Résumé : Les corrosions par trous d'épingle sur les pipelines de pétrole et de gaz sont difficiles à détecter et à dimensionner et, par conséquent, posent un défi important à la pratique de gestion de l'intégrité des pipelines. Cette étude développe deux modèles de réseaux neuronaux convolutifs (CNN) pour identifier les trous d'épingle et prédire la taille et l'emplacement des corrosions par trous d'épingle en fonction des signaux de fuite de flux magnétique générés à l'aide de l'analyse magnétostatique par éléments finis. De nombreux cas d'analyse paramétrique par éléments finis en trois dimensions sont générés pour entraîner et valider les deux modèles CNN. De plus, une analyse complète des algorithmes évalue les performances du modèle, fournissant ainsi un aperçu de l'application pratique des modèles CNN dans la gestion de l'intégrité des pipelines. Le modèle de classification CNN proposé s'avère très précis dans la classification des trous d'épingle et des défauts de corrosion en général. Le modèle de régression CNN proposé s'avère très précis pour prédire l'emplacement du trou d'épingle et obtient une précision raisonnablement élevée dans l'estimation de la profondeur et du diamètre du trou d'épingle, même en présence de bruits de mesure. Cette étude indique l'efficacité de l'utilisation d'algorithmes d'apprentissage profond pour améliorer les pratiques de gestion de l'intégrité des pipelines corrodés.

Mots-clés : algorithme d'apprentissage profond ; réseau neuronal convolutif ; pipeline; corrosion; sténopé; analyse par éléments finis ; signal de fuite de flux magnétique

check for updates

Citation : Shen, Y. ; Zhou, W. Classification et régression des corrosions par sténopé sur les pipelines sur la base de signaux de fuite de flux magnétique à l'aide de réseaux de neurones convolutifs. Algorithmes 2024, 17, 347. https://doi.org/10.3390/a17080347

Rédacteurs académiques : Chih-Lung Lin, Bor-Jiunn Hwang, Shaou-Gang Miaou et Chi-Hung Chuang

Reçu : 24 juin 2024
Révisé : 1er août 2024
Accepté : 6 août 2024
Publié : 8 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.

Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Les pipelines sont des moyens économiques et sûrs pour transporter et distribuer de grands volumes de produits pétroliers et gaziers sur de grandes distances [1]. Cependant, les pannes de pipelines peuvent entraîner de graves conséquences telles que des décès, des pertes économiques et des dommages environnementaux [2]. Parmi les différents mécanismes de défaillance menaçant l'intégrité structurelle des pipelines, la corrosion constitue la principale menace [3]. L'outil d'inspection en ligne (ILI) est largement utilisé pour détecter et mesurer les corrosions afin d'évaluer l'intégrité des pipelines [4], et les fuites de flux magnétique (MFL) sont la technologie ILI prédominante pour identifier les défauts de corrosion internes et externes dans les pipelines de liquides et de gaz. [5]. Le Pipeline Operators Forum (POF) suggère qu'une seule anomalie de corrosion sur la surface interne ou externe d'un pipeline peut être classée dans l'une des sept classes basées sur la longueur et la largeur de corrosion signalées par l'ILI, comme illustré dans la figure 1 [6], c'est-à-dire , rainurage axial, rainurage circonférentiel, rainurage axial, rainurage circonférentiel, général, sténopé et piqûre, pour mieux quantifier la précision des outils ILI et faciliter la sélection de modèles appropriés pour évaluer la capacité d'éclatement des pipelines contenant des défauts de corrosion [6,7].

L'estimation de la taille des anomalies de corrosion sur la base des signaux MFL relève du domaine de la modélisation inverse. L'industrie des pipelines utilise fréquemment le modèle inverse itératif en raison de sa précision remarquable [8-10]. Cependant, le modèle inverse itératif est coûteux en termes de calcul. Les modèles inverses basés sur l'apprentissage automatique offrent des alternatives prometteuses au dimensionnement des défauts de corrosion à partir des signaux MFL avec une grande précision, mais sont plus efficaces en termes de calcul que les modèles inverses itératifs [11-13]. Par

Machine Translated by Google

Kandroodi et coll. [14] ont introduit la largeur et la longueur du défaut estimées par l'algorithme, ainsi que les valeurs crête à crête extraites des signaux MFL, dans un réseau neuronal à fonction de base radiale gaussienne pour prédire la profondeur de la corrosion. Feng et coll. [15] ont proposé une méthodologie d'ajustement d'erreur pour reconstruire les profils de corrosion à partir de signaux MFL basée sur un réseau neuronal à fonction de base radiale. Cependant, ces modèles présentent certaines limites : la formation de ces modèles repose sur des caractéristiques choisies manuellement à partir des signaux MFL, omettant potentiellement des données précieuses nécessaires pour prédire avec précision le profil de corrosi De plus, la majorité de ces techniques se concentrent uniquement sur une dimension du signal MFL, négligeant les informations critiques présentes dans les deux autres dimensions.

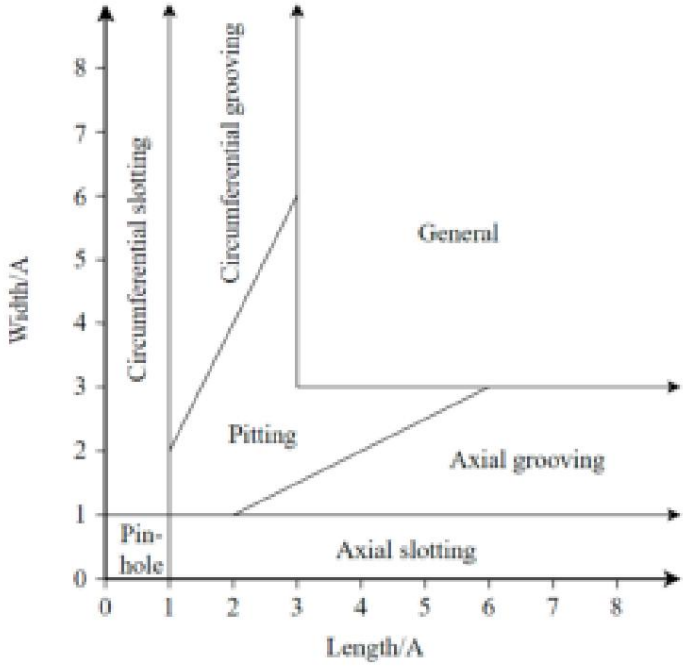


Figure 1. Sept catégories d'anomalies de corrosion sont basées sur la longueur et la largeur de l'anomalie. Remarque : A = max{10 mm, épaisseur de paroi du tuyau}.

Au cours de la dernière décennie, les réseaux de neurones convolutifs (CNN) sont devenus un algorithme d'apprentissage profond révolutionnaire, donnant des résultats remarquables dans divers domaines de la reconnaissance de formes, notamment le traitement de l'image et de la voix [16,17]. L'outil CNN peut extraire automatiquement des fonctionnalités approfondies des données d'entrée à la place des extracteurs de fonctionnalités conçus manuellement ; par conséquent, il est bien adapté pour résoudre la modélisation inverse basée sur les signaux MFL. L'application de CNN dans l'analyse inverse des signaux MFL a été rapportée dans la littérature. Lu et coll. [18] ont introduit la transformation visuelle CNN, qui prédit la longueur, la largeur et la profondeur des anomalies en prenant en compte deux composantes (circonférentielle et radiale) du signal MFL tout en ignorant la composante longitudinale. Shen et Zhou [19] ont utilisé CNN pour estimer les emplacements et les dimensions des corrosions sur les pipelines en acier sur la base de trois composantes des signaux MFL. Wang et Chen [20] ont proposé une architecture CNN comprenant deux modules principaux : un pour la classification des anomalies et un autre pour la régression de la taille des anomalies. L'entrée du module de classification comprend les trois composantes du signal MFL et les sorties sont sept catégories d'anomalies, comme illustré dans la figure 1. Le module de régression intègre sept CNN distincts, chacun adapté à un type d'anomalie spécifique, prédisant des dimensions telles que largeur, longueur et profondeur du défaut.

Selon un rapport du European Gas Pipeline Incident Data Group (EGIG) [21], la corrosion par sténopé est identifiée comme un contributeur majeur aux défaillances des pipelines en raison de son potentiel d'évolution d'un petit volume initial de perte de métal à une perte de métal plus grave, compte tenu du défi de la détection via la surveillance de la pression, en particulier lorsque les produits qui fuient ne dépassent pas le niveau du sol [22,23]. Malgré l'impact des piqûres sur l'intégrité des pipelines, des étud

Machine Translated by Google

qui sont spécifiquement axés sur la détection et le dimensionnement des corrosions par trous d'épingle sont limités dans la littérature, créant ainsi un manque de connaissances. De plus, de petites anomalies de corrosion telles que des trous d'épingle peuvent parfois se chevaucher avec une zone de corrosion relativement grande [23-25], créant un défaut complexe connu dans cette étude sous le nom de défaut de corrosion par trou d'épingle en général (PIC). Il existe également peu d'études dans la littérature portant sur la détection et le dimensionnement des défauts PIC.

L'objectif de cette étude est de développer des algorithmes d'apprentissage profond, c'est-à-dire des modèles CNN, pour classer et dimensionner les trous d'épingle et les défauts PIC sur la base d'un large ensemble de signaux MFL. Le modèle de classification CNN traite six scénarios, à savoir les trous d'épingle sur la surface interne ou externe du tuyau, les corrosions générales sur la surface interne ou externe et les défauts PIC sur la surface interne ou externe. Les signaux pour les trous d'épingle et les défauts PIC identifiés par le modèle de classification sont ensuite introduits dans le modèle de régression CNN pour prédire les tailles et l'emplacement des défauts de trou d'épingle et PIC. En pratique, il n'est pas nécessaire d'appliquer les deux modèles CNN séquentiellement. En fonction de l'application spécifique, les données peuvent être introduites directement dans le modèle de régression si la classification a déjà été effectuée par d'autres moyens et si l'objectif principal est de prédire la taille du défaut. Étant donné que les signaux MFL réels sont généralement propriétaires et non accessibles au public, les signaux MFL utilisés dans cette étude sont générés numériquement en appliquant les analyses magnétostatiques par éléments finis (FEA) tridimensionnelles (3D) à l'aide du logiciel commercial COMSOL (version 5.6). Les cartes de signaux MFL tridimensionnelles (c'est-à-dire longitudinales, circonférentielles et radiales) servent d'entrées pour la formation et la validation des modèles CNN afin de classer les défauts et de prédire les tailles, ainsi que les positions circonférentielles et longitudinales, des trous d'épingle sur le pipeline. L'impact des bruits de mesure associés au signal MFL sur la précision du modèle de régression CNN est également étudié. Cette recherche comble un manque de connaissances dans la littérature en appliquant le CNN pour détecter, localiser et dimensionner les trous d'épingle et les défauts de corrosion. Il est à noter qu'il est crucial d'identifier avec précision la position du trou d'épingle dans la corrosion générale, car la position relative du trou d'épingle dans la corrosion générale a un effet marqué sur la capacité d'éclatement du défaut PIC et a donc de fortes implications sur la précision. prédiction de la capacité d'éclatement. Cette étude démontre l'efficacité du CNN pour la détection et le dimensionnement de la corrosion par trous d'épingle et la viabilité du CNN pour les applications dans les pratiques de gestion de l'intégrité des pipelines.

Le reste du document est structuré comme suit. La section 2 présente les principes de la technologie MFL et les informations générales essentielles sur la relation entre les signaux MFL et les caractéristiques des anomalies. La section 3 présente le modèle FEA proposé et les paramètres de simulation des signaux MFL. La section 4 décrit la structure des modèles CNN développés et leur application à la classification et à la régression des trous d'épingle. La précision prédictive correspondante des modèles CNN est également discutée dans la section 4. La section 5 résume les remarques finales de l'étude.

2. Principes de la technique MFL

La FEA magnétostatique est utilisée pour générer numériquement les signaux MFL dans l' étude. L'équation directrice correspondante dérivée des équations de Maxwell est présentée dans l'équation (1) :

$$\times \quad \perp \quad \times A = 0\ \mu \tag{1}$$

où × désigne l'opération de curl ; μ représente la perméabilité magnétique de l'échantillon ferromagnétique (N/A2), et A est le vecteur potentiel magnétique. L'équation (1) est résolue dans les conditions aux limites spécifiées pour obtenir A, qui est ensuite utilisée pour obtenir la densité de flux magnétique B comme suit [26,27] :

$$B = \quad \times UNE \tag{2}$$

La technique MFL repose sur le principe de mesure selon lequel un matériau ferromagnétique démontre une MFL au voisinage d'un défaut, comme illustré sur la figure 2 [18]. La procédure de mesure est la suivante : lors du processus ILI, l'outil MFL est propulsé

Machine Translated by Google

à travers le pipeline et les aimants à l'intérieur de l'outil génèrent un champ magnétique. Dans un tuyau parfait, tout le flux magnétique reste à l'intérieur de la paroi du tuyau [28], ce qui entraîne une distribution uniforme du champ magnétique [29]. Cependant, en présence de défauts, une partie du flux magnétique « s'échappe » à l'extérieur du tuyau, formant un champ de fuite magnétique proche du défaut [30]. Ce champ de fuite est détecté par des capteurs Hall, générant des signaux électriques [29]. L'intensité et la répartition des fuites de flux varient en fonction de la géométrie du défaut sur le pipeline. En capturant le signal MFL, qui consiste généralement en un champ vectoriel de densité de flux magnétique (B) tridimensionnel , représenté par Bx dans la direction circonférentielle du tuyau, By dans la direction radiale et Bz dans la direction longitudinale (axiale), il est possible d'estimer les caractéristiques de taille et de localisation du défaut de corrosion.

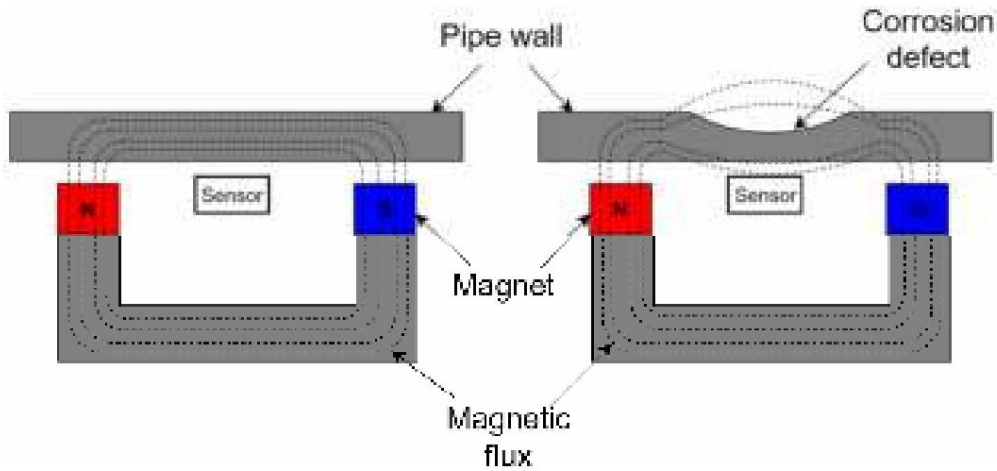


Figure 2. Principe de la technique MFL.

Les outils MFL axiaux (AMFL) et circonférentiels (CMFL) sont les deux types d'outils MFL classés par la disposition des aimants à l'intérieur de l'outil MFL. Dans les outils AMFL, les aimants permanents sont orientés parallèlement à l'axe longitudinal du tuyau. Ces outils excellent dans la détection et la mesure des défauts dans la direction circonférentielle du tuyau mais sont moins efficaces dans le dimensionnement des caractéristiques orientées axialement [31]. À l'inverse, les outils CMFL possèdent des aimants positionnés dans la direction circonférentielle du pipeline. Ils sont plus précis dans la mesure des caractéristiques alignées dans la direction longitudinale par rapport aux o L'outil CMFL est au centre de cette étude car il est plus précis que l' outil AMFL pour dimensionner les anomalies de corrosion orientées longitudinalement, qui sont perpendiculaires à la contrainte circonférentielle résultant de la pression interne de la canalisation et peuvent donc avoir une grande influence sur la capacité d'éclatement du pipeline.

3. Simulation des signaux MFL à l'aide de FEA

3.1. Paramètres de simulation

L'outil CMFL considéré dans cette étude comprend trois paires d'aimants permanents. NdFeB est sélectionné comme matériau magnétique avec une force coercitive élevée de 895 000 A/m, des propriétés magnétiques stables et une énergie magnétique élevée [32]. L'acier utilisé pour simuler le tuyau est supposé être de nuance X52, caractérisé par une densité de flux rémanent (Br) de 1T, une conductivité électrique de 5,882 × 106 S/m, une intensité de champ coercitif (Hc) de 415 A/m et densité de flux de saturation (Bs) de 1,9T. La figure 3 montre les schémas d'une coupe transversale de l' outil CMFL et de la section de tuyau, et les propriétés géométriques et matérielles de chaque composant sont résumées dans le tableau 1. Il est à noter que la valeur de décollement est supposée être de 5 mm, ce qui est la distance entre le capteur et la surface interne du tuyau [33].

Machine Translated by Google

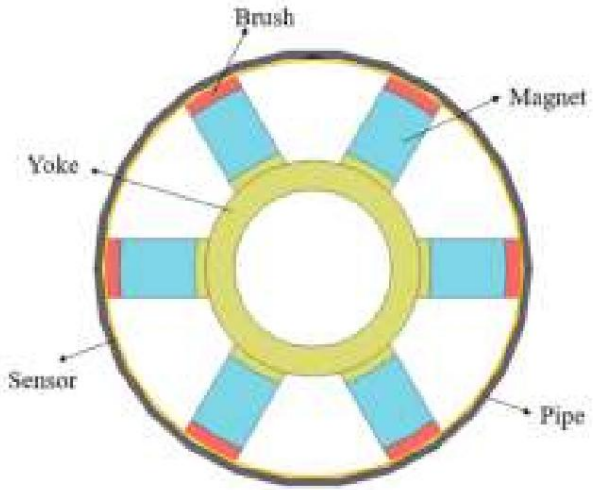


Figure 3. Illustration de l'outil CMFL.

Tableau 1. Géométrie de chaque composant dans le modèle CMFL 3D proposé.

Éléments	Section de tuyau	Aimants	Pinceaux	Joug
géométrie (mm)	600 × 610 × 10 [a]	Longueur ×	b 300 × 81 × 105	81 × 20	b 300 × 300 × 40 300 × [un]

[a] diamètre extérieur × épaisseur de paroi. [b] Longueur × largeur × épaisseur.

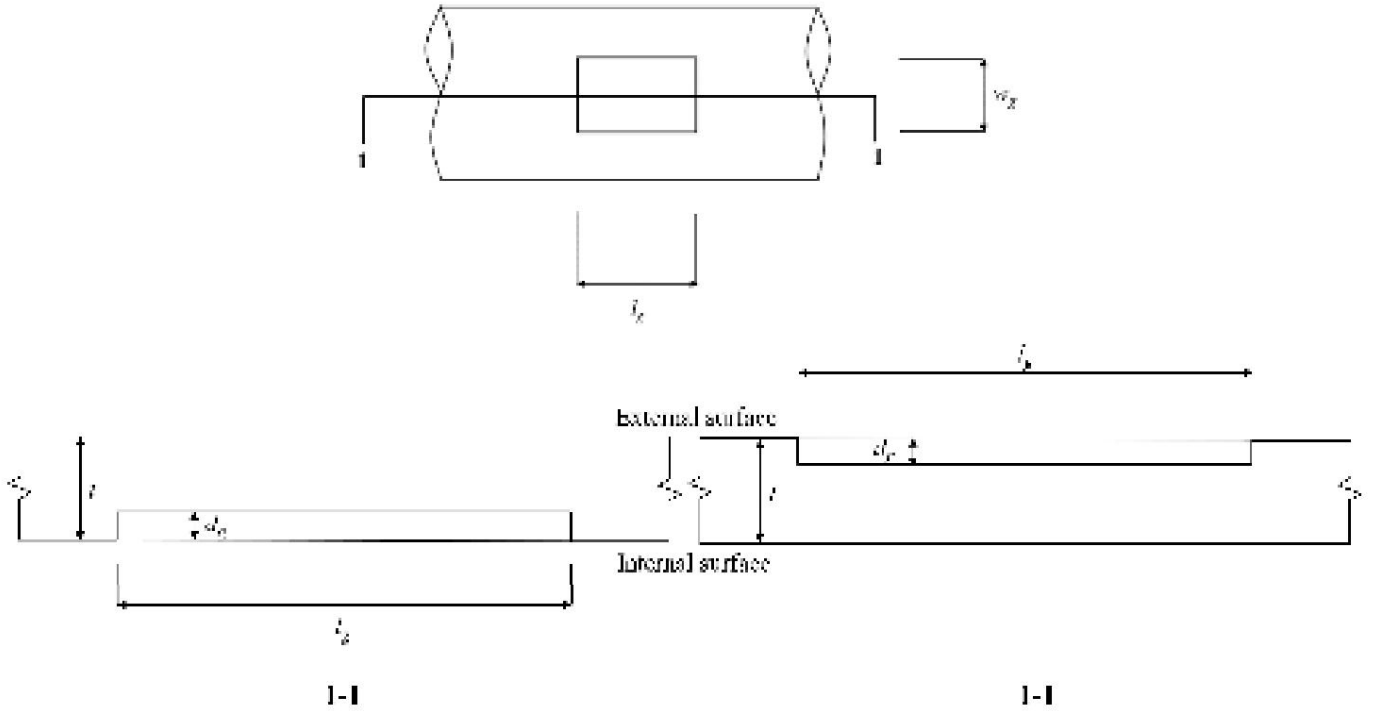
3.2. Cas FEA

Une série de modèles FEA sont créés et analysés avec les mêmes conditions de magnétisation et attributs de canalisation, mais ils varient en termes de types de corrosion, de paramètres géométriques et emplacements dans cette étude. La simulation porte sur 9600 cas d'analyse, avec 1600 cas alloués à chacun des six scénarios comme illustré sur la figure 4, à savoir la corrosion générale interne. (Gin), corrosion générale externe (Gex), sténopé interne (Pin), sténopé externe (Pex), interne Défaut PIC (PICin) et défaut PIC externe (PICex). La corrosion générale est idéalisée comme de forme cuboïde, définie par la largeur (wg), la profondeur (dg) et la longueur (lg), et le trou d'épingle est idéalisé comme une forme cylindrique, définie par le rayon (rp) et la profondeur (dp). Notez le centre du trou d'épingle dans un défaut PIC est supposé coïncider avec celui de la corrosion générale, qui est défini par deux coordonnées, c'est-à-dire son emplacement longitudinal (h) et circonférentiel emplacement (φ), dans une zone désignée de 160 × 160 mm, comme le montre la figure 5. Cette zone est marqué en bleu clair, avec son centre spécifié par les coordonnées (r, = 0, h = 0), où r représente le rayon intérieur du tuyau. Cette zone carrée spécifique de 160 × 160 mm sur le φ La paroi intérieure du tuyau a été sélectionnée car elle englobe la plus grande surface, montrant une profil de densité de flux magnétique entre les aimants voisins.

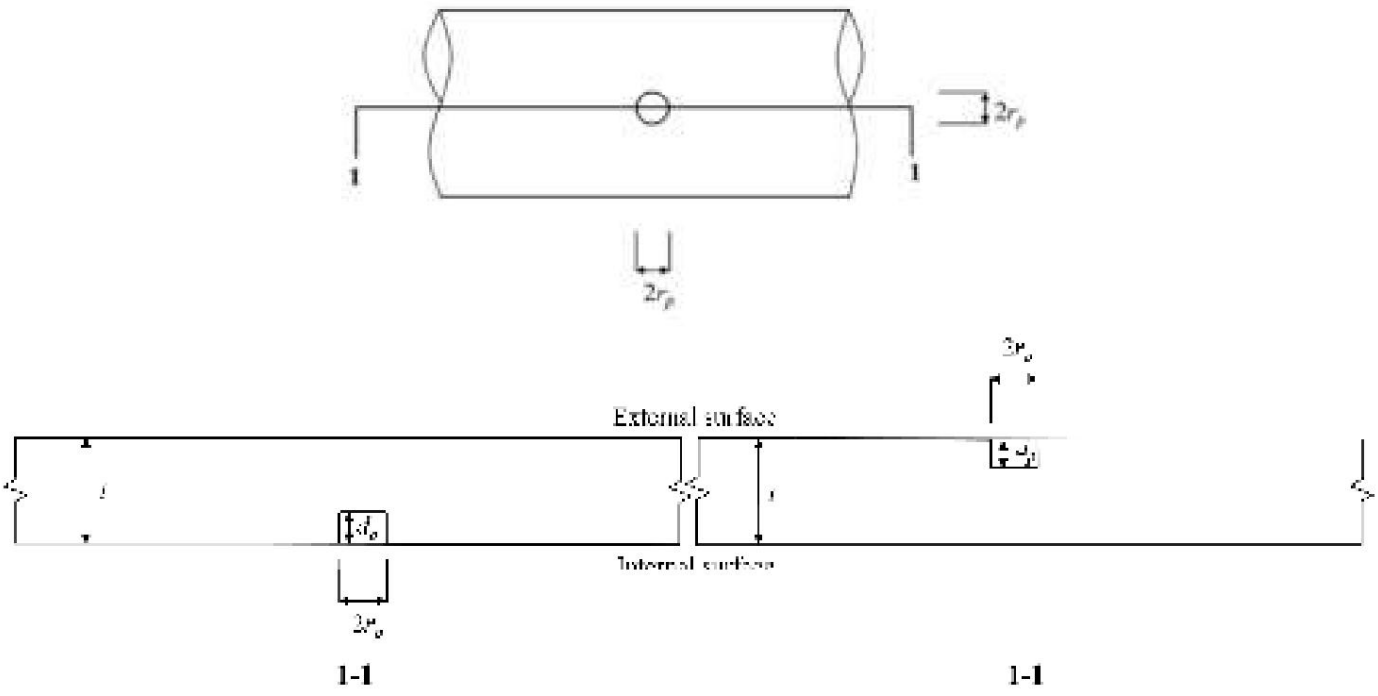
Les cas d'analyse portant uniquement sur la corrosion générale, c'est-à-dire Gin et Gex, comprennent quatre valeurs de wg, c'est-à-dire 30, 50, 70 et 90 mm, quatre valeurs de dg/t, c'est-à-dire 20, 40, 60 et 80 %, où t désigne l'épaisseur de la paroi du tuyau et quatre valeurs de lg, c'est-à-dire 30, 50, 70 et 90 mm. Cela résulte dans 64 corrosions générales au total avec des tailles variables. En termes de localisation de la corrosion, cinq valeurs de h (−50, −25, 0, 25 et 50 mm) et cinq valeurs de φ (−10, −5, 0, 5 et 10 degrés) sont considérés. Les permutations de 64 défauts et 25 emplacements conduisent à 1600 cas pour chacun de Gin et Gex.

Les cas d'analyse impliquant uniquement la corrosion par piqûre, c'est-à-dire Pin et Pex, comprennent huit valeurs de rp, c'est-à-dire 1, 1,5, 2, 2,5, 3, 3,5, 4 et 4,5 mm, et huit valeurs de dp/t, c'est-à-dire 10, 20, 30, 40, 50, 60, 70 et 80 %. Il en résulte au total 64 corrosions par trous d'épingle de différentes tailles. En termes de localisation de la corrosion, cinq valeurs de h (−50, −25, 0, 25 et 50 mm) et cinq les valeurs de φ (−10, −5, 0, 5 et 10 degrés) sont prises en compte. La permutation de 64 défauts et 25 emplacements conduisent également à 1 600 caisses pour chacun des Pin et Pex.

Machine Translated by Google



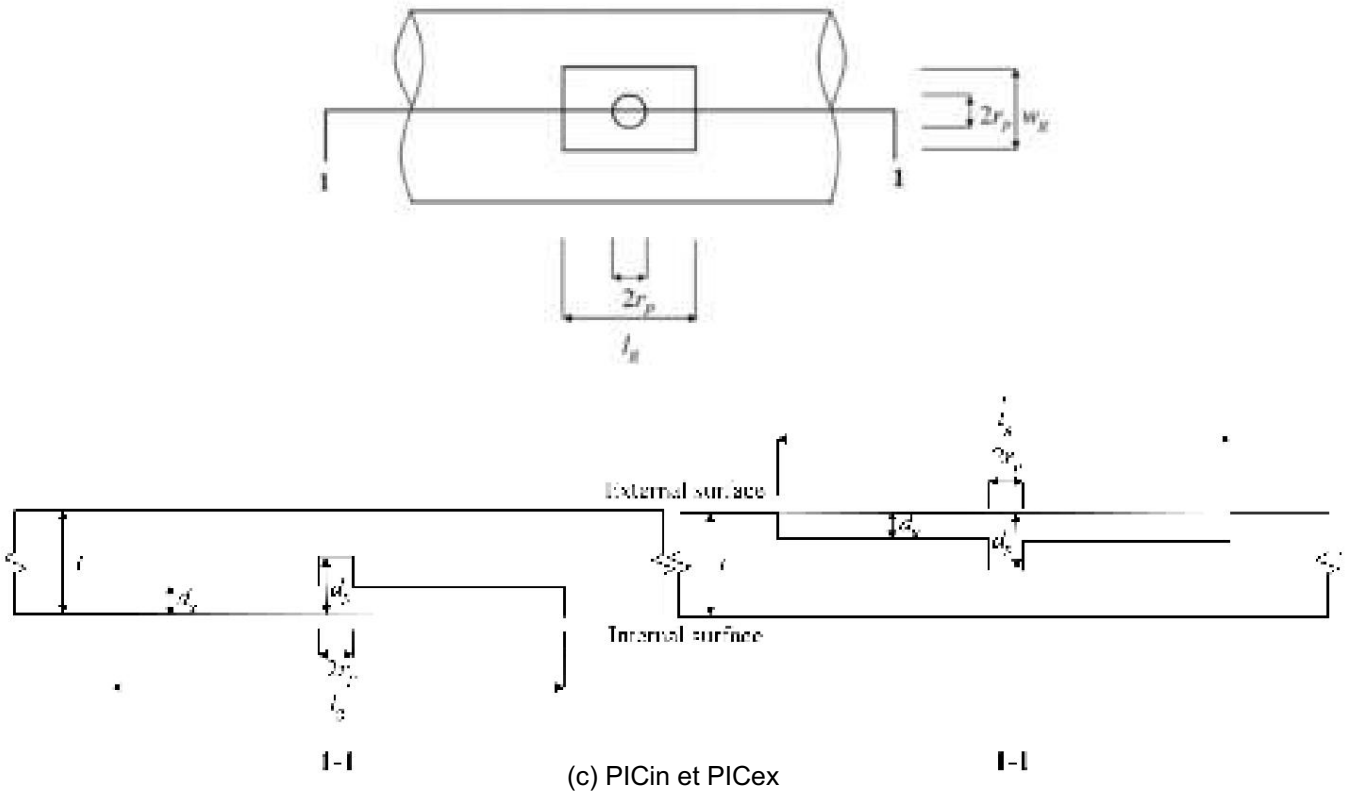
(a) Gin et Gex



(b) Broche et Pex

Figure 4. Suite.

Machine Translated by Google



(c) PICin et PICex

Figure 4. Six types de situations de corrosion : (a) Gin et Gex ; (b) Épingle et Pex ; (c) PICin et PICex.

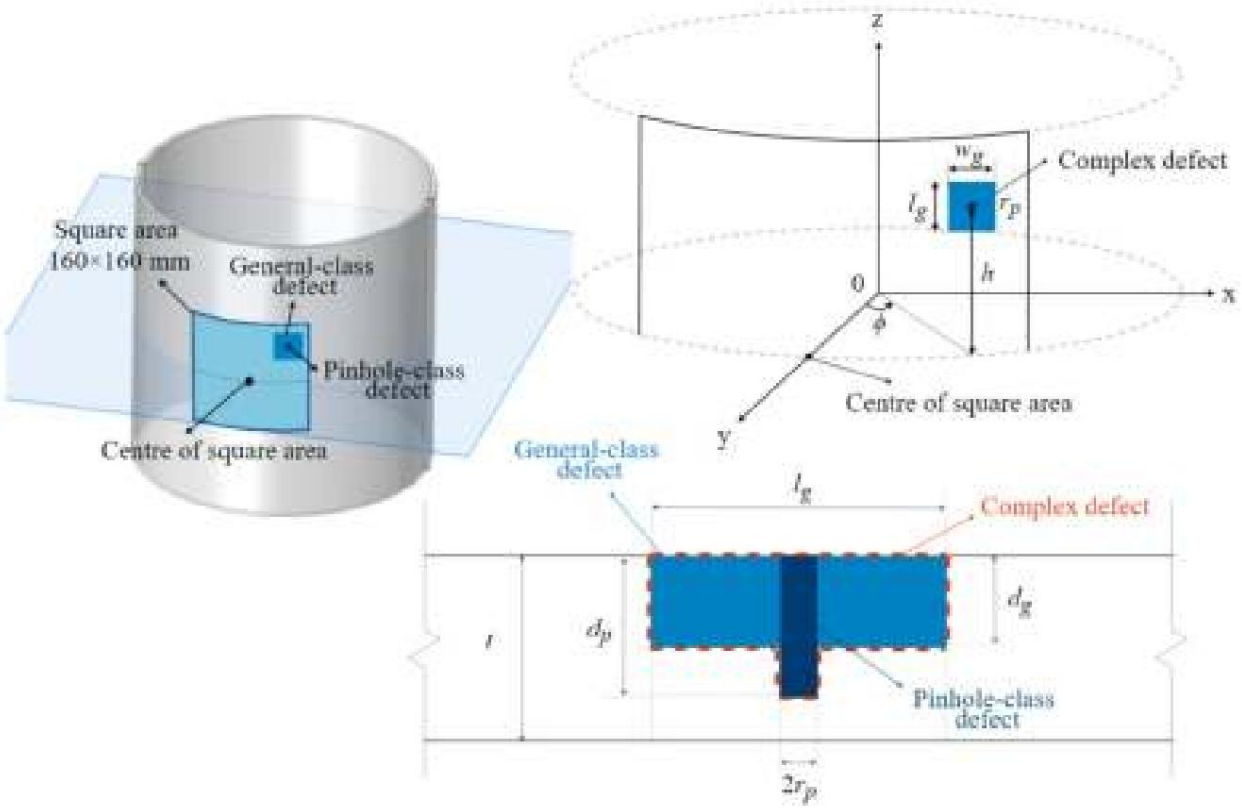


Figure 5. Système de coordonnées cylindrique définissant les paramètres de localisation h et φ.

Machine Translated by Google

Les cas d'analyse impliquant des défauts PIC, c'est-à-dire PICin et PICex, incluent wg = lg = 50 mm et quatre valeurs de rp, c'est-à-dire 1, 2, 3 et 4 mm. En termes d'ampleur des corrosions générales et des corrosions par trous d'épingle, trois groupes sont considérés pour capturer un plus large éventail de relations relatives entre les corrosions générales et les corrosions par trous d'épingle. Le premier comprend dg/t égal à 20 % et dp/t égal à 30, 40, 50, 60, 70 ou 80 %. Le deuxième groupe comprend dg/t égal à 30 % et dp/t égal à 40, 50, 55, 60, 70 ou 80 %. Le dernier groupe comprend dg/t égal à 40 % et dp/t égal à 50, 60, 70 ou 80 %. Il convient de noter que dp/t est supposé être au moins 10 % supérieur à dg/t dans les trois groupes. Cela donne au total 64 défauts PIC de tailles variables. En termes de localisation de la corrosion, cinq valeurs de h (−50, −25, 0, 25 et 50 mm) et cinq valeurs de φ (−10, −5, 0, 5 et 10 degrés) sont prises en compte. La permutation de 64 défauts et 25 emplacements conduit à 1 600 cas pour chacun des PICin et PICex. Il s'ensuit qu'un ensemble de données composé de 9 600 cas (1 600 × 6 = 9 600) est généré dans cette étude. Ce vaste ensemble de données est destiné à faciliter la formation et la validation du modèle CNN, permettant une analyse complète et un développement de modèle robuste.

Le modèle FEA développé maille à l'aide d'éléments tétraédriques à 4 nœuds, où la taille minimale est de 4 mm et la taille maximale du maillage est de 55 mm après une étude de convergence. Plus précisément, les grilles situées à l'intérieur de la zone de 160 mm × 160 mm sont finement résolues à intervalles de 1 mm. A chaque point de la grille, la valeur de B est calculée par interpolation entre deux nœuds du maillage. Le système d'équations non linéaires est résolu à l'aide de la méthode de Newton-Raphson. À titre d'illustration, la figure 6 représente les informations de densité de flux magnétique 3D (c'est-à-dire Bx, By et Bz) dans la zone de 160 mm × 160 mm d'un modèle FE sélectionné. Le modèle FEA proposé a été validé en comparant les résultats de simulation avec les données expérimentales d'un outil CMFL rapporté par Ireland et Torres [34] et d'un outil AMFL rapporté par Azizzadeh et Safizadeh [35]. Les lecteurs sont renvoyés à Shen et Zhou [19] pour plus de détails sur la validation.

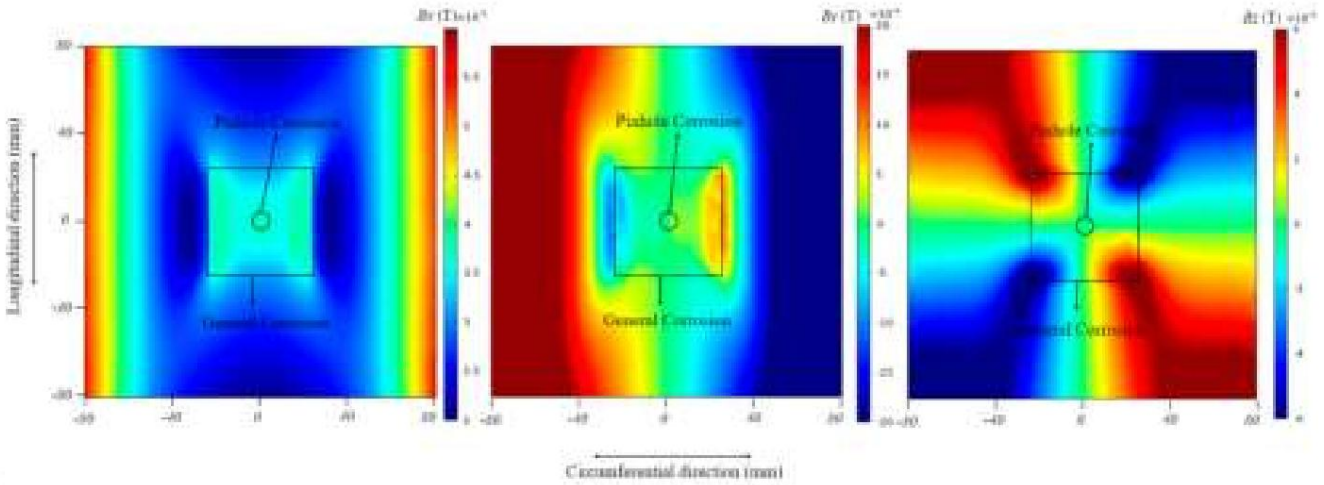


Figure 6. Bx, By et Bz obtenus par FEA correspondant à un défaut PIC sur la surface interne du tuyau (wg = 50 mm, lg = 50 mm, dg/t = 40 %, rp = 4 mm, dp/t = 80 %, h = 0 mm et φ = 0 degré).

Les signaux MFL dans les applications pratiques incluent inévitablement du bruit provenant de facteurs tels qu'une épaisseur de paroi de tuyau non uniforme et des différences dans les valeurs de décollage des capteurs [36]. Ces bruits peuvent avoir un impact sur la précision du modèle inverse, nécessitant l'inclusion de bruits dans les signaux MFL simulés par FEA. Le rapport signal/bruit (SNR) constitue un paramètre important pour quantifier le bruit dans le signal MFL tridimensionnel. L'équation (3) définit le SNR dans cette étude [37] :

$$SNR = 10\log 10 \frac{\sum_{n=1}^{160}\sum_{m=1}^{160} Bx(n,m)^2 + Par(n,m)^2 + Bz(n,m)^2}{\sum_{n=1}^{160}\sum_{m=1}^{160} wx(n,m)^2 + Wy(n,m)^2 + wz(n,m)^2} \quad (3)$$

Machine Translated by Google

où wx, wy et wz désignent les bruits dans trois directions du tuyau, à savoir respectivement les directions circonférentielles, radiales et longitudinales, et (n, m) (où n, m = 1, 2, ... , 160) représente la coordonnée de chaque grille (avec une résolution de 1 mm) dans la zone 160 mm × 160 mm, avec un total de 25 600 grilles. En considérant une valeur représentative de SNR égale à 20 [38], trois matrices 160 × 160 de bruits blancs à distribution gaussienne, supposées indépendantes les unes des autres, sont simulées puis appliquées respectivement à Bx, By et Bz. .

Chaque matrice suppose une valeur moyenne de zéro, avec des écarts types ajustés par essais et erreurs jusqu'à atteindre le rapport signal/bruit prédéfini. Soit σn l'écart type de wx ; les écarts types de wy et wz sont supposés être respectivement de 0,75σn et 0,25σn . La mise à l'échelle de σn est basée sur les grandeurs relatives de Bx, By et Bz dérivées de l'analyse par éléments finis : pour un cas d'analyse donné, Bx a systématiquement la grandeur la plus élevée, suivie par By et Bz, respectivement. L'entrée bruyante finale du modèle CNN est constituée des trois matrices Bx, By et Bz, chacune combinée avec les matrices bruyantes respectives wx, wy et wz, ce qui donne lieu à des représentations bruitées des signaux MFL.

Il est souligné que les matrices de signaux MFL bruitées ne sont introduites dans le modèle de régression CNN que dans le cadre d'une partie étendue de la section 4.2.4. Fan et coll. [39] ont rapporté que l'effet du bruit gaussien sur la précision du modèle de classification CNN est mineur. Acharya et coll. [40] ont indiqué que la suppression du bruit n'est pas nécessaire pour la classification des images avec des algorithmes d'apprentissage profond [41]. Par conséquent, nous considérons uniquement l'impact des bruits sur la précision du modèle de régression dans la présente étude.

4. Réseau neuronal convolutif 4.1.

Classement CNN 4.1.1.

Informations d'entrée L'entrée

d'un modèle CNN est généralement représentée sous la forme d'une image couleur RVB, comprenant trois canaux de couleur correspondant respectivement au vert, au rouge et au bleu. Cette entrée est compatible avec les signaux MFL tridimensionnels. Pour un modèle FE donné, les informations de densité de flux magnétique 3D (c'est-à-dire Bx, By et Bz) dans la zone carrée désignée sont exportées à partir du logiciel COMSOL sur les grilles de 160 mm × 160 mm avec une résolution de 1 mm.

Les trois matrices 160 × 160 générées de signaux MFL sans bruit (une matrice pour chaque dimension) sont désignées comme l'entrée du modèle CNN proposé, et les six types de corrosion (c'est-à-dire Gin, Gex, Pin, Pex, PICin, et PICex) sont la sortie.

4.1.2. Structure proposée Le

modèle de classification CNN proposé dans cette étude comprend un total de 20 couches : les couches 1 à 12 sont dédiées à l'extraction de caractéristiques, tandis que les couches 13 à 20 sont axées sur la classification. La couche 20 sert de couche de sortie, comprenant six composants qui représentent les six types de corrosion. Les informations détaillées sur chaque couche sont résumées dans le tabl
Comme le montre le tableau 2, la taille du filtre dans la couche de convolution est sélectionnée pour être (5 × 5), ce qui est une taille typique pour les grands filtres [42]. La taille du filtre dans la couche maxpooling est (2 × 2), cohérente avec celles couramment utilisées dans la littérature existante [43,44]. Plusieurs couches d'abandon qui suppriment aléatoirement un certain pourcentage de neurones du réseau neuronal pendant le processus de formation sont incluses dans le modèle pour éviter le surajustement. Le taux d'abandon employé est de 0,2, c'est-à-dire que 20 % des neurones du réseau neuronal sont désactivés de manière aléatoire. La figure 7 représente l'architecture du modèle de classification proposé avec la dimension de chaque couche, où les couches d'entrée, de convolution et de pooling maximum sont représentées respectivement en rose, bleu et vert. Notez que les couches de suppression ne sont pas illustrées dans la figure 7 car elles n'ont pas d'impact sur la dimensionnalité des données. La ligne blanche épaisse représente la couche aplatie, les couches entièrement connectées sont représentées en orange et les couches de sortie sont représentées en jaune dans la figure 7.

Machine Translated by Google

Tableau 2. Informations sur la couche du modèle de classification CNN.

Partie	Numéro de couche	Nom du calque	Paramètres
Extraction de fonctionnalités	1	Convolution	64 filtres de taille égale à (5 × 5)
	2	Pooling maximum	La taille du pooling est égale à (2 × 2)
	3	Abandonner	Taux = 0,2
	4	Convolution	128 filtres de taille égale à (5 × 5)
	5	Pooling maximum	La taille du pooling est égale à (2 × 2)
	6	Abandonner	Taux = 0,2
	7	Convolution	256 filtres de taille égale à (5 × 5)
	8	Pooling maximum	La taille du pooling est égale à (2 × 2)
	9	Abandonner	Taux = 0,2
	10	Convolution	512 filtres de taille égale à (5 × 5)
	11	Pooling maximum	La taille du pooling est égale à (2 × 2)
	12	Abandonner	Taux = 0,2
Classification	13	Aplatir le calque	Le numéro d'unité est égal à 51 200
	14	Couche entièrement connectée	Le numéro d'unité est égal à 128
	15	Abandonner	Taux = 0,2
	16	Couche entièrement connectée	Le numéro d'unité est égal à 64
	17	Abandonner	Taux = 0,2
	18	Couche entièrement connectée	Le numéro d'unité est égal à 32
	19	Abandonner	Taux = 0,2
	20	Sortie 1	Le numéro d'unité est égal à 1
		Sortie 2	Le numéro d'unité est égal à 1
		Sortie 3	Le numéro d'unité est égal à 1
		Sortie 4	Le numéro d'unité est égal à 1
		Sortie 5	Le numéro d'unité est égal à 1
		Sortie 6	Le numéro d'unité est égal à 1

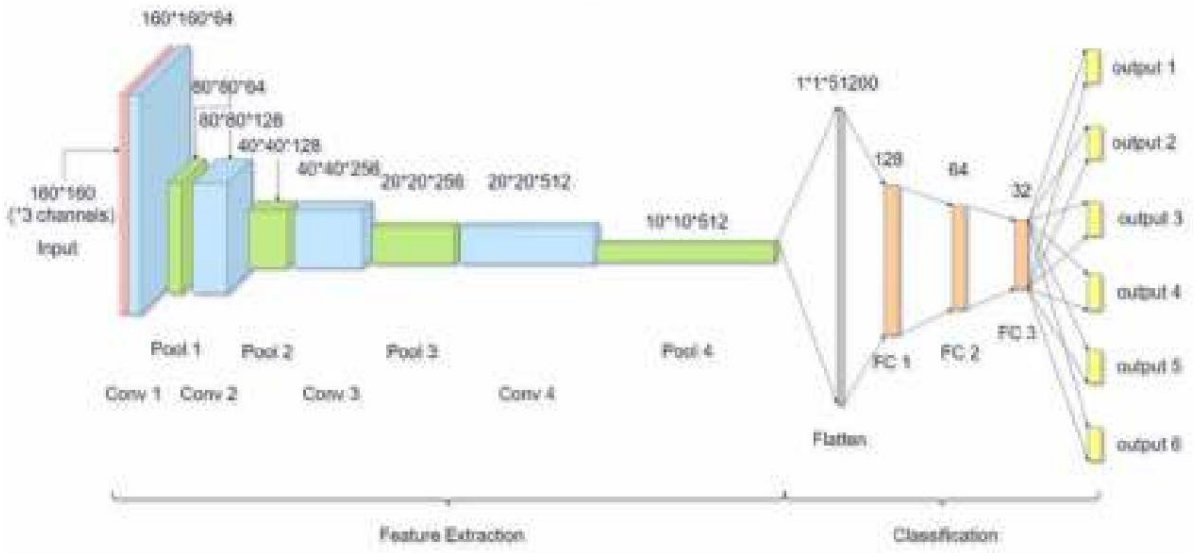


Figure 7. Structure du modèle de classification CNN proposé.

Machine Translated by Google

4.1.3. Résultats

En adoptant un ratio données de test de formation de 80 à 20 %, il existe 1 280 cas de formation et 320 cas de tests pour chacun des six types de corrosion. Les résultats de classification obtenus à partir du Le modèle CNN proposé est résumé dans le tableau 3. Le tableau 3 démontre que le modèle CNN proposé Le modèle atteint systématiquement des précisions de prédiction supérieures à 0,93 dans les deux formations et les ensembles de données de test, avec la plus faible précision de 0,934 observée dans l'ensemble de données d'entraînement pour Gex. Il est à noter que le modèle fonctionne mieux dans l'ensemble de données de test que dans l'ensemble de données d'entraînement, avec des taux de précision de chaque type de corrosion dans l'ensemble de données de test supérieurs ou égaux au taux dans l'ensemble de données de formation. Cela peut être dû aux techniques de régularisation, telles que abandon, adopté dans le modèle CNN. Bien que ces techniques puissent empêcher le surapprentissage, ils peuvent également introduire du bruit ou de l'incertitude pendant l'entraînement, ce qui pourrait temporairement dégrader les performances du modèle sur les données d'entraînement tout en améliorant la généralisation et les performances sur les données de test [45].

Tableau 3. Résultats du modèle de classification CNN.

Catégorie N°	Catégorie	Total Train		Test	Erreur de prédiction dans le train	Erreur de prédiction en test	Précision du test de précision du train	
0	Gin	1600	1280	320	7	0	0,978	1.000
1	Gex	1600	1280	320	21	0	0,934	1.000
2	Épingle	1600	1280	320	0	0	1.000	1.000
3	PEX	1600	1280	320	2	0	0,994	1.000
4	PIcin	1600	1280	320	1	0	0,997	1.000
5	PICex	1600	1280	320	1	1	0,997	0,997

Notamment, 32 cas dans l'ensemble de données de formation et un cas dans l'ensemble de données de test ont été mal classé, ce qui entraîne un taux de prédiction erronée de 0,34 % (33/9 600 = 0,34 %), démontrant l'excellente précision du modèle développé. La figure 8 illustre la distribution de erreurs de prédiction dans l'ensemble de données d'entraînement dans différentes catégories. Par exemple, 65,6 % (c.-à-d. 21) des cas de formation mal classés appartiennent à la catégorie Gex . Parmi ceux-ci, quatre cas sont classés à tort comme Pin, et 17 cas sont classés à tort comme Pex. Par ailleurs, six des 33 cas ont des emplacements mal classés (surface interne/externe de la canalisation), soit deux Pex cas mal classés comme Pin, et quatre cas Gex mal classés comme Pin.

Notez que le modèle atteint une meilleure précision de classification pour Pin, Pex, PICin et PICex. que pour le Gin et le Gex. Les valeurs de dg/t incluses dans l'ensemble des données pour les six types de les corrosions sont de 20, 30, 40, 60 et 80 %, tandis que les valeurs de dp/t incluses dans l'ensemble des données sont 20, 30, 40, 50, 55, 60, 70 et 80 %. La plage plus large de profondeur de corrosion pour les boîtiers impliquant des trous d'épingle (c'est-à-dire Pin, Pex, PICin et PICex) par rapport aux cas impliquant des la corrosion (Gin et Gex) permet au modèle d'apprendre un ensemble plus riche de caractéristiques et de variations associé à la corrosion par sténopé, ce qui rend le modèle plus apte à distinguer différents types et sévérité de corrosion par piqûre. Les 28 cas mal classés dans le Les catégories Gin et Gex impliquent des corrosions relativement peu profondes, parmi lesquelles 26 cas ont dg/t = 20 % et deux cas ont dg/t = 40 %. La répartition des 28 cas mal classés par wg/lg est illustré à la figure 9. La figure 9 indique que le modèle de classification CNN est moins précis pour les corrosions générales larges avec wg/lg supérieur à 1,67. Il est également à noter que le modèle développé est moins précis pour les cas contenant des corrosions externes (c'est-à-dire Gex, Pex, et PICex) que pour les cas impliquant des corrosions internes (c'est-à-dire Gin, Pin et PICin). Un potentiel L'explication est que l'entrée du modèle (c'est-à-dire la carte du signal MFL) pour les défauts externes a moins caractéristiques prononcées car la variation de B due au défaut de perte de métal sur la face externe La surface du tuyau est plus petite que celle provoquée par le défaut de perte de métal interne comme l'externe Le défaut est plus éloigné du capteur que le défaut interne.

Machine Translated by Google

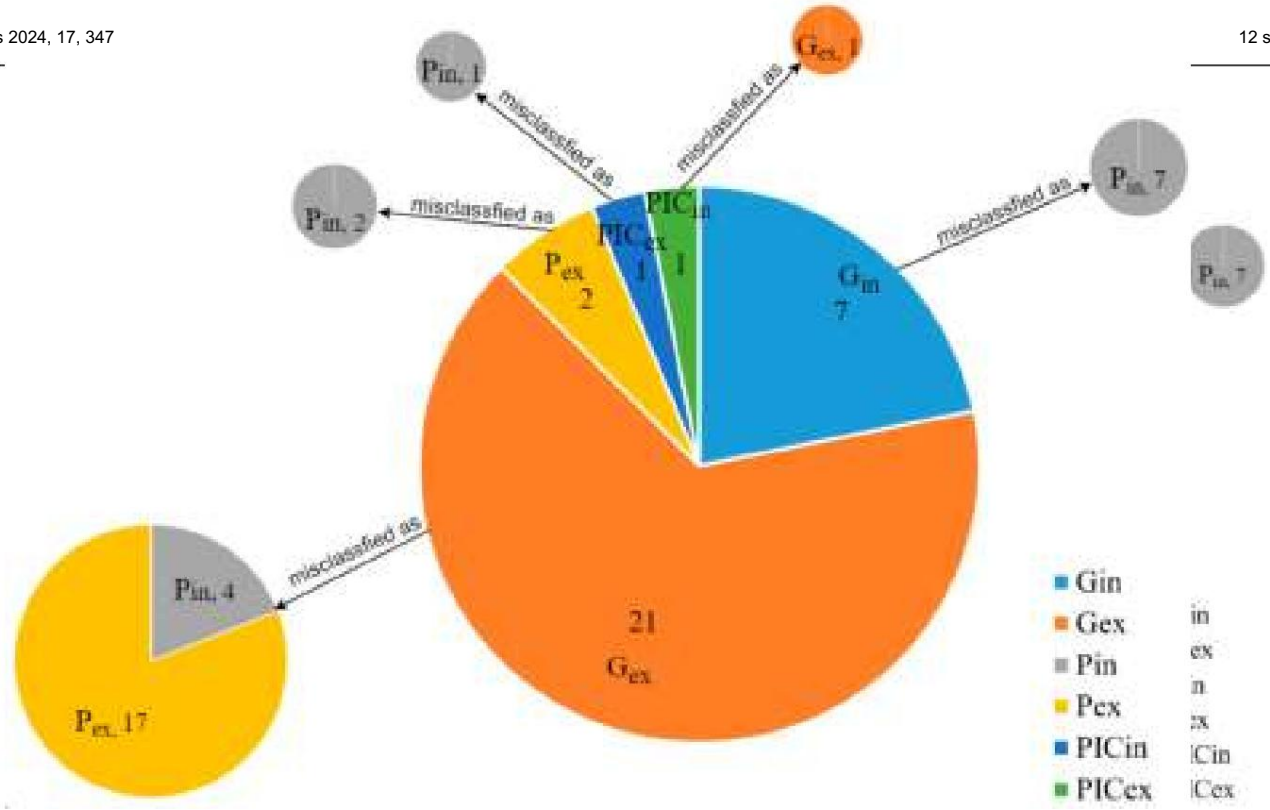


Figure 8. Répartition des erreurs de classification dans l'ensemble de données de formation.

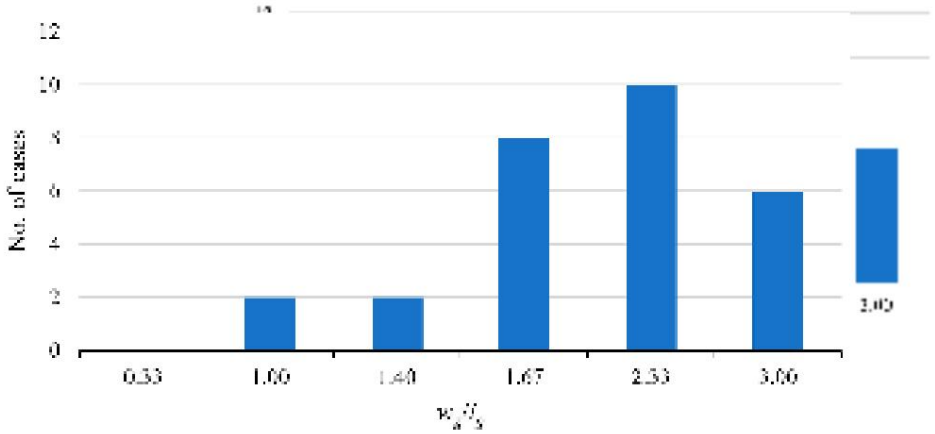


Figure 9. Répartition des cas mal classés à Gin et Gex par wg/lg.

4.2. Régression CNN

4.2.1. Informations d'entrée

Les résultats du modèle de classification CNN sont utilisés pour extraire les cas correctement classés pour Pin, Pex, PICin et PICex, à l'exclusion de quatre erreurs de classification dans l'ensemble de données d'entraînement et d'une erreur de classification dans l'ensemble de données de test, résultant en 1 600, 1 598, 1 599, et 1 598 cas d'entrée pour Pin, Pex, PICin et PICex, respectivement, pour entraîner et valider le modèle de régression CNN . Les trois matrices 160 × 160 générées de signaux MFL sans bruit pour chacun des 6395 (=1600 + 1598 + 1599 + 1598) constituent l'entrée du modèle de régression développé, et les rp, dp/t, h et Les valeurs φ pour chaque cas sont la sortie. Les données correctement classées sont adoptées pour entraîner le modèle de régression CNN car l'inclusion de données mal classées (telles qu'une anomalie générale de corrosion classée à tort comme un trou d'épingle) dans la formation est problématique, par exemple, définir une fonction d'erreur applicable à toutes les données d'entrée, y compris les données mal classées. . Alternativement, la formation du modèle de régression CNN peut être complètement séparée du modèle de classification ; c'est-à-dire que l'ensemble des 6 400 cas d'analyse paramétrique par éléments finis impliquant des trous d'épingle peuvent être utilisés pour entraîner le modèle de régression. Dans cette étude, l'ensemble de données correctement classé, qui comprend 6 395 cas et est presque identique à l' ensemble de données complet, est utilisé pour entraîner et valider le modèle de régression

Machine Translated by Google

4.2.2. Structure proposée

Le modèle de régression CNN proposé dans la présente étude comprend un total de 23 couches : les couches 1 à 16 sont dédiées à l'extraction de caractéristiques, tandis que les couches 17 à 23 se concentrent sur régression. La couche 23 est la couche de sortie, qui comprend quatre composants représentant les valeurs rp, dp/t, h et φ prévues , respectivement. Le modèle de régression CNN détaillé les informations sur les couches sont présentées dans le tableau 4. Il est à noter qu'en plus des couches abandonnées introduit dans la section 4.1.2, plusieurs couches de normalisation par lots sont incorporées dans les couches CNN typiques pour éviter le surajustement [46–48]. Comme le montre le tableau 4, le réglage hyperparamètres dans le processus d'extraction de caractéristiques du modèle de régression, tels que la taille du filtre de la couche de convolution, la taille du pooling et le taux d'abandon sont les mêmes que ceux du modèle de classification CNN, tandis que le taux d'abandon du travail dans la régression Le processus du modèle de régression CNN est de 0,5. La figure 10 représente l'architecture du développé un modèle de régression avec les dimensions de différentes couches. Notez que l'abandon et les couches de normalisation par lots sont exclues du diagramme car elles n'ont pas d'impact sur le dimensionnalité des données. La palette de couleurs utilisée pour distinguer les différentes couches de la figure 10 est cohérents avec ceux utilisés dans la figure 7.

Tableau 4. Informations sur la couche du modèle de régression CNN.

Partie	Numéro de couche	Nom du calque	Paramètres
Extraction de fonctionnalités	1	Convolution	64 filtres de taille égale à (5 × 5)
	2	Normalisation par lots	
	3	Pooling maximum	La taille du pooling est égale à (2 × 2)
	4	Abandonner	Taux = 0,2
	5	Convolution	128 filtres de taille égale à (5 × 5)
	6	Normalisation par lots	
	7	Pooling maximum	La taille du pooling est égale à (2 × 2)
	8	Abandonner	Taux = 0,2
	9	Convolution	256 filtres de taille égale à (5 × 5)
	10	Normalisation par lots	
	11	Pooling maximum	La taille du pooling est égale à (2 × 2)
	12	Abandonner	Taux = 0,2
	13	Convolution	512 filtres de taille égale à (5 × 5)
	14	Normalisation par lots	
	15	Pooling maximum	La taille du pooling est égale à (2 × 2)
	16	Abandonner	Taux = 0,2
Régression	17	Aplatir le calque	Le numéro d'unité est égal à 51 200
	18	Abandonner	Taux = 0,5
	19	Couche entièrement connectée	Le numéro d'unité est égal à 64
	20	Couche entièrement connectée	Le numéro d'unité est égal à 128
	21	Couche entièrement connectée	Le numéro d'unité est égal à 256
	22	Couche entièrement connectée	Le numéro d'unité est égal à 512
	23	Sortie 1	Le numéro d'unité est égal à 1
		Sortie 2	Le numéro d'unité est égal à 1
		Sortie 3	Le numéro d'unité est égal à 1
		Sortie 4	Le numéro d'unité est égal à 1

Machine Translated by Google

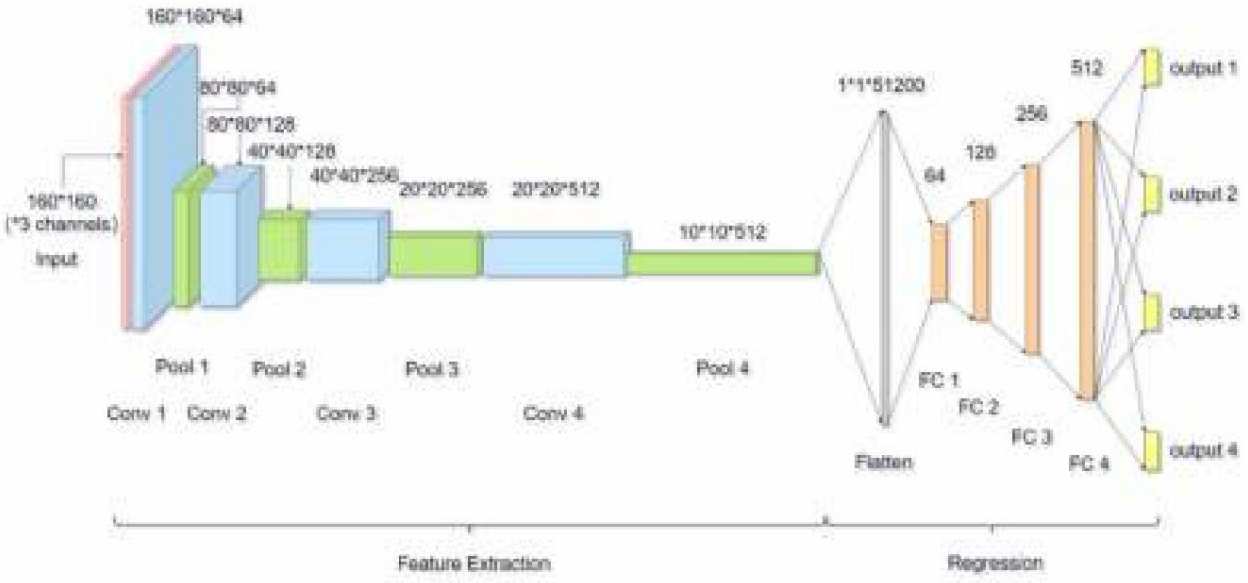


Figure 10. Structure du modèle de régression CNN proposé.

4.2.3. Résultats

Le coefficient de détermination (c'est-à-dire R ffi 2) est utilisé pour mesurer la précision de chaque prédiction de paramètre (c'est-à-dire rp, dp/t, h et φ), définie par l'équation suivante.

$$R^2 \triangleq 1 - \frac{\sum_{je=1}^{N} (Yvrai,i - Ypred,i)^2}{\sum_{je=1}^{N} (Yvrai,i - \overline{Yvrai})^2} \quad (4)$$

où Ytrue,i et Ypred,i représentent respectivement les valeurs vraies et prédites de chaque paramètre prédit pour le ième (i = 1, 2,. . ., N) point dans un ensemble de données donné contenant N points de données , et Ytrue désigne la vraie valeur moyenne de N points de données associés à un paramètre prédit. Le tableau 5 résume le R inclus dans le modèle de régression. Le R 2 les valeurs des quatre paramètres de l'ensemble de données de test sur la figure 11. Les valeurs réelles des quatre 2 sont également tracées dans ce que l'on appelle un graphique radar paramètres rp, dp/t, h et φ comparées aux valeurs estimées par le CNN pour les corrosions dans l'ensemble de test sont représentées dans la figure 12. Cela ressort clairement du tableau 5 que les prédictions du modèle CNN concernant les paramètres de localisation de la corrosion (c'est-à-dire h et φ) montrent un fort accord avec les valeurs réelles, comme en témoignent les valeurs R de 1,0 pour les quatre catégories de corrosion. En termes de prédictions de taille (c'est-à-dire rp et dp/ t), le modèle est plus précis pour les corrosions internes (Pin et PICin) que pour les corrosions externes (Pex et PICex). Ces résultats indiquent également que les performances du modèle sont meilleures pour les cas contenant uniquement des trous d'épingle que pour les cas contenant du PIC. Ceci est attendu car le PIC introduit une complexité supplémentaire, notamment la présence d'une corrosion générale, qui peut avoir un impact sur la précision du modèle. Dans l'ensemble, la précision R obtenue dans la prédiction des paramètres de taille, avec une précision plus élevée observée pour les corrosions sur la 2 valeurs démontrent le modèle CNN surface interne et les trous d'épingle. Cependant, il convient de noter que bien que les prédictions pour rp et dp/t soient, en général, exactes, la figure 12 révèle des bandes d'erreur relativement importantes dans les prédictions de rp et dp/t . Par exemple, les valeurs prédites de rp vont de 1,14 mm à 3,45 mm pour (rp)true égal à 2 mm. Les grandes bandes d'erreur peuvent s'expliquer par les petites dimensions de la corrosion en trou d'épingle et la résolution spatiale limitée du modèle. En outre, la présence de caractéristiques de corrosion composites peut introduire des complexités et des incertitudes supplémentaires, conduisant à des marges d'erreur plus larges dans les prévisions. Par conséquent, des recherches plus approfondies sont nécessaires pour explorer les facteurs contribuant à la réduction des erreurs de régression pour les corrosions par piqûres.

Machine Translated by Google

Tableau 5. Résultats du modèle de régression CNN sans bruit.

Numéro de catégorie	Catégorie	Cas			Test R 2			
		Total	Former	Test	rp	dp/t	h	φ
0	Épingle	1600	1280	320	0,91	0,95	1h00	1h00
1	PEX	1598	1277	321	0,88	0,92	1h00	1h00
2	PIcin	1599	1280	319	0,89	0,85	1h00	1h00
3	PICex	1598	1279	319	0,86	0,77	1h00	1h00
Dans l'ensemble		6395	5116	1279	0,89	0,91	1h00	1h00

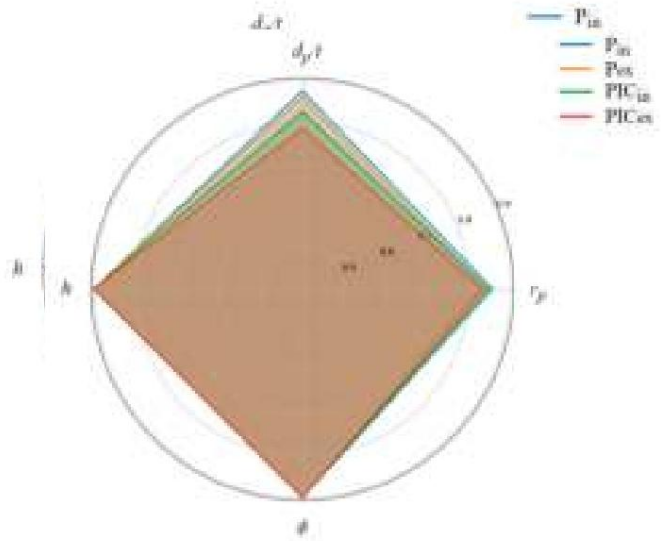


Figure 11. Carte radar pour R 2 de métriques (rp, dp/t, h et φ) pour les corrosions Pin, Pex, PICin et PICex dans l'ensemble de données de test.

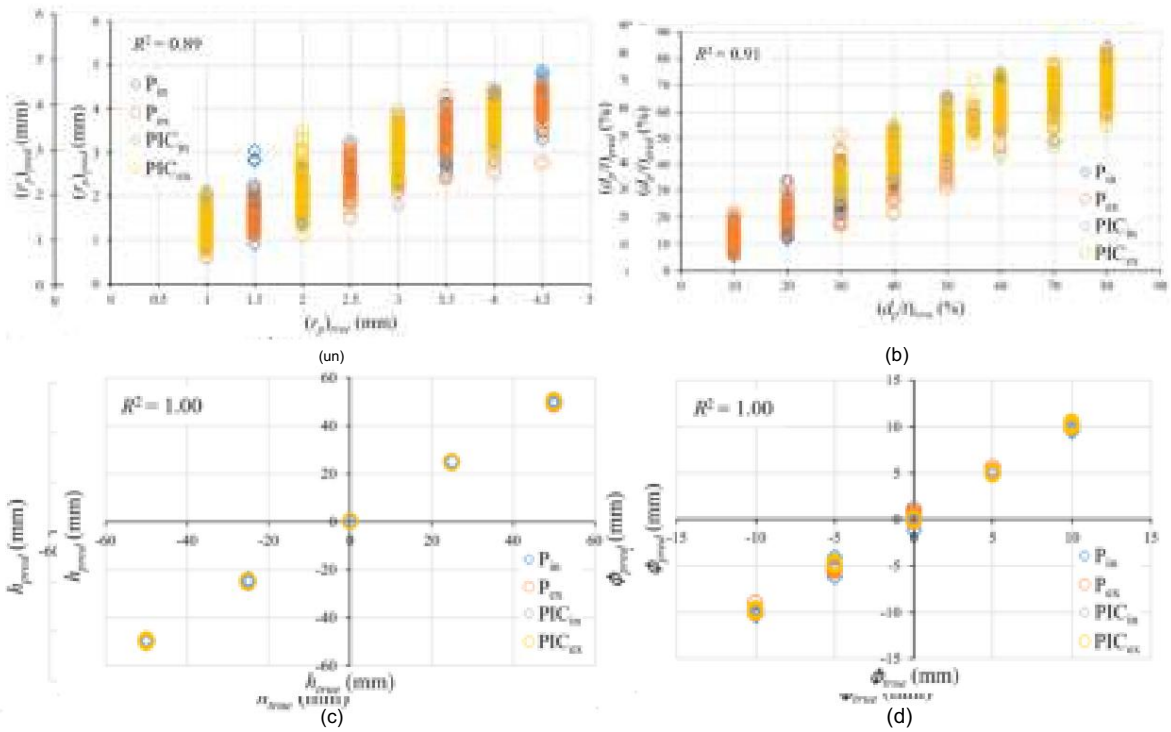


Figure 12. Comparaison des valeurs vraies et prédites de rp, dp/t, h et φ pour les anomalies dans le Ensemble de données de test de régression sans bruit.

Machine Translated by Google

4.2.4. Influence du bruit

La figure 13 compare les résultats du modèle de régression CNN pour l'entrée bruyante (SNR = 20) avec ceux du scénario sans bruit décrit à la section 4.2.3. Avec la présence de bruit, 2 les valeurs R pour prédire rp et dp/t diminuent quelque peu par rapport aux valeurs ff sans bruit et sont plus importantes pour les cas Pex et PICex , indiquant que le modèle est plus affecté par le bruit lors de la prévision des tailles de corrosion externe. tt
scénario. La diminution de R^2

Malgré la diminution de la précision dans la prévision des tailles de corrosion, le modèle maintient 2 R = 1 pour les paramètres de localisation (h et φ) pour tous les types de corrosion à l'exception du Pex, indiquant sa robustesse dans la capture des informations spatiales même en présence de bruit. . ff

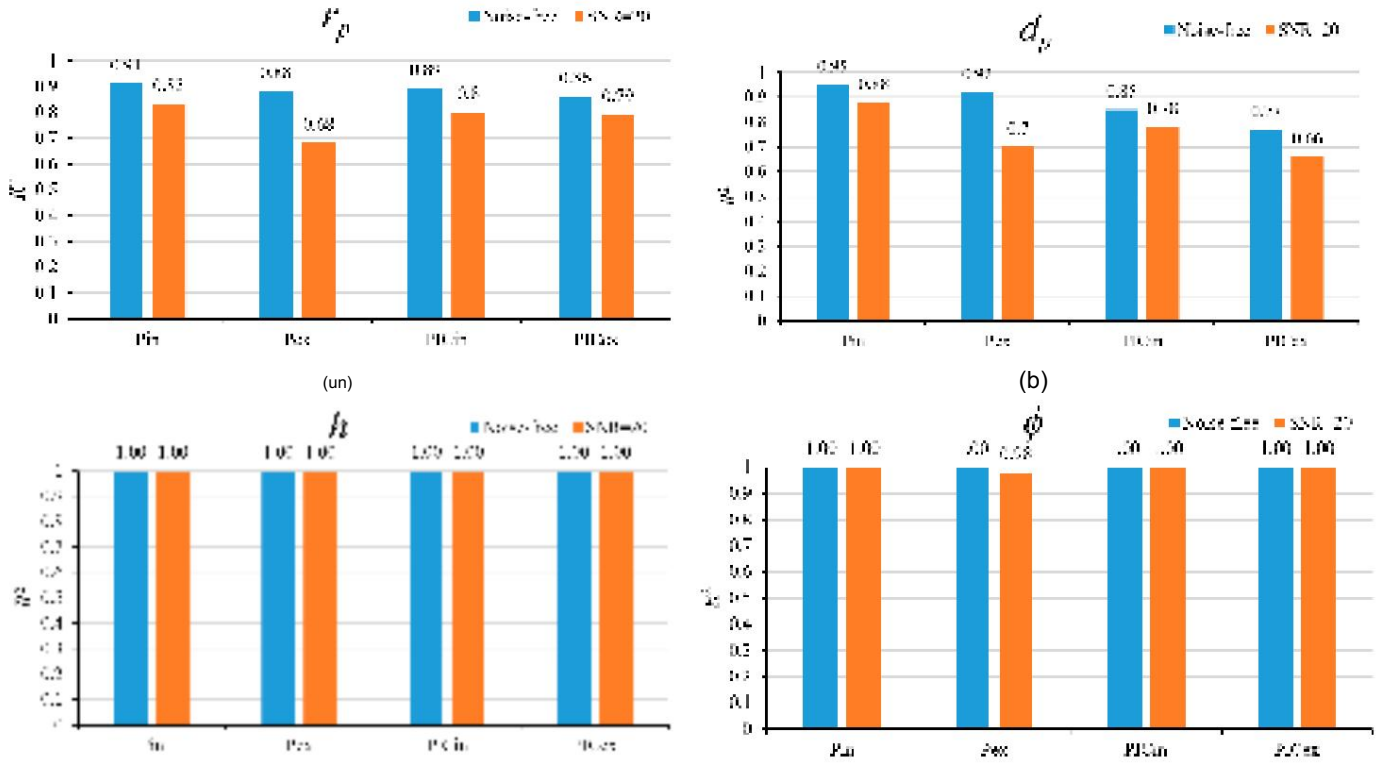


Figure 13. Comparaison de R^2 entre sans bruit et SNR = 20 scénarios pour différents types de corrosion. ff

Enfin, il est important de souligner que les modèles de classification et de régression CNN développés dans cette étude sont spécifiquement adaptés à certains attributs des pipelines et des outils MFL, tels que l'épaisseur de la paroi du tuyau, les propriétés des aimants, les valeurs de décollement et les formes des défauts. Par conséquent , des recherches plus approfondies sont nécessaires pour déterminer dans quelle mesure le modèle CNN fonctionne avec différentes épaisseurs de paroi de canalisation et différents paramètres d'outil MFL. Des études supplémentaires sont également nécessaires pour prendre en compte l'évolution de l'outil MFL dans les modèles de prédiction. ff

5. Conclusions

Cet article rend compte d'une nouvelle étude axée sur la classification et le dimensionnement des corrosions par trous d'épingle sur les pipelines à l'aide d'algorithmes d'apprentissage profond. Nous proposons un modèle de classification CNN pour classer six types différents de corrosions (c'est-à-dire Gin, Gex, Pin, Pex, PICin et PICex) sur les pipelines et un modèle de régression CNN pour estimer la taille et l'emplacement des défauts de sténopé sur la base des signaux MFL. généré à l'aide de FEA magnétostatique. De nombreux cas FEA paramétriques 3D impliquant des corrosions générales en forme de boîte et des trous d'épingle en forme de cylindre sont utilisés pour simuler les signaux MFL 3D en faisant varier la profondeur, la longueur, la largeur et les emplacements longitudinaux et circonférentiels du défaut. Les modèles de classification et de régression CNN sont ensuite entraînés et validés à l'aide des signaux MFL simulés.

Le modèle de classification proposé s'avère avoir une excellente précision dans la classification de six types de défauts de corrosion avec un taux d'erreur de prédiction de 0,34 % (33/9 600 = 0,34 %). Le

Machine Translated by Google

Le modèle de régression proposé est très précis pour prédire l'emplacement du trou d'épingle et atteint une bonne précision dans la prévision de la profondeur et du diamètre du trou d'épingle. Grâce à une analyse comparative du modèle de régression CNN avec des signaux sans bruit et bruyants (SNR = 20) en entrée, il est observé que l'impact du bruit sur la précision prédictive du modèle de régression est modéré. Cette étude démontre l'application d' algorithmes d'apprentissage profond pour faciliter la gestion de l'intégrité des pipelines contenant des défauts de corrosion de forme complexe .

Contributions de l'auteur : YS : enquête, visualisation, analyse formelle, méthodologie, rédaction – version originale . WZ : Acquisition de financement, supervision, conceptualisation, méthodologie, rédaction – révision et édition. Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Les auteurs reconnaissent le soutien financier reçu du Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) (subvention n° RGPIN-2019-05160) et de la Faculté de génie de l'Université Western Ontario.

Déclaration de disponibilité des données : Les données sont disponibles sur demande auprès des auteurs.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Références

1. Zhou, Q. ; Wu, W. ; Liu, D. ; Li, K. ; Qiao, Q. Estimation de la probabilité de défaillance par corrosion des oléoducs et gazoducs basée sur une approche de logique floue. Ing. Échouer. Anal. 2016, 70, 48-55. [Référence croisée]
2. Murphy, JF Pannes de pipeline Nightmare, planification fantastique, cygnes noirs et gestion de l'intégrité – Une revue. Processus Saf. Programme. 2015, 34, 207. [Réf. croisée]
3. Shen, Y. ; Zhou, W. Une comparaison des statistiques sur les incidents liés aux pipelines de transport de pétrole et de gaz terrestres au Canada et aux États-Unis. Int. J.Crit. Infrastructurer. Prot. 2024, 45, 100679. [Réf. croisée]
4. Vanaei, RH ; Eslami, A. ; Egbewande, A. Une revue de la corrosion des pipelines, de l'inspection en ligne (ILI) et des modèles de taux de croissance de la corrosion. Int. J. Presse. Navires Pip. 2017, 149, 43-54. [Référence croisée]
5. Chanson, H. ; Yang, L. ; Liu, G. ; Tian, G. ; Ona, DI ; Chanson, Y. ; Li, S. Analyse comparative des équipements d'inspection en ligne et technologies. IOP Conf. Ser. Maître. Sci. Ing. 2018, 382, 032021. [Réf. croisée]
6. Forum des exploitants de pipelines. Spécifications et exigences pour l'inspection en ligne des pipelines. Version 2021. Disponible en ligne : https://pipelineoperators.org/documents (consulté le 18 mai 2022).
7. Sutherland, J. ; Bluck, M. ; Pearce, J. ; Quick, E. La validation de la technologie d'inspection en ligne MFL de dernière génération conduit à une détection améliorée et à des spécifications de dimensionnement pour les trous d'épingle, les piqûres, les rainures axiales et les fentes axiales. Dans les actes de la 8e conférence internationale sur les pipelines de l'ASME 2010, Calgary, AB, Canada, 27 septembre-1er octobre 2010.
8. Peng, X. ; Anyaoha, États-Unis ; Liu, Z. ; Tsukada, K. Analyse des données de fuite de flux magnétique (MFL) pour l'évaluation de la corrosion des pipelines. IEEETrans . Magné. 2020, 56, 1–15. [Référence croisée]
9. Han, W. ; Yang, P. ; Xia, F. ; Xue, Y. Inversion du signal de fuite de flux magnétique de défauts corrosifs basée sur une recherche locale génétique modifiée algorithme. J.Shanghai Jiaotong Univ. (Sci.) 2009, 14, 168-172. [Référence croisée]
10. Priewald, RH; Magèle, C. ; Grand livre, PD ; Pearson, NR; Mason, JSD Inversion rapide du signal de fuite de flux magnétique pour la reconstruction de profils de défauts arbitraires dans l'acier à l'aide d'éléments finis. IEEETrans. Magné. 2012, 49, 506-516. [Référence croisée]
11. Hwang, K. ; Mandayam, S. ; Udpa, SS Caractérisation des signaux d'inspection de gazoducs à l'aide de la fonction neuronale à base d'ondelettes réseaux. NDT E Int. 2000, 33, 531-545. [Référence croisée]
12. Ramuhalli, P. ; Udpa, L. ; Udpa, SS Algorithmes d'inversion basés sur les réseaux neuronaux dans l'évaluation non destructive des fuites de flux magnétique. J.Appl. Phys. 2003, 93, 82748276. [Réf. croisée]
13. Han, WH; Que, reconstruction de défauts PW 2-D à partir de signaux MFL basée sur un algorithme d'optimisation génétique. Dans les actes du Conférence internationale IEEE 2005 sur la technologie industrielle, Hong Kong, Chine, 14-17 décembre 2005.
14. Kandroodi, M. ; Arabi, BN ; Bassiri, MM; Ahmadabadi, MN Estimation de la profondeur et de la longueur des défauts à partir de mesures de fuite de flux magnétique : vérification avec des simulations, des expériences et des données de raclage. IEEETrans. Magné. 2016, 53, 1–10. [Référence croisée]
15. Feng, J. ; Li, F. ; Lu, S. ; Liu, J. Reconstruction rapide des profils de défauts à partir de mesures de fuite de flux magnétique à l'aide d'une méthodologie d'ajustement d'erreur basée sur RBFNN. IET Sci. Mesures. Technologie. 2017, 11, 262-269. [Référence croisée]
16. Yao, G. ; Lei, T. ; Zhong, J. Une revue de la reconnaissance d'action basée sur les réseaux neuronaux convolutifs. Reconnaissance de modèles. Lett. 2019, 118, 14-22. [Référence croisée]
17. Taye, MM Compréhension théorique du réseau neuronal convolutif : concepts, architectures, applications, orientations futures. Calcul 2023, 11, 52. [CrossRef]
18. Lu, S. ; Feng, J. ; Zhang, H. ; Liu, J. ; Wu, Z. Une méthode d'estimation de la taille des défauts à partir d'une image MFL utilisant la transformation visuelle réseau neuronal convolutif. IEEETrans. Ind. Informer. 2018, 15, 213-224. [Référence croisée]

Machine Translated by Google

19. Shen, Y. ; Zhou, W. Développement d'un modèle de réseau neuronal convolutif pour prédire la taille et l'emplacement des défauts de corrosion sur pipelines basés sur des signaux de fuite de flux magnétique. Int. J. Presse. Navires Pip. 2023, 207, 105-123. [Référence croisée]
20. Wang, HA; Chen, G. Méthode d'estimation de la taille des défauts pour les signaux de fuite de flux magnétique à l'aide de réseaux neuronaux convolutifs. Aperçu 2020, 62, 86-91. [Référence croisée]
21. EGIG. 10e rapport du Groupe européen de données sur les incidents de gazoducs (période 1970-2016) ; EGIG : Groningue, Pays-Bas, 2018 ; Doc. numéro VA 17.R.0395.
22. Zhang, H. ; Sha, S. ; Willis, C. ; Qingshan, F. ; Chen, P. Étude de faisabilité de l'inspection des sténopés via une fuite de flux magnétique et essais hydrostatiques dans les oléoducs et gazoducs. IOP Conf. Ser. Maître. Sci. Ing. 2021, 1043, 022053.
23. Feng, Q. ; Yan, B. ; Chen, P. ; Shirazi, SA Analyse des défaillances et modèle de simulation de la corrosion par sténopé de l'oléoduc raffiné. Ing. Échouer. Anal. 2019, 106, 1-28. [Référence croisée]
24. Subramanian, C. Corrosion par piqûre localisée des tuyaux API 5L de qualité A utilisés dans les applications industrielles de tuyauterie d'eau d'incendie. Ing. Échouer. Anal. 2018, 92, 405-417. [Référence croisée]
25. Askari, M. ; Aliofkhazraei, M. ; Afroukhteh, S. Une étude complète sur la corrosion interne et la fissuration des oléoducs et des gazoducs. J. Nat. Gaz Sci. Ing. 2019, 71, 102971. [Réf. croisée]
26. Kadhim, KN; Al-Rufaye, AHR Les effets d'un champ magnétique transversal uniforme sur le profil de débit et de vitesse local. Int. J.Civ. Ing. Technologie. 2016, 7, 140-151.
27. Ji, F. ; Wang, C. ; Soleil, S. ; Wang, W. Application du FEM 3D dans l'analyse de simulation des signaux MFL. Aperçu 2009, 51, 32-35. [Référence croisée]
28. Bubenik, T. Méthodes électromagnétiques de détection de la corrosion dans les pipelines souterrains : fuite de flux magnétique (MFL) ; Souterrain Corrosion des pipelines ; Woodhead Publishing : Cambridgeshire, Royaume-Uni, 2014 ; pp. 215-226.
29. Shi, Y. ; Zhang, C. ; Li, R. ; Cai, M. ; Jia, G. Théorie et application de la détection des pipelines de fuite de flux magnétique. Capteurs 2015, 15, 31036-31055. [Référence croisée]
30. Zhang, Y. ; Oui, Z. ; Wang, C. Une méthode rapide pour la reconstruction de la taille des fissures rectangulaires dans les tests de fuite de flux magnétique. NDT E Int. 2009, 42, 369-375. [Référence croisée]
31. Walker, J. Inspection en ligne des pipelines : technologies avancées pour une exploitation économique et sûre des oléoducs et des gazoducs ; Verlag Moderne Industrie : Landsberg am Lech, Allemagne, 2010.
32. Liu, Y. ; Gao, X. ; Wang, Y. ; Yang, X. Optimisation des paramètres sensibles du mécanisme de support à aimant permanent. J. Mech. Sci. Technologie. 2014, 28, 2707-2714. [Référence croisée]
33. Yang, L. ; Zhang, G. ; Liu, G. ; Gao, S. Effet du décollage sur l'inspection des fuites de flux magnétique des pipelines. Dans les actes du 17e 2008 Conférence mondiale sur les essais non destructifs, Shanghai, Chine, 25-28 octobre 2008.
34. Irlande, RC ; Torres, CR Modélisation par éléments finis d'un magnétiseur circonférentiel. Sens. Actionneurs A Phys. 2006, 129, 197-202. [Référence croisée]
35. Azizzadeh, T. ; Safizadeh, MS Éléments finis tridimensionnels et simulation expérimentale de CND de type fuite de flux magnétique pour la détection des corrosions par piqûres. Dans Actes de la 4e Conférence internationale iranienne sur les CND (IRNDT) 2017, Téhéran, Iran, 21-22 février 2017.
36. Chen, L. ; Li, X. ; Qin, G. ; Lu, Q. Traitement du signal d'un défaut de surface de fuite de flux magnétique inspecté dans l'acier des pipelines. Russe. J. Non-destruct. Test. 2008, 44, 859-867. [Référence croisée]
37. Piao, G. ; Guo, J. ; Cabane.; Leung, H. L'effet des courants de Foucault induits par le mouvement sur les fuites de flux magnétique (MFL) à grande vitesse inspection des tuyaux en acier à paroi épaisse. Rés. Non-destruct. Évaluer. 2020, 31, 48-67. [Référence croisée]
38. Li, F. ; Feng, J. ; Zhang, H. ; Liu, J. ; Lu, S. ; Ma, D. Reconstruction rapide de profils de défauts de pipeline arbitraires à partir de mesures MFL en utilisant un algorithme de recherche d'harmonie modifié. IEEETrans. Instrument. Mesures. 2018, 67, 9. [Réf. croisée]
39. Ventilateur, X. ; Dai, M. ; Liu, C. ; Wu, F. ; Yan, X. ; Feng, Y. ; Feng, Y. ; Su, B. Effet du bruit d'image sur la classification des lésions cutanées à l'aide réseaux de neurones convolutifs profonds. Tsinghua Sci. Technologie. 2019, 25, 425-434. [Référence croisée]
40. Acharya, UR ; Oh, SL ; Hagiwara, Y. ; Tan, JH; Adam, M. ; Gertych, A. ; San Tan, R. Un modèle de réseau neuronal convolutif profond pour classer les battements cardiaques. Calculer. Bouillir. Méd. 2017, 89, 389-396. [Référence croisée] [Pub Med]
41. Tesfai, H. ; Saleh, H. ; Al-Qutayri, M. ; Mohammad, MB; Tekeste, T. ; Khandoker, A. ; Mohammad, B. Shufflenet léger basé sur CNN pour la classification des arythmies. Accès IEEE. 2022, 10, 111842-111854. [Référence croisée]
42. Ouest, NE ; O'shea, T. Architectures profondes pour la reconnaissance de modulation. Dans les actes du symposium international IEEE 2017 sur les réseaux d'accès dynamique au spectre (DySPAN), Baltimore, MD, États-Unis, 6-9 mars 2017.
43. Baranwal, S. ; Khandelwal, S. ; Arora, A. Réseau neuronal convolutif d'apprentissage profond pour la détection de la maladie des feuilles de pommier. Dans Actes de la Conférence internationale 2019 sur l'informatique durable en science, technologie et gestion (SUSCOM), Amity University Rajasthan, Jaipur, Inde, 26-28 février 2019.
44. Virupakshappa, K. ; Marin, M. ; Oruklu, E. Une architecture de réseau neuronal convolutif multi-résolution pour la détection de défauts par ultrasons. Dans les actes du Symposium international sur les ultrasons (IUS) 2018 de l'IEEE, Kobe, Japon, 22-25 octobre 2018.
45. Bilmes, J. Sous-ajustement et surajustement dans l'apprentissage automatique. Notes de cours UW ECE. 2020. Disponible en ligne : https://people.ece. uw.edu/bilmes/classes/ee511/ee511_spring_2020/overfitting_underfitting.pdf (consulté le 20 mars 2023).
46. Ioffé, S. ; Szegedy, C. Normalisation par lots : accélérer la formation approfondie du réseau en réduisant le changement de covariable interne. Dans les actes de la Conférence internationale 2015 sur l'apprentissage automatique, Lille, France, 6-11 juillet 2015.

Machine Translated by Google

47. Thakkar, V. ; Tewary, S. ; Chakraborty, C. Normalisation par lots dans les réseaux de neurones convolutifs — Une étude comparative avec les données CIFAR-10. Dans Actes de la cinquième conférence internationale IEEE 2018 sur les applications émergentes des technologies de l'information (EAIT), IIEST Shibpur, Bengale occidental, Inde, 12-13 janvier 2018.
48. Björck, N. ; Gomes, CP ; Selman, B. ; Weinberger, KQ Comprendre la normalisation des lots. Av. Informations neuronales. Processus. Système. 2018, 31, 1 à 12.

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.

Machine Translated by Google

algorithmes

MDPI

Article

Le problème de planification des machines parallèles avec différentes vitesses et temps de libération dans les opérations de transport de minerai

Luis Tarazona-Torres, Ciro Amaya*, Alvaro Paipilla, Camilo Gomez et David Alvarez-Martinez

Département de génie industriel, Université de Los Andes, Cra 1 N18A12, Bogota 111711, Colombie ; le.tarazona@uniandes.edu.co (LT-T.); a.paipilla@uniandes.edu.co (AP) ; gomez.ch@uniandes.edu.co (CG) ; d.alvarezm@uniandes.edu.co (DA-M.)
* Correspondance : ca.amaya@uniandes.edu.co

Résumé : Les opérations de transport de minerai sont cruciales au sein de l'industrie minière car elles fournissent des minéraux essentiels aux usines de production. Réalisées avec des équipements opérationnels sophistiqués et coûteux, ces opérations nécessitent une planification méticuleuse pour garantir que les objectifs de production sont atteints tout en optimisant l'utilisation des équipements. Dans cette étude, nous présentons un algorithme pour déterminer la quantité minimale d'équipement de transport requise pour atteindre l'objectif de transport du minerai. Pour y parvenir, un modèle mathématique a été développé, le considérant comme un problème de planification de machines parallèles avec différentes vitesses et temps de libération, en se concentrant sur la minimisation à la fois du temps de réalisation et des coûts associés à l'utilisation de l'équipement. De plus, un autre algorithme a été développé pour permettre l'évaluation tactique de ces deux variables. Ces procédures et le modèle contribuent de manière significative aux décideurs en fournissant une approche systématique de l'allocation des ressources, garantissant que les équipements de chargement et de transport sont utilisés à leur plein potentiel tout en respectant les contraintes budgétaires et les calendriers opérationnels. Cette approche optimise l'utilisation des ressources et améliore l'efficacité opérationnelle, facilitant ainsi l'amélioration continue des opérations minières.

Mots-clés : problème d'ordonnancement de machines parallèles ; équipement de transport de minerai ; modèle mathématique; industrie minière

check for updates

Citation : Tarazona-Torres, L. ; Amaya, C. ; Paipilla, A. ; Gomez, C. ; Alvarez-Martinez, D. Le problème de planification des machines parallèles avec différentes vitesses et temps de libération dans l'opération de transport de minerai. Algorithmes 2024, 17, 348. https://doi.org/10.3390/a17080348

Rédacteurs académiques : Alexandre Dolgui, David Lemoine, María I. Restrepo et Franck Werner

Reçu : 2 mai 2024
Révisé : 24 juillet 2024
Accepté : 30 juillet 2024
Publié : 8 août 2024

Copyright : © 2024 par les auteurs. Licencié MDPI, Bâle, Suisse.
Cet article est un article en libre accès distribué selon les termes et conditions des Creative Commons Licence d'attribution (CC BY) (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

L'une des activités les plus essentielles des opérations minières est le transport du minerai ; ce processus consiste à déplacer la roche fragmentée du site de dynamitage à l'aide d'équipements de chargement et de transport [1]. Les roches minéralisées sont ensuite transportées vers le processus de concassage pour réduire leur taille avant d'être envoyées à l'usine de récupération, où des métaux précieux tels que le cuivre, l'or, le molybdène et l'aluminium, entre autres, sont extraits [2-4]. La figure 1 illustre le schéma du processus décrit. Une mauvaise planification des opérations de transport de minerai peut diminuer les niveaux de productivité, affecter la réalisation des objectifs de production dans les délais établis et, par conséquent, avoir un impact sur les coûts opérationnels. Plusieurs facteurs doivent être pris en compte pour planifier efficacement les opérations de transport de minerai, tels que la quantité d'équipements de chargement et de transport disponibles et leur entretien programmé [5]. Il est également essentiel de tenir compte des objectifs de production établis, des horaires de travail et des niveaux topographiques. D'autres aspects importants incluent la teneur du minerai, le nombre de rampes de chargement disponibles pour les opérations de chargement et l'impact environnemental [6].

Les sociétés minières disposent d'une quantité définie d'équipements de chargement et de transport allouée en fonction des objectifs de production fixés pour une période spécifique. Il est crucial de déterminer la quantité nécessaire de ces équipements pour transporter efficacement les minerais afin de contrôler les coûts associés, qui peuvent être importants en fonction du temps nécessaire pour atteindre ces objectifs [7]. Connaître la quantité minimale d'équipement nécessaire optimise les opérations et facilite l'analyse tactique. Cette analyse permet une meilleure prise de décision face aux aléas, garantissant l'atteinte des objectifs de production au moindre coût possible.

Machine Translated by Google

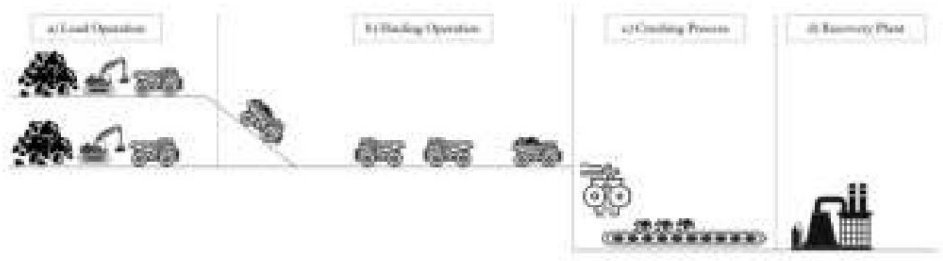


Figure 1. Schéma de l'opération de transport du minerai.

Comme le montre la figure 1, le processus de transport du minerai implique l'affectation de l'équipement de transport à l'équipement de chargement. Ceux-ci ne fonctionnent pas forcément à la même vitesse de chargement car certains équipements ont des capacités de chargement différentes. Le chargement est effectué simultanément et est décrit comme un système de machines parallèles à différentes vitesses. Le nombre d'équipements de transport étant limité, ils sont utilisés autant de fois que nécessaire pour atteindre l' objectif de production. Cependant, ils ne sont pas disponibles pendant le transport du minerai vers le processus de concassage, ayant un « temps de libération » avant d'être à nouveau disponibles pour être rechargés. Ainsi, ce scénario peut être adapté à un problème de machines parallèles avec des vitesses et des temps de déclenchement différents.

La nouveauté de ce travail réside dans l'intégration de différentes vitesses et temps de relâchement dans le modèle mathématique pour la planification parallèle des machines dans les opérations de transport de minéraux (PMSPOHO). Ces fonctionnalités permettent une représentation plus réaliste et précise des conditions opérationnelles dans l'environnement minier. La vitesse de l'équipement de chargement joue un rôle crucial dans ce modèle. Les variations des vitesses de chargement peuvent avoir un impact significatif sur l' efficacité des opérations de transport du minerai. En prenant en compte différentes vitesses de chargement, le modèle peut allouer plus précisément les ressources pour correspondre à la nature dynamique des opér
Cela garantit que les équipements à chargement plus rapide sont utilisés au maximum de leur potentiel tandis que les équipements plus lents se voient attribuer des tâches mieux adaptées à leurs capacités, optimisant ainsi le flux de travail global.

L'intégration de la vitesse de l'équipement de chargement dans l'algorithme de planification maximise l'utilisation de l'équipement et minimise le temps d'inactivité de l'équipement de transport. Cela améliore l'efficacité opérationnelle en réduisant les retards et en garantissant un flux continu de matériaux.
Le modèle mathématique développé aborde ces variables critiques, en optimisant l'allocation des équipements de transport, en maximisant leur utilisation et en réduisant les temps d'inactivité. Cette approche garantit une plus grande efficacité opérationnelle en alignant les capacités des équipements sur les exigences du processus minier. L'allocation stratégique des ressources, guidée par la vitesse et la disponibilité de chaque machine de chargement, contribue à un fonctionnement plus rationalisé et plus rentable , soulignant l'importance de ces variables pour obtenir des performances optimales.

Le modèle mathématique contribue de manière significative à plusieurs aspects essentiels. Il facilite une planification et un ordonnancement plus efficaces des opérations de transport, permettant des ajustements dynamiques en fonction des conditions changeantes. De plus, il permet une analyse comparative entre les coûts et les délais de réalisation, optimisant ainsi les décisions économiques et de productivité.
En réduisant les temps d'arrêt et en améliorant l'utilisation des ressources, le modèle répond efficacement aux défis logistiques et financiers des opérations minières et s'adapte à divers scénarios opérationnels, augmentant ainsi son applicabilité et sa pertinence dans le secteur. Le reste de cet article est structuré comme suit : La section 2 présente les travaux connexes. La section 3 détaille la méthodologie. Les expériences informatiques et l'analyse des résultats sont discutées dans la section 4.

Enfin, les conclusions tirées de l'étude sont présentées dans la section 5.

2. Travaux connexes

Le tableau 1 présente un aperçu des travaux connexes traitant du problème. Il détaille les caractéristiques considérées dans notre domaine d'intérêt, telles que les temps de déclenchement (ri) et les différentes vitesses (Qm). Les fonctions objectives utilisées sont également analysées, telles que le délai de réalisation (Cmax) et les coûts (k), entre autres. De plus, les méthodes de résolution utilisées sont examinées et classées comme exactes ou approximatives, ainsi que le nombre de machines utilisées et

Machine Translated by Google

les approches adoptées dans chaque cas. Cette comparaison permet de visualiser diverses méthodologies et techniques mises en œuvre dans la littérature, mettant en avant leur différences et similitudes concernant l'analyse actuelle.

La caractéristique ri n'a pas encore été suffisamment explorée, notamment dans la recherche minière. où la disponibilité des équipements de transport est cruciale, comme le montre [8,9]. En revanche, l'étude de Qm a attiré davantage d'attention académique, même si l'accent reste largement mis sur machines identiques. Concernant la fonction objectif, les études se sont généralement concentrées sur minimiser les variables telles que k et Cmax de manière isolée. Meza [2] illustre un modèle plus complexe intégration, qui intègre une analyse du temps et des coûts des équipements de chargement et de transport grâce à un modèle de simulation. Hong et Lin [10] se sont concentrés sur la minimisation des retards maximaux (Lmax), tandis que Koryagin et Voronov [11] et Oliskevych et al. [12] a envisagé la réunion P comme essentiel pour maintenir la continuité opérationnelle des usines de production.

En comparaison, nos travaux élargissent considérablement le niveau d'analyse en incluant à la fois Qm et ri , nous permettant de modéliser les environnements de production de manière plus réaliste, notamment dans les opérations de transport de minerai. Ceci est particulièrement pertinent dans les configurations où le nombre de machines dépasse cinq, un scénario dans lequel de nombreuses études préfèrent les méthodes approximatives en raison à leur efficacité informatique. De plus, notre approche comprend une étude itérative qui permet une analyse au niveau tactique pour définir l'utilisation optimale de l'équipement, complétée par un support analytique solide pour la prise de décision afin de maximiser l'efficacité opérationnelle. Notre approche intègre également une analyse de complexité qui n'a pas encore été explorée [10,13,14]. Ainsi, nos travaux comblent une lacune importante dans la littérature et fournissent des informations critiques qui faciliter les décisions stratégiques et améliorer l'efficacité opérationnelle dans l'industrie minière.

Tableau 1. Travaux connexes sur le problème de planification de machines parallèles.

	Caractéristiques		Objectif Fonction			Méthodes		Nombre. Machines			
Référence	Qm	ri	Cmax	k	Autre	Exact	Approximatif	m ≤ 5	m ≥ 6	Complexité Analyse	Approches
[2]	x		x	x		x		x			Modèle de simulation FIFO
[15]			x			x		x			Algorithme de branche et de prix avec serveur unique
[16]			x			x	x	x	x		Formulation MIP et hybride algorithme heuristique avec serveur unique
[17]			x			x	x	x			Formulation MIP et recherche taboue avec un seul serveur
[18]			x			x		x			Formulation MIP avec des tailles de travail
[19]	x			x		x		x			Formulation MIP
[20]	x			x		x		x			Formulation MILP
[21]	x			x		x	x	x	x		Formulation et insertion du MILP Heuristique sous tarifs temporels d'utilisation
[22]	x			x		x		x			Formulation MILP sous tarifs selon l'heure d'utilisation
[8]		x	x			x	x	x	x		Formulation MILP et Algorithme GA
[11]	x				P.		x	x			Algorithme heuristique
[23]			x			x		x			Formulation MIP avec temps de préparation et serveur unique
[10]			x		Lmax		x	x		x	Algorithmes constructifs dynamiques
[24]			x			x	x				Modèle de programmation par contraintes sous contraintes d'indisponibilité et modifier LPT
[13]			x			x	x	x		x	Formulation MIP et recherche taboue

Machine Translated by Google

Tableau 1. Suite

Référence	Caractéristiques Objectif Fonction Qm	ri	Cmax	k	Autre	Méthodes Exact	Approximatif	Nombre. Machines m ≤ 5	m ≥ 6	Complexité Analyse	Approches
[25]				x			x		x		Formulation MIP
[14]			x				x	x	x	x	Formulation MIP avec emplois indépendants
[26]	x		x				x	x			Formulation MIP avec contrainte de fonction
[12]	x				P.		x	x			Formulation MIP
[27]				x				x	x		Modèle de simulation de Monte-Carlo
[9]		x	x					x	x		Algorithmes VNS et ACO
[28]	x		x				x	x		x	Algorithme MIP et TSP
[29]			x				x	x			Modèle de programmation par contraintes
[30]				x			x	x	x		Formulation MIP et Algorithme MOBSO
[31]	x			x			x	x	x		Formulation MILP avec heure d'utilisation et SPT, MDPC, MDEC
[32]	x		x				x	x		x	Formulation MIP et GA – GWO algorithme hybride
Notre papier	x	x	x	x			x	x	x	x	Formulation itérative du MIP

Considérant les études connexes sur l'optimisation de l'opération de transport du minerai, il a été ont découvert que Meza [2] a développé un modèle pour l'application des concepts de simulation d'un système de chargement et de transport pour les opérations minières à ciel ouvert, en tenant compte du type de chargement et équipement de transport, heure d'arrivée, temps d'attente, temps de positionnement, temps de chargement, aller l'heure, l'heure de prise, l'heure de déchargement et l'heure de retour, en plus du nombre de voyages en tant que objectif de production, et en envisageant un système FIFO pour le chargement et le déchargement. De l'autre part, Vasquez et al. [3] ont réalisé un modèle de programmation en nombres entiers pour minimiser le total temps de travail d'une flotte de chargeuses à profil bas dans une mine souterraine, en plus de développer un algorithme optimal en temps polynomial intégré au processus de prise de décision, obtenir des résultats assez proches du temps total optimal. Eivazy et Askari-Nasab [33] ont développé un modèle de programmation linéaire en nombres entiers mixtes (MILP) multi-destination pour minimiser les coûts d'exploitation dans l'exploitation minière à ciel ouvert ; dans le modèle, ils ont pris en compte les stocks et les pieux de mélange en plus des systèmes de chargement dans le sens horizontal et de la prise de décision en matière de rampe. Anjomshoa et coll. [34] ont développé un modèle de simulation et de programmation en nombres entiers mixtes (MIP) pour optimiser les quais de dépassement, le dimensionnement et la planification du parc d'équipements de transport en exploitation minière souterraine.

Par ailleurs, Gligoric [35] a développé une enquête sur les aires de dépassement pour le halage équipements dans l'exploitation minière souterraine, dans lesquels ils ont cherché à optimiser la quantité nécessaire de baies et l'emplacement optimal pour minimiser les temps d'attente. Gonen et coll. [36] a mené une étude pour l'exploitation minière souterraine où ils ont évalué les capacités de production annuelles avec différents systèmes de transport, en tenant compte des coûts de transport unitaires et de la profondeur de la mine. Eivazy et Askari-Nasab [37] a présenté une méthodologie basée sur des hiérarchies de planification de production pour l'exploitation minière à ciel ouvert, à moyen terme, développer un MILP pour minimiser les coûts d'exploitation ; la planification comprenait les stocks, les usines de transformation et les décharges, ainsi qu'une sélection de routes et de rampes. Tom-Socarras [38] a généré un outil de gestion des coûts basé sur un arbre de décision pour le transport de matériaux dans les mines à ciel ouvert. Uribe [39] a développé une étude qui a analysé la fonctionnalité et les performances du fonctionnement semi-automatique en chargement et le matériel de transport utilisé dans l'exploitation minière souterraine. Wi ¸ecek [40] a proposé une procédé de contrôle du chargement et du transport de minerai dans une mine avec pieux et

Machine Translated by Google

système de chambres. Wiest [41] a développé un modèle de planification pour l'exploitation minière à ciel ouvert, axé sur la réduction des coûts.

Solomon [42] a modélisé les systèmes de chargement et de transport du minerai dans l'exploitation minière à ciel ouvert, en développant un modèle prenant en compte le front de taille, les itinéraires, l'équipement et les destinations du minerai. Gaspar et Jhasmin [43] ont analysé les indicateurs d'utilisation et de disponibilité pour l'optimisation des coûts de transport du minerai. Li et coll. [44], dans leurs recherches sur l'exploitation minière souterraine, ont développé un algorithme d'essaim d'abeilles pour maximiser le revenu total au cours d'une période de planification ; en outre, ils ont créé un algorithme génétique pour la planification de l'expédition des équipements de transport afin de minimiser le temps d'attente de ces équipements. Elijah [45] a développé un modèle mathématique pour calculer le taux d'arrivée et de service de différentes quantités d'équipements de transport impliqués dans le transport de matériaux. Huisa [46] a amélioré l'utilisation des équipements de chargement et de transport pour réduire le coût de ces opérations.

Choi et coll. [47] ont développé un système intelligent sans supervision pour prédire les performances du système d'équipement de transport dans l'exploitation minière à ciel ouvert en utilisant une combinaison de différents modèles d'optimisation. Li et coll. [48] se sont concentrés sur un modèle mathématique pour la réordonnancement des opérations prenant en compte les exigences des espaces de production, de l'environnement d'exploitation et de l'usure des équipements de production, tout en cherchant à obtenir le taux d'achèvement de la planification maximal et la plus faible fluctuation de la teneur du minerai. Abolghasemian et al. [49] ont présenté une optimisation multi-objectifs du système de mine à ciel ouvert, recherchant la quantité maximale d'extraction et minimisant le temps de transport, en tenant compte de la capacité de stockage, de l'équipement de transport et du budget. Shamsi et coll. [50] ont réalisé une IP avec des restrictions opérationnelles pour planifier le système de transport dans l'exploitation minière à ciel ouvert, en considérant la meilleure valeur actuelle nette comme objectif. Parichehreh et coll. [51] ont abordé le problème de planification de machines parallèles sans rapport avec l'efficacité énergétique, en intégrant la détérioration des tâches et les effets d'apprentissage des opérateurs, fournissant ainsi des informations précieuses sur l'optimisation multi-objectifs pour les environnements de production complexes.

Notre étude introduit une approche nouvelle et significative pour optimiser les opérations de chargement et de transport de minerai en modélisant un problème tactique à l'aide d'un modèle de planification généralement utilisé pour les décisions opérationnelles. Ce modèle représente des machines parallèles avec des vitesses différentes, ajoutant un plus grand réalisme. Plus précisément, la nature du problème se reflète dans la disponibilité dynamique des équipements de chargement, représentée par des temps de libération dynamiques, où les tâches ne sont pas connues à l'avance mais sont révélées au fur et à mesure de leur exécution. Cette approche permet des analyses plus approfondies et plus détaillées, mieux adaptées aux complexités des environnements de production minière. La mise en œuvre de ce modèle peut être intégrée aux stratégies tactiques et opérationnelles pour les rendre plus efficientes et efficaces, en soutenant la prise de décision dans les sociétés minières en améliorant considérablement l'efficacité opérationnelle, en réduisant les coûts et en augmentant la productivité. De plus, cette étude contribuera à la fois au développement théorique et à l'application pratique en génie minier, améliorant ainsi les opérations essentielles du secteur.

3. Méthodologie Cette

section décrit la méthodologie utilisée dans cette étude. Il est structuré comme suit : La section 3.1 définit le problème et définit le contexte de la recherche. La section 3.2 détaille les hypothèses de modélisation et présente la formulation mathématique du problème, y compris les fonctions objectives et les contraintes. La section 3.3 décrit les algorithmes utilisés pour déterminer et évaluer la quantité optimale d'équipement de transport de minerai. Cela comprend la procédure permettant de déterminer la quantité minimale d'équipement de transport et l' évaluation des compromis coût-délai. Les sous-sections suivantes abordent systématiquement les objectifs de recherche et fournissent une compréhension complète des méthodes appliquées.

3.1. Définition du problème

En raison de l'importance cruciale du transport du minerai pour la production, cette recherche vise à déterminer la quantité minimale d'équipement de transport nécessaire pour minimiser à la fois les coûts associés à l'utilisation de l'équipement et le temps d'achèvement de l'objectif de production.

Machine Translated by Google

dans un horizon temporel défini. Les opérations de transport de minerai sont essentielles à l'approvisionnement continu des usines de production en minéraux essentiels. Cependant, les décisions techniques concernant l'équipement requis sont souvent prises sur la base de l'expérience et sous pression de temps.
Même si ces décisions peuvent être rapides, elles peuvent parfois seulement être optimales, ce qui peut entraîner une augmentation des coûts ou des retards ayant un impact sur la production des usines.

Cette recherche aborde l'inefficacité et le sous-optimisme potentiel des processus décisionnels traditionnels pour déterminer la quantité minimale d'équipement de transport nécessaire pour atteindre l'objectif de production. Cela peut entraîner une surestimation ou une sous- estimation des équipements de transport, entraînant des dépenses inutiles ou des goulots d'étranglement lors du transport du minerai. En développant un modèle mathématique qui prend en compte le problème de planification de machines parallèles avec différentes vitesses et temps de libération, nous visons à fournir une approche systématique et basée sur les données pour optimiser l'allocation des équipements.

Une analyse détaillée de la relation entre les coûts et le délai d'exécution pourrait révéler des scénarios précieux , permettant des décisions plus éclairées et appropriées en cas d'imprévu dans l'opération de transport. Ce modèle permettra aux décideurs d'équilibrer les compromis entre les coûts d'équipement et l'efficacité opérationnelle, en garantissant que la bonne quantité d' équipement est déployée pour atteindre les objectifs de production sans engager de dépenses inutiles.

Le problème peut être conceptualisé comme un problème de planification de machines parallèles avec différentes vitesses et temps de libération. Alors que l'objectif principal est une décision tactique visant à déterminer la quantité minimale d'équipement de transport et à analyser les coûts liés au temps d'exécution, une adaptation à ce problème de planification dans le cadre d'un modèle de programmation entier est exécutée. Selon la notation établie par Graham [52], un problème d'ordonnancement d'opérations est défini par trois champs (α| β|γ) : α représente la configuration du système, β les contraintes et caractéristiques du système, et γ la fonction objectif. En utilisant ce cadre et en suivant la nomenclature décrite par Pinedo et Hadavi [53], la configuration du système correspond à des machines parallèles avec différentes vitesses (Qm), où ces machines sont les équipements de chargement avec des temps de traitement variables. Une caractéristique importante du système comprend les temps de largage (ri) incorporés dans le temps de transport. La fonction objectif vise à minimiser à la fois les coûts et les délais de réalisation. La figure 2 illustre le calendrier de l' opération de transport du minerai conceptualisée comme le problème mentionné précédemment.

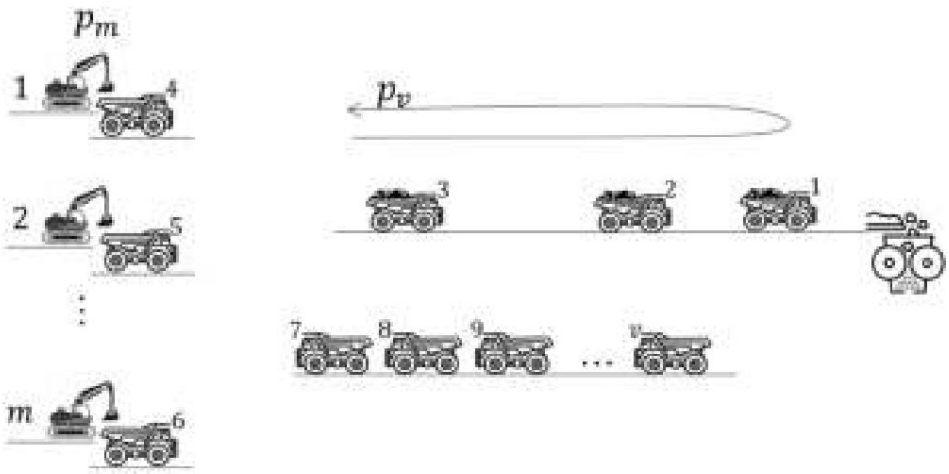


Figure 2. Calendrier considérant les opérations de transport de minerai.

3.2. Modélisation

Hypothèses du modèle. Les hypothèses du modèle sont les suivantes, basées sur les travaux présentés par Meza [2] : • Un seul point de

déchargement du minerai a été considéré. • Le temps de
ravitaillement est négligeable. • Les équipements
de chargement et de transport sont disponibles à tout moment. • Le temps d'attente
au déchargement du minerai est insignifiant.

Machine Translated by Google

Notation. Le tableau 2 fournit la notation du modèle proposé, montrant le type et définition, respectivement.

Tableau 2. Notation du modèle proposé.

Var	Taper	Définition
M	Ensemble	Ensemble d'équipement de chargement M = {1, 2, . . . , m} .
V	Ensemble	Ensemble d'équipements de transport V = {1, 2, . . . , v}.
T	Ensemble	Ensemble d'unités de temps : T = {1, 2, . . . , h}.
p_j	Paramètre	Temps de chargement du matériel de chargement j.
p_i	Paramètre	Temps de transport pour le transport de l'équipement i.
P	Paramètre	L'objectif de production est le nombre total de déplacements à effectuer.
h	Paramètre	Horizon temporel pour atteindre l'objectif de production.
k_j	Paramètre	Coût associé à l'utilisation de l'équipement de chargement.
k_i	Paramètre	Coût associé à l'utilisation du matériel de transport.
xijt	Variable	Variable de décision principale, prend la valeur 1 si l'équipement de transport i utilise l'équipement de chargement j au temps t, 0 sinon.
yjt	Variable	La variable auxiliaire de décision prend la valeur 1 si chargement de l'équipement j utilise au temps t, 0, sinon.
bouton	Variable	La variable auxiliaire de décision prend la valeur 1 si l'équipement de transport i utilise au temps t, 0, sinon.
Cmaxk	Variable	Délai d'achèvement de l'objectif de production
	Variable	Coûts associés à l'utilisation de l'équipement

Fonction objectif. La première fonction objectif (équation (1)) cherche à minimiser k associés à l'utilisation d'équipements. La deuxième fonction objectif (équation (2)) cherche à minimiser Cmax de tous les déplacements.

$$\min k \tag{1}$$

$$\min Cmax \tag{2}$$

Contraintes. En ce qui concerne les contraintes, l'ensemble des contraintes de l'équation (3) garantit l'affectation d'un seul véhicule de transport i à un véhicule de chargement j à l'instant t. L'ensemble de Les contraintes de l'équation (4) garantissent qu'un seul équipement de chargement j par équipement de transport i au temps t.

$$\sum_{j \in V} xijt \le 1 \quad j \quad M, t \quad T \tag{3}$$

$$\sum_{j \in M} xijt \le 1 \quad i \quad V, t \quad T \tag{4}$$

L'ensemble des contraintes des équations (5) et (6) garantit que le chargement et le transport les équipements sont utilisés pendant la durée requise lorsqu'ils sont attribués.

$$\sum_{\tau=t+1}^{t+p_j} yj\tau \ge (p_j - 1) \cdot xijt \quad i \quad V, j \quad M, t \quad T \setminus (t \le h - p_j) \tag{5}$$

$$\sum_{\tau=t+1}^{t+p_j+p_i} zi\tau \ge (p_i + p_j - 1) \cdot xijt \quad i \quad V, j \quad M, t \quad T \setminus (t \le h - p_i - p_j) \tag{6}$$

L'ensemble des contraintes des équations (7) et (8) garantit que le chargement et le transport les équipements ne sont pas réservés en double et sont utilisés uniquement lorsqu'ils sont disponibles.

$$xijt \le 1 - yjt \quad i \quad V, j \quad M, t \quad T \tag{7}$$

$$xijt \le 1 - zit \quad i \quad V, j \quad M, t \quad T \tag{8}$$

Machine Translated by Google

L'ensemble des contraintes de l'équation (9) garantit Cmax.

$$Cmax \geq t \cdot xijt + p_{i}^{h} + p_{j} \quad i \quad V, j \quad M, t \quad T \tag{9}$$

L'ensemble des contraintes de l'équation (10) garantit que P est satisfait.

$$\sum_{i}^{V} \sum_{j}^{M} \sum_{t}^{T} xijt \geq P \quad i \quad V, j \quad M, t \quad T \tag{10}$$

L'équation (11) calcule k.

$$k = \sum_{i}^{V} \sum_{t}^{T} h_{k} \cdot zit + je \quad \sum_{j}^{M} \sum_{t}^{T} l_{kj} \cdot yjt \quad i \quad V, j \quad M, t \quad T \tag{11}$$

L'ensemble des contraintes des équations (12) et (13) garantit la nature des variables.

$$xijt, yjt, zit \quad \{0, 1\} \quad i \quad V, j \quad M, t \quad T \tag{12}$$

$$Cmax, k \geq 0 \tag{13}$$

Le modèle proposé peut être utilisé dans les processus décisionnels tactiques et opérationnels, car il spécifie avec précision les deux décisions. Cependant, la complexité de ce modèle de programmation linéaire à nombres entiers mixtes en termes de quantité de variables entières est i·j·t + j·t + i·j, et il comporte deux autres variables continues. Si nous analysons le nombre de contraintes, nous trouvons un total de 7·i.j·t + j·t + i·j. Le fait que le modèle s'appuie sur des vecteurs binaires (pour indiquer les heures de démarrage des camions et les périodes d'activité) augmente intrinsèquement la complexité mais ouvre également la possibilité (i) d'inclure des inégalités valides pour écarter les solutions qui ne correspondent pas à la structure du problème ou (ii) des techniques de décomposition, telles que l'utilisation de coupes combinatoires de Benders, ou l'adoption d'un cadre d'horizon glissant [54], pour tirer parti des stratégies diviser pour régner ; ces deux éléments font tous deux partie de recherches en cours. On pourrait envisager de réduire les délais, mais cela empêcherait d'utiliser ces modèles dans des schémas de contrôle comme ceux fournis par un jumeau numérique, une technique répandue dans la gestion des opérations minières [55].

3.3. Algorithmes pour déterminer et évaluer la quantité optimale d'équipement de transport de minerai

Pour intégrer le double objectif du modèle de minimiser à la fois le temps et le coût d'achèvement, une approche itérative a été utilisée dans laquelle les modèles sont créés en utilisant une seule fonction objectif. Comme le montrent les algorithmes 1 et 2, le processus itératif permet d'explorer le compromis entre le délai de réalisation et les coûts, fournissant ainsi un ensemble de solutions qui équilibrent les deux objectifs.

Compte tenu de l'analyse au niveau tactique, la procédure suivante a été proposée pour déterminer la quantité optimale d'équipement de transport, comme détaillé dans l'algorithme 1. L'algorithme hl commence par définir les données d'entrée, qui incluent M, V, T , p i , p j , h et P. Ensuite, définir les données de sortie, qui incluent le nombre minimum d'équipements de transport (vopt) et l' affectation de l'équipement de transport à l'équipement de chargement (Schedule), y compris Cmax.

Ensuite, attribuez v à la grandeur de V, le nombre total d'équipements de transport disponibles (ligne 1). La variable Continue est définie sur true pour lancer la boucle (ligne 2). Tant que cela reste vrai, un modèle est créé (en considérant les équations (2) à (13) du modèle PMSPOHO, les données d'entrée et v) et un résultat qui minimise le Cmax est obtenu (lignes 3 à 5). Une évaluation de faisabilité est menée, où si le résultat est réalisable, v est réduit d'une unité (en essayant de résoudre le problème avec moins d'équipement de transport) (lignes 6 à 8). Si cela n'est pas réalisable, vopt est augmenté d'une unité et la boucle se termine (lignes 9 à 13). Cet algorithme renvoie vopt et Schedule (ligne 14).

Machine Translated by Google

Algorithme 1 Déterminateur optimal de l'équipement de transport.

Entrée : M, V, T , p $^{h}_{je}$, p$^{»}_{j}$, h, P;

Sortie : vopt : matériel de transport minimum, Calendrier : affectation du matériel de transport aux équipements de chargement dont le Cmax.

```
1 : v ← |V|
2 : Continuer ← vrai
3 : pendant que Continuer, faites
4 :     modèle ← Créer un modèle (équations (2) – (13), entrée, v)
5 :     Résultat ← Résoudre (modèle, Réduire)
6 :     si EstFaisable (Résultat) alors
7 :         v ← v − 1
8 :         Résultat, Calendrier
9 :     autre
10 :        vopt ← v + 1
11 :        Continuer ← faux
12 : fin si
13 : fin pendant
14 : retour vopt, Calendrier
```

La procédure suivante a été proposée pour une analyse de la relation entre k et Cmax, détaillés dans l'algorithme 2. Ce pseudocode permet d'observer le comportement de k à mesure que la Cmax augmente. L'algorithme commence par définir les données d'entrée, qui comprennent les mêmes facteurs que dans l'algorithme 1 : k $^{h}_{je}$, $^{kl}_{j}$, et v. Ensuite, la définition des données de sortie comprend un liste des coûts et liste des délais de résiliation (Cmax). Définissez v (ligne 1). Le modèlelb est créé (en considérant les équations (2) à (13) du modèle PMSPOHO et les données d'entrée) (ligne 2), et un résultat lb de la minimisation de Cmaxlb est obtenu (ligne 3), en extrayant ce résultat (ligne 4). Un modélisme est créé (en considérant les équations (1), (3) – (13) du modèle PMSPOHO et les données d'entrée) (ligne 5), et un Resultub de la minimisation de k est obtenu (ligne 6), en extrayant Cmaxub pour cela résultat (ligne 7).

Algorithme 2 Évaluateur coût-temps progressif

Entrée : M, V, T , p $^{h}_{je}$, p$^{»}_{j}$, k $^{h}_{je}$, $^{kl}_{j}$, h, P et v ;

Sortie : ListPair(Cmax, k) : Liste des coûts par rapport à Cmax.

```
1 : Définir v
2 : modellb ← Créer un modèle (équations (2) – (13), entrée)
3 : Résultatlb ← Résoudre (modellb, Réduire)
4 : Cmaxlb ← ExtraireCmax(Résultatlb)
5 : modelub ← Créer un modèle (équations (1), (3) – (13), entrée)
6 : Resultub ← Résoudre (modelub, Réduire)
7 : Cmaxub ← ExtraireCmax (Résultat)
8 : pour Cmaxre f dans Cmaxlb à Cmaxub faire
9 :     modelre f ← Créer un modèle (équations (1), (3) – (8), (10) – (13), entrée)
10 :    Résultat f ← Résoudre (modelre f , Réduire)
11 :    Coût ← ExtractCost (Résultat f)
12 :    ListePaire(Cmax, k) ← AddToList(Cmaxre f , Coût)
13 : fin pour
14 : retour Cmax, k
```

Une évaluation progressive est effectuée qui itère Cmaxre f par 1 de Cmaxlb à Cmaxub (ligne 8). Pour chaque valeur, un modèle f est créé (en considérant les équations (1), (3) – (8), et (10) – (13) du modèle PMSPOHO et des données d'entrée) (ligne 9). Le résultat f est obtenu en minimisant k à chaque itération (ligne 10) ; k est extrait et ajouté à une liste (Cost) pour chaque Cmaxre f ,k) correspondant (lignes 14), (lignes 11-12). Cet algorithme renvoie la liste des paires (Cmaxre f

Machine Translated by Google

fournir une évaluation détaillée de la façon dont k varie en fonction des changements $_{f}$. Dans l'algorithme, de Cmaxre. Cmaxre sert de valeur de référence intermédiaire pour évaluer de manière itérative le compromis entre la minimisation de Cmax et de k. Cette évaluation itérative donne un aperçu du rapport coût-temps compromis dans le modèle.

4. Expériences informatiques et analyse des résultats

Cette section présente les expériences informatiques pour évaluer les performances de le modèle et les algorithmes PMSPOHO proposés. Toutes les expériences ont été menées sur un ordinateur avec un système d'exploitation Windows 11, Intel(R) Core(TM) i5-12450H de 12e génération, Processeur 2,50 GHz et 8 Go de RAM. Le langage de programmation Python 3.11 a été utilisé pour le développement. L'optimiseur commercial Gurobi 10.0.1 et l'optimiseur gratuit CBC ont été utilisé avec un temps limite de recherche de 3600 s et un GAP de 0,01.

Basé sur des données réelles fournies dans les études de Meza [2], Vásquez [3] et Anchiraico et Rojas [4], qui incluent le nombre d'équipements de chargement et de transport, les les temps de transport, les horizons temporels et les objectifs de production, 34 instances expérimentales ont été généré. Ces instances sont conçues pour simuler des scénarios du monde réel, permettant ainsi une évaluation de nos méthodologies proposées dans diverses conditions. Cette approche garantit que notre configuration expérimentale est basée sur des paramètres opérationnels réalistes, améliorant ainsi la l'applicabilité et la pertinence de nos résultats. Il est important de noter qu'une comparaison directe avec les études des auteurs ci-dessus n'a pas été menée, car leurs méthodologies de recherche diffèrent significativement de notre approche proposée.

4.1. Réglage des valeurs des paramètres de l'optimiseur

Gurobi Optimizer version 10.0.1 propose une large gamme de paramètres conçus pour affiner les performances de l'optimiseur. Compte tenu de la vaste gamme d'options possibles configurations, identifiant la combinaison optimale de paramètres qui améliorent l'ensemble la performance du modèle pose un défi important. Le TuningAPI de Gurobi offre un outil précieux pour explorer divers paramètres pouvant améliorer les performances du modèle dans des cas spécifiques [56].

Compte tenu de la procédure indiquée par Cuellar-Usaquén et al. [57], un échantillon de sept instances ont été créées pour identifier les paramètres appropriés pour notre modèle à l'aide de API de réglage. Cet échantillon inclut les cas où le temps de calcul obtenu par le modèle PMSPOHO utilisant les paramètres par défaut (PMSPOHO-D) dépasse le délai de sa classe respective par deux écarts types.

Le tableau 3 montre les résultats du réglage des paramètres de Gurobi sur sept instances de problèmes différentes . Chaque ligne du tableau représente une instance, avec des colonnes indiquant le Gurobi spécifique paramètres ajustés pour cette instance. Les paramètres incluent Symmetry, DegenMove, Heuristiques, GomoryPasses, PreDepRow, MIPFocus, Méthode, CutPasses et BranchDir.

Tableau 3. Configurations des paramètres trouvées pour le modèle mathématique.

	Résumé			Paramètres de valeur										
Inst.	vm	P	Symétrie	DegenMoves	Heuristique	GomoryPasses	Méthode	PreDepRow	MIPFocus	CutPasses	BranchDir	Tc-D	Tc-T	Variation
1	3	2	20	2								0,8	0,79	1,25%
2	6	2	20	2								1,22	1,15	5,74%
3	8	4	81		0	1	1					69,59	55,03	20,92%
4	12	4	81				2					46,54	44,4	4,60%
5	15	4	81					0				32,53	31,64	2,74%
6	8	4	90			1					1	139,22	120,68	13,32%
7	20	4	100						1			81,62	79,9	2,11%

Le troisième cas se démarque considérablement car il présente la variation la plus élevée (20,92 %) entre le temps de calcul par défaut (Tc-D) et le temps de calcul ajusté temps (Tc-T), indiquant une amélioration significative des performances avec les paramètres ajustés. Dans ce cas, les paramètres Heuristics, GomoryPasses et PreDepRow ont été ajustés, ce qui a entraîné une réduction substantielle du temps de calcul de 69,59 à 55,03 unités de temps. Cet ajustement a réduit le temps de calcul global et optimisé le utilisation de ressources informatiques, rendant le processus plus efficace et efficient. Le con-

Machine Translated by Google

La différence considérable entre Tc-D et Tc-T souligne l'importance de la personnalisation et du réglage fin des paramètres du modèle pour obtenir une efficacité maximale dans la résolution de problèmes mathématiques complexes.

Dans ce cas, le paramètre heuristique ajusté apparaît comme un acteur clé dans l'accélération du processus de résolution. En augmentant l'utilisation de méthodes heuristiques, qui sont des stratégies de résolution utilisant des règles simples et rapides pour générer rapidement des solutions satisfaisantes, le modèle a exploré et évalué les solutions potentielles plus efficacement. Cette approche a considérablement réduit le temps nécessaire pour parvenir à une solution réalisable, démontrant ainsi son caractère pratique et son efficacité. L'application de l'heuristique est un outil précieux, en particulier dans les scénarios où la vitesse est critique et où une certaine précision peut être sacrifiée au profit d'une vitesse de résolution améliorée.

GomoryPasses, un autre paramètre ajusté, contrôle le nombre de passes de coupe Gomory. Ces passes sont des techniques d'optimisation qui renforcent la relaxation linéaire du modèle, conduisant à une convergence plus rapide. L'ajustement du nombre de ces passes peut avoir un impact significatif sur l'efficacité de la solution, en équilibrant la nécessité d'améliorer la solution avec le temps supplémentaire requis pour chaque coupe. PreDepRow, qui gère le prétraitement des lignes dépendantes, a également fondamentalement amélioré l'efficacité en réduisant la taille du problème et en améliorant la stabilité numérique. Un prétraitement efficace peut simplifier le modèle en éliminant les redondances et en réduisant la complexité globale. Dans le troisième cas, l'ajustement de PreDepRow a considérablement réduit le temps de calcul, soulignant l'importance d'une bonne gestion du modèle. Ces ajustements soulignent l'importance du réglage des paramètres pour améliorer l'efficacité du solveur, permettant ainsi des résultats plus rapides et plus précis.

4.2. Résultats de l'analyse comparative pour les optimiseurs

Le tableau 4 présente les résultats de l'analyse comparative des optimiseurs dans 34 instances. Les instances ont été organisées en fonction de l'objectif de production, car une augmentation du temps de calcul (Tc) a été constatée à mesure que cet objectif augmentait. Les résultats ont été comparés entre le modèle par défaut (PMSPOHO-D), le modèle optimisé (PMSPOHO-T) et le modèle avec l'optimiseur gratuit (PMSPOHO-CBC). Les paramètres critiques pour la comparaison étaient Tc et GAP, qui mesurent le modèle de complexité et le solveur d'efficacité. Dans cette étude, la valeur GAP est fournie directement par l'optimiseur Gurobi dans le cadre de la sortie de sa solution.

Le modèle optimisé (PMSPOHO-T) gère systématiquement des temps de calcul inférieurs à ceux des autres modèles. Cela indique que le réglage des paramètres a considérablement amélioré l' efficacité du solveur. Le GAP, qui est la différence entre la solution la plus connue et la meilleure borne inférieure trouvée par le solveur, est de 0,00 % pour PMSPOHO-D et PMSPOHO-T, ce qui indique que les solutions trouvées sont optimales. Cependant, PMSPOHO-CBC présente un GAP plus élevé dans plusieurs cas, reflétant une précision moindre et soulignant les limites d'un optimiseur commercial.

PMSPOHO-T se distingue par les meilleurs résultats en termes de Tc, résolvant les instances plus rapidement tout en maintenant la précision de la solution. Cela contraste fortement avec le modèle avec l' optimiseur gratuit (PMSPOHO-CBC), qui montre dans de nombreux cas un GAP plus élevé, soulignant ainsi ses limites. Cela démontre également que la complexité du problème est NP-Hard, ce qui signifie qu'à mesure que les instances augmentent, des résultats optimaux pourraient être obtenus après des délais beaucoup plus longs, faisant d'un optimiseur gratuit une option inefficace pour les applications sur le terrain.

La figure 3 compare le temps de calcul (Tc) et le Cmax pour les optimiseurs PMSPOHO-D, PMSPOHO-T et PMSPOHO-CBC, par exemple 34. PMSPOHO-D et PMSPOHO-T atteignent le même Cmax de 183, mais PMSPOHO-T est plus efficace en termes de temps de calcul, avec 671,8 contre 798,8 pour PMSPOHO-D. D'autre part, PMSPOHO-CBC présente les pires performances, avec un Cmax de 229 et un temps de calcul de 3600. Cela indique que PMSPOHO-T est le plus efficace parmi les trois optimiseurs évalués, atteignant un bon équilibre entre le temps de calcul et le Cmax. .

Machine Translated by Google

Tableau 4. Résultats de l'analyse comparative pour les optimiseurs par instance.

Exemple	P vm Cmax	PMSPOHO-D			PMSPOHO-T			PMSPOHO-CBC		
			Tc	ÉCART Cmax		Tc	ÉCART Cmax			ÉCART TC
	20 20 2 20 16 2 20	29	14,52	0,00% 9,51	29	9,73	0,00% 5,92	30	3600	4,58%
1	10 2 20 8 2 20 6 2	29	0,00%	3,82 0,00%	29	0,00%	2,91 0,00%	30	3600	4,64%
2	20 5 2 20 4 2 20 3	29	2,51	0,00% 1,6	29	1,85	0,00% 1,58	31	3600	5,21%
3	2 81 20 4 81 16 4	29	0,00%	1,41 0,00%	29	0,00%	1,03 0,00%	30	3600	4,82%
4	81 10 4 81 8 4 81	31	1,13	0,00% 0,9	31	1,01	0,00% 1,82	32	3600	3,77%
5	6 4 81 4 4 81 20 6	34	0,00%	28,25 0,00%	34	0,00%	26,74 0,00%	35	3600	2,36%
6	81 16 6 81 10 6 81	40	18,34	0,00% 49,23	40	18,61	0,00% 46,00	41	3600	2,56%
7	8 6 81 6 6 90 5 3	51	69,59	0,00% 66,54	51	%	50,86 0,00%	51	1846	0,00%
8	90 4 3 90 3 3 200	55	0,00%	76,29 0,00%	55	63,01	0,00% 59,78	60	3600	8,52%
9	19 4 200 18 4 200	55	42,12	0,00% 318,67	55	0,00%	42,26 0,00%	60	3600	8,20%
10	17 4 200 16 4 200	64	0,00%	94,3 0,00 %	64	130,11	0,00% 76,82	68	3600	6,45%
11	15 4 200 14 4 200	78	100,46	0,00%	78	0,00 %	100,13 0,00%	82	3600	5,50%
12	13 4 200 12 4 200	100	145,77	0,00% 42,98	100	90,32	0,00% 26,41	106	3600	5,72%
13	11 4 200 10 4 200	147	0,00%	83,77 0,00%	147	0,00%	40,36 0,00%	152	3600	3,27%
14	9 4 200 8 4	38	44,52	0,00%	38	30,78	0,00% 1627,21	42	3600	10,25%
15		42	1784,21	0,00%	42	0,00%	1437,64 0,00%	46	3600	9,86%
16		63	1693,64	0,00%	63	1463,63	0,00%	70	3600	10,52%
17		77	1634,63	0,00%	77	1400,24	0,00% 1437	85	3600	10,73%
18		98	1640,24	0,00% 158	98	.04	0,00% 1291,43	110	3600	12,72%
19		130	4,04	0,00% 1554,43	130	0,00%	1349,23	163	3600	25,32%
20		161	0,00%	1532,23	161	0,00%	1372,64	202	3600	25,47%
21		212	0,00%	1524,64	212	0,00%	1019,45	259	3600	22,23%
22		125	0,00%	1279,45	125	0,00%	746,6 0,00%	161	3600	29,03%
23		125	0,00%	1017,6 0,00%	125	615,6	0,00% 671,8	161	3600	28,53%
24		125	865,6	0,00% 798,8	125	0,00%		161	3600	28,93%
25		125	0,00%		125			161	3600	28,73%
26 27		125			125			164	3600	31,45%
28		125			125			159	3600	27,32%
29		125			125			161	3600	29,05%
30		151			151			194	3600	28,63%
31		151			151			189	3600	25,43%
32		151			151			186	3600	23,23%
33		166			166			203	3600	22,01%
34		183			183			229	3600	24,93%

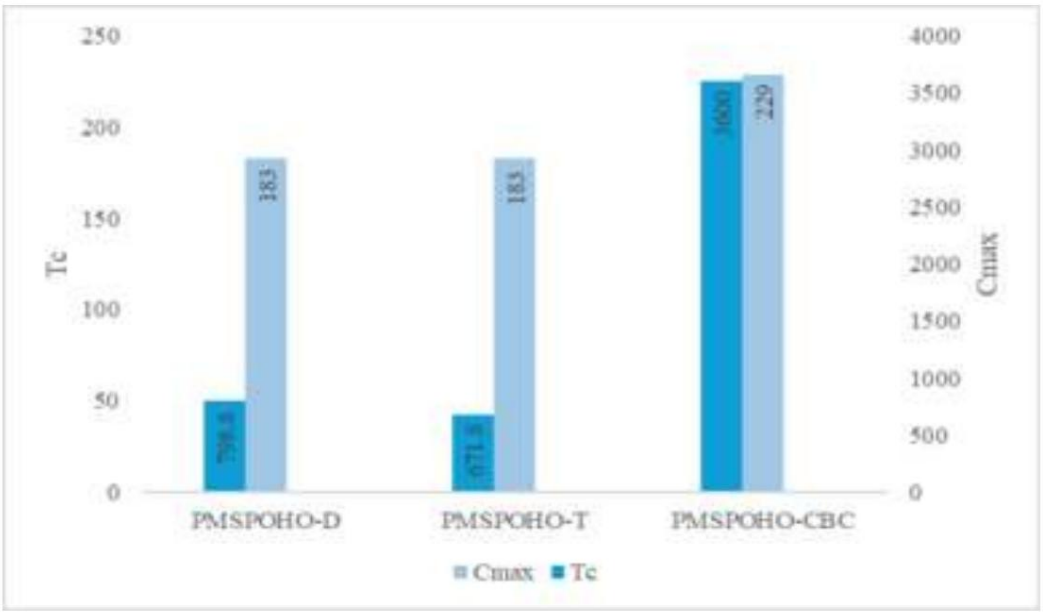


Figure 3. Comparaison du temps de calcul et de Cmax par exemple 34.

4.3. Informations sur l'approche tactique

Le tableau 5 montre l'efficacité des algorithmes appliqués. L'algorithme 1 calcule le nombre minimum d'équipements de transport, soit huit équipements. De l'autre D'autre part, l'algorithme 2 effectue une évaluation k-Cmax , permettant une analyse plus appropriée entre ces deux variables. Cela améliore la prise de décision tactique en montrant comment la variation de Cmax a un impact sur k et vice versa. La minimisation de la Cmax entraîne une réduction significative valeurs (entre 33 et 49), bien qu'une augmentation de k permette aux planificateurs d'équilibrer le temps l'efficacité et les coûts opérationnels. L'algorithme 2 fournit une approche plus complète et équilibrée

Machine Translated by Google

perspective pour la prise de décision tactique et est soutenu par l'algorithme 1, qui définit le nombre de ressources nécessaires pour atteindre l'objectif de production.

Tableau 5. Résultats de l'optimisation des algorithmes proposés.

			Réduire k		Minimiser la Cmax		
v	m	P.	Cmaxubk	Tc	Cmaxlb	k	Tc
19	4	200	51	2350 151,46 2350	33	6998	54,97
18	4	200	51	219,68 2350	33	6648	46.31
17	4	200	51	117,08 2350	33	6298	38.57
16	4	200	51	99,44 2350	33	5948	33.81
15	4	200	51	125,16 2350	33	5598	31.16
14	4	200	51	162,80 2350 75,34	33	5248	26.78
13	4	200	51		34	4898	24.63
12	4	200	51	2350 58,55	35	4548	20.73
11	4	200	51	2350 111.22	37	4198	21h40
10	4	200	51	2350 103h30	40	3848	32,76
9	4	200	50	2350 44,62	44	3498	19h40
8	4	200	51	2350 76,94	49	3148	15.93

À mesure que les délais d'exécution raccourcissent, les coûts augmentent de manière disproportionnée, ce qui indique un compromis entre le temps et la rentabilité des opérations de transport de minerai. Les courbes plus raides en bas les délais de réalisation impliquent des augmentations de coûts significatives pour des opérations plus rapides, tandis que les délais plus plats les pentes à des moments plus élevés suggèrent un seuil de coût minimum. L'éventail plus large des délais de réalisation avec davantage d'équipements de transport signifie que, bien que des équipements supplémentaires puissent accélérer les opérations, cela entraîne également des coûts plus élevés, ce qui souligne l'importance de l'équilibrage contraintes budgétaires et urgence opérationnelle dans la programmation des opérations minières.

La figure 4 montre la relation entre les coûts et le délai de réalisation pour différents nombre d'équipements de transport. Il y a une diminution marquée des coûts avec seulement une légère augmentation dans le temps d'achèvement. Une approche tactique pour déterminer la quantité minimale de transport l'équipement nécessaire pour atteindre les objectifs de production ne doit pas seulement prendre en compte le coût et l'achèvement le temps mais aussi la pente de chaque droite, qui supporte l'application des algorithmes proposées dans ce document. Par exemple, passer de neuf à huit unités de transport augmente le temps de réalisation tout en réduisant considérablement les coûts. Cela suggère que le fonctionnement avec huit unités de transport pourrait être une décision tactique efficace si le temps supplémentaire ne suffisait pas. ne compromettra pas les autres priorités de l'opération de transport du minerai.

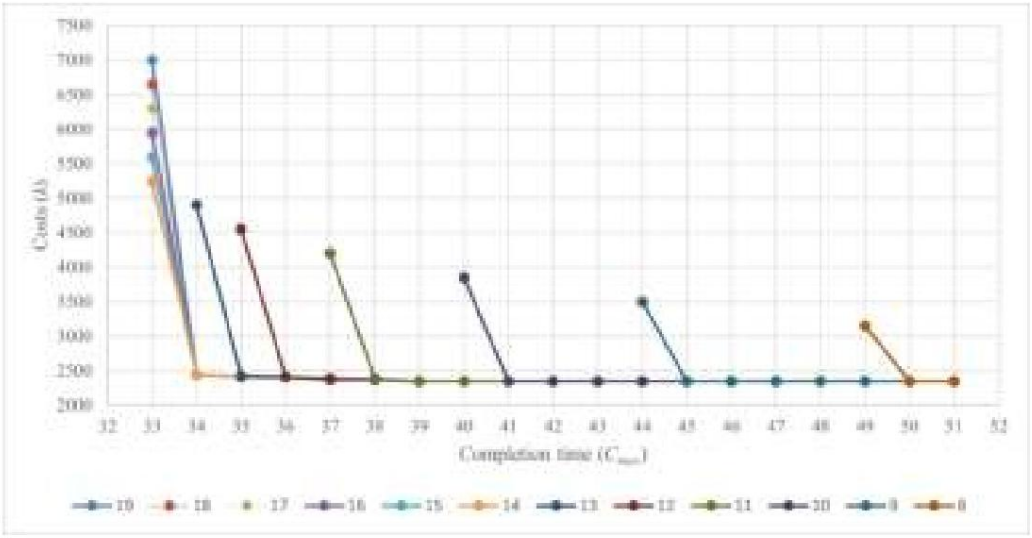


Figure 4. Coûts k par rapport au temps de réalisation Cmax pour les instances proposées dans le tableau 5.

Machine Translated by Google

Les décideurs doivent mettre en balance l'augmentation des coûts associée à l'utilisation d'un plus grand nombre d'équipements et la réduction du temps de transport. Au-delà d'un certain point, l'ajout d'unités de transport supplémentaires diminue l'avantage de réduire les délais de réalisation, mais entraîne des coûts considérablement plus élevés. Identifier le moment où le surcoût ne justifie plus la diminution du temps de transport est essentiel pour optimiser les opérations et gérer efficacement les ressources.

4.4. Aperçus de l'approche opérationnelle

La figure 5 montre le programme, par exemple 34, qui considère huit équipements de transport et quatre équipements de chargement. Celui-ci visualise le programme de chargement du minerai pour une opération minière au cours des 50 premiers voyages de la cible de production. Il est intéressant de voir comment l'allocation est organisée pour garantir que la quantité minimale nécessaire d'équipement de transport soit maintenue en mouvement constant, ce qui indique une évaluation minutieuse des temps de chargement et des intervalles entre les tâches. Cette stratégie garantit qu'aucun équipement de chargement ne reste inutilisé et que l' équipement de transport est efficacement distribué pour atteindre l'objectif de production.

En minimisant la quantité d'équipement de transport, l'opération peut devenir plus rentable et améliorer la logistique en réduisant la congestion et les goulots d'étranglement potentiels dans l'opération de transport de minerai. Cette approche équilibrée entre efficience et efficacité en matière de planification est cruciale pour une opération réussie qui s'adapte aux exigences dynamiques et souvent imprévisibles de l'exploitation minière.

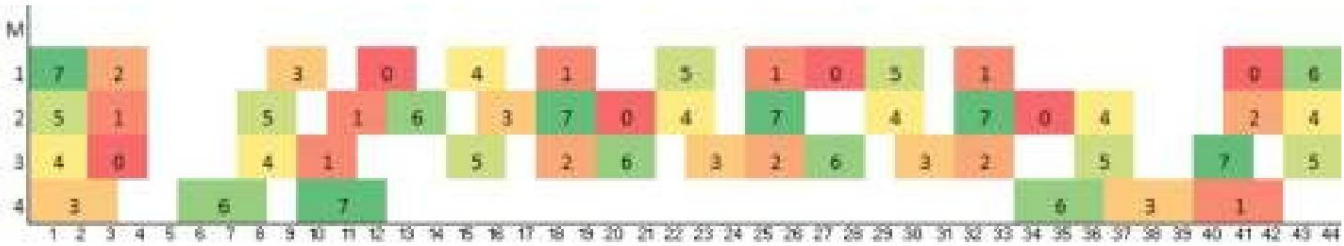


Figure 5. Programme par exemple 34 en considérant le temps de réalisation minimum Cmax.

Le tableau 6 fournit une répartition détaillée de l'utilisation des équipements de chargement et de transport dans une opération minière, mettant en évidence l'allocation et l'utilisation de ces ressources dans la production . On observe que les équipements de chargement deux et trois sont les plus utilisés, avec des pourcentages d'utilisation de 32 % et 30 %, respectivement, ce qui indique une forte demande et une utilisation fréquente. En revanche, le véhicule de chargement quatre est le moins utilisé, avec un taux d'utilisation de seulement 12 Cela suggère que les ressources ne sont pas entièrement optimisées, d'autant plus que le chargement du véhicule quatre pourrait être le plus coûteux à exploiter, car les équipements moins fréquemment utilisés nécessitent généralement un entretien régulier et occupent un espace précieux qui pourrait être mieux utilisé.

La sous-utilisation des équipements de chargement coûteux soulève d'importantes questions de gestion des coûts et d'efficacité opérationnelle. Le coût élevé de cet équipement peut être dû à des fonctionnalités spécialisées qui nécessitent une maintenance plus sophistiquée ou un investissement initial plus élevé. Cependant, son utilisation sous-optimale implique que cet investissement doit produire les rendements escomptés, ce qui est essentiel dans les opérations visant à maximiser l'efficacité et la rentabilité. Cette situation souligne le besoin urgent de réévaluer la répartition des tâches et l'allocation des ressources pour garantir que tous les équipements de chargement sont utilisés d'une manière qui justifie leur coût, optimisant ainsi l'ensemble de l'opération et réduisant le gaspillage potentiel des ressources.

Les résultats démontrent que le modèle proposé est efficace en termes de temps de calcul et que les algorithmes proposés sont des outils tactiques utiles pour une prise de décision plus efficace concernant l'allocation des équipements de transport nécessaires pour atteindre les objectifs fixés dans les opérations de transport de minerai. Cependant, il est essentiel de reconnaître que l'analyse d'instances plus significatives en tant que modèle mathématique pourrait compliquer le temps requis pour obtenir des résultats, même avec les paramètres de l'optimiseur finement réglés. Il sera donc essentiel d'étendre les délais de recherche de solutions et de développer des méthodes approximatives pour réduire ces délais et parvenir à des solutions pratiques et exécutables.

Machine Translated by Google

Tableau 6. Utilisation des équipements de chargement et de transport.

	Équipement de chargement					
Équipement de transport	1	2	3	4	Total	Utilisation
0	3	2	1	0	6	12%
1	3	2	1	1	7	14%
2	1	1	3	0	5	10%
3	1	1	2	2	6	12%
4	1	4	2	0	7	14%
5	2	2	3	0	7	14%
6	1	1	2	2	6	12%
7	1	3	1	1	6	12%
Total	13	16	15	6	50	
Utilisation	26%	32%	30%	12%		

5. Conclusions

Le modèle proposé démontre une efficacité significative en termes de calcul
temps, équilibrant efficacement le coût et le temps d'achèvement des opérations de transport de minerai. Candidature
les deux algorithmes ont fourni des outils tactiques qui améliorent la prise de décision concernant le
attribution du matériel de transport. L'efficacité du modèle est évidente dans le calcul
résultats, montrant que le réglage des paramètres peut améliorer considérablement les performances du solveur.
Le modèle optimisé a systématiquement géré des temps de calcul inférieurs aux paramètres par défaut,
soulignant l'importance d'ajustements minutieux des paramètres. De plus, en intégrant
les temps de largage dynamiques et les différentes vitesses des équipements de transport dans le modèle ajoutent une couche
de réalisme, garantissant que la planification reflète les conditions opérationnelles réelles. Cet aspect
optimise davantage le processus d'allocation, rendant le modèle robuste sur le plan informatique.

Malgré les résultats, le modèle présente des limites. Une des principales limites est son évolutivité.
À mesure que la taille des instances augmente, le temps nécessaire pour obtenir des résultats peut devenir
prohibitif, même avec des paramètres optimisés. Ce problème est inhérent aux modèles mathématiques
traiter des problèmes complexes et NP-difficiles. De plus, le fait que le modèle s'appuie sur des données exactes
les méthodes d'optimisation ne peuvent être utiles que parfois pour la prise de décision en temps réel dans
des environnements dynamiques où des réponses rapides sont essentielles.

Les travaux futurs devraient se concentrer sur l'allongement des périodes de recherche de solutions et sur le
développement de méthodes approximatives pour remédier à ces limitations. Ces méthodes, notamment heuristiques ou
Les approches métaheuristiques peuvent réduire les temps de calcul et fournir des solutions pratiques
et exécutables quasi optimales dans des scénarios du monde réel. De plus, l'intégration de la machine
des techniques d'apprentissage pour prédire et ajuster dynamiquement les paramètres pourraient améliorer la
performances et adaptabilité du modèle [58].

Contributions des auteurs : Conceptualisation, LT-T., CA et DA-M. ; méthodologie, LT-T. et CA ;
logiciel, LT-T. et CG ; analyse formelle, LT-T. et AP ; enquête, LT-T. et AP ; données
curation, LT-T.; rédaction – préparation du projet original, LT-T. ; rédaction – révision et édition, LT-T.,
CA et DA-M. ; visualisation, LT-T.; supervision, CA et DA-M.; administration de projet, LT-T.
Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : L'APC a été financé par l'Université de Los Andes.

Déclaration du comité d'examen institutionnel : sans objet.

Déclaration de consentement éclairé : sans objet.

Déclaration de disponibilité des données : Les données présentées dans cette étude sont disponibles sur : https://github.com/LuisTarazonaTorres/Problème-de-planification-de-machines-parallèles---Hualing-Equipment.git (consulté sur 29 juillet 2024).

Remerciements : Les auteurs remercient grandement tous les participants qui ont contribué au
conduite de cette étude.

Machine Translated by Google

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Références

1. Déroute, M. ; Parida, CK Optimisation des paramètres de dynamitage dans les mines à ciel ouvert. doctorat Thèse, Institut National de Technologie Rourkela, Odisha, Inde, 2007.
2. Meza Castro, JE Développe un modèle pour l'application de simulation sur un système de chargement et de transport de Desmonte dans une opération minière sur le tajo ouvert. Thèse de licence, Université pontificale catholique du Pérou, San Miguel, Pérou, 2011.
3. Vasquez, OC; Sepúlveda, JM; Córdova, F. Modélisation et optimisation des opérations de véhicules dans les mines souterraines de cuivre. Rem Rév. De Minas 2011, 64, 365-371. [Référence croisée]
4. Anchiraico, A. ; Rojas, K. Optimización del Sistema de Acarreo y Transporte en Labores de Preparación de las Zonas de Profundización Medianta la Metodología Six Sigma Operada por la ECM Zicsa en la Unidad Minera Inmaculada. Thèse de licence , Universidad Peruana de Ciencias Aplicadas, Santiago de Surco, Pérou, 2023.
5. Soofastaei, A. ; Karimpour, E. ; Chevaliers, P. ; Kizil, M. Opérations de chargement et de transport économes en énergie ; Springer : Cham, Suisse, 2018 ; pp. 121-146. [Référence croisée]
6. Choi, Y. ; Baek, J. ; Park, S. Examen des applications basées sur les SIG pour l'exploitation minière : planification, exploitation et gestion environnementale. Appl. Sci. 2020, 10, 2266. [Réf. croisée]
7. Owolabi, A. Sélection des équipements de chargement et de transport pour une production optimale dans une carrière de granit. Int. J. Min. Sci. 2019, 5, 35-40.
8. Marte Collado, JM Programación de la Producción en Máquinas Paralelas Sujeto a Adelantos, Retrasos y Fechas límite. Mémoire de maîtrise, Universitat Politècnica de València, Valence, Espagne, 2017.
9. Salazar-Hornig, E. ; Soto-Gavilán, G. Minimisation de la durée de vie du problème des machines parallèles non liées aux temps de configuration dépendant de la sécurité au moyen d'un algorithme hybride VNS/ACO. Révérend Ing. Univ. De Medellín 2021, 20, 171-184. [Référence croisée]
10. Hong, Caroline du Sud ; Lin, BM Planification de machines dédiées parallèles avec graphiques de conflits. Calculer. Ind. Ing. 2018, 124, 316-321. [Référence croisée]
11. Koryaguine, M. ; Voronov, A. Améliorer l'organisation des systèmes pelle-camion dans les mines de charbon à ciel ouvert. Transp. Problème. 2017, 12, 113-122. [Référence croisée]
12. Oliskevych, M. ; Kovalyshyn, S. ; Magats, M. ; Chevtchouk, V. ; Sukach, O. L'optimisation du planning de la flotte de camions compte tenu de leur interaction et des restrictions de l'accord européen de travail des équipages. Transp. Problème. 2020, 15, 157-170. [Référence croisée]
13. Wu, L. ; Wang, S. Méthodes exactes et heuristiques pour résoudre le problème de planification de machines parallèles avec des tâches multiprocesseurs. Int. J.Prod. Écon. 2018, 201, 26-40. [Référence croisée]
14. Elidrissi, A. ; Benmansour, R. ; Benbrahim, M. ; Duvivier, D. Formulations mathématiques pour l'ordonnancement de machines parallèles problème avec un seul serveur. Int. J.Prod. Rés. 2021, 59, 6166-6184. [Référence croisée]
15. Gan, HS; Wirth, A. ; Abdekhodaee, A. Un algorithme de branchement et de prix pour le cas général de planification de machines parallèles avec un serveur unique. Calculer. Opéra. Rés. 2012, 39, 2242-2247. [Référence croisée]
16. Kim, MA ; Lee, YH Modèles MIP et algorithme hybride pour minimiser la durée de vie du problème de planification des machines parallèles avec un seul serveur. Calculer. Opéra. Rés. 2012, 39, 2457-2468. [Référence croisée]
17. Hasani, K. ; Kravchenko, SA ; Werner, F. Minimiser les interférences pour la planification de deux machines parallèles avec un seul serveur. Int. J. Prod. Rés. 2014, 52, 7148-7158. [Référence croisée]
18. Kosch, S. ; Beck, JC Un nouveau modèle MIP pour la planification de lots parallèles avec des tailles de tâches non identiques. Dans Actes de l'intégration des techniques d'IA et de RO dans la programmation par contraintes : 11e conférence internationale (CPAIOR 2014), Cork, Irlande, 19-23 mai 2014 ; Actes 11 ; Springer : Berlin/Heidelberg, Allemagne, 2014 ; pp. 55-70.
19. Sharma, A. ; Zhao, F. ; Sutherland, JW Planification économique d'une entreprise manufacturière fonctionnant selon un tarif d'électricité en fonction de l'heure de consommation . J. Nettoyer. Prod. 2015, 108, 256-270. [Référence croisée]
20. Bajany, D. ; Xia, X. ; Zhang, L. Un modèle MILP pour la planification des camions-pelles afin de minimiser la consommation de carburant. Procédure énergétique 2017, 105, 2739-2745. [Référence croisée]
21. Che, A. ; Zhang, S. ; Wu, X. Planification de machines parallèles sans rapport avec l'énergie et soucieuse de l'énergie selon les tarifs de l'électricité en fonction de l'heure d'utilisation. J. Nettoyer. Prod. 2017, 156, 688-697. [Référence croisée]
22. Cheng, J. ; Chu, F. ; Zhou, M. Un modèle amélioré pour la planification de machines parallèles en fonction du prix de l'électricité en fonction de l'heure d'utilisation. IEEETrans. Automatique. Sci. Ing. 2017, 15, 896-899. [Référence croisée]
23. Elidrissi, A. ; Benmansour, R. ; Benbrahim, M. ; Duvivier, D. Formulations MIP pour un problème de planification de machines parallèles identiques avec un seul serveur. Dans Actes de la 4e Conférence internationale sur l'optimisation et les applications (ICOA) 2018, Mohammedia, Maroc, 26-27 avril 2018 ; IEEE : Piscataway, New Jersey, États-Unis, 2018 ; p. 1 à 6.
24. Kaabi, J. ; Harrath, Y. Planification sur des machines parallèles uniformes avec des contraintes d'indisponibilité périodiques. Int. J.Prod. Rés. 2019, 57, 216-227. [Référence croisée]
25. Quiroz Silva, EE Programme d'objets pour les dimensions et l'attribution d'une flotte de camions à un Empresa Minera. Interfaces 2021, 87-112. [Référence croisée]
26. Hsu, CL; Liao, JR Deux problèmes d'ordonnancement de machines parallèles avec contrainte de fonction. Discret. Dynam. Nat. Soc. 2020, 2020, 1–6. [Référence croisée]

Machine Translated by Google

27. Tapia, E. ; Salazar Araya, A. ; Saavedra, N. ; Nehring, M. ; Mora, J. Une analyse des stratégies de camion plein ou de godet plein dans les opérations de chargement et de transport des mines à ciel ouvert. Int. J. Min. Récupérez. Environ. 2021, 35, 1–11. [Référence croisée]
28. Tigreros Cardenas, ID Propuesta Para la Programación de Producción en Máquinas Paralelas con Tiempos de Alistamiento en unea Empresa del Sector Agroindustrial. 2021. Disponible en ligne : https://repositorio.uceva.edu.co/bitstream/handle/20.500.129 93/2744/T00031983.pdf?sequence=1&isAllowed=y (consulté le 29 juillet 2024).
29. Yunusoglu, P. ; Topaloglu Yildiz, S. Approche de programmation par contraintes pour un problème de planification de machines parallèles sans rapport avec des contraintes multi-ressources avec des temps de configuration dépendants de la séquence. Int. J.Prod. Rés. 2022, 60, 2212-2229. [Référence croisée]
30. Ma, X. ; Fu, Y. ; Gao, K. ; Zhu, L. ; Sadollah, A. Un problème de planification et de routage multi-objectifs pour les services de soins de santé à domicile via l'optimisation du brainstorming. Système complexe. Modèle. Simul. 2023, 3, 32-46. [Référence croisée]
31. Feng, L. ; Chen, G. ; Zhou, S. ; Zhou, X. ; Jin, M. Un problème de planification de machines parallèles non liées à l'efficacité énergétique avec le traitement par lots et les prix de l'électricité en fonction de l'heure d'utilisation. Mathématiques 2024, 12, 376. [CrossRef]
32. Hu, B. ; Xiong, Z. ; Soleil, A. ; Yuan, Y. Planification des véhicules de transport de conteneurs dans les mines de charbon à ciel ouvert basée sur l' algorithme hybride GA-GWO . Appl. Sci. 2024, 14, 3986. [Réf. croisée]
33. Eivazy, H. ; Askari-Nasab, H. Un modèle de programmation linéaire mixte en nombres entiers pour la planification de la production à court terme des mines à ciel ouvert. Min. Technologie. 2012, 121, 97-108. [Référence croisée]
34. Anjomshoa, H. ; Albrecht, A. ; Lee, D. ; Pudney, P. Transport efficace de véhicules dans les mines souterraines. Min. Technologie. 2012, 121, 83-90. [Référence croisée]
35. Gligoric, Z. Optimisation du système de transport dans une mine souterraine à petite échelle à l'aide d'ensembles flous. Conception de systèmes. p. 63. Disponible en ligne : https://www.researchgate.net/publication/287397565_Optimization_of_haulage_system_in_an_underground_ small_scale_mine_using_fuzzy_sets (consulté le 29 juillet 2024).
36. Gonen, A. ; Malli, T. ; Kose, H. Sélection d'un système de transport de minerai pour une mine souterraine métallifère/Dobór systemu transportu rud w kopalni podziemnej rud metalu. Cambre. Min. Sci. 2012, 57, 779-785. [Référence croisée]
37. Eivazy, H. ; Askari-Nasab, H. Un modèle hiérarchique d'optimisation de la planification de la production des mines à ciel ouvert. Int. J. Min. Mineur. Ing. 2012, 4, 89-115. [Référence croisée]
38. Tom Socarras, JN Árbol de Decisión para la Gestión y Control de Costo de Transporte Interior Mina. Thèse de licence, Université du Chili, Santiago, Chili, 2014.
39. Uribe Neira, PA Analyse de la fonctionnalité et du développement de l'opération semi-automatique sur les équipements de chargement ; Acarreo y Descarga en Minería Subterránea Load Haul-Dump Vehicle (LHD). Thèse de licence, Université du Chili, Santiago, Chili, 2014.
40. Wi'ecek, D. ; Burduk, A. ; Kuric, I. L'utilisation de l'ANN pour améliorer l'efficacité et assurer la stabilité de l'extraction du minerai de cuivre processus. Acta Montán. Slovaquie 2019, 24, 1–14.
41. Wiest Goyeneche, JR Modèle de programmation d'opérations dans une mine du ciel ouvert : application dans l'organisation Couronne. Mémoire de maîtrise, Universidad de La Sabana, Chía, Colombie, 2019.
42. Salomon, L. ; Ortiz, A. Modèle conceptuel de systèmes de chargement et de transport de minerai à ciel ouvert. Univ. Cien. Et Tecnol. 2020, 24, 41-50.
43. Gaspar, S. ; Jhasmin, C. Meilleure productivité en équipements de transport et de transport de minéraux et desmonte en la Veta Gavia–Niveau 100, Unidad Minera Huarón. Thèse de licence, Universidad Continental, Los Olivos, Pérou, 2020.
44. Li, N. ; Feng, S. ; Oui, H. ; Wang, Q. ; Jia, M. ; Wang, L. ; Zhao, S. ; Chen, D. Modèle d'optimisation de la répartition pour les équipements de transport entre les chantiers sur la base de la planification des ressources minières à court terme. Métaux 2021, 11, 1848. [CrossRef]
45. Elie, K. ; Joseph, G. ; Samuel, M. ; Mauti, D. Optimisation du système de transport par camion-pelle dans une mine à ciel ouvert en utilisant une approche de mise en file d'attente. Arabe. J. Geosci. 2021, 14, 973. [CrossRef]
46. Huisa Supho, A. Optimización de Equipos de Cargío y Acarreo en el Tramo Botaderos-Trituradora Thyssen Krupp, pour l' augmentation de la production dans une entreprise cimentière. Thèse de licence Universidad Tecnológica Del Perú, Lima, Pérou, 2021.
47. Choi, Y. ; Nguyen, H. ; Bui, XN ; Nguyen-Thoi, T. Optimisation des performances du système de camions de transport pour la production de minerai dans les mines à ciel ouvert à l'aide de méthodes basées sur le Big Data et l'apprentissage automatique. Ressource. Politique 2022, 75, 102522. [CrossRef]
48. Li, N. ; Feng, S. ; Lei, T. ; Oui, H. ; Wang, Q. ; Wang, L. ; Jia, M. Optimisation du plan de rééchelonnement des équipements de transport minier souterrain basé sur une simulation de panne aléatoire. Durabilité 2022, 14, 3448. [CrossRef]
49. Abolghasemian, M. ; Kanafi, AG; Daneshmand-Mehr, M. Optimisation multiobjectif basée sur la simulation du système de transport minier à ciel ouvert : une méthode NBI modifiée et une approche de méta-modélisation. Complexité 2022, 2022, 3540736. [CrossRef]
50. Shamsi, M. ; Pourrahimian, Y. ; Rahmanpour, M. Optimisation de la planification de la production des mines à ciel ouvert en tenant compte d' un système de transport optimal entre le transport par camion et le concassage et le transport semi-mobiles dans la fosse. Int. J. Min. Récupérez. Environ. 2022, 36, 142-158. [Référence croisée]
51. Parichehreh, M. ; Gholizadeh, H. ; Fathollahi-Fard, A. ; Wong, K. Un problème de planification de machines parallèles sans rapport avec l'efficacité énergétique avec effet d'apprentissage des opérateurs et détérioration des emplois. Int. J. Environ. Sci. Technologie. 2024, 1-26. [Référence croisée]
52. Graham, RL; Lawler, EL; Lenstra, JK; Kan, AR Optimisation et approximation dans le séquençage et l'ordonnancement déterministes : une enquête. Dans Annales de mathématiques discrètes ; Elsevier : Amsterdam, Pays-Bas, 1979 ; Volume 5, pages 287 à 326.
53. Pinedo, M. ; Hadavi, K. Scheduling : Théorie, algorithmes et développement de systèmes. Dans Actes des actes de recherche opérationnelle 1991 : documents de la 20e réunion annuelle/Vorträge der 20. Jahrestagung ; Springer : Berlin/Heidelberg, Allemagne, 1992 ; p. 35-42.
54. Yu, K. ; Yang, modèle J. MILP et algorithme d'horizon roulant pour la planification des grues dans un terminal à conteneurs de stockage hybride. Mathématiques. Problème. Ing. 2019, 2019, 4739376. [Réf. croisée]

Machine Translated by Google

55. Tarazona-Torres, LE; Amaya-Guio, Californie ; Álvarez-Martínez, D. Équipe asynchrone pour le problème de planification de Flow Shop : une étude de cas. Dans Actes de la 4e Conférence internationale sud-américaine sur l'ingénierie industrielle et la gestion des opérations, Lima, Pérou, 9-11 mai 2023.
56. Gurobi Optimisation, LLC. Paramètres de l'optimiseur Gurobi version 10.0.2 ; Gurobi Optimization, LLC. : Beaverton, OR, États-Unis, 2024.
57. Cuellar-Usaquén, D. ; Palacio, A. ; Ospina, E. ; Botero, M. ; Álvarez-Martínez, D. Modélisation et résolution du problème de coupe du point final. Int. Trans. Opéra. Rés. 2023, 30, 800-830. [Référence croisée]
58. Hou, Y. ; Wang, H. ; Huang, X. Un algorithme évolutif multi-objectif basé sur le Q-learning pour la production verte intégrée et problèmes de planification de la distribution. Ing. Appl. Artif. Intell. 2024, 127, 107434. [Réf. croisée]

Avis de non-responsabilité/Note de l'éditeur : Les déclarations, opinions et données contenues dans toutes les publications sont uniquement celles du ou des auteurs et contributeurs individuels et non de MDPI et/ou du ou des éditeurs. MDPI et/ou le(s) éditeur(s) déclinent toute responsabilité pour tout préjudice corporel ou matériel résultant des idées, méthodes, instructions ou produits mentionnés dans le contenu.