Machine Translated by Google

Enzyklopädie

Autor und Herausgeber

Elzein Salah Elzein

IEEE-Mitglied

- Beiträge zum Thema Algorithmen

IEEE Publisher, Großbritannien



Dieser Inhalt ist vertraulich

Machine Translated by Google

Algorithmen

MDPI

Artikel

Auf Convolutional Neural Networks basierendes digitales Diagnosetool zur Erkennung psychosomatischer Erkrankungen

Marta Narigina, Andrejs Romanovs* und Yuri Merkuryev

Abteilung für Modellierung und Simulation, Technische Universität Riga, Kipsalas-Straße 6A, LV-1048 Riga, Lettland; marta.narigina@rtu.lv (MN); jurijs.merkurjevs@rtu.lv (YM)
*** Korrespondenz: andrejs.romanovs@rtu.lv**

Zusammenfassung: Dieser Artikel bewertet die Fähigkeiten von Convolutional Neural Network (CNN)-Modellen bei der Emotionserkennung anhand von Gesichtsausdrücken und soll die Diagnose psychosomatischer Erkrankungen unterstützen, die typischerweise in klinischen Umgebungen gestellt wird. Unter Verwendung des FER-2013-Datensatzes wurden zwei CNN-Modelle entwickelt, die sechs Emotionen mit einer Genauigkeit von 64 % erkennen – wenn auch nicht gleichmäßig verteilt; sie zeigten eine höhere Effektivität bei der Identifizierung von „glücklich“ und „überrascht“. Die Bewertung wurde anhand mehrerer Leistungsmetriken durchgeführt – Genauigkeit, Präzision, Erinnerung und F1-Werte – sowie einer weiteren Validierung mit einer zusätzlichen simulierten klinischen Umgebung zur Überprüfung der Praktikabilität. Obwohl diese Untersuchung vielversprechende Werte für die zukünftige Verwendung zeigt , unterstreicht sie die Ne Diese Forschung unterstreicht den potenziellen Wert der KI als Ergänzung zu herkömmlichen Diagnoseansätzen und konzentriert sich dabei auf einen größeren Anwendungsbereich (breitere Datensätze) sowie Schwerpunktbereiche (multimodale Integration), die bei den Empfehlungen künftiger Studien berücksichtigt werden sollen. Diese Studie unterstreicht die Bedeutung von CNN-Modellen bei der Entwicklung psychosomatischer Diagnostik und der Förderung zukünftiger Entwicklungen auf der Grundlage **von Ethik und Patientenversorgung.**

Schlüsselwörter: Faltungsneuronale Netze; künstliche Intelligenz; Emotionserkennung; psychosomatische Erkrankungen; Gesichtsausdrucksanalyse; Echtzeit-Emotionserkennung

check for updates

Zitat: Narigina, M.; Romanovs, A.; Merkuryev, Y. Digitale Diagnostik auf Basis von Faltungs-Neuralnetzen Werkzeug zur Identifizierung von Psychosomatische Erkrankungen. Algorithmen **2024,** 17, **329. https://doi.org/10.3390/a17080329**

Wissenschaftliche Redakteurin: Gabriella Trucco

Empfangen: 29. Juni 2024
Überarbeitet: 23. Juli 2024
Akzeptiert: 23. Juli 2024
Veröffentlicht: 30. Juli 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.
Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/ 4.0/).

1. Einleitung

Die genaue Diagnose psychosomatischer Erkrankungen stellt eine große Herausforderung für die moderne Gesundheitsversorgung dar. Psychosomatische Erkrankungen, bei denen psychologische Faktoren die physiologischen Funktionen erheblich beeinflussen, äußern sich häufig in Symptomen, die denen rein körperlicher Erkrankungen ähneln. Diese Überschneidung kann zu Fehldiagnosen und in der Folge zu ungeeigneten Behandlungsstrategien führen **[1–3]. Mit der zunehmenden Integration künstlicher Intelligenz in Diagnoseprozesse** ergeben sich immer mehr Möglichkeiten, die Diagnosegenauigkeit mithilfe **technologischer Hilfsmittel zu verbessern [4,5].**

Jüngste Fortschritte im Bereich des maschinellen Lernens, insbesondere Convolutional Neural Networks (CNNs), haben vielversprechende Ergebnisse in verschiedenen medizinischen Anwendungen gezeigt, darunter bildbasierte Diagnostik und Verhaltensvorhersagen [6]. Die Fähigkeit dieser Technologie, subtile Muster in Gesichtsausdrücken zu analysieren und zu interpretieren, bietet einen neuartigen Ansatz zur Identifizierung emotionaler Hinweise, die auf psychosomatische Erkrankungen hinweisen [7].

Ziel dieser Studie ist es, die Lücke zwischen psychologischen Untersuchungen und traditionellen Diagnosemethoden zu schließen, indem ein CNN-basiertes Tool zur Analyse von Gesichtsausdrücken integriert wird, das ein Hilfsmittel zur Diagnose psychosomatischer Erkrankungen bietet. Diese Forschung dreht sich um mehrere

zentrale Fragen, darunter die folgenden: 1. Welche aktuellen Einschränkungen gibt es bei der Diagnose psy
nationale Methoden [8]?
2. Kann ein CNN-Modell, das darauf trainiert ist, emotionale Ausdrücke zu erkennen, die diagnostische Genauigkeit bei psychosomatischen Erkrankungen verbessern **[9]?**

Machine Translated by Google

3. Wie effektiv kann ein solches Modell in klinische Arbeitsabläufe integriert werden, um medizinisches Fachpersonal zu unterstützen, ohne dessen Expertenurteil zu ersetzen?

Indem dieser Artikel diese Fragen beantwortet, untersucht er die Machbarkeit und Wirksamkeit des Einsatzes eines CNN-basierten Diagnosehilfsmittels im klinischen Umfeld und verbessert so die allgemeine Genauigkeit und Effizienz der Diagnose psychosomatischer Erkrankungen [10].

Die Abschnitte sind sorgfältig strukturiert, um eine detaillierte Analyse der verschiedenen Aspekte zu bieten, die bei der Erstellung CNN-basierter Diagnosetools eine Rolle spielen. Jeder Teil ist so konzipiert, dass er jede Phase des Entwicklungsprozesses eingehend behandelt.

Abschnitt 2, Verwandte Arbeiten, untersucht verschiedene Arbeiten, die Deep-Learning- Methoden zur Verbesserung der Gesichtsemotionserkennung eingesetzt haben. Diese Techniken wurden anhand des FER-2013-Datensatzes ausgewertet und haben bemerkenswerte Verbesserungen bei der Genauigkeit der Emotionserkennung gezeigt.

Abschnitt 3, Methodik, beschreibt detailliert den Datensatz, die CNN-Architektur, den Trainingsprozess, die Vorverarbeitungsschritte und die multimodale Integration von Gesichtsanalyse und Pulsdaten zur Vorhersage psychischer Störungen.

Abschnitt 4, Bewertung von CNN-Modellen, beurteilt die Zuverlässigkeit und Wirksamkeit von CNN-Modellen anhand verschiedener Leistungsmesswerte und Kreuzvalidierungstechniken in simulierten klinischen Umgebungen.

Abschnitt 5, Ergebnisse und Analyse, präsentiert Leistungsmesswerte von CNN-Modellen, Stärken, Bereiche mit Verbesserungsbedarf und statistische Validierung durch Konfusionsmatrizen und Genauigkeitsdiagramme .

Abschnitt 6, Diskussion, interpretiert die Bedeutung der Ergebnisse, mögliche Verzerrungen, Einschränkungen, ethische Überlegungen und praktische Auswirkungen der Integration von KI-Tools im Gesundheitswesen.

Abschnitt 7, Schlussfolgerungen, fasst die wichtigsten Erkenntnisse zusammen und empfiehlt die Erweiterung von Datensätzen, die Durchführung von Validierungen in der realen Welt und die Kombination multimodaler Daten, um die Diagnosemöglichkeiten zu verbessern. Dabei werden ethische Überlegungen und die potenziellen Auswirkungen der KI auf die Verbesserung der Patientenversorgung hervorgehoben.

Jeder Abschnitt baut auf dem vorherigen auf – alles mit dem Ziel, eine umfassende Diskussion aufzubauen , die das enorme Potenzial der CNNs bei der Veränderung der **Diagnosemethoden in Gesundheitseinrichtungen unterstreicht .**

2. Verwandte Werke

Jüngste Fortschritte im Bereich Deep Learning und Computer Vision haben das Feld der Gesichtsemotionserkennung (FER) erheblich beeinflusst. In diesem Abschnitt werden verschiedene Studien besprochen, in denen Deep-Learning-Techniken eingesetzt wurden, um die Leistung von FER-Systemen zu verbessern, insbesondere solche, die mit dem FER-2013-Datensatz ausgewertet wurden.

In Referenz [11] haben Georgescu et al. einen Ansatz vorgestellt, bei dem Merkmale, die automatisch von einer Reihe von CNN-Architekturen gelernt wurden, mit handgefertigten Merkmalen über das Bag-of-Visual-Words-Modell (BOVW) kombiniert werden. Die Klassifizierung von Emotionen wird durch diese hybride Methodik erreicht, die ein lokales Lernframework verwendet und durch die kombinierte Verwendung sowohl automatischer als auch **handgefertigter Merkmalssätze** eine bessere **Leistung in Bezug auf die Genauigkeit** aufweist.

Pecoraro et al. [12] entwickelten ein Local Multi-Head Channel (LHC) Self-Attention-Modul, das als komplementäre Komponente zum Standard Convolutional Neural Network (CNN) bezeichnet wird. Dieser Mechanismus vermeidet die allgemeinen Aufmerksamkeitsaktionen und konzentriert sich stattdessen auf die Teile des Gesichts, die Schlüsselakteure im emotionalen Ausdruck sind. Es wurde gezeigt, dass man durch die Einführung von LHC-Modulen die Genauigkeit der Gesichtsausdruckserkennung steigern kann, indem man kritische Bereiche im Gesicht lokalisiert.

Fard et al. [13] führten den Adaptive Correlation (Ad-Corre) Loss ein, der aus drei Elementen besteht: Feature Discriminator, Mean Discriminator und Embedding Discriminator. Ad-Corre Loss wird in Verbindung mit dem üblichen Cross-Entropy Loss verwendet, um die Unterscheidungsfähigkeit von Merkmalsvektoren zu erhöhen, was folglich zu

Machine Translated by Google

bessere Genauigkeit der Klassifizierung. Diese neue Verlustfunktion wurde auf Xception getestet als sowie Resnet50-Modelle, und die Ergebnisse zeigten erhebliche Leistungsverbesserungen.

Ein neuartiges FER-Modell wurde von Vignesh et al. [14] vorgeschlagen, das eine Fusion von U-Net- basierten Segmentierungsblöcken in die VGG-19-Architektur darstellt. Dieser einzigartige Ansatz, gen Segmentierung VGG-19 legt den Schwerpunkt auf signifikante Merkmale aus einer Feature-Map, die in führt wiederum zu besseren Ergebnissen: erweiterte Merkmalsextraktion und höhere Genauigkeit in Emotionen erkennen.

Mukhopadhyay et al. [15] präsentierten eine Deep-Learning-basierte FER-Methode, die Texturmerkmale wie lokale binäre Muster (LBPs), lokale ternäre Muster (LTPs) und abgeschlossene lokale binäre Muster (CLBPs). Durch das Trainieren eines CNN auf diese Texturmerkmale, die Das Modell erreichte eine verbesserte Genauigkeit bei der Erkennung von Gesichtsemotionen.

Shahzad et al. [16] schlugen die ZFER-Methode vor, die auf Zonierung basiert und die Erkennung von Emotionen im Gesicht durch die Bereitstellung zusätzlicher Orientierungspunkte, einschließlich der Augen, Augenbrauen, Nase, Stirn und Mund. Dieser Ansatz verwendet den Hybrid-VGG-16 Modell, das auf der Grundlage zonenbasierter Orientierungspunkte lernt, um die Genauigkeit der Klassifizierung zu erhöhen Gesichtsausdrücke.

El Boudouri und Bohi [17] schlugen EmoNeXt vor, ein modifiziertes ConvNeXt-Design speziell für FER. EmoNeXt beinhaltet ein STN (Spatial Transformer Network) zur Konzentration auf funktionsreichen Bereichen im Gesicht neben Squeeze-and-Excitation-Blöcken, die es ermöglichen Erfassung kanalweiser Abhängigkeiten. Darüber hinaus fügten sie eine Selbstaufmerksamkeitsregularisierung hinzu Begriff, der darauf abzielt, Merkmalsvektoren effizient durch Komprimierung zu generieren. Ergebnisse basierend auf Experimente mit dem FER-2013-Datensatz zeigten, dass EmoNeXt besser ist als alle bestehenden hochmodernen Modellen, wenn es um die Genauigkeit der Emotionsklassifizierung geht.

Der Datensatz FER-2013, der in diesen Arbeiten häufig verwendet wird, umfasst 35.887 Graustufenbilder die in sieben emotionale Klassen eingeteilt werden: Wut, Ekel, Angst, Glück, Neutralität, Traurigkeit und Überraschung. Sein großes Volumen und seine Vielfalt an Daten machen diesen Datensatz zu einem Standardreferenzpunkt für die FER-Systembewertung [18].

Die Kombination dieser fortschrittlichen Deep Learning-Methoden (einschließlich Self-Attention Mechanismen und Hybridmodelle, die sowohl automatische als auch handgefertigte Funktionen haben) plus neuartige Verlustfunktionen haben die Leistung von Systemen, die für die Gesichtserkennung entwickelt wurden, erheblich verbessert Ausdruckserkennung. Diese Änderungen unterstreichen das Potenzial von Deep Learning bei der Erstellung Starke und zuverlässige FER-Modelle – EmoNeXt stellt auf diesem Gebiet einen deutlichen Fortschritt dar.

3. Methodik

Der primäre Datensatz, der in dieser Studie verwendet wurde, war der FER-2013-Datensatz [19] , der Die in Tabelle 1 bereitgestellte Abbildung umfasst 35.887 Bilder von menschlichen Gesichtern, jedes mit einem von sieben emotionale Ausdrücke wie folgt: Wut, Ekel, Angst, Glück, Neutralität, Traurigkeit und Überraschung. Dieser Datensatz ist im Bereich der Computervision für seine Vielfalt weithin anerkannt und Komplexität, indem es Gesichtsausdrücke enthält, die in einer Vielzahl realer Situationen erfasst wurden.

Tabelle 1. Test- und Trainingsbilder des FER-2013-Datensatzes.

Emotion	Trainingsbilder	Testbilder
Wütend	3995	958
Ekel	436	111
Furcht	4097	1024
Glücklich	7215	1774
Neutral	4965	1233
Traurig	4830	1247
Überraschung	3171	831

Vor der Verarbeitung wurden die Bilder auf eine einheitliche Größe von 48 × 48 Pixel skaliert. Die Pixelwerte wurden auf den Bereich [0, 1] normalisiert, um eine einheitliche Eingabe für das neuronale Netzwerk, das die Stabilisierung des Lernprozesses und die Verbesserung des Modells ermöglicht Konvergenz [20].

Machine Translated by Google

Um die Simulation verschiedener Gesichtsausdrücke und Zustände, die dem Modell im wirklichen Leben begegnen können, zu erweitern , wurden **neben Rotation plus Zoom plus horizontalem Spiegeln und Verschieben verschiedene Datenerweiterungsmethoden eingesetzt .** Dieser Schritt zielt **darauf ab, mehr Generalisierungsmöglichkeiten für das Modell zu erreichen.**

Obwohl der FER-2013-Datensatz detailliert und vollständig war, wurde die Emotion „Ekel“ aus dieser Studie herausgelassen, da sie eine viel geringere Anzahl von Beispielen als die anderen Emotionen enthielt. Im Trainingsset gab es 436 Bilder, die „Ekel“ ausdrückten, und im Testset waren es 111. Diese Asymmetrie innerhalb des Datensatzes könnte dazu führen, dass nicht ausreichend für diese spezielle Emotion trainiert wurde, was zu einem unzuverlässigen Modell mit einer Tendenz zu den Kategorien führt, die viele emotionale Beispiele enthalten. Unsere Entscheidung, „Ekel“ auszuschließen, zielte darauf ab, während des Modelltrainings einen ausgewogeneren Ansatz beizubehalten, damit die Vorhersagen unseres Modells für alle anderen E

3.1. CNN-Architektur und -Entwicklung

Der Trainingsprozess umfasste mehrere Epochen, wobei das Training aufgrund von Validierungsverlusten vorzeitig beendet wurde, um ein Übertraining zu verhindern. Ein benutzerdefinierter Scheduler wurde eingesetzt, um die Lernrate dynamisch zu regulieren. Mit zunehmender Anzahl von Epochen wurde die Lernrate reduziert, was zur Verfeinerung der Modellgewichte

Die Experimente wurden auf einem MSI Summit B15 A11 MT-Laptop durchgeführt, der mit einem Intel Core i7-Prozessor und 64 GB RAM ausgestattet war und auf einem Linux-Betriebssystem lie Jede Trainingssitzung, die 100 Epochen umfasste, dauerte ungefähr 10 Stunden.

Die Architektur des Convolutional Neural Network wurde entwickelt, um schrittweise detailliertere Merkmale aus den Gesichtsbildern

zu extrahieren. • Die ersten Schichten des Convolutional Neural Network sind so konzipiert, dass sie schrittweise detaillierte Merkmale aus den Gesichtsbildern extrahieren. Das Netzwerk besteht aus zwei Convolutional-Schichten mit jeweils 64 Filtern, gefolgt von einer Max-Pooling-Schicht. Diese Konfiguration ist darauf ausgelegt, grundlegende Gesichtsmerkmale wie Kanten und **einfache Texturen zu erfassen .**

• Die Zwischenschichten bestehen aus Sätzen von Faltungsschichten mit einer zunehmenden Anzahl von Filtern, mit 128 und 256 Filtern pro Satz. Auf jeden Satz folgt dann eine Max-Pooling-Schicht. Die oben genannten Schichten sind darauf ausgelegt, komplexere Merkmale zu erkennen, wie etwa die Bestandteile von Gesichtsausdrücken

[3]. • Die letzten Schichten des Netzwerks werden als „erweiterte Schichten“ bezeichnet. Ein letzter Satz tiefer Faltungsschichten mit 512 Filtern wird verwendet, um hochrangige Merkmale zu extrahieren , die für die Erkennung subtiler Nuancen in

Ausdrücken entscheidend sind. • Das Netzwerk wird mit vollständig verbundenen Schichten abgeschlossen, die als Klassifikatoren fungieren, d. h. Klassifizierungsschichten . Um Überanpassung zu vermeiden, verwendet das Modell eine Dropout-Regularisierung , während eine Batch-Normalisierung verwendet wird, um den mittleren Output während des gesamten

Im Trainingsprozess wurde der Adam-Optimierer [21] eingesetzt, der für seine Effektivität bei der Handhabung spärlicher Gradienten und seine anpassbaren Lernraten bekannt ist . Der **Trainingsprozess umfasste mehrere Epochen, wobei das Training aufgrund von Validierungsverlusten vorzeitig beendet wurde , um ein Übertraining zu verhindern. Ein benutzerdefinierter Scheduler wurde eingesetzt, um die Lernrate dynamisch zu regulieren. Mit zunehmender Anzahl von Epochen wurde die Lernrate reduziert, was zur Verfeinerung der Modellgewichte**

Das zweite Modell folgt einer ähnlichen, aber erweiterten Architektur und integriert zusätzliche Schichten und Techniken zur Leistungssteigerung wie folgt:

- **Erste Schichten: Ähnlich wie beim ersten Modell umfassen die ersten Schichten zwei Faltungsschichten mit jeweils 64 Filtern, gefolgt von Batch-Normalisierung und LeakyReLU-Aktivierung [23], abschließend eine Max-Pooling-Schicht und Dropout zur**

Regularisierung. • Zwischenschichten: Die Zwischenschichten bestehen aus zwei Sätzen Faltungsschichten mit 128 und 256 Filtern pro Satz, jeweils gefolgt von Batch-Normalisierung, LeakyReLU- Aktivierung, Max-Pooling- und Dropout-Schichten.

Machine Translated by Google

• Erweiterte Ebenen: Dieses Modell enthält einen zusätzlichen Satz tiefer Faltungsebenen mit 256 Filtern, gefolgt von Batch-Normalisierung und LeakyReLU-Aktivierung, bevor Max Pooling und Dropout angewendet
werden. • Letzte Ebenen: Nach dem Abflachen der Feature-Maps enthält das Modell eine dichte Ebene mit 256 Einheiten, Batch-Normalisierung, LeakyReLU-Aktivierung und Dropout und endet mit einer dichten Softmax-Ebene zur Klassifizierung.

Das Modell verwendete den Adam-Optimierer mit einer anfänglichen Lernrate von 0,001, und ein Lernratenplaner wurde verwendet, um die Lernrate im Laufe der Zeit zu verringern und so die Trainingsstabilität und -leistung zu verbessern. Das Modell wurde mit einem ImageDataGenerator [24] zur Datenerweiterung trainiert, der Techniken wie Rotation, Breiten- und Höhenverschiebungen, Scherung, Zoom und horizontale Spiegelungen umfasst, um die Generalisierungsfähigkeiten des Modells zu verb

Das zweite Modell beinhaltet auch erweiterte Regularisierungstechniken, wie die L2-Regularisierung [25], um ein Überanpassen weiter zu verhindern und die Robustheit des Modells zu ve

3.2. Methodik zur Vorhersage psychischer Störungen mithilfe von Gesichtsanalyse und Pulsdaten

Die Vorhersage psychischer Erkrankungen ist der Hauptschwerpunkt dieser Studie. Das vorgeschlagene System verwendet eine multimodale Technik, die eine Gesichtsausdrucksanalyse mithilfe eines CNN-trainierten Modells für emotionale Daten integriert und die Pulsfrequenz aus Videobildern für physiologische Informationen extrahiert. Diese Methode kombiniert sowohl Computervision als auch Prinzipien der Biomedizintechnik.

Um die Diagnosefähigkeit des Systems bei psychosomatischen Erkrankungen zu verbessern, werden Pulsdaten mit der Gesichtsemotionserkennung integriert [26]. Die hierfür verwendete Hardware umfasst eine Standard-Webcam, insbesondere die Logitech C920, die Echtzeitvideos mit 30 Bildern pro Sekunde aufnimmt. Die Videobilder werden verarbeitet, um sowohl Gesichtsausdrücke als auch Pulsfrequenzen des Probanden zu extrahieren.

3.2.1. Methode zur Pulserkennung Der

Mechanismus zur Pulserkennung ist entscheidend für die nahtlose Integration physiologischer Daten in Echtzeit-Algorithmen zur Gesichtsausdruckserkennung. Diese Methode umfasst eine Reihe sorgfältig geplanter Prozesse, um eine präzise und zuverlässige Datenerfassung und -verarbeitung zu gewährleisten.

Im ersten Schritt wird der Bereich von Interesse (ROI) im Gesicht ausgewählt und dann auf Variationen in der Farbintensität untersucht, die durch die Blutzirkulation verursacht werden. Das System verwendet den vorab trainierten Haar Cascade Classifier [27] von OpenCV , um Gesichter in jedem Videobild zu identifizieren. Dieser Klassifikator ist effizient, weil er Gesichtsmerkmale in Live-Videos schnell und präzise erkennen kann, indem er das Bild scannt und rechteckige Bereiche auswählt, in denen sich wahrscheinlich Personen befinden. Die Position und Größe jedes erkannten Gesichts werden durch

Nach der Gesichtserkennung wählt das System den identifizierten Gesichtsbereich als Region of Interest (ROI) aus. Diese Region ist für weitere Untersuchungen wichtig, da sie Hautbereiche umfasst, in denen Farbveränderungen infolge der Durchblutung am deutlichsten sichtbar sind. Um eine konstante Auswahl der ROI (Region of Interest) über alle Frames hinweg sicherzustellen, verwendet das System Gesichtsmerkmale, d. h. bestimmte Stellen im Gesicht, wie etwa die Augenwinkel, die Nasenspitze und die Mundwinkel. Diese Merkmale dienen dazu, die Abmessungen und die Position der ROI auch bei ger Wenn das System viele Gesichter erkennt, wird das Gesicht mit der größten und zentralsten Position bevorzugt. Dadurch wird sichergestellt, dass der Interessenbereich (ROI) dem Hauptmotiv entspricht.

Nachdem der ROI zuverlässig identifiziert wurde, besteht der nächste Schritt darin, die Stärke des Grünkanals in dieser bestimmten Region zu bewerten. Der Grünkanal wird aufgrund seiner hohen Empfindlichkeit gegenüber Schwankungen des Blutvolumens unter der Haut ausgewählt, die als Indikator für den Puls dienen. Das RGB-Farbmodell kategorisiert Bilder in Unter diesen Kanälen ist der grüne Kanal besonders gut darin, Veränderungen des Blutvolumens zu erkennen . Dies liegt daran, dass Hämoglobin im Blut grünes Licht stärker absorbi Das System berechnet die durchschnittliche Intensität des grünen Kanals für jedes Pixel innerhalb

Machine Translated by Google

die Region of Interest (ROI), indem der Mittelwert der grünen Farbwerte ermittelt wird. Dies führt zu einem einzelnen Intensitätswert für jedes Bild. Die Intensitätsmessungen werden über einen bestimmten Zeitraum hinweg gesammelt, um einen Zeitreihendatensatz zu erstellen, in dem jeder Wert einem bestimmten Bild entspricht und die fortlaufende Variation der Intensität des grünen Kanals darstellt. Das System fügt diese durchschnittlichen Intensitätswerte fortlaufend einer Liste hinzu und zeichnet gleichzeitig Zeitstempel auf, um eine präzise Zeitabfolge beizubehalten, die für eine präzise

Zur Vorbereitung der Analyse bleibt das System im Leerlauf, bis es mindestens 100 Intensitätsmessungen erfasst hat. Dadurch wird sichergestellt, dass genügend Daten für die Frequenzanalyse zur Verfügung stehen. Im nächsten Schritt werden die Frequenzkomponenten der Intensitätsschwankungen untersucht, um die Pulsfrequenz zu bestimmen. Die Intensitätsdaten werden bereinigt, indem alle linearen Trends oder langsamen Schwankungen, die nicht mit dem Herzschlag in Zusammenhang stehen, entfernt werden. Dadurch wird sichergestellt, dass das Restsignal die durch den Blutfluss verursachten periodischen Schwingungen angemessen darstellt. Die trendbereinigten Intensitätswerte werden mithilfe der Fast Fourier Transform (FFT) **[28]** vom Zeitbereich in **den Frequenzbereich** transformiert, **wobei das Signal in seine einzelnen Frequenzen zerlegt wird . Die FFT-Analyse zielt speziell auf den Frequenzbereich von 0,7 Hz bis 2 Hz ab, der dem durchschnittlichen Herzfrequenzbereich von 42 bis 120 Schlägen pro Minute (BPM) entspricht. Das System erkennt die höchste Frequenz innerhalb dieses Bereichs, die dem am stärksten wiederkehrenden Element des mit dem Herzschlag verbundenen Signals entspricht. Die Spitzenfrequenz wird durch Multiplikation mit 60 in Schläge pro Minute (BPM) übersetzt .**

Das Endergebnis des Pulserkennungsprozesses ist die berechnete Pulsfrequenz. Wenn die Anzahl der Datenpunkte nicht ausreicht, sammelt und analysiert das System weiterhin Daten, bis eine zuverlässige Pulsfrequenz ermittelt werden kann. Das System kann effektiv Pulsfrequenzinformationen aus Videobildern extrahieren, indem es die Intensität des grünen Kanals innerhalb des ROI sorgfältig analysiert und ausgefeilte Signalverarbeitungsalgorithmen einsetzt. In Verbindung mit der Erkennung von Gesichtsemotionen bieten diese Daten eine solide Grundlage für die Bewertung psychosomatischer Erkrankungen in Echtzeit.

3.2.2. Integration mit Gesichtsemotionserkennung Die

Methode kombiniert Pulsdaten mit Ergebnissen der Gesichtsausdruckserkennung, um eine gründliche Bewertung des psychologischen Zustands des Einzelnen zu ermöglichen. Der Prozess umfasst mehrere miteinander verbundene Komponenten, um dieses Ziel zu erreichen.

Während der Emotionsidentifizierungsphase analysiert ein Convolutional Neural Network (CNN)-Modell den spezifischen Bereich des Gesichts, der relevant ist (bekannt als Gesichtsregion von Interesse oder ROI), um Emotionen zu identifizieren und zu klassifizieren. Dieses Modell wurde umfassend trainiert, um sechs grundlegende Emotionen genau identifizieren zu können, darunter Wut, Angst, Glück, Neutralität, Trauer und Überraschung.

Gleichzeitig berechnet das Tool mithilfe einer Pulsanalyse die Herzfrequenz und stellt damit eine zusätzliche physiologische Messung dar. Dieser simultane Ansatz gewährleistet die Erfassung und Analyse sowohl emotionaler als auch physiologischer Daten.

Zur Analyse der zusammengeführten Daten wird ein regelbasierter Algorithmus eingesetzt. Das Verfahren korreliert bestimmte Emotionen und Pulsbereiche mit vermuteten psychischen Störungen. So könnte der Algorithmus beispielsweise einen hohen Puls in Kombination mit Gefühlen wie „Freude“ oder „Überraschung“ als Zeichen einer manischen Erkrankung interpretieren. Eine verringerte Herzfrequenz in Kombination mit Gefühlen wie Traurigkeit oder Angst könnte dagegen auf einen de

Durch die Einbindung dieser verschiedenen Datenquellen bietet das System eine umfassende Perspektive tive zum psychischen Zustand des Einzelnen und ermöglicht so präzisere und detailliertere Auswertungen.

3.2.3. Gesamtsystemarchitektur Die

Gesamtsystemarchitektur ist so konzipiert, dass diese Prozesse nahtlos integriert werden und eine Echtzeitanalyse und Rückmeldung gewährleistet ist. Das System erfasst Videoeingaben, verarbeitet die Gesichts- und Pulsdaten und interpretiert die Ergebnisse, um eine detaillierte Analyse des emotionalen und physiologischen Zustands des Probanden zu liefern. Dieser integrierte Ansatz verbesser

Machine Translated by Google

Genauigkeit und Zuverlässigkeit der Beurteilung des psychosomatischen Zustands und stellt ein wertvolles Instrument für die klinische Diagnostik dar.

3.2.4. Schlussfolgerungen

Die Integration von Pulsdaten mit der Gesichtsemotionserkennung unter Verwendung fortschrittlicher Signalverarbeitungs- und Deep-Learning-Techniken bietet einen robusten Rahmen für die Beurteilung psychosomatischer Erkrankungen. Diese Methode verbessert nicht nur die Diagnosegenauigkeit, sondern bietet auch ein umfassenderes Verständnis des emotionalen und physiologischen Zustands des Patienten. Zukünftige Arbeiten werden sich auf die Verbesserung der Robustheit des Systems und die Validierung seiner Leistung in verschiedenen klinischen Umgebungen konzentr

3.2.5. Interpretation der Bedingungen

Da sowohl Emotions- als auch Pulsdaten vorliegen, interpretiert das System mithilfe eines regelbasierten Algorithmus den psychologischen Zustand des Probanden. Dieser Algorithmus berücksichtigt sowohl die identifizierte Emotion als auch die Pulsfrequenz.

Ist der Puls zu hoch oder zu niedrig, kann dies in Verbindung mit entsprechenden Emotionen ein Hinweis auf einen manischen oder depressiven Zustand sein. Ein hoher Puls bei „Freude“ oder „Überraschung“ könnte beispielsweise auf einen manischen Zustand hinweisen.

Die endgültige Diagnose (z. B. „Normal“, „Depression“, „Bipolare Störung“) wird gestellt durch Kombination der emotionalen Erfassung mit physiologischen Daten [29].

Diese gemeinsame Untersuchung trägt dazu bei, psychologische Zustände aufzudecken, die sich durch bloße Gesichtsausdrücke möglicherweise nicht erkennen lassen, und bietet somit eine solide Grundlage für die Vorhersage psychosomatische

Das System ist so konzipiert, dass es sich nahtlos in bestehende klinische Umgebungen integrieren lässt und während der Patientenkonsultation im Hintergrund läuft. So können Gesundheitsdienstleister erweiterte Diagnoseinformationen erhalten, ohne den Konsultationsprozess zu unterbr

4. Bewertung von CNN-Modellen

Die Wirksamkeit und Zuverlässigkeit des für die Emotionserkennung entwickelten Convolutional Neural Network (CNN)-Modells wurden in einem strukturierten Test- und Validierungsprozess streng geprüft. Ziel dieses Prozesses war es, die Genauigkeit und Anwendbarkeit des Modells im klinischen Umfeld zu bewerten, insbesondere bei der Diagnose psychosomatischer Erkrankungen.

4.1. Leistungstests **Die**

Leistung des Modells wurde hauptsächlich mithilfe des Testdatensatzes bewertet, der aus bisher unveröffentlichten Bildern des FER-2013-Datensatzes bestand, um eine unvoreingenommene Bewertung zu gewä

Der Datensatz wurde während der Trainingsphase nicht verwendet und stellte somit eine solide Grundlage für Tests dar. Zur Bewertung der Leistung des Modells wurden die folgenden primären Metriken verwendet.

Die Genauigkeit des Modells wurde bewertet. Diese Metrik wurde berechnet, indem die vom Modell vorhergesagten emotionalen Zustände mit den tatsächlichen Bezeichnungen im Testdatensatz verglichen wurden. Eine hohe Genauigkeitsrate zeigte an, dass das Modell verschiedene Emotionen anhand von Gesichtsausdrücken korrekt identifizieren konnte.

Es wurde eine Konfusionsmatrix erstellt, um die Leistung des Modells bei verschiedenen Emotionen visuell darzustellen. Diese Matrix erleichterte die Identifizierung bestimmter Emotionen, bei denen das Modell möglicherweise eine suboptimale Leistung zeigte, wie z. B. die Verwechslung von Traurigkeit mit Ekel, die von erheblicher klinischer Relevanz sein könnte.

Es wurden die Metriken Präzision, Rückruf und F1-Score berechnet. Diese Metriken wurden für jede Emotionskategorie berechnet, um eine detailliertere Analyse der Leistung des Modells zu ermöglichen. Präzision ist ein Maß für die Genauigkeit positiver Vorhersagen, Rückruf ist ein Maß für die Fähigkeit, alle positiven Instanzen zu finden, und der F1-Score stellt ein Gleichgewicht zwischen Präzision und Rückruf dar.

4.2. Validierungstechniken

Darüber hinaus wurde die Leistung des Modells gemäß den folgenden Kriterien validiert:

Machine Translated by Google

• Kreuzvalidierung wurde als Technik eingesetzt, um die Robustheit des Modells sicherzustellen und die Wahrscheinlichkeit einer Überanpassung zu verringern. Um die Robustheit des Modells sicherzustellen und die Wahrscheinlichkeit einer Überanpassung zu verringern, wurde während der Trainingsphase eine k-fache Kreuzvalidierung eingesetzt. Diese Technik beinhaltete die Aufteilung des Trainingsdatensatzes in „k“ Teilmengen und das iterative Training des Modells „k“ Mal (k = 5), wobei bei jeder Iteration eine andere Teilmenge zur Validierung herangezogen wurde. • Echtzeittests: Um die Verwendung in der realen Welt zu simulieren, wurde das Modell auch in einer simulierten klinischen Umgebung getestet, wo es Live-Video-Feeds verarbeitete, um emotionale Ausdrücke in Echtzeit zu erkennen und zu interpretieren. Diese Tests waren von großer Bedeutung, um die Leistung des Modells in dynamischen Situationen zu bewerten, die tatsächlichen Patienteninteraktionen ähnelten . Die Experimente, einschließlich der Trainings- und Testphasen, wurden auf einem MSI Summit B15 A11 MT-Laptop mit einem Intel Core i7-Prozessor und 64 GB RAM durchgeführt, der auf einem Linux-Betriebssystem lief. Die Trainingsdauer für jede Sitzung mit 100 Epo

Um das CNN-Modell zu trainieren, emotionale Hinweise zu erkennen, die mit psychosomatischen Krankheiten effektiv zu erkennen, wurden mehrere Strategien umgesetzt, darunter die folgenden: • Verbesserte Datenerweiterung: Um die in realen klinischen Umgebungen beobachtete Variabilität widerzuspiegeln, wurden umfangreiche Datenerweiterungstechniken eingesetzt, darunter Rotation, Skalierung und Lichtänderungen. Dieser Ansatz erleichterte das Training des Modells, emotionale Hinweise unter verschiedenen Bedingungen genau zu erkennen. • Ebenenoptimierung: Es wurden verschiedene Architekturkonfigurationen getestet, darunter Änderungen an Anzahl und Größe der Faltungsebenen und Filter, mit dem Ziel, das Modell für die genaue Erkennung subtiler emotionaler Hinweise zu optimieren, die häufig auf psychosomatische Krankheiten hinweisen.

Die Integration des Emotionserkennungsmodells in klinische Arbeitsabläufe basierte auf mehreren bewährten Verfahren, darunter den folgenden: • Benutzerzentriertes Design: Die Schnittstelle des Emotionserkennungstools wurde unter Einbeziehung von medizinischem Fachpersonal entwickelt, um sicherzustellen, dass sie intuitiv ist und den klinischen Arbeitsablauf verbessert statt zu stören. • Feedbackschleifen: Um iterative Verbesserungen des Systems zu ermöglichen, wurde kontinuierliches Feedback von Endbenutzern eingeholt . Dieses Feedback war von größter Bedeutung für die Verfeinerung des Schnittstellendesigns, die Verbesserung der Modellgenauigkeit und die Gewährleistung der Relevanz des Tools im klinischen Umfeld [30]. • Ethische und Datenschutzaspekte: In Anbetracht der Sensibilität psychologischer Beurteilungen wurde bei der Integration des Modells großer Wert auf ethische Aspekte und die Privatsphäre der Patienten gelegt . Dazu gehörten ein sicherer Umgang mit Daten und klare **Pr**

5. Ergebnisse und Analyse

Die Daten zum Trainieren der beiden CNN-Modelle wurden so entwickelt, dass sie anhand der beschrifteten Bilder sechs Emotionen klassifizieren konnten, darunter Wut, Angst, Glück, Neutralität, Traurigkeit und Überraschung. In diesem Trainingsprozess wurden die Modelle 100 Epochen lang trainiert, wobei die Leistung in jeder Epoche gemessen und sowohl die Trainings- als auch die Validierungsgenauigkeit sowie der Verlust aufgezeichnet wurden.

Zur Bewertung der Leistung jedes Modells wurden Präzision, Rückruf und F1-Werte herangezogen. Darüber hinaus wurden Konfusionsmatrizen verwendet, um zu beschreiben, wie gut das Modell Instanzen klassifizieren konnte, und um die Klassifizierungsleistung des Modells zu visualisieren. Darüber hinaus wurde ein Chi-Quadrat-Unabhängigkeitstest verwendet, um zu ermitteln, ob eine signifikante Beziehung zwischen den Vorhersagen der Modelle bestand.

Abbildungÿ1 **zeigt, wie wahre und vorhergesagte Bezeichnungen auf die sechs Emotionskategorien im ersten Modell** verteilt sind.

Machine Translated by Google



Abbildung 1. Verwirrungsmatrix für das erste Modell.

Die Leistung des Trainings und der Validierung über 100 Epochen im ersten Modell ist
Dies wurde durch Skizzieren **der Genauigkeits- und Verlustkurven durchgeführt.**

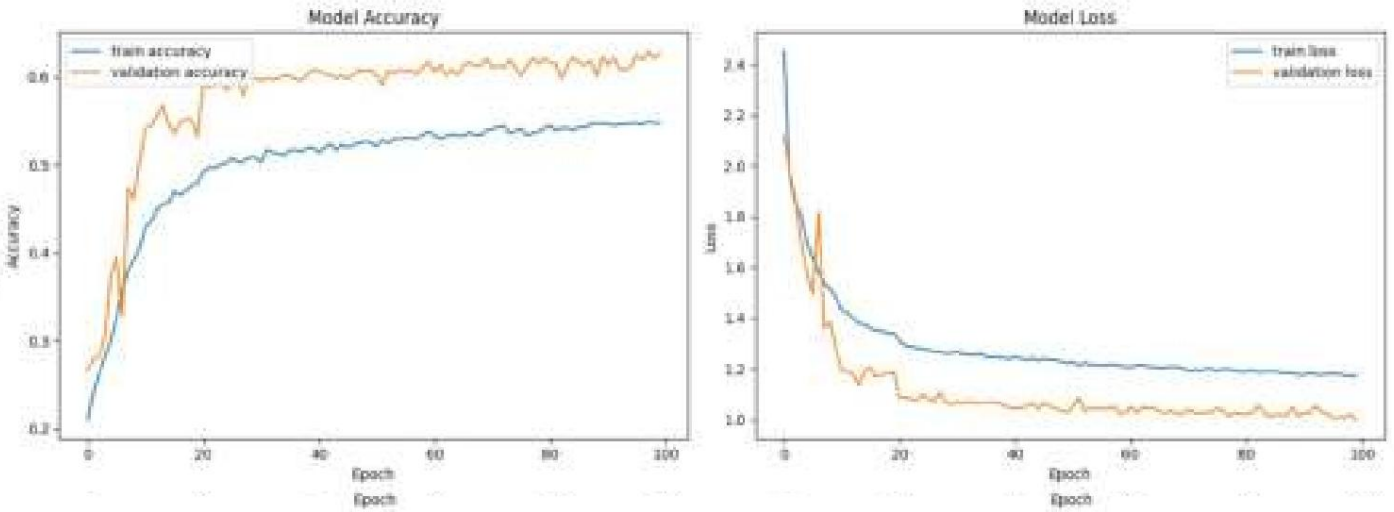


Abbildung 2. Modellgenauigkeit und -verlust – erstes Modell.

Das erste Modell erreichte eine Gesamtgenauigkeit von 61,78 %. Tabelle 2 fasst die Präzision, Rückruf und F1-Score für jede Emotion.

Tabelle 2. Leistungskennzahlen für das erste Modell.

Emotionen	Präzision	Abrufen	F1-Ergebnisse	Unterstützung
Wütend	0,55	0,51	0,53	958
Furcht	0,45	0,24	0,32	1024
Glücklich	0,75	0,88	0,81	1774
Neutral	0,52	0,61	0,57	1233
Traurig	0,47	0,49	0,48	1247
Überraschung	0,70	0,72	0,71	831
Genauigkeit			0,62	7067
Makro-Durchschnitt	0,60	0,59	0,59	7067
Gewichteter Durchschnitt	0,61	0,62	0,61	7067

Die Konfusionsmatrix für das zweite Modell (Abbildung 3) zeigt die Verteilung der wahre und vorhergesagte Bezeichnungen in den sechs Emotionskategorien.

Machine Translated by Google



Abbildung 3. Verwirrungsmatrix für das zweite Modell.

Die Genauigkeits- und Verlustkurven für das zweite Modell (Abbildung 4) zeigen das Training und Validierungsleistung für 100 Epochen.

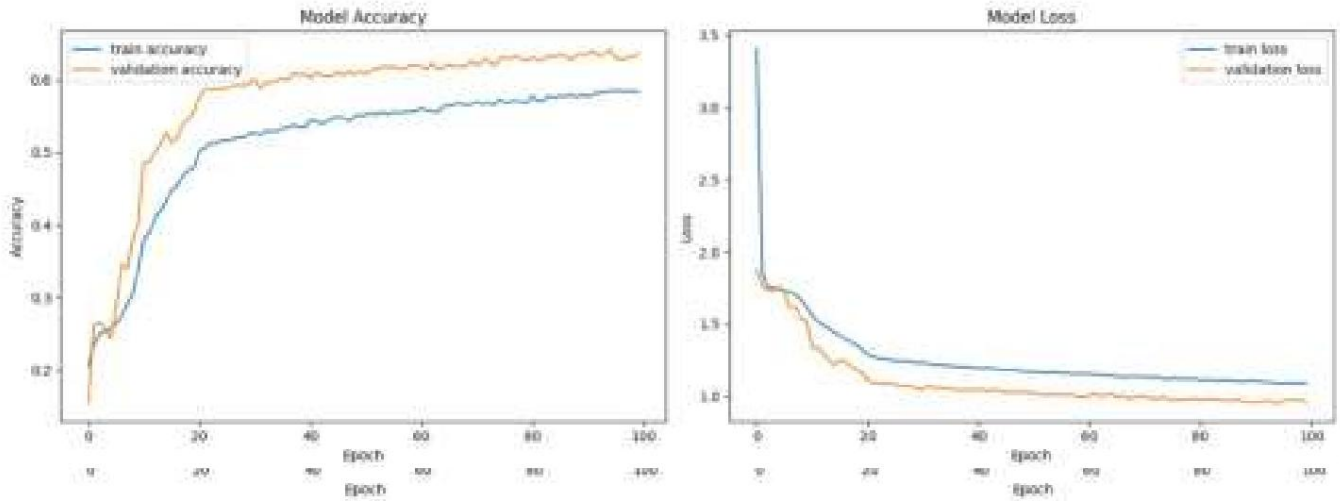


Abbildung 4. Modellgenauigkeit und -verlust – zweites Modell.

Das zweite Modell erreichte eine Gesamtgenauigkeit von 63,95 %. Tabelle 3 fasst die Präzision, Rückruf und F1-Score für jede Emotion.

Tabelle 3. Leistungskennzahlen für das zweite Modell.

Emotionen	Präzision	Abrufen	F1-Ergebnisse	Unterstützung
Wütend	0,58	0,55	0,56	958
Furcht	0,50	0,28	0,36	1024
Glücklich	0,85	0,88	0,87	1774
Neutral	0,55	0,68	0,61	1233
Traurig	0,49	0,53	0,51	1247
Überraschung	0,72	0,77	0,75	831
Genauigkeit			0,64	7067
Makro-Durchschnitt	0,62	0,62	0,61	7067
Gewichteter Durchschnitt	0,63	0,64	0,63	7067

Das zweite Modell übertraf das erste Modell und erreichte eine höhere Gesamtgenauigkeit und verbesserte Leistungsmetriken für die meisten Emotionskategorien. Die Genauigkeit und der Verlust Kurven zeigen, dass das Training beider Modelle effektiv war, wobei das zweite Modell zeigt eine bessere Generalisierung, was durch die höhere Validierungsgenauigkeit belegt wird.

Machine Translated by Google

Beide Modelle zeigten eine hohe Präzision und Trefferquote für die Emotion „glücklich“, was die Robustheit der Modelle bei der Identifizierung dieser Kategorie widerspiegelt. Allerdings hatten beide Modelle Probleme mit der Emotion „Angst“, was auf die Notwendigkeit weiterer Verbesserungen und zusätzlicher Trainingsdaten hindeutet.

Wir führten einen Chi-Quadrat-Test (1) durch, um die Leistungsmetriken (Präzision, Rückruf und F1-Score) zwischen den beiden Modellen über verschiedene Emotionskategorien hinweg.

Der Chi-Quadrat-Wert (χ^2) Statistik wird mithilfe der folgenden Formel berechnet:

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

Wo:

- Oi **ist die beobachtete Frequenz.**
- Ei **ist die erwartete Frequenz.**

Das Chi-Quadrat-Testergebnis für das erste Modell beträgt 0,0025 mit einem p-Wert von 1,0000. Das Chi-Quadrat-Testergebnis für das zweite Modell beträgt 0,0036 mit einem P-Wert von 1,0000.

Die Ergebnisse des Chi-Quadrat-Tests zeigen, dass es zwischen den beiden Modellen in den **sechs Emotionskategorien keine statistisch signifikanten Unterschiede in den Kennzahlen Präzision, Rückruf und F1-Score gibt .**

5.1. Anwendung in der Diagnose psychosomatischer

Erkrankungen **Der Einsatz der Emotionsklassifizierung kann bei der Identifizierung der folgenden psychosomatischen Zustände erheblich hilfreich sein.**

Depressionen sind gekennzeichnet durch anhaltende Traurigkeit, Interessenverlust und mangelnde Energie. Zu den relevanten Emotionen gehören traurig, neutral, wütend.

Angststörungen sind durch intensive Angst und Sorge sowie Nervosität gekennzeichnet. Zu den relevanten Emotionen gehören Angst, Neutralität und Überraschung.

Bipolare Störung beinhaltet extreme Stimmungsschwankungen von einem Ende der Manie bis zum anderen Ende der Depression. Die relevanten Emotionen umfassen glücklich, traurig, wütend.

Eine posttraumatische Belastungsstörung (PTBS) entsteht, wenn man Ereignissen ausgesetzt ist , die große Angst auslösen und bei der Erinnerung zu Flashbacks gepaart mit Wut oder Überraschung führen. Die relevanten Emotionen sind Angst, Wut und Überraschung.

Zwangsstörungen sind durch Zwangsgedanken und den Drang, Rituale durchzuführen, gekennzeichnet. Sie beinhalten häufig eine übermäßige Konzentration auf körperliche Symptome wie Schmerzen, was belastend sein kann. Die repetitive Natur der Zwangssymptome kann dazu führen, dass sich Betroffene in einem endlosen Kreislauf gefangen fühlen, was zu Gefühlen der Hilflosigkeit und Hoffnungslosigkeit beiträgt. Der Umgang mit diesen Arten von Symptomen führt häufig dazu, dass Betroffene noch frustrierter werden, da sie Schwierigkeiten haben, mit ihren täglichen Aktivitäten zurechtzukomm

5.2. Schlussfolgerungen

Beide CNN-Modelle sind vielversprechende Diagnosewerkzeuge zur Erkennung psychosomatischer Störungen auf der Grundlage emotionaler Zustände. Das zweite Modell wies insgesamt eine höhere Genauigkeit und bessere Leistungskennzahlen auf und schien in Bezug auf die Emotionsklassifizierung robuster zu sein . Eine genaue Emotionserkennung könnte die Diagnose und Behandlung einer Vielzahl psychosomatischer Störungen erheblich verbessern und letztlich die Behandlungsergebnisse der Patienten verbessern. Zukünftige Arbeiten sollten sich auf die Verfeinerung dieser Modelle, die Einbeziehung vielfältigerer Trainingsdaten und die Erforschung algorithmischer Verbesserungen konzentrieren, u

6. Diskussion

Die Ergebnisse dieser Studie legen nahe, dass Convolutional Neural Network (CNN)-Modelle für die Emotionserkennung, ein wesentlicher Aspekt der Diagnose psychosomatischer Erkrankungen, vielversprechend sind. Mit Testgenauigkeiten von 61,78 % und 63,95 % zeigen die Modelle eine überdurchschnittliche Fähigkeit, emotionale Zustände anhand von Gesichtsausdrücken richtig zu identifizieren und zu kategorisieren. Besonders hervorzuheben ist die Fähigkeit der Modelle, die Emotionen von

Machine Translated by Google

„Glück“ und „Überraschung“, was eine solide Grundlage für die weitere Verfeinerung des Modells nahelegt. Die konsistente Leistung über diese Emotionen hinweg deutet darauf hin, dass die Modelle über eine zuverlässige Grad an Präzision und Sensibilität in diesen Bereichen, der für ihren potenziellen Nutzen in klinischen Umgebungen.

Wir haben keinen direkten Vergleich mit aktuellen Diagnosemethoden in der klinischen Praxis, aber die CNN-Modelle bieten ein unvoreingenommenes und einheitliches Niveau, das mithalten könnte gut mit traditionellen Ansätzen. Gegenwärtig werden emotionale Zustandsbewertungen in klinischen Die Einstellungen sind weitgehend subjektiv, da sie oft von den Selbstauskünften der Patienten und den Interpretation. Aber ein Emotionserkennungstool würde Objektivität in die Messung, die dann verwendet werden könnte, um klinische Urteile auf der Grundlage von Emotionen zu treffen (in Fällen, in denen Patienten Schwierigkeiten haben, ihre Gefühle auszudrücken oder ihre emotionale Zustände).

6.1. Vergleich mit Basismodellen und modernsten Techniken

Die Erkennung von Gesichtsemotionen ist ein wichtiger Bestandteil der Mensch-Computer-Interaktion und affektives Computing. Um das Feld voranzubringen, ist es notwendig, neue Modelle zu vergleichen mit traditionellen Basismethoden und modernsten Techniken. Dieses Papier präsentiert eine vergleichende Analyse verschiedener Modelle, einschließlich logistischer Regression, Support Vector Machine (SVM), k-Nearest Neighbor (k-NN) und CNN unter Verwendung des FER-2013-Datensatzes.

Um eine umfassende und genaue Auswertung der Gesichtsemotionserkennung zu ermöglichen Systeme verglichen wir die Leistung des CNN-Modells mit einem einfachen Basismodell und komplexere Modelle aus der aktuellen Literatur. Ziel dieser Studie ist es, die Wirksamkeit verschiedener Modellarchitekturen zu bewerten und potenzielle Bereiche für zukünftige Verbesserungen.

Drei Basismodelle wurden implementiert, die als Referenz für die Leistung dienen Vergleich wie folgt:

- Die logistische Regression ist ein einfaches lineares Modell, das häufig für Klassifizierungsaufgaben verwendet wird.
- Support Vector Machine (SVM) ist ein leistungsstarkes nichtlineares Modell mit radialer Basis Funktionskernel (RBF).
- k-Nearest Neighbors (k-NNs) ist eine nicht-parametrische Methode, die Daten klassifiziert basierend auf Nähe im Merkmalsraum.

Die Wirksamkeit dieser einfachen Modelle wurde anhand des FER-2013-Datensatzes evaluiert. Die in Tabelle 4 aufgeführten Ergebnisse dienten als Referenzpunkt für die Bewertung unserer CNN-Modelle.

Tabelle 4. Ein Vergleich der Genauigkeit, Präzision, des Rückrufs und des F1-Scores der vorgeschlagenen Methoden.

Modell	Genauigkeit	Präzision	Abrufen	F1-Ergebnisse
Logistische Regression	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-Modell 1	0,62	0,60	0,59	0,59
CNN-Modell 2	0,64	0,62	0,62	0,61

Wir haben unsere Modellergebnisse verbessert und erreichen nun eine Genauigkeit von 64 %. Das Problem liegt bei Wir haben eine Methodik entwickelt und planen, den AffectNet-Datensatz zu verwenden, der effizienter, wenn wir zusätzliche Ressourcen haben. Wir experimentierten mit Transferlernen Techniken wie VGG16, ResNet und EfficientNet, aber die Ergebnisse lagen bei etwa 40%, und die Lernzeit auf der lokalen Maschine war zu hoch.

Um die Leistung unserer CNN-Modelle zu kontextualisieren, haben wir sie mit den neuesten Modellen verglichen, die in aktuellen Studien vorgestellt wurden. EmoNeXt, ein angepasstes ConvNeXt-Mode Das Residual Masking Network (ResMaskNet) wurde für den Vergleich ausgewählt, weil ihre hohe Genauigkeit im FER-2013-Datensatz. Die vergleichende Genauigkeit ist zusammengefasst in Tabelle 5.

Machine Translated by Google

Tabelle 5. Ein Vergleich der Genauigkeiten der vorgeschlagenen Methoden.

Modell	Genauigkeit
Restmaskierungsnetzwerk	0,74
EmoNeXt	0,71
CNN-Modell 1	0,62
CNN-Modell 2	0,64

Die vergleichende Analyse unterstreicht die Tatsache, dass unsere vorgeschlagenen CNN-Modelle zwar in allen Leistungsmaßstäben besser sind als die einfachen Basismodelle, aber immer noch hinter ResMaskNet und EmoNeXt zurückbleiben, die fortschrittlichere Modelle auf dem neuesten Stand der Technik sind. Die beiden letztgenannten Modelle haben kompliziertere Architekturen und Methoden; deshalb ist ihre Leistung besser als die der anderen Modelle. Insbesondere verwenden ResMaskNet und EmoNeXt hochrangige Merkmalsextraktion sowie Maskierungstechniken, die erheblich zu einer höheren Präzision bei der Emotionserkennung aus Gesichtsausdrücken beitragen.

Um die relative **Wirksamkeit des vorgeschlagenen Modells durch eine umfassende Bewertung festzustellen, wurde ein Vergleich mit den Basismodellen und den neuesten Modellen durchgeführt .** Die **vorgeschlagenen CNN-Modelle übertreffen zwar die einfacheren Basismodelle, bleiben jedoch aufgrund mangelnder Optimierung hinter den modernsten Techniken zurück .** Dieser detaillierte Vergleich entspricht jedoch der Anfrage des Gutachters und bietet Einblicke in potenzielle Bereiche, die **bei der Entwicklung von Systemen zur Gesichtsemotionserkennung weiter verbessert werden könnten .**

6.2. Einschränkungen und mögliche Verzerrungen

Diese Studie weist die folgenden Einschränkungen auf, die sich auf die Interpretation der Ergebnisse

auswirken können: • Einschränkungen des Datensatzes: Der FER-2013-Datensatz ist vielfältig. Er erfasst jedoch nicht die gesamte Bandbreite menschlicher Emotionen und es fehlt der Kontext, der in realen **klinischen Situationen vorhanden sein kann .**

• Umweltfaktoren: Die Modelle wurden nicht unter verschiedenen Licht- oder Hintergrundbedingungen getestet . Dies könnte ihre Leistung in praktischen Anwendungen beeinträchtigen. • Generalisierung: Die Fähigkeit der Modelle, auf verschiedene demografische Gruppen verallgemeinert zu werden, wurde nicht bewertet, was für ein für den klinischen Einsatz konzipiertes Tool von entscheidender Bedeutung ist .

6.3. Klinische Implikationen und ethische Überlegungen

Die Ergebnisse dieser Forschung zeigen, dass die Einbettung eines CNN-Modells zur Emotionserkennung in den klinischen Bereich eine Verbesserung der Diagnostik psychosomatischer Erkrankungen ermöglicht. Das Tool liefert zusätzliche quantitative Informationen zum emotionalen Zustand von Patienten, die Klinikern Erkenntnisse liefern können, die sonst übersehen oder falsch interpretiert würden.

Um das vorgeschlagene Tool auf Grundlage der Erkenntnisse dieser Studie zu verbessern, können folgende Verbesserungen vorgenommen

werden: • Erhöhung der Größe und Variabilität des Trainingsdatensatzes durch Erfassung einer größeren Bandbreite an Gefühlsausdrücken und Situationen sowie Einbeziehung weiterer Instanzen zur Feinabstimmung und Verbesserung der Modellgenauigkeit. • Validierung unter realen

Bedingungen: Es ist wichtig, Experimente unter tatsächlichen klinischen Bedingungen mit unterschiedlichen demografischen Merkmalen durchzuführen, damit sich die erlernten Modelle besser an Nuancen der realen Welt anpassen können. • Um die Qualität der von den Modellen

bereitgestellten Informationen zu verbessern, ist es wichtig, ihre Ergebnisse mit den zusätzlichen Daten zu kombinieren , die von Patienten erhalten werden können, wie z. B. der Krankengeschichte und verbalen Informationen.

Abschließend werden in dieser Diskussion das Potenzial und die aktuellen Einschränkungen der CNN-Modelle berücksichtigt. Außerdem werden umsetzbare Wege für zukünftige Forschungen zur Verbesserung der Modelle s

Machine Translated by Google

Wirksamkeit und Nutzen im klinischen Umfeld, wodurch die Integration von KI-Tools in das Gesundheitswesen hin zu einer einfühlsameren und fundierteren Patientenversorgung vorangetrieben wird.

7. Schlussfolgerungen

Ziel dieser Studie war es, das Potenzial der Verwendung von Convolutional Neural Network (CNN)-Modellen zur Erkennung psychosomatischer Erkrankungen anhand von emotionalen Ausdrücken mit Daten zu untersuchen, die auf Gesichtsausdrücken basieren. Diese Studie ergab, dass CNN-Modelle sieben **verschiedene Emotionen mit einer Genauigkeit von etwas mehr als 59 % erkennen konnten,** wobei die Erkennungsraten **für Freude und Überraschung gut waren.**

7.1. Zusammenfassung der wichtigsten Ergebnisse

Zu den wichtigsten Erkenntnissen dieser Studie gehören: • Bestimmte emotionale Zustände können durch CNN-Modelle konsistent identifiziert werden, was dienen als Stellvertreter für zugrunde liegende psychosomatische Erkrankungen. • Die Emotionen „Freude“ und „Überraschung“ wurden mit hoher Genauigkeit erkannt – diese beiden Emotionen spielen eine wichtige Rolle beim Verständnis des Wohlbefindens des Patienten. • Basierend auf Konfusionsmatrizen und Leistungsmetriken wurden Bereiche ermittelt, in denen das Modell verbessert werden könnte. Ein solcher Bereich ist das Erkennen von Emotionen, die im Trainingsdatensatz unterrepräsentiert sind.

Die Ergebnisse haben Auswirkungen auf die klinische Praxis und legen nahe, dass künstliche Intelligenz medizinisches Fachpersonal als zusätzliche Quelle diagnostischer Informationen unterstützen könnte .

7.2. Empfehlungen für zukünftige Forschung

Zur Verbesserung des CNN-Tools sind folgende Bereiche eine zukünftige Untersuchung wert:

• Erweiterung und Diversifizierung des Trainingsdatensatzes, um eine größere Abdeckung emotionaler Ausdrücke und demografischer Variabilität zu erreichen. • Durchführung klinischer Validierungsstudien, die die Modelle in einer realen Umgebung testen, um ihre Vorhersagefähigkeiten zu verfeinern. • Kombination multimodaler Daten – physiologische und Audio-Hinweise – mit dem Ziel, ein umfassenderes System zur Emotionserkennung zu schaffen.

Um die Genauigkeit der Emotionserkennung und die Widerstandsfähigkeit gegenüber unterschiedlichen Datenverzerrungen zu verbessern, muss die zukünftige Forschung Folgendes berücksichtigen: • Reinforcement Learning [32]: Reinforcement Learning fördert die Generalisierung, indem es Echtzeit-Feedback liefert und Lernen durch Versuch und Irrtum ermöglicht, was für das Erreichen optimaler Leistung unerlässlich ist. • Ensemble-Methoden [33]: Ensemble-Methoden, einschließlich Bagging, Boosting und Stacking , helfen dabei, Modelle für eine verbesserte Leistung zu kombinieren. • GANs [34]: GANs generieren synthetische Daten, um Trainingsdatensätze zu erweitern, was eine wichtige Komponente zur Verbesserung der Gesamtqualität eines Modells ist. • RNNs [35]: RNNs werden verwendet, um zeitliche Muster zu analysieren und so zu einem besseren Verständnis zu verhelfen . stehende emotionale Zustände. • Hybridmodelle: Die Zusammenführung von CNNs mit SVMs oder Entscheidungsbäumen ermöglicht die Handhabung unterschiedlicher Datentypen, insbesondere wenn es um Komplexität geht. • Neural Architecture Search (NAS) [36]: Bei der Neural Architecture Search wird der Entdeckungsprozess automatisiert, um eine optimale neuronale Netzwerkstruktur zu erreichen, die je nach Art des vorliegenden Datensatzes am besten geeignet ist.

Der Einsatz dieser Methoden kann Modelle zur Emotionserkennung erheblich verbessern und zur Schaffung eines wesentlich besseren Instruments für die klinische Diagnose führen.

Ethische Implikationen und Patientenakzeptanz könnten auch Gegenstand künftiger Studien zu KI-Tools sein, die in der klinischen Diagnostik eingesetzt werden. Dadurch wird sichergestellt, dass solche Technologien das menschliche Urteilsvermögen ergänzen und nicht ersetzen.

Machine Translated by Google

Der Einsatz von CNN-Modellen bei der Diagnose psychosomatischer Erkrankungen trägt zu dem bedeutenden Schritt bei, der an der Schnittstelle zwischen Technologie und Gesundheit gemacht werde Andere technologische Durchbrüche wie das im Rahmen dieser Forschung entwickelte Instrument zur Emotionserkennung haben potenziell sogar noch größeren Wert, da sie zu präziseren Diagnosen und effektiveren Behandlungen führen können, was zu besseren Behandlungsergebnissen führt und ein tieferes Verständnis davon ermöglicht, wie emotionale Zustände mit der körperlichen Gesundheit zusammenhängen (was wiederum Informationen über ihre gegenseitige Abhängigkeit liefert). Wir müssen jedoch sicherstellen, dass diese Instrumente **unter angemessener Berücksichtigung ethischer Grundsätze, der Wahrung der Patientenvertraulichkeit und – nicht zu vergessen – Mechanismen entwickelt werden ,** die **die Arzt-Patienten-Bindung stär**

Beiträge der Autoren: Konzeptualisierung, MN und AR; Methodik, YM; Validierung, MN und YM; Untersuchung, MN; Schreiben – Originalentwurf, MN; Schreiben – Überprüfung und Bearbeitung, AR und YM; Überwachung, AR. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimme

Finanzierung: Diese Forschung erhielt keine externe Finanzierung.

Datenverfügbarkeitserklärung: Die in diesem Artikel präsentierten Datensätze sind nicht ohne weiteres verfügbar, da sie Teil einer laufenden Doktorarbeit sind. Anfragen zum Zugriff auf die Datensätze sind an Marta Narigina zu richten.

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Verweise

1. **Hong, AJ; DiStefano, D.; Dua, S. Können CNNs menschliche Emotionen genau klassifizieren? Eine vergleichende Studie.** arXiv **2023,** **arXiv:2310.09473. Online verfügbar: https://arxiv.org/pdf/2310.09473 (abgerufen am 1. Juni 2024).**
2. **Romanovs, A.; Sultanovs, E.; Buss, E.; Merkuryev, Y.; Majore, G. Herausforderungen und Lösungen für belastbare Telemedizindienste. In Proceedings des 2020 IEEE 8th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), Vilnius, Litauen, 22.–24. April 2021; S. 1–7. [CrossRef]**
3. **Navakauskiene, R.; Treigyte, G.; Borutinskaite, V.-V.; Matuzevicius, D.; Navakauskas, D.; Magnusson, K.-E. Alpha-Dystrobrevin und seine assoziierten Proteine in menschlichen Promyelozytenleukämiezellen, die zur Apoptose induziert wurden.** J. Proteom. **2012,** 75, **3291 [Querverweis] [PubMed]**
4. **Narigina, M.; Kempelis, A.; Romanovs, A. Maschinelles Lernen als Grundlage für die Prognose von Sensordaten zur Verbesserung der Umweltsensorik .** WSEAS Trans. Syst. **2023,** 22, **543–555. [CrossRef]**
5. **Valiuliene, G.; Treigyte, G.; Savickiene, J.; Matuzevicius, D.; Alksne, M.; Jarasiene-Burinskaja, R.; Bukelskiene, V.; Navakauskas, D.; Navakauskiene, R. Histonmodifikationsmuster in Geweben und Tumoren aus einem Xenograft- Modell für akute Promyelozytenleukämie als Reaktion auf eine kombinierte epigenetische Therapie.** Biomed. Pharmacother. **2016,** 79, **62–70. [CrossRef] [PubMed]**
6. **Narigina, M.; Osadcijs, E.; Romanovs, A. Analyse medizinischer Datenverarbeitungstechnologien. In Proceedings der 63. Internationalen Wissenschaftlichen Konferenz für Informationstechnologie und Managementwissenschaften der Technischen Universität Riga (ITMS) 2022, Riga, Lettland, 6.–7. Oktober 2022; S. 1–6. [CrossRef]**
7. **Sultanovs, E.; Strebko, J.; Romanovs, A.; Lektauers, A. Die Informationstechnologien im Kontrollmechanismus medizinischer Prozesse. In Proceedings der 61. Internationalen Wissenschaftlichen Konferenz zu Informationstechnologie und Managementwissenschaften der Technischen Universität Riga (ITMS), Riga, Lettland, 15.–16. Oktober 2020; S. 1–5. [CrossRef]**
8. **Wei, Z. Eine neuartige Methode zur Gesichtsausdruckserkennung für Echtzeitanwendungen. In** der IOP-Konferenzreihe: Materialwissenschaft und -technik; **IOP Publishing: Bristol, Großbritannien, 2019; Band 612, S. 062048. Online verfügbar: https://iopscience.iop.org/article/10.1 088/1757-899X/612/5/052048/pdf (abgerufen am 1. Juni 2024).**
9. **Qu, D.; Zheng, Y.; Li, H.; Wang, J.; Chen, X. Gesichtsemotionserkennung mit CNN in PyTorch.** arXiv **2023, arXiv:2312.10818. Online verfügbar: https://arxiv.org/pdf/2312.10818 (abgerufen am 22. Juli 2024).**
10. **Verkruysse, W.; Svaasand, LO; Nelson, JS Remote photoplethysmographic imaging using ambient light.** Opt. Express **2008,** 16, **21434–21445. [CrossRef] [PubMed]**
11. **Georgescu, M.-I.; Ionescu, RT; Popescu, M. Lokales Lernen mit tiefen und handgefertigten Funktionen zur Gesichtsausdruckserkennung.** IEEE Access **2019, 7, 64827–64836. [CrossRef]**
12. **Pecoraro, R.; Basile, V.; Bono, V. Lokale Mehrkopfkanal-Selbstaufmerksamkeit zur Gesichtsausdruckserkennung.** Informationen **2022,** 13, **419. [CrossRef]**
13. **Fard, AP; Mahoor, MH Ad-Corre: Adaptiver korrelationsbasierter Verlust für die Gesichtsausdruckserkennung in der freien Wildbahn.** IEEE-Zugriff **2022,** 10, **26756–26768. [CrossRef]**
14. **Vignesh, S.; Savithadevi, M.; Sridevi, M.; Sridhar, R. Ein neuartiges Gesichtsemotionserkennungsmodell mittels Segmentierung VGG-19 Architektur.** Int. J. Inf. Technol. **2023,** 15, **1777–1787. [CrossRef]**
15. **Mukhopadhyay, M.; Dey, A.; Kahali, S. Eine auf Deep Learning basierende Methode zur Gesichtsausdruckserkennung unter Verwendung von Texturmerkmalen.** Neural Comput. Appl. **2023,** 35, **6499–6514. [CrossRef]**

Machine Translated by Google

16. Shahzad, T.; Iqbal, K.; Khan, MA; Iqbal, N. Rolle der Zonierung im Gesichtsausdruck mittels Deep Learning. IEEE Access **2023,** 11, **16493–16508.** [CrossRef]
17. El Boudouri, Y.; Bohi, A. EmoNeXt: Ein angepasster ConvNeXt zur Gesichtsemotionserkennung. In Proceedings der 2023 IEEE 25th Internationaler Workshop zur Multimedia-Signalverarbeitung (MMSP), Poitiers, Frankreich, 27.–29. September 2023. [CrossRef]
18. Goodfellow, IJ; Erhan, D.; Carrier, PL; Courville, A.; Mirza, M.; Hamner, B.; Cukierski, W.; Tang, Y.; Thaler, D.; Lee, D.-H. Herausforderungen beim Repräsentationslernen: Ein Bericht über drei Wettbewerbe im Bereich maschinelles Lernen. In Proceedings of the Neural Information Processing: 20. Internationale Konferenz, ICONIP 2013, Daegu, Republik Korea, 3.–7. November 2013; Proceedings, Springer: Berlin/Heidelberg, Deutschland, 2013; S. 117–124. [CrossRef]
19. Sambare, M. FER 2013: Facial Expression Recognition Dataset [Datensatz]. Kaggle. Online verfügbar: https://www.kaggle.com/ Datensätze/msambare/fer2013/Daten (abgerufen am 1. Juni 2024).
20. Chand, HV; Chrisanthus, A.; Thampi, AK Eine Übersicht über verschiedene CNN-basierte Ansätze zur Gesichtsemotionserkennung. In Proceedings des International Congress on Information and Communication Technology, Lalitpur, Nepal, 26.–28. April 2023; Online verfügbar: https://ieeexplore.ieee.org/document/10133947 (abgerufen am 1. Juni 2024). 21. ¸Sen, SY; Özkurt, N. Convolutional
Neural Network Hyperparameter Tuning mit Adam Optimizer zur EKG-Klassifizierung. In Proceedings der 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), Istanbul, Türkei, 15.–17. Oktober 2020; S. 1–6. [CrossRef]
22. Ozdemir, MA; Elagoz, B.; Alkan, A. Echtzeit-Emotionserkennung aus Gesichtsausdrücken mithilfe von CNN. In Proceedings des Medical Technologies Congress 2019 (TIPTEKNO), Izmir, Türkei, 3.–5. Oktober 2019; IEEE: Piscataway, NJ, USA, 2019. Online verfügbar: https://ieeexplore.ieee.org/document/8895215 (abgerufen am 1. Juni 2024).
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Empirische Evaluation gleichgerichteter Aktivierungen in Faltungsnetzwerken. arXiv 2015. [CrossRef]
24. Chowdhury, S.; Chowdhury, S.; Ifty, JT; Khan, R. Fahrzeugerkennung und -klassifizierung mithilfe tiefer neuronaler Netzwerke. In Proceedings der 2022 International Conference on Electrical and Information Technology (IEIT), Malang, Indonesien, 15.–16. September 2022; S. 95–100. [CrossRef]
25. Wang, Y.; Bian, Z.-P.; Hou, J.; Chau, L.-P. Faltungsneuronale Netzwerke mit dynamischer Regularisierung. IEEE Trans. Neural Netw. Lernen. Syst. **2021,** 32, **2299–2304.** [CrossRef]
26. Li, X.; Chen, J.; Zhao, G.; Pietikäinen, M. Fernmessung der Herzfrequenz anhand von Gesichtsvideos in realistischen Situationen. In Proceedings der 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23.–28. Juni 2014; S. 4264–4271. [CrossRef]
27. Pradeep, V.; Madhushree, B.; Sumukha, BS; Richards, GR; Prashant, SP Erkennung von Gesichtsemotionen mit CNN und OpenCV. In Proceedings der 2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS), Bengaluru, Indien, 25.–26. April 2024; S. 1–6. [CrossRef]
28. Yang, Y.; Liu, F.; Li, M.; Jin, J.; Weber, E.; Liu, Q.; Crozier, S. Pseudopolare Fourier-Transformation – Compressed Sensing MRI auf Basis der virtuellen Realität. IEEE Trans. Biomed. Eng. **2017,** 64, **816–825.** [CrossRef]
29. Tarassenko, L.; Villarroel, M.; Guazzi, A.; Jorge, J.; Clifton, DA; Pugh, C. Berührungslose videobasierte Vitalzeichenüberwachung mittels Umgebungslicht und autoregressive Modelle. Physiol. Meas. **2014,** 35, **807–831.** [CrossRef]
30. De Haan, G.; Jeanne, V. Robuste Pulsfrequenz aus chrominanzbasiertem rPPG. IEEE Trans. Biomed. Eng. **2013,** 60, **2878–2886.** [CrossRef]
31. Poh, M.-Z.; McDuff, DJ; Picard, RW Berührungslose, automatisierte Herzpulsmessungen mittels Videobildgebung und blinder Quellentrennung . Opt. Express **2010,** 18, **10762–10774.** [CrossRef]
32. Song, C.; Chen, C.; Li, Y.; Wu, X. Anwendung von Deep Reinforcement Learning bei der Klassifizierung von Elektromyographiedaten. In Proceedings der 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS), Shenzhen, China, 25.–27. Oktober 2018; S. 505–510. [Querverweis]
33. Gang, Z.; Jia, C.; Guo, C.; Li, P.; Gao, J.; Zhao, L. Vorhersage der chronisch obstruktiven Lungenerkrankung basierend auf einem mehrstufigen Composite Ensemble Learning Framework. In Proceedings der 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Istanbul, Türkei, 5.–8. Dezember 2023; S. 1921–1924. [CrossRef]
34. Chang, J.; Zhang, Z.; Wang, Z.; Li, J.; Meng, L.; Lin, P. Generative Listener EEG zur Sprachemotionserkennung unter Verwendung generativer kontradiktorischer Netzwerke mit Compressed Sensing. IEEE J. Biomed. Health Inform. **2024,** 28, **2025–2036.** [CrossRef] [PubMed]
35. Le, MD; Singh Rathour, V.; Truong, QS; Mai, Q.; Brijesh, P.; Le, N. Multimodulares rekurrentes Convolutional Neural Network mit Transformer Encoder zur EKG-Arrhythmie-Klassifizierung. In Proceedings der 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Athen, Griechenland, 27.–30. Juli 2021; S. 1–5. [CrossRef]
36. Zhao, F.; Nie, J.; Ma, M.; Chen, X.; He, X.; Wang, B.; Hou, Y. Bewertung der Rolle verschiedener heterogener Regionen in der DCE-MRT zur Vorhersage molekularer Subtypen von Brustkrebs basierend auf Network Architecture Search und Vision Transformer. In Proceedings der 45. jährlichen internationalen Konferenz der IEEE Engineering in Medicine & Biology Society (EMBC) 2023, Sydney, Australien, 24.–27. Juli 2023; S. 1–4. [CrossRef]

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultieren.

Machine Translated by Google

Algorithmen

Artikel

Lester: Rotoskopie-Animation durch Videoobjekt Segmentierung und Tracking

Ruben Tous

Abteilung für Computerarchitektur, Universitat Politecnica de Catalunya, 08034 Barcelona, Spanien; ruben.tous@upc.edu

Zusammenfassung: Dieser Artikel stellt Lester vor, eine neuartige Methode zur automatischen Synthese von 2D-Animationen im Retro-Stil aus Videos. Die Methode nähert sich der Herausforderung hauptsächlich als Problem der Objektsegmentierung und -verfolgung. Videobilder werden mit dem Segment Anything Model (SAM) verarbeitet und die resultierenden Masken werden durch nachfolgende Bilder mit DeAOT verfolgt, einer Methode der hierarchischen Ausbreitung für die halbüberwachte Segmentierung von Videoobjekten. Die Geometrie der Konturen der Masken wird mit dem Douglas-Peucker-Algorithmus vereinfacht. Schließlich können optional Gesichtsmerkmale, Pixelierung und ein grundlegender Randlichteffekt hinzugefügt werden. Die Ergebnisse zeigen, dass die Methode eine ausgezeichnete zeitliche Konsistenz aufweist und Videos mit unterschiedlichen Posen und Erscheinungsbildern, dynamischen Aufnahmen, Teilaufnahmen und unterschiedlichen Hintergründen korrekt verarbeiten kann. Die vorgeschlagene Methode bietet einen einfacheren und deterministischeren Ansatz als auf Diffusionsmodellen basierende Video-zu-Video-Übersetzungspipelines, die unter zeitlichen Konsistenzproblemen leiden und mit pixeligen und schematischen Ausgaben nicht gut zurechtkommen. Darüber hinaus ist die Methode praktikabler als Techniken, die auf der dreidimensionalen Schätzung der menschlichen Körperhaltung basieren, individuelle, handgefertigte 3D-Modelle erfordern und hinsichtlich der Art der Szenen, die sie vera

Schlüsselwörter: Animation; Rotoskopie; Segmentierung; Computergrafik; Video; künstliche Intelligenz

check for updates

Zitat: Tous, R. Lester: Rotoskopie-Animation durch Videoobjektsegmentierung und -verfolgung. Algorithmen **2024,** 17, **330. https://doi.org/10.3390/a17080330**

Wissenschaftlicher Redakteur: Arslan Munir

Empfangen: 24. Juni 2024
Überarbeitet: 23. Juli 2024
Akzeptiert: 25. Juli 2024
Veröffentlicht: 30. Juli 2024

Copyright: © 2024 beim Autor. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/4.0/).

1. Einleitung

In Branchen wie Kino oder Videospielen wird Rotoskopie eingesetzt, um die Informationen aus Live-Action-Filmmaterial für die Erstellung von 2D-Animationen zu nutzen. Diese Technik trug zur Entstehung einer einzigartigen visuellen Ästhetik in bestimmten Kult-Videospielen der 80er und 90er Jahre bei, wie „Prince of Persia“ (1989), „Another World“ (1991) oder „Flashback“ (1992), und sie wird weiterhin in modernen Titeln wie „Lunark“ (2023) eingesetzt. Wie im Fall von Lunark erfordert diese Technik jedoch immer noch manuelle Arbeit auf Frame-Ebene, was sie sehr teuer macht.

Jüngste Fortschritte in der künstlichen Intelligenz (KI) haben die Möglichkeit eröffnet, diesen Prozess bis zu einem gewissen Grad zu automatisieren. Die jüngsten Fortschritte bei bedingten generativen Modellen für die Video-zu-Video-Konvertierung bieten einen vielversprechenden Ansatz zur Lösung dieses Problems. Diese Methoden sind jedoch noch nicht in der Lage, durchgängig ein ausgefeiltes, professionelles Ergebnis mit einheitlichem Stil und räumlicher und zeitlicher Konsistenz zu liefern . Ihre Leistung variiert erheblich je nach verschiedenen Aspekten des Eingabevideos, wie Szene und Auflösung. Darüber hinaus sind sie langsam und erfordern enorme Rechenressourcen. Daher bleibt der Bedarf an einer effizienten und deterministischen automatisierten Lösung ein ungelöstes Problem.

Dieser Artikel beschreibt eine neuartige Methode namens Lester, mit der dieser Prozess automatisiert wird. Der Benutzer gibt ein Video und eine Farbpalette ein und das System generiert automatisch eine 2D-Animation im Retro-Stil in einem Format, das direkt in einem Videospiel oder einem Animationsfilm verwendet werden kann (siehe Abbildung 1 für ein Beispiel).

Die Methode nähert sich der Herausforderung hauptsächlich als ein Problem der Objektsegmentierung. Aus einem Bild, das eine Person darstellt, wird mit dem Segment Anything Model (SAM) [1] eine Maske für jedes unterschiedliche visuelle Merkmal (Haare, Haut, **Kleidung usw.)** generiert . **Um die Masken eindeutig zu identifizie**

Machine Translated by Google

In allen Einzelbildern wird ein Video-Objekt-Tracking-Algorithmus nach dem Ansatz von [2] eingesetzt. Die Masken werden bereinigt, um Segmentierungsfehler zu beheben, und ihre Geometrie wird mit dem Douglas-Peucker-Algorithmus [3] vereinfacht , um den gewünschten visuellen Stil zu erreichen. Abschließend können optional noch Gesichtszüge, Pixelierung und ein einfacher Randlichteffekt hinzugefügt werden.





Abbildung 1. Beispiel eines synthetisierten Animationsbildes (rechts) und des zugehörigen Originalbildes (links, Gesicht wurde anonymisiert) aus einem Beispiel-Eingabevideo.

2. Verwandte Arbeiten

Max Fleischers Erfindung des Rotoskops im Jahr 1915 revolutionierte die Möglichkeit, Zeichentrickfilme mit Stand- oder Bewegtbildern zu erstellen. Varianten dieser Technik werden bis heute in Animationsfilmen (z. B. „Der dunkle Schirm“ (2006)), Fernsehserien (z. B. „Undone“ (2019)) und Videospielen (z. B. „Prince of Persia“ (1989), „Another World“ (1991), „Flashback“ (1992), „Lunark“ (2023)) verwendet. Heutige Produktionen integrieren Bildverarbeitungsalgorithmen, um die Aufgabe zu erleichtern, aber sie basieren hauptsächlich auf kostspieliger Handarbeit . Eine vollständige Automatisierung mit einem ausgefeilten professionellen Ergebnis (einheitlicher Stil, räumliche und zeitliche Konsistenz) ist derzeit noch nicht realisierbar. Jüngste Fortschritte im Bereich der künstlichen Intelligenz (KI) erö

Nach unserem besten Wissen gibt es zwei Hauptstrategien zur Lösung dieses Problems. Einerseits kann diese Herausforderung mit bedingten generativen Modellen angegangen werden, die heutzutage bei Bildübersetzungsaufgaben weit verbreitet sind. Die ersten Methoden (z. B. CartoonGAN [4], AnimeGAN [5] oder die Arbeit in [6]) verwendeten bedingte generative kontradiktorische Netzwerke (GAN), um Fotos von realen Szenen in Bilder im Cartoon-Stil umzuwandeln. Diese Methoden wurden kürzlich von denen übertroffen, die auf Diffusions Obwohl die meisten dieser Methoden für die Text-zu-Bild-Erzeugung entwickelt wurden, wurden sie mit Techniken wie SDEdit [9] oder ControlNet [10] erfolgreich für die Bild-zu-Bild-Übersetzung angepasst.

Es gibt auch reine Bild-zu-Bild-Übersetzungsmethoden mit Diffusionsmodellen (z. B. SR3 [11] und Palette [12]), jedoch mit einem überwachten Aufbau, der nur auf Situationen angewendet werden kann, in denen eine Grundwahrheit von Bildpaaren (Original, Ziel) wie Superauflösung oder Kolorierung vorliegt . Selbst wenn wir ein gutes bedingtes generatives Modell haben, besteht in jedem Fall immer noch die Herausforderung, es auf die Einzelbilder eines Videos anzuwenden und dabei die zeitliche Konsistenz zu wahren. Die neuesten Ansätze (z. B. [13]) kombinieren Tricks (z. B. gekachelte Eingaben), um die Ausgabe des generativen Modells zu vereinheitlichen, mit videogesteuerten Stilinterpolationstechniken (z. B. EbSynth Die Ergebnisse dieser Methoden haben sich stark verbessert, und für einige Stile sind die Ergebnisse so gut, dass sie bereits in unabhängigen Produktionen verwendet werden. Für die meisten digitalen Kunststile erzwingen diese Methoden jedoch einen unerwünschten Kompromiss zwischen Stiltreue und zeitlicher Konsistenz. Unter den neuesten Arbeiten dieser Art ist Sora [15] von OpenAI besonders hervorzuheben. Die Analyse der bereitgestellten Beispiele zeigt, dass Sora die physische Welt hervorragend erfasst und über die grundlegenden Elemente eines Weltmodells verfügt. Es hat das Potenzial, sich in Zukunft zu einer Physik-Engine zu entwickeln [16].

Ein völlig anderer Ansatz besteht dagegen darin, einen 3D-Algorithmus zur Schätzung der menschlichen Pose anzuwenden, um die 3D-Pose durch die verschiedenen Frames zu verfolgen. Die 3D-Pose wird dann auf ein 3D-Charaktermodell mit dem gewünschten Aussehen (Haare, Kleidung usw.) angewendet. Normalerweise wird dieses 3D-Charaktermodell von Hand mit einer speziellen Software erstellt. Schließlich werden eine geeignete Shader- und Rendering-Konfiguration eingerichtet, um den gewünschten visuellen Stil zu erhalten. Dies ist der Ansatz, der in [17–21] verfolgt wird. Dieser Ansatz hat viele Nachteile. Erstens hängt er vom Stand der Technik der Schätzung der 3D-Menschenpose ab, der zwar wei

Machine Translated by Google

Situationen (niedrige Auflösung, Kürzungen, ungewöhnliche Posen), hauptsächlich aufgrund der Schwierigkeit, genügend Grunddaten zu erhalten. Zweitens erfordert es die Entwicklung eines geeigneten 3D-Modells für jeden einzelnen Charakter. Drittens basiert die 3D-Pose auf einem vorgegebenen Mod Um es auf ein anderes Modell anzuwenden, ist ein Schritt zur Neuausrichtung der Pose erforderlich. Die Gelenkrotationen können leicht neu ausgerichtet werden, die SMPL-Blendformen jedoch nicht, was zu unrealistischen Körperverformungen führt . Aus diesem Grund werden Demos dieser Art von Arbeit normalerweise mit Robotern durchgeführt, da diese keine verformbaren Teile haben. Ein Vorteil dieses Ansatzes im Vergleich zu generativen Modellen besteht darin, dass er eine gute zeitliche Konsistenz bietet, da Inkonsistenzen aus dem Schritt zur Schätzung der 3D-Menschenpose durch Interpolation, wie **in** [20] **beschrieben, ordnungsgemäß gelöst werden können .**

In dieser Arbeit schlagen wir einen dritten Ansatz vor, der hauptsächlich auf Segmentierung und Tracking basiert. Im Gegensatz zur zuerst erwähnten Strategie (generative Modelle) ist der vorgeschlagene Ansatz vorhersehbarer und bietet eine viel bessere zeitliche Konsistenz. Im Gegensatz zur zweiten erwähnten Strategie (3D-Schätzung der menschlichen Pose) ist der vorgeschlagene Ansatz robuster und viel praktischer, da er kein benutzerdefiniertes 3D-Modell für jeden neuen Charakter erfordert.

Dies ist nicht das erste Mal, dass diese Art von Problem auf diese Weise angegangen wurde. In [22] wurde Segmentierung verwendet, um ein Porträtvideo anhand eines künstlerischen Beispielbildes zu stilisieren, wobei bessere Ergebnisse erzielt wurden als **mit dem damals vorherrschenden neuronalen Stiltransfer |**

3. Methodik 3.1. Überblick

Der Hauptinput für die Methode (der Code ist verfügbar unter https://github.com/rtous/ lester-code (abgerufen am 20. Juli 2024); der Datensatz und die Ergebnisse der Experimente sind öffentlich verfügbar unter https://github.com/rtous/lester (abgerufen am 20. Juli 2024)) ist eine Zielvideosequenz T **einer menschlichen Performance. Es gibt keine allgemeinen Einschränkungen hinsichtlich der Merkmale der abgebildeten Person (Geschlecht, Alter, Kleidung usw.), Pose, Ausrichtung, Kamerabewegungen , Hintergrund, Beleuchtung, Okklusionen, Ausschnitte (Teilkörperaufnahmen) oder Auflösun Es ist jedoch bekannt, dass bestimmte Szenarien die Ergebnisse negativ beeinflussen können (dies wird in Abschnitt 4 behandelt). Während die Methode in erster Linie für Videos mit einer einzelnen Person entwickelt wurde, wird ihre theoretische Anwendung mit mehreren Personen in Abschnitt 4 erörtert. Beim Betrieb ohne zusätzliche Eingaben verwendet die Methode standardmäßig bestimmte Einstellungen für Segmentierung, Farben und andere Attribute. Es ist jedoch davon auszugehen, dass der Benutzer möglicherweise benutzerdefinierte Konfigurationen vorgibt, was zusätzliche Eingaben wie Segmentierungsaufforderungen, Farbpaletten usw. erforderlich macht, wie in den entsprechenden Unterabschnitten weiter unten beschrieben. Die Ausgabe der Methode ist ein Satz von PNG-Bildern, die die resultierende animierte Figur und keinen Hintergrund zeigen (wobei der Alphakanal auf Null gesetzt ist).**

Die Methode besteht aus drei unabhängigen Verarbeitungsschritten. Der wichtigste ist die Segmentierung und Verfolgung der verschiedenen visuellen Merkmale (Haare, Haut, Kleidung usw.) der in den Videobildern dargestellten Person. Der zweite Schritt konzentriert sich auf die Behebung von Segmentierungsfehlern durch Maskenbereinigung und Vereinfachung ihrer polygonalen Konturen. Der dritte und letzte Schritt umfasst das Anwenden anpassbarer Endbearbeitungsdetails, einschließlich Farben, künstlichem Randlicht, Gesichtszügen und Pixelierung. Abbildung 2 zeigt den gesamten Arbeitsablauf der Methode. Jede Komponente wird in den folgenden Unterabschnitten näher erläutert.

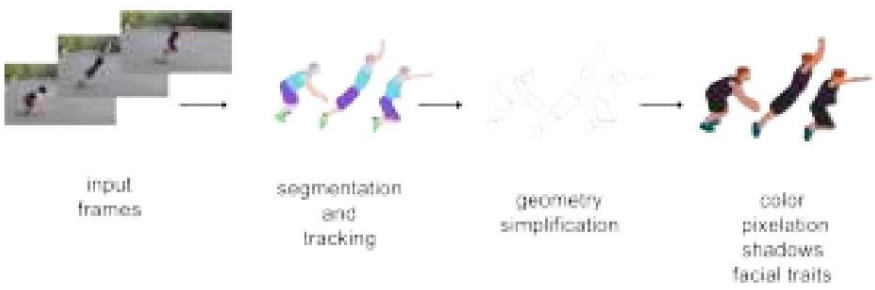


Abbildung 2. Übersicht über den Arbeitsablauf des Systems (Eingaben links, Ausgaben rechts).

Machine Translated by Google

3.2. Segmentierung und Tracking

Zur Segmentierung der verschiedenen visuellen Merkmale des Charakters wird das Segment Anything Model (SAM) [1] verwendet. SAM ist ein auf Eingabeaufforderungen basierendes Segmentierungsmodell, das sich als Basismodell zur Lösung einer Vielzahl von Aufgaben etabliert. Aus einem Eingangsbild mit drei Kanälen (Höhe × Breite × 3) und einer Eingabeaufforderung, die räumliche oder Textinformationen enthalten kann, wird eine Maske (Höhe × Gewicht × 3) generiert, wobei jedem Pixel eine RGB-Farbe zugewiesen wird. Jede Farbe repräsentiert eine Objektbezeichnung aus dem Objektbezeichnungsraum L = l1, l2, ... , die aus der Eingabeaufforderung abgeleitet wurde und dem vorhergesagten Der Prompt kann jede beliebige Information sein, die angibt, was in einem Bild segmentiert werden soll (z. B. eine Reihe von Vordergrund-/Hintergrundpunkten, eine grobe Box oder Maske oder ein frei formulierter Text). SAM besteht aus einem Bildencoder (basierend auf MAE [24], einem vorab trainierten Vision Transformer [25]), der eine Bildeinbettung berechnet , einem Prompt-Encoder (der den frei formulierten Text basierend auf CLIP [26] darstellt), der Prompts einbettet, und dann werden die beiden Informationsquellen in einem Maskendecoder (basierend auf Transformer-Segmentierungsmodellen [27]) kombinier

Die Anwendung von SAM auf Frame-Ebene würde zu unterschiedlichen Labels für dasselbe Objekt in unterschiedlichen Frames führen. SAM geht nicht von einer geschlossenen Liste von Objektlabels aus, sondern erstellt und vergibt Labels während der Ausführung. Daher gibt es keine Garantie dafür, dass in jedem Frame demselben Objekt dieselben Labels zugewiesen werden. Um dies konsistent zu machen, verw Die SAM-Ergebnisse des ersten Frames werden mit DeAOT [28], einer Methode der hierarchischen Ausbreitung zur halbüberwachten Segmentierung von Videoobjekten, mit denselben Beschriftungen an den zweiten Frame weitergegeben. Die hierarchische Ausbreitung kann Informationen aus früheren Frames schrittweise an den aktuellen Frame weitergeben und die Funktion des aktuellen Frames von objektagnostisch auf objektspezifisch übertragen. SAM wird auf nachfolgende Frames angewendet, um neue Objekte zu erkennen, aber die Segmentierung bereits gesehener Objekte wird von DeAOT durchgeführt.

Wir haben uns für schlüsselwortbasierte Eingabeaufforderungen entschieden, um die gewünschte Segmentierung anzugeben. Wir bieten einige Standardeingabeaufforderungen für gängige Szenarien (z. B. „Haare“, „Haut“, „Schuhe“, „T-Shirt“, „Hose“), aber es ist davon auszugehen, dass der Benutzer in vielen Szenarien eine Alternative angeben muss . Diese Eingabeaufforderungen werden aufgrund ihrer Einschränkungen nicht mit den Freitextfunktionen von SAM verarbeitet , sondern mit Grounding-DINO [29], einem Objektdetektor für offene Mengen , der sie in Objektboxen übersetzt, die als Box-Eingabeaufforderungen an SAM gesendet werden. Ein Problem, das wir identifiziert haben, besteht darin, dass bestimmte Objektkombinationen zu falschen Masken führen, wenn sie in derselben Eingabeaufforderung angegeben werden. Um dieses Problem zu lösen, ermöglichen wir die Angabe mehrerer Eingabeaufforderungen , die unabhängig voneinander verarbeitet werden. Abbildung 3 zeigt den Arbeitsablauf der Segmentierungs- und Verfolgung

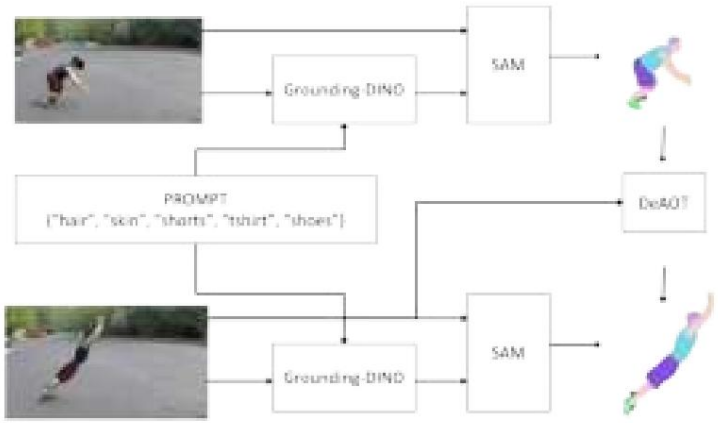


Abbildung 3. Arbeitsablauf der Segmentierungs- und Tracking-Phase für die ersten beiden Frames (Eingaben links, Ausgabe rechts).

3.3. Vereinfachung der Konturen

Eine der gewünschten Optionen für den visuellen Zielstil ist die Möglichkeit, sehr einfache Geometrien anzuzeigen, d. h. Konturen mit kurzen polygonalen Ketten, um den visuellen Aspekt von z. B. „Another World“ nachzuahmen. Um dies zu erreichen, werden eine Reihe von Schritten ausgeführt.

Machine Translated by Google

Für jedes Bild ist das Ergebnis der Segmentierung und Verfolgung eine Maske mit jedem Pixel eine RGB-Farbe zugewiesen, die eine Objektbeschriftung darstellt. Die Maske wird zunächst in Untermasken unterteilt, eine für jede Farbe/Beschriftung. Jede Untermaske wird wiederum in Konturen unterteilt mit dem Algorithmus aus [30]. Jede einzelne Kontur wird durch einen Vektor der Scheitelpunktkoordinaten dargestellt von die Begrenzungsliniensegmente. Die Anzahl der Liniensegmente jeder Kontur wird reduziert mit der Douglas–Peucker-Algorithmus [3]. Der Algorithmus beginnt mit der Verbindung der Endpunkte der Kontur mit einer geraden Linie. Eckpunkte, die näher an der Linie liegen als ein gegebener Toleranzwert t, **sind entfernt. Anschließend wird die Linie an dem Punkt geteilt, der am weitesten von der Referenzlinie entfernt ist. Dadurch entstehen zwei neue Linien. Der Algorithmus arbeitet rekursiv, bis alle Eckpunkte innerhalb der Toleranz entfernt sind. Die Toleranz** t **ist einer der optionalen Parameter der vorgeschlagenen Methode und ermöglicht dem Benutzer steuern den Grad der Vereinfachung. Da jede Kontur unabhängig vereinfacht wird, sind unerwünschte Zwischen den Geometrien angrenzender Konturen können Lücken entstehen, die zuvor perfekt ausgerichtet. Um dies zu vermeiden, werden vor dem Vereinfachungsschritt alle Konturen erweitert Faltung der Untermaske mit einem 4 × 4-Kernel. Es ist erwähnenswert, dass die Vereinfachung Algorithmus kann für bestimmte Konturgeometrien und Werte von t fehlschlagen. In solchen Situationen (normalerweise sehr kleine Konturen) bleibt die ursprüngliche Kontur erhalten. Siehe Algorithmus 1 und Abbildung 4.**



(A) **(B)** **(C)** **(D)** **e)**

Abbildung 4. Grafische Übersicht über die Konturvereinfachungsphase eines Frames: (a) Ergebnis des Segmentierungs- und Tracking-Schritts; (b) Konturerkennung; (c) Konturerweiterung; (d) Konturvereinfachu (e) resultierender farbiger Rahmen.

Algorithmus 1 Konturvereinfachungsalgorithmus

Gefordert wird: M **eine Liste mit Untermasken,** ÿ **eine minimale Konturfläche,** t **ein Toleranzwert für Vereinfachung**

Stellen Sie sicher, dass: R **eine Liste von Konturen ist, wobei Ri,j die vereinfachte Kontur** j **der Teilmaske** i ist

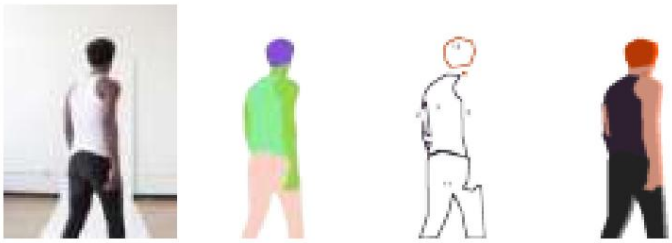
Funktion SIMPLIFY(M, ÿ, t)
  für i **ÿ 0 bis** Länge(M) **ÿ 1 mache**
    Ci **ÿ** Konturen abrufen (Mi)
    für j **ÿ 0 bis** Länge(Ci) **ÿ 1 do**
      wenn FlächeO f(Ci,j) > ÿ **dann**
        Di,j ÿ dilateContour(Ci,j)
        if contextCanBeSimpli f ied(Di,j , t) **then**
          Ri,j ÿ simpli f yContour(Di,j , t)
        anders
          Ri,j ÿ Di,j
        Ende, wenn
      Ende, wenn
    Ende für
Ende für
  Rückgabe R
Funktion beenden

contourCanBeSimpli f ied **gibt false zurück, wenn der Douglas-Peucker-Algorithmus fehlschlägt, wie bei bestimmten Konturgeometrien und t-Werten auftreten.**

Machine Translated by Google

3.4. Details zur

Fertigstellung **Der dritte und letzte Schritt der Methode umfasst das Anwenden anpassbarer Details zur Fertigstellung, die den Benutzern Flexibilität bei der Gestaltung des visuellen Stils des Ergebnisses geben. Der wichtigste Aspekt ist die Farbe. Die Methode generiert eine visuelle Anleitung mit eindeutigen Kennungen für jedes Farbsegment, wie in Abbildung 5 dargestellt. Einer der Inputs der Methode ist eine Farbpalette, die die RGB-Werte angibt, die auf jede Segmentkennun**



(A) **(B)** **(C)** **(D)**

Abbildung 5. Farbgebungsprozess: (a) Eingabebild; (b) Segmentierung; (c) visuelle Anleitung mit eindeutigen Kennungen für jedes Farbsegment; (d) Endbild.

Ein weiteres optionales Detail zur Verfeinerung ist die Möglichkeit, einige schematisch wirkende Gesichtsmerkmale hinzuzufügen. Die Merkmale werden aus den 68 Gesichtsmerkmalen generiert, die mit dem in **[31] beschriebenen Frontalgesichtsdetektor gewonnen werden** können, **der ein Histogramm orientierter Gradienten (HOG) [32] in Kombination mit einem linearen Klassifikator, einer Bildpyramide und einem Gleitfenster-Erkennungsschema verwendet. Eine weitere Option ist die Möglichkeit, einen sehr einfachen Randlichteffekt hinzuzufügen, indem dasselbe Bild mit unterschiedlicher Helligkeit überlagert und horizontal verschoben wird. Schließlich wird ein Pixeleffekt erzielt, indem das Bild mithilfe einer bilinearen Interpolation verkleinert ur Abbildung 6 zeigt, wie Randlicht, Gesichtszüge und Pixelung schrittweise auf ein Beispielbild angewendet werden.**



(A) **(B)** **(C)** **(D)**

Abbildung 6. Optionale Details zur Endbearbeitung: (a) Ergebnis der Flat-Methode (b) Randlicht (c) Gesichtszüge (d) Pixelung.

4. Experimente und Ergebnisse

Um die Leistungsfähigkeit der Methode zu bewerten, haben wir einen Satz von 25 kurzen Videos mit einer Dauer von 5 bis 20 Sekunden zusammengestellt. Die Videos zeigen eine einzelne Person, die verschiedene Kleidungsstile zeigt und verschiedene Aktionen ausführt. Die Videos stammen aus dem UCF101 Human Actions-Datensatz [33], dem Fashion Video Dataset [34], dem Kinetics-Datensatz [35] und YouTube. Der zusammengestellte Datensatz enthält sowohl Videos mit niedriger als auch hoher Auflösung sowie sowohl Videos aus der freien Natur als auch aus kontrollierten Umgebungen. Der Datensatz und die Ergebnisse der Experimente finden Sie in [36]. Die Abbildungen 7 und 8 zeigen einige Beispielergebnisse.

Um die Wahrnehmungsqualität der generierten Animationen zu bewerten, wurde eine Benutzerstudie durchgeführt. Es wurde eine Umfrage mit Fragen zu drei visuellen Aspekten erstellt: Formqualität , Farbkonsistenz und zeitliche Konsistenz sowie Gesamtpunktzahl. Die Bewertungsoptionen basierten auf der in der ITU-T-Empfehlung P.910 **[37] beschriebenen Kategoriebewertung (ACR) auf einer Skala von 1 = schlecht bis 5 = ausgezeichnet. Die subjektive Qualitätsbewertung der Ergebnisse ist in Tabelle 1 zusammengefasst. Wir berichten über den MOS (Mean Opinion Score, beschrieben in ITU-T P.800.1 [38]). Die Videos sind in drei Kategorien unterte**

Machine Translated by Google

(ITW-LR), in freier Wildbahn hohe Auflösung (ITW-HR) und kontrollierte Umgebung hohe Auflösung (C-HR).

Tabelle 1. Subjektive Qualitätsbewertung MOS-Werte des ACR (1 = Schlecht bis 5 = Ausgezeichnet) von Form Qualität, Farbkonsistenz, zeitliche Konsistenz und Gesamtqualität. [A] **ITW-LR ÿ in freier Wildbahn niedrig Auflösung;** [B] **ITW-HR ÿ hohe Auflösung in freier Wildbahn;** [C] **C-HR ÿ Kontrollierte Einstellung hoher Auflösung.**

Teilmenge	#	Form	Farbe	Zeitlich	Gesamt
ITW-LR [A]	5	3.20	3.24	3,84	3.36
ITW-HR [B]	14	3.81	4.03	4.33	3,99
C-HR [C]	6	4,70	4,60	4,67	4,67
alle		3,90	4.01	4.31	4.02

Stärken: Die Ergebnisse zeigen, dass die Methode in der Lage ist, qualitativ hochwertige Animationen zu synthetisieren von verschiedenen Eingangsvideos (ein durchschnittlicher MOS 4,02 über 5). Die Leistung deutlich verbessert sich (4,67) bei hochauflösenden Videos, die unter kontrollierten Bedingungen aufgenommen wurden (Videos 2, 11, 12, 13, 14 und 15). Die Methode weist eine sehr gute zeitliche Konsistenz auf und kann Verarbeiten Sie Videos mit verschiedenen Posen und Erscheinungen, dynamische Aufnahmen (z. B. Video 5), Videos mit Teilaufnahmen (z. B. die halbnahe Aufnahme in Video 7) und unterschiedlichen Hintergründen.

Es ist anzumerken, dass die Methode, obwohl sie nicht der Hauptschwerpunkt dieser Arbeit ist, Videos mit mehreren Personen verarbeiten. Objekte, die verschiedenen Personen zugeordnet sind, unabhängig voneinander verfolgt werden, jedes mit einer eindeutigen Kennung, selbst wenn die Objekte vom gleichen Typ sind. Dadurch können diese Objekte unterschiedlichen Farben zugeordnet werden, wodurch die gesamte Pipeline auf Aufnahmen mit mehreren Personen angewendet. Abbildung 9 zeigt ein Beispielergebnis aus einem Bild mit mehreren Personen.

Einschränkungen: Die Segmentierungs- und Tracking-Leistung nimmt bei Videos mit niedrige Auflösung (alle Videos aus dem UCF101 Human Actions-Datensatz), komplexe Hintergründe (Videos 17 und 20), kontrastarme Farben (Videos 8, 10, 22 und 23), unangemessene Beleuchtung (Videos 14 und 25) und Okklusionen (Video 23). Abbildung 10 zeigt vier Beispielbilder mit Fehler aus vier verschiedenen Videos (Videos 22, 23, 25 und 8 aus dem Testdatensatz).

Rechenaufwand: Die Methode wurde mit der Programmiersprache Python implementiert . Die Experimente wurden auf einem High-End-Server mit einem Quadcore ausgeführt. Intel i7-3820 mit 3,6 GHz mit 64 GB DDR3 RAM und 4 NVIDIA Tesla K40 GPU-Karten mit jeweils **12 GB GDDR5. Die maximale RAM-Auslastung der Methode wurde mit 3,1 GB aufgezeichnet während der Ausführung der Segmentierungsaufgabe, die auch einen Spitzenwert für den GPU-RAM ausmacht Nutzung von 2,1 GB. Für ein einzelnes Bild dauert Schritt 1 (Segmentierung und Tracking) 0,56 s, Schritt 2 (Vereinfachung der Konturen) dauert 0,43 s und Schritt 3 (Fertigstellung der Details, hauptsächlich der Gesichtszüge) dauert 1,32 s. Insgesamt arbeitet die Methode mit 0,43 FPS (1 FPS ohne Gesichtsmerkmale).**

Vergleich mit Bild-zu-Bild-generativen Modellen: Obwohl wir nicht wissen, Andere Methoden, die versucht haben, das Problem für diesen spezifischen visuellen Stil zu lösen, könnten als Alternative bedingte generative Modelle verwendet werden. Eine mögliche Pipeline, beschrieben in [13] wird Stable Diffusion [7] für die Bild-zu-Bild-Übersetzung unter Verwendung von ControlNet [10] adaptiert, und verwendet gekachelte Eingänge zusammen mit videogeführter Stilinterpolation (EbSynth [14]), um visuelle Konsistenz über mehrere Frames hinweg beibehalten. Wir haben diese Verarbeitungspipeline mit einigen Videos aus dem Testdatensatz, um die Durchführbarkeit dieses Ansatzes zu bewerten. Um das zu erreichen Um den gewünschten Stil zu erreichen, haben wir Stable Diffusion mit Dreambooth [39] feinjustiert . Abbildung 11 zeigt vier Beispielbilder, die mit diesem Ansatz erhalten wurden, aber um die zeitliche Konsistenz zu beurteilen, Ergebnisse finden Sie in unserem Ergebnisrepository in [36]. Die Ergebnisse zeigen mehrere Nachteile im Vergleich zur vorgeschlagenen Methode. Erstens weisen sie eine schlechte zeitliche Konsistenz über Frames. Zweitens erfordert die Methode die Angabe des Grads der Treue zum gewünschten Stil, wo eine höhere Wiedergabetreue die ursprüngliche Rahmenstruktur und Details beeinträchtigt, was zu einer unerwünschter Kompromiss. Drittens variiert die Leistung erheblich je nach verschiedenen Aspekte des Eingangsvideos, wie Szene und Auflösung. Schließlich arbeitet die Methode deutlich langsamer als die vorgeschlagene Methode, mit einer Verarbeitungsgeschwindigkeit von nur 0,3 FPS.

Machine Translated by Google

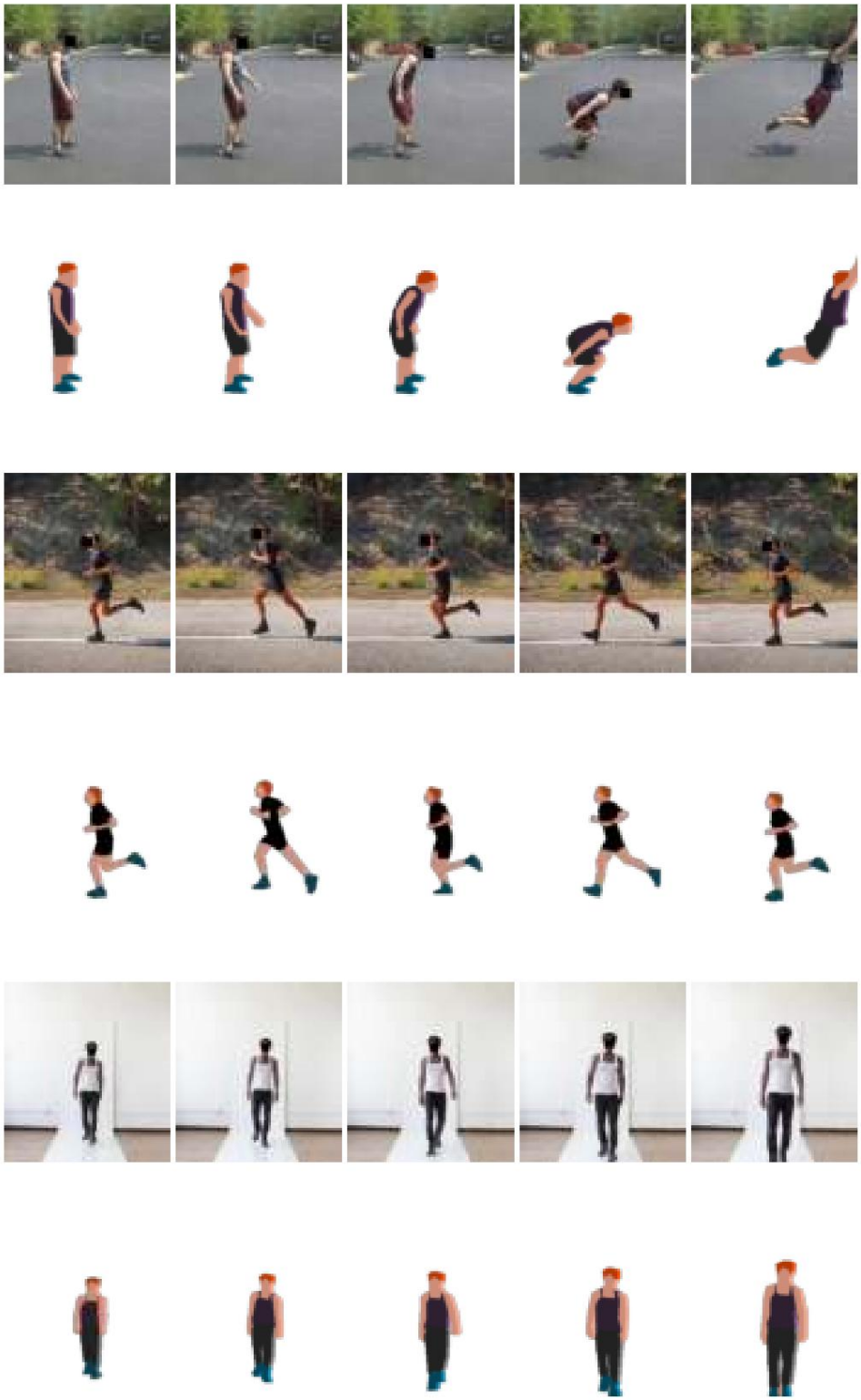


Abbildung 7. Beispielergebnisse. Ungerade Zeilen: Eingabevideos (Gesichter wurden anonymisiert). Gerade Zeilen: Methodenergebnisse.

Machine Translated by Google

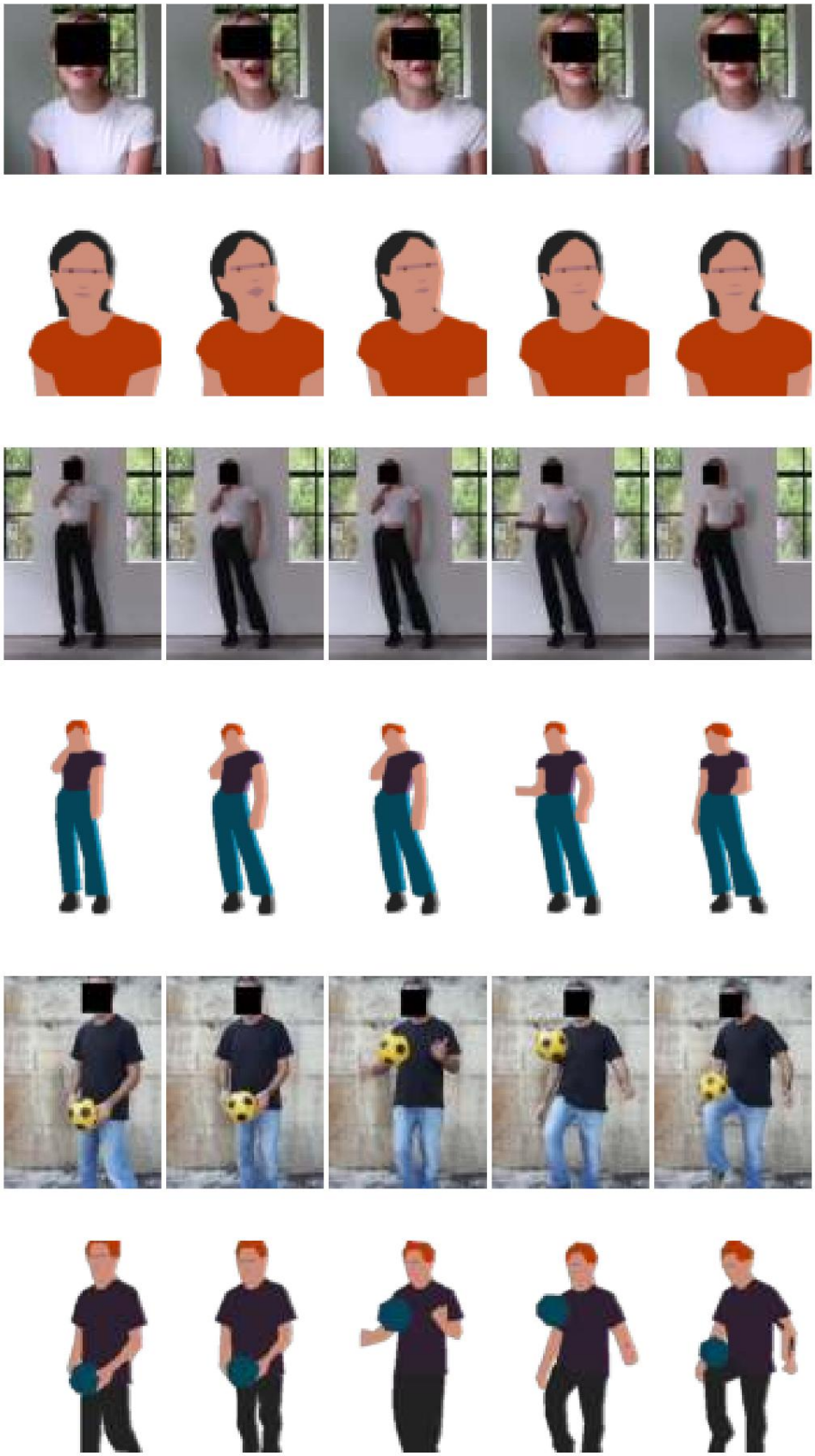


Abbildung 8. Beispielergebnisse. Ungerade Zeilen: Eingabevideos (Gesichter wurden anonymisiert). Gerade Zeilen: Methodenergebnisse.

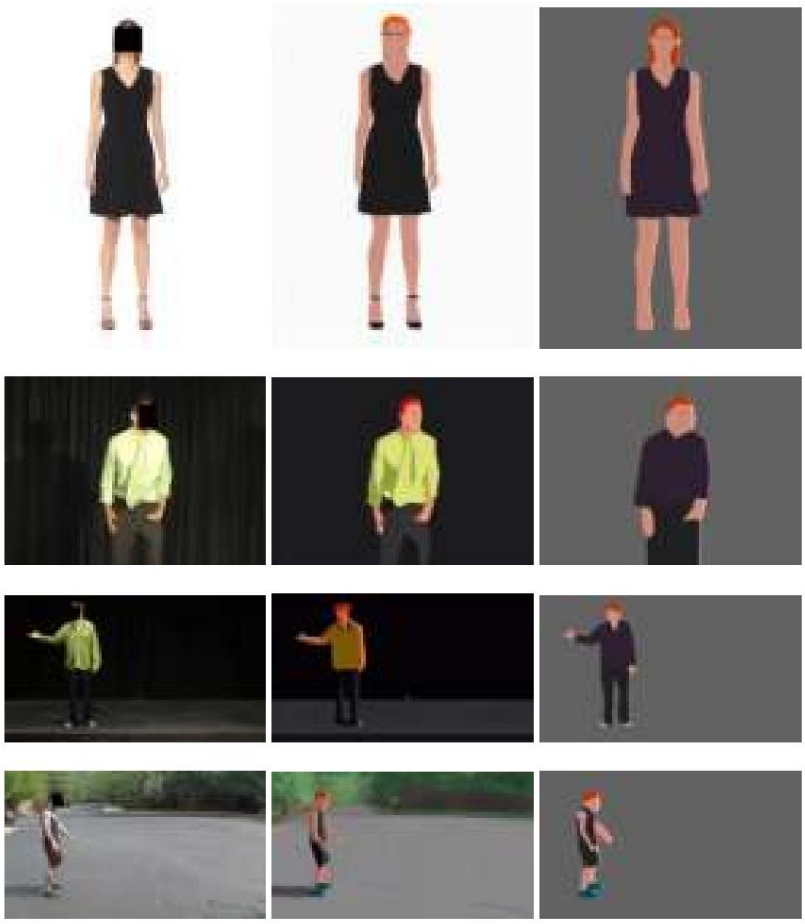
Machine Translated by Google



Abbildung 9. Beispiel eines synthetisierten Animationsbildes (rechts) und des zugehörigen Originalbildes aus einem Beispiel
Geben Sie ein Mehrpersonenvideo ein.



Abbildung 10. Beispiele für Fehlerfälle. Obere Reihe: Eingabevideos (Videos 22, 23, 25 und 8 aus dem Testdatensatz, Gesichter wurden anonymisiert); untere Reihe: Ergebnisse.



(A) **(B)** **(C)**

Abbildung 11. Vergleich mit einer generativen Pipeline: (a) Eingaberahmen; (b) Beschreibung der generativen Pipeline in [13]; (c) Lester (vorgeschlagene Methode).

Machine Translated by Google

Diskussion: Die erzielten Ergebnisse sind sehr zufriedenstellend und bestätigen, dass die neuesten Fortschritte bei Segmentierung und Tracking neue Anwendungen ermöglichen, auch im Bereich der Bildsynthese. Die vorgeschlagene Methode hat sich als effektive Methode zur Generierung rotoskopischer Animationen aus Videos erwiesen. Die festgestellten Einschränkungen stellen kein Handicap dar, da davon ausgegangen wird, dass die Methode in ausdrücklich ausgewählten oder erstellten Videos verwendet wird. Auch der Rechenaufwand ist nicht von Belang, da davon ausgegangen wird, dass die Methode bei Bearbeitungsaufgaben verwendet wird, die keine Echtzeitausführung erfordern. Die Methode ist in erster Linie für Videos mit einer einzelnen Person konzipiert , könnte aber durch Verwendung mehrerer **Parametrisierungen für Segmentierungsaufforderungen und Farbpaletten an Videos mit mehreren Personen angepas**

5. Schlussfolgerungen

In dieser Arbeit präsentieren wir Lester, eine Methode zur automatischen Synthese von 2D- animierten Figuren im Retro-Stil aus Videos. Anstatt die Herausforderung mit einem Bild-zu- Bild-generativen Modell oder mit Werkzeugen zur Einschätzung menschlicher Posen anzugehen, wird sie hauptsächlich als Problem der Objektsegmentierung und -verfolgung behandelt. Eine subjektive Qualitätsbewertung zeigt, dass der Ansatz machbar ist und dass die Methode hochwertige Animationen aus Videos mit **unterschiedlichen Posen und Erscheinungen, dynamischen Aufnahmen, Teilaufnahmen und unterschiedlichen Hintergründen generier Obwohl dies nicht das Ziel ist, ist es erwähnenswert, dass die Segmentierungs- und Tracking- Leistung der Methode in der Praxis weniger konsistent ist, insbesondere bei Videos mit niedriger Auflösung. Es wird jedoch erwartet, dass die Methode bei ausgewählten oder maßgeschneiderten Videos in halbkontrollierten Umgebungen verwendet wird.**

Aufgrund seiner Fähigkeiten stellt die Methode ein wirksames Werkzeug dar, das die Kosten für die Generierung rotoskopischer Animationen für Videospiele oder Animationsfilme erheblich senkt. Aufgrund seiner Eigenschaften ist es einfacher und weniger deterministisch als auf generativen Modellen basierende Pipelines und praktischer und flexibler als Techniken, die auf der Schätzung menschliche

Finanzierung: Diese Forschung wird teilweise vom spanischen Ministerium für Wissenschaft und Innovation im Rahmen des Vertrags PID2019-107255GB und durch das SGR-Programm 2021-SGR-00478 der katalanischen Regierung fina

Datenverfügbarkeitserklärung: Der Datensatz und die Ergebnisse der Experimente sind öffentlich verfügbar unter https://github.com/rtous/lester (abgerufen am 20. Juli 2024). Der Code ist unter https://github verfügbar. com/rtous/lester-code (abgerufen am 20. Juli 2024).

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Abkürzungen

In diesem Manuskript werden die folgenden Abkürzungen verwendet:

ACR Absolute Kategoriebewertung
Generative Adversarial Networks (GAN)
HOG-Histogramm orientierter Gradienten
MOS Mittlerer Meinungswert
SAM Segment Anything-Modell

Verweise

1. **Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, AC; Siehe, WY; et al. Segmentieren Sie alles.** arXiv **2023, arXiv:2304.02643.**
2. **Cheng, Y.; Li, L.; Xu, Y.; Li, X.; Yang, Z.; Wang, W.; Yang, Y. Segmentieren und verfolgen Sie alles.** arXiv **2023, arXiv:2305.06558.**
3. **Douglas, DH; Peucker, TK Algorithmen zur Reduzierung der Anzahl von Punkten, die zur Darstellung einer digitalisierten Linie oder ihrer Karikatur erforderlich sind. In** Classics in Cartography: Reflections on Influential Articles from Cartographica; **John Wiley & Sons: Hoboken, NJ, USA, 2011; S. 15–28. [CrossRef]**
4. **Chen, Y.; Lai, YK; Liu, YJ CartoonGAN: Generative Adversarial Networks für die Foto-Cartoonisierung. In Proceedings of the IEEE Computer Society, Salt Lake City, UT, USA, 18.–23. Juni 2018; S. 9465–9474.**
5. **Chen, J.; Liu, G.; Chen, X. AnimeGAN: Ein neuartiges leichtes GAN für Fotoanimation. In** Artificial Intelligence Algorithms and Applications, Proceedings des 11. Internationalen Symposiums, ISICA 2019, Guangzhou, China, 16.–17. November 2019; **Springer: Berlin/Heidelberg, Deutschland, 2020; S. 242–256. [CrossRef]**

Machine Translated by Google

6. **Liu, Z.; Li, L.; Jiang, H.; Jin, X.; Tu, D.; Wang, S.; Zha, Z. Unüberwachte kohärente Video-Cartoonisierung mit perzeptueller Bewegungskonsistenz . In Proceedings der 36. AAAI-Konferenz über künstliche Intelligenz (AAAI-22), virtuell, 22. Februar–1. März 2022. [CrossRef]**

7. **Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. Hochauflösende Bildsynthese mit latenten Diffusionsmodellen.** arXiv **2021, arXiv:2112.10752.**
8. **Esser, P.; Kulal, S.; Blattmann, A.; Entezari, R.; Muller, J.; Saini, H.; Levi, Y.; Lorenz, D.; Sauer, A.; Boesel, F.; et al. Skalierung gleichgerichteter Flusstransformatoren für die hochauflösende Bildsynthese.** arXiv **2024, arXiv:2403.03206.**
9. **Meng, C.; He, Y.; Song, Y.; Song, J.; Wu, J.; Zhu, JY; Ermon, S. SDEdit: Geführte Bildsynthese und -bearbeitung mit stochastischen Differentialgleichungen. In Proceedings der International Conference on Learning Representations, virtuell, 25.–29. April 2022.**
10. **Zhang, L.; Rao, A.; Agrawala, M. Hinzufügen einer bedingten Kontrolle zu Text-zu-Bild-Diffusionsmodellen. In Proceedings of the 2023 IEEE/CVF Internationale Konferenz für Computer Vision (ICCV), Paris, Frankreich, 2.–6. Oktober 2023.**
11. **Saharia, C.; Ho, J.; Chan, W.; Salimans, T.; Fleet, DJ; Norouzi, M. Bild-Superauflösung durch iterative Verfeinerung.** arXiv **2021, arXiv:2104.07636.**
12. **Saharia, C.; Chan, W.; Chang, H.; Lee, C.; Ho, J.; Salimans, T.; Fleet, D.; Norouzi, M. Palette: Bild-zu-Bild-Diffusionsmodelle. In Proceedings der ACM SIGGRAPH 2022 Conference Proceedings, New York, NY, USA, 7.–11. August 2022; SIGGRAPH '22. [Querverweis]**
13. **Yang, S.; Zhou, Y.; Liu, Z.; Loy, CC Rerender A Video: Zero-Shot Text-geführte Video-zu-Video-Übersetzung. In Proceedings of the ACM SIGGRAPH Asia Conference Proceedings, Sydney, NSW, Australien, 12.–15. Dezember 2023.**
14. **Jamriska, O. ebsynth: Fast Example-Based Image Synthesis and Style Transfer. 2018. Online verfügbar: https://github.com/jamriska/ebsynth (abgerufen am 20. Juli 2024).**
15. **Brooks, T.; Peebles, B.; Holmes, C.; DePue, W.; Guo, Y.; Jing, L.; Schnurr, D.; Taylor, J.; Luhman, T.; Luhman, E.; et al. Video Generation Models as World Simulators. 2024. Online verfügbar: https://openai.com/index/video-generation-models-as-world- simulators/ (abgerufen am 20. Juli 2024).**
16. **Chen, Z.; Li, S.; Haque, M. Ein Überblick über Sora von OpenAI und sein Potenzial für freie Spiele mit Physik-Engines und virtuelle Realität.** EAI Endorsed Trans. Robot. **2024, 3. [CrossRef]**
17. **Liang, D.; Liu, Y.; Huang, Q.; Zhu, G.; Jiang, S.; Zhang, Z.; Gao, W. Video2Cartoon: Erstellen von 3D-Cartoons aus Fußballübertragungen Video. In Proceedings of the MULTIMEDIA '05, Singapur, 11. November 2005.**
18. **Ngo, V.; Cai, J. Konvertieren von 2D-Fußballvideos in 3D-Cartoons. In Proceedings der 10. Internationalen Konferenz über Kontrolle 2008, Automation, Robotics and Vision, Hanoi, Vietnam, 17.–20. Dezember 2008; S. 103–107. [CrossRef]**
19. **Weng, CY; Curless, B.; Kemelmacher-Shlizerman, I. Photo Wake-Up: 3D-Charakteranimation aus einem einzigen Foto. In Proceedings der IEEE/CVF-Konferenz zu Computer Vision und Mustererkennung (CVPR), Long Beach, CA, USA, 15.–20. Juni 2019.**
20. **Tous, R. Pictonaut: Filmcartoonisierung mithilfe von 3D-Menschenhaltungsschätzung und GANs.** Multimed. Tools Appl. **2023,** 82, **1–15. [Querverweis]**
21. **Ruben Tous, JN; Igual, L. Vervollständigung menschlicher Posen in Teilkörperaufnahmen mit der Kamera.** J. Exp. Theor. Artif. Intell. **2023, 1–11. [CrossRef]**
22. **Fischer, J.; Jamriška, O.; Simons, D.; Shechtman, E.; Lu, J.; Asente, P.; Lukáÿc, M.; Sýkora, D. Beispielbasierte Synthese stilisierter Gesichtsanimationen.** ACM Trans. Graph. **2017,** 36, **1–11. [CrossRef]**
23. **Gatys, LA; Ecker, AS; Bethge, M. Ein neuronaler Algorithmus des künstlerischen Stils.** arXiv **2015, arXiv:1508.06576.**
24. **He, K.; Chen, X.; Xie, S.; Li, Y.; Dollar, P.; Girshick, R. Maskierte Autoencoder sind skalierbare Vision-Lerner. In Proceedings der 2022 IEEE/CVF-Konferenz über Computer Vision und Mustererkennung (CVPR), New Orleans, LA, USA, 18.–24. Juni 2022; S. 15979–15988. [CrossRef]**
25. **Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. Ein Bild sagt mehr als 16 x 16 Worte: Transformatoren für die Bilderkennung im großen Maßstab. In Proceedings der 9. Internationalen Konferenz zum Lernen von Repräsentationen, ICLR 2021, virtuell, 3.–7. Mai 2021. Online verfügbar: https://openreview.net/ (abgerufen am 20. Juli 2024).**
26. **Radford, A.; Kim, JW; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Übertragbare visuelle Modelle durch natürliche Sprachüberwachung lernen. In Proceedings der 38. Internationalen Konferenz für maschinelles Lernen, virtuell, 18.–24. Juli 2021; S. 8748–8763.**
27. **Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End-Objekterkennung mit Transformatoren. In Proceedings of the Computer Vision–ECCV 2020: 16. Europäische Konferenz, Glasgow, Großbritannien, 23.–28. August 2020; S. 213–229. [Querverweis]**
28. **Yang, Z.; Yang, Y. Entkopplungsfunktionen bei der hierarchischen Ausbreitung zur Segmentierung von Videoobjekten. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), New Orleans, LA, USA, 28. November–9. Dezember 2022.**
29. **Liu, S.; Zeng, Z.; Ren, T.; Li, F.; Zhang, H.; Yang, J.; Li, C.; Yang, J.; Su, H.; Zhu, J.; et al. Grounding dino: Dino mit geerdetem Vortraining zur Erkennung offener Objekte verbinden.** arXiv **2023, arXiv:2303.05499.**
30. **Suzuki, S.; be, K. Topologische Strukturanalyse digitalisierter Binärbilder durch Randfolgen.** Comput. Vision Graph. Bild Prozess. **1985,** 30, **32–46. [CrossRef]**
31. **Kazemi, V.; Sullivan, J. Gesichtsausrichtung in einer Millisekunde mit einem Ensemble von Regressionsbäumen. In Proceedings der 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23.–28. Juni 2014; S. 1867–1874. [CrossRef]**

Machine Translated by Google

32. Dalal, N.; Triggs, B. Histogramme orientierter Gradienten zur Erkennung von Personen. In Proceedings der 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20.–26. Juni 2005; Band 1, Seiten 886–893. [Querverweis]
33. Soomro, K.; Zamir, AR; Shah, M. UCF101: Ein Datensatz von 101 menschlichen Handlungsklassen aus Videos in freier Wildbahn. arXiv 2012, arXiv:1212.0402.
34. Zablotskaia, P.; Siarohin, A.; Zhao, B.; Sigal, L. DwNet: Dichtes Warp-basiertes Netzwerk für posengesteuerte menschliche Videogenerierung. In Proceedings der 30. British Machine Vision Conference 2019, BMVC 2019, Cardiff, Großbritannien, 9.–12. September 2019; S. 51.
35. Kay, W.; Carreira, J.; Simonyan, K.; Zhang, B.; Hillier, C.; Vijayanarasimhan, S.; Viola, F.; Green, T.; Back, T.; Natsev, P.; et al. Die Videodatensatz zur Kinetik menschlicher Aktionen. arXiv 2017, arXiv:1705.06950.
36. Tous, R. Lester Projekt-Homepage. Online verfügbar: https://github.com/rtous/lester (abgerufen am 20. Juli 2024).
37. ITU. Subjektive Methoden zur Beurteilung der Videoqualität für Multimedia-Anwendungen. ITU-T-Empfehlung P.910. 2008. Verfügbar online: https://www.itu.int/rec/T-REC-P.910 (abgerufen am 20. Juli 2024).
38. ITU. Mean Opinion Score (MOS)-Terminologie. ITU-T-Empfehlung P.800.1 Methoden zur objektiven und subjektiven Beurteilung der Sprach- und Videoqualität. 2016. Online verfügbar: https://www.itu.int/rec/T-REC-P.800.1 (abgerufen am 20. Juli 2024).
39. Ruiz, N.; Li, Y.; Jampani, V.; Pritch, Y.; Rubinstein, M.; Aberman, K. DreamBooth: Feinabstimmung von Text-zu-Bild-Diffusionsmodellen für die themengesteuerte Generierung. arXiv 2023, arXiv:2208.12242.

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Artikel

Eine Schwarmintelligenzlösung für den profitablen Einsatz mehrerer Fahrzeuge Abhol- und Lieferproblem

Abeer I. Alhujaylan [1,*] **und Manar I. Hosny** [2]

[1] Abteilung für Informationstechnologie, College of Computer, Qassim University, Buraydah 51452,
[2] Saudi-Arabien Abteilung für Informatik, College of Computer and Information Sciences, King Saud University, Riyadh 11362, Saudi-Arabien; mifawzi@ksu.edu.sa
* Korrespondenz: a.alhujaylan@qu.edu.sa

Zusammenfassung: Liefer-Apps verzeichnen ein starkes Wachstum und erfordern effiziente Algorithmen, um den Transport zu koordinieren und Gewinne zu erzielen. Ein Problem, das die Ziele von Liefer-Apps berücksichtigt, ist das profitable Abhol- und Lieferproblem mit mehreren Fahrzeugen (MVPPDP). In diesem Artikel schlagen wir acht neue Metaheuristiken vor, um die anfänglichen Lösungen für das MVPPDP basierend auf dem bekannten Schwarmintelligenzalgorithmus Artificial Bee Colony (ABC) zu verbessern: K-means-GRASP-ABC(C)S1, K-means- GRASP-ABC(C)S2, Modified K-means-GRASP-ABC(C)S1, Modified K-means-GRASP-ABC(C)S2, ACO-GRASP-ABC(C)S1, ACO-GRASP-ABC(C)S2, ABC(S1) und ABC(S2). Alle Methoden erzielten in den meisten Fällen eine bessere Leistung in Bezug auf die Verarbeitungszeit. So betrugen beispielsweise die Durchschnittszeiten der Algorithmen für 250 Kunden 75,9, 72,86, 79,17, 73,85, 76,60, 66,29, 177,07 und 196,09 und waren damit schneller als die modernsten Methoden, die 300 s brauchten. Darüber hinaus zeigten alle vorgeschlagenen Algorithmen bei kleinen Instanzen eine gute Leistung im Hinblick auf den Gewinn, indem sie dreizehn neue beste Lösungen und fünf gleichwertige Lösungen zu den besten bekannten Lösungen erzielten. Bei mittelgroßen und großen Instanzen hinken die Algorithmen jedoch leicht hinterher, was auf die gierige Randomisierungsstrategie und GRASP zurückzuführen ist, die in der Scout-Bee-Phase verwendet wurden. Darüber hinaus priorisieren unsere Algorithmen minimale Lösungen und Iterationen für schnelle Verarbeitungszeiten in alltäglichen M-Commerce-Apps,

Schlüsselwörter: profitables Abhol- und Lieferproblem mit mehreren Fahrzeugen; künstliche Bienenkolonie; Schwarmintelligenz ; metaheuristische Algorithmen

check for updates

Zitat: Alhujaylan, AI; Hosny, MI Eine Schwarmintelligenzlösung für das profitable Abhol- und Lieferproblem mehrerer Fahrzeuge. Algorithmen **2024,** 17, **331. https://doi.org/10.3390/a17080331**

Wissenschaftlicher Herausgeber: Vittorio Maniezzo

Erhalten: 2. Juli 2024
Überarbeitet: 17. Juli 2024
Akzeptiert: 25. Juli 2024
Veröffentlicht: 31. Juli 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.
Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/4.0/).

1. Einleitung

In den letzten fünf Jahren haben Liefer-Apps ein beträchtliches Wachstum erlebt, insbesondere Apps zur Lieferung von Essen. Hier hat COVID-19 der Branche einen Schub um einige Jahre beschert, da Millionen von Menschen während der Ausgangssperre online Essen und andere wichtige Dienstleistungen bestellte Die Lebensmittellieferplattform Instacart meldete, dass ihre Ziele für 2022 bereits in der dritten Lockdown-Woche erreicht wurden. Im Jahr 2022 entwickelte sich China zum größten Markt für Lebensmittellieferungen mit einem atemberaubenden Marktwert von 42,5 Milliarden USD. Bis 2029 wird erwartet, dass die Marktgröße der gesamten Branche der Lebensmittelliefer-Apps 165 Milliarde Darüber hinaus bestellen 85 % der Amerikaner ihre Speisen online über Apps. Darüber hinaus bieten 80 % der Restaurants in den USA ihren Kunden die Möglichkeit, bequem per Handy zu bestellen [2]. In Saudi-Arabien bieten Apps für die Essenslieferung ebenfalls sofortige Dienste und ein optimiertes Benutzererlebnis, damit die Menschen ihr Essen genießen können. Infolgedessen entstehen regelmäßig neue Anwendungen, die darum wetteifern, über 35 Millionen Menschen Dienste, Bequemlichkeit und Komfort zu bieten. Beispiele für diese Anwendungen, die in Saudi-Arabien sehr beliebt geworden sind, sind Mrsool, Uber Eats, Talabat, Hunger Station, Ninja, Jahez, To You, The Chifz, Shgardi und NANA [3]. Die steigende Nachfrage nach Anwendungen zur Essenslieferung hat gleichzeitig den Druck auf die L

Machine Translated by Google

Schnelle Algorithmen zur Koordination des Warentransports, um Gewinne zu erzielen und die Zufriedenheit der Kunden zu gewährleisten. Eines der in der Literatur vorgestellten Probleme, das das Ziel von Apps zur Essenslieferung nachahmt, nämlich Gewinne zu steigern, die Zufriedenheit der Benutzer zu erreichen und die Kosten zu senken, ist das Problem der profitablen Abholung und Lieferung

Gansterer et al. [4] führten das MVPPDP ein, einen Untertyp des selektiven Abhol- und Lieferproblems (SPDP) und eine neuartige Variante des bekannten Abhol- und Lieferproblems (PDP). Innerhalb der Beschränkungen eines eintägigen Planungshorizonts versucht das MVPPDP, die rentabelsten Routen zu ermitteln, indem die Reisekosten minimiert und die Einnahmen maximiert werden. Das MVPPDP ist ein statisches Problem, das durch einen standardisierten Fuhrpark , ein zentrales Depot, eine Gruppe von Kunden und eine vorgegebene Menge an Anfragen und Produkten gekennzeichnet ist. Ein Kundenpaar, bestehend aus dem Abholkunden und dem Lieferkunden , definiert jede Anfrage. Auf diese Weise werden die Produkte von einem Abholkunden zu einem Lieferkunden befördert, um einen Einnahmestrom für den Betrieb zu generieren.
Zu den realen Anwendungen von MVPPDP gehören mobile Anwendungen für Essenslieferdienste und Lieferunternehmen.

Das MVPPDP ist eine der Anwendungen zur kombinatorischen Optimierung (CO), die bei Forschern in der Informatik und im Operations Research besondere Aufmerksamkeit erregt.
Die Lösung eines CO-Problems erfordert hauptsächlich die Suche nach der optimalen oder bestmöglichen Lösung aus einer begrenzten Anzahl möglicher Lösungsvorschläge. Zur Lösung von CO-Problemen können zahlreiche Algorithmen verwendet werden, die in zwei Kategorien eingeteilt werden können: exakte Algorithmen, die garantiert optimale Lösungen in angemessener Zeit liefern , und Näherungsalgorithmen, die normalerweise als Heuristiken oder Metaheuristiken bezeichnet werden und zwar nicht garantiert eine optimale Lösung liefern, aber in angemessener Zeit eine gute Lösung bieten.
Die Wörter Heuristik und Meta stammen aus dem Griechischen, wobei Heuristik von „finden“ oder „heuriskein“ abgeleitet ist , was bedeutet , neue Methoden und **Strategien zu entdecken**, die darüber hinausgehen, auf eine höhere Ebene. **Probleme lösen, während das Präfix „meta“** „x“ bedeutet . **Metaheuristik bedeutet „entdecken“.**
lassen sich anhand der Anzahl der gleichzeitig verwendeten Lösungen in zwei Kategorien einteilen: auf einer einzigen Lösung basierende Metaheuristiken (S-Metaheuristiken) und populationsbasierte Metaheuristiken (P-Metaheuristiken). Bei der S-Metaheuristik liegt der Schwerpunkt auf der Modifizierung und Verbesserung einer einzigen Kandidatenlösung. Bei der P-Metaheuristik hingegen werden mehrere Kandidatenlösungen statt nur einer einzigen Lösung erhalten und verbessert. Sie basiert auf der iterativen Verbesserung von Lösungen in einer Population und der Weiterentwicklung der gesamten Population, bis ein Satz guter Lösungen erreicht ist. Dank ihrer explorationsorientierten Eigenschaften sind die P-Metaheuristiken besser darin, den Suchraum zu diversifizieren als die S-Metaheuristik. Die bekanntesten Beispiele für P-Metaheuristiken sind von natürlichem Verhalten inspirierte Algorithmen, wie genetische Algorithmen und Schwarmintelligenz (SI) [5].

Im Allgemeinen gibt es zwei Phasen, die angewendet werden sollten, um eine gute Problemlösung zu erhalten : Lösungskonstruktion und Lösungsverbesserung. Bei der Lösungskonstruktion werden eine oder mehrere anfängliche mögliche Lösungen erstellt, die als Ausgangspunkt für die Erkundung des Suchraums betrachtet werden können. Die Lösungsverbesserung basiert auf der Modifizierung der Ausgangslösungen , um sie gemäß vordefinierter Strategien zu verbessern, bis eine qualitativ hochwertige Lösung erreicht ist oder Abbruchkriterien erfüllt sind (z. B. eine maximale Anzahl von Iterationen oder eine bestimmte Zeitspanne

Da es sich beim MVPPDP um ein NP-schweres Problem handelt (da es sich um einen Untertyp des SPDP handelt, das wiederum ein Sonderfall des PDP ist) [4], können metaheuristische Algorithmen eingesetzt werden, um optimale Lösungen in angemessener Zeit effizient zu finden, insbesondere für Probleme mittlerer bis großer Größe. In diesem Artikel stellen wir neue Algorithmen zur Lösung dieses Problems vor, mit dem Ziel, die in **[6,7]** **vorgeschlagene ursprüngliche Lösung zu verbessern . Unser Ansatz basiert auf der Verwendung des Artificial Bee Colony (ABC)-Algorithmus, einem der weithin anerkannten Algorithmen für künstliche Intelligenz, die als SI verschiedene Strategien implementiert, von denen jede die Wirksamkeit bei der Beschleunigung des Suchvorgangs und der Verkürzung der Verarbeitungszeit bewies.** bekannt sind. Mit ABC wurden

Die Hauptziele dieses Dokuments sind (1) die Unterstützung von Essenslieferdiensten und Lieferunternehmen bei der Optimierung ihrer Dienstleistungen durch beschleunigte Kundenauswahl und

Machine Translated by Google

Routenplanung, um sowohl den Gewinn als auch die Kundenzufriedenheit zu maximieren; (2) effiziente Transportplanung, die dazu beitragen kann, Energieressourcen zu schonen, indem die umweltschädlichen Auswirkungen von Gasemissionen verringert werden; und dass (3) die vorgeschlagenen Ansätze in einigen getesteten Fällen Ergebnisse erzielen, die die in der **Literatur für das MVPPDP dokumentierten hochmodernen Methoden übertreffen .**

Der Rest dieses Dokuments ist wie folgt gegliedert: Abschnitt 2 enthält einen Überblick über einige verwandte Arbeiten. Abschnitt 3 führt in die formale Problemdefinition ein. Die vorgeschlagene Methode wird in Abschnitt 4 ausführlich beschrieben . Abschnitt 5 veranschaulicht die experimentellen Ergebnisse. Abschließend werden in Abschnitt 6 Schlussfolgerungen und zukünftige Arbeiten vorgestellt.

2. Verwandte Studien

Das MVPPDP wird als eine Art SPDP klassifiziert und hat zum Ziel, den Gewinn zu maximieren und gleichzeitig die Reisekosten zu minimieren, indem nur profitable Kundenpaare besucht werden. Das Erreichen einer günstigen Route, die die Differenz zwischen dem Gesamtumsatz und den gesamten Reisekosten maximiert, ist das Ziel sowohl des MVPPDP als auch des profitablen Tourproblems (PTP). Der Hauptunterschied besteht darin, dass die Fahrzeugroute beim PTP keinen Beschränkungen unterliegt, einschließlich, aber nicht beschränkt auf die maximale Fahrtzeit, Fahrzeugkapazitätsgrenzen oder Vorrangbeschränkungen [8]. In der wissenschaftlichen **Literatur gibt es nur wenig Forschung zum PTP [9]. Daher beginnen wir mit der Präsentation von Forschungsergebnissen zum MVPPDP.**

Gansterer et al. [4] führten die ersten vier Algorithmen ein. Die Metaheuristik der variablen Nachbarschaftssuche (VNS) wurde in [10] implementiert. Während der Konstruktionsphase der VNS wurden **die gierige Konstruktionsheuristik (C1) und die zweistufige billigste Einfügungsheuristik (C2) verwendet . Während der Verbesserungsphase wurden die nachfolgenden Nachbarschaftsoperatoren implementiert: der paarweise Rückwärtsaustausch, der paarweise Vorwärtsaustausch, die Vorwärtseinfügung, das Verschieben von Paaren, die Rückwärtseinfügung, der Inter-Tour-Austausch, die Inter-Tour-Einfügung, die Inter-Dummy-Einfügung, der Inter-Dummy-Austausch und der Schwerkraftzentrumsaustausch. Die Autoren implementierten zwei Ansätze zur Anwendung der Nachbarschaftsoperatoren: eine sequentielle (Seq) Methode, die eine vorgegebene Sequenz von Nachbarschaften verwendet, und eine selbstadaptive (Sea) Methode, die sich basierend auf dem Suchstatus anpasst, um die Sequenz der Nachbarschaften zu bestimmen. Um die vorgeschlagene Methode mit einem alternativen Algorithmus zu vergleichen, wurde so Eine Evaluierung des vorgeschlagenen Algorithmus wurde an 36 randomisierten Instanzen unterschiedlicher Größe durchgeführt. In Bezug auf die Lösungsqualität zeigten die experimentellen Ergebnisse, dass beide Varianten des VNS das GLS übertrafen.**

Haddad [11] führte einen neuen Algorithmus zur Lösung des MVPPDP ein. In Erwartung einer Lösung für das profitable Abhol- und Lieferproblem schlugen sie einen Algorithmus namens IPPD vor, der aus einer Kombination aus iterierter lokaler Suche und zufällig variablem Nachbarschaftsabstieg abgeleitet wurde. Während der Suche war der IPPD-Algorithmus nicht darauf beschränkt, nur praktikable Lösungen zu akzeptieren. Für die Konstruktion der anfänglichen Lösung wurde die **von Gansterer et al. [4]** verwendete **C1-**Heuristik **implementiert. Um die Lösung zu verbessern, wurden mehrere Nachbarschaftsbewegungen implementiert: Paartausch oder Paarverschiebung, Blocktausch oder Blockverschiebung, Abhol- oder Lieferverschiebung, Interblocktausch oder Interblockverschiebung, Schwerpunktaustausch, Einfügeoperator und Entfernungsoperator sowie Interpaartausch oder Interpaarverschiebung. Anhand der von Gansterer et al. [4] vorgeschlagenen Benchmark-Instanzen wurde der vorgeschlagene Algorithmus evaluiert. Er demonstrierte seine Wirksamkeit bei der Handhabung kleiner und mittlerer Fälle, indem er für sechs Fälle neue optimale Lösungen identifizierte. Bei großen Instanzen war die Leistung des vorgeschlagenen Algorithmus jedoch unzureichend.**

In Alhujaylan und Hosny [6] wurde die Konstruktionsphase des weithin anerkannten Greedy Randomized Adaptive Search-Verfahrens (GRASP) verwendet, um vorläufige Lösungen für das MVPPDP zu entwickeln. Sie verwendeten denselben Datensatz, der von [4] erstellt wurde. Sie führten einen Leistungsvergleich zwischen der Methode und zwei Konstruktionsheuristiken durch, die zuvor in der Literatur zur Konstruktion erster Lösungen für das MVPPDP dokumentiert worden waren . Die experimentellen Ergebnisse zeigten, dass die vor

Machine Translated by Google

alternative Konstruktionsheuristiken, insbesondere bei Instanzen mit geringer Größe, bei denen acht neue beste Anfangslösungen für das Problem gefunden wurden.

In [7] stellten Alhujaylan und Hosny neue Heuristiken zur Konstruktion von Anfangslösungen für das MVPPDP vor. Die Heuristiken basieren auf der Clusterung des Suchraums des MVPPDP unter Verwendung von drei Clustermethoden: K-Means, adaptive K-Means und Ant Colony Optimization (ACO). Anschließend wurde eine modifizierte Version von GRASP verwendet, um die anfängliche MVPPDP-Lösung basierend auf den Ergebnissen jedes Clusteralgorithmus zu konstruieren. Die vorgeschlagene Lösung wurde auf den Datensatz angewendet, der von den Forschern in [4] vorgestellt wurde . Sie verglichen ihre Ergebnisse mit denen der anderen Konstruktionsheuristiken, die zuvor für das MVPPDP verwendet wurden. Die experimentellen Ergebnisse bewiesen die Wirksamkeit der vorgeschlagenen Algorithmen sowohl in Bezug auf die Lösungsqualität als auch auf die Verarbeitungszeit.

In [12] integrierten die Autoren Zeitfenster mit unsicheren Reisezeiten in das MVPPDP. Die Forscher nutzten eine Teilmenge der in [4] erstellten Datenbank, die zwanzig und fünfzig Anfragen umfasste, was zwei bzw. drei Fahrzeugen entsprach.

Das Ziel des Kunden ist es, eine Lösung zu finden, die den Nettogewinn optimiert, der sich aus der Differenz zwischen den erzielten Einnahmen, den Kosten der Route und den Kosten aufgrund von Zeitfensterverletzungen ergibt. Sie haben breite Zeitfenster und Standardzeitfenster berücksichtigt. Unter der Annahme, dass Reisezeiten unabhängige gammaverteilte Zufallsvariablen sind und mit einer risikoneutralen Perspektive haben sie gezeigt, dass die erwarteten Kosten für Pünktlichkeit und frühe Ankunft in geschlossener Form berechnet werden können . Die Autoren schlugen vor, einen Graphensuchansatz als Lösungsmethode zu verwenden . Experimente haben gezeigt, dass in **angemessener Zeit präzise Lösungen erreicht werden können .**

3. Problemdefinition

Das MVPPDP ist ein vordefiniertes statisches Problem mit mehreren Bestandteilen. Es wird angenommen, dass ein zentrales Depot für den Empfang von Kundenanfragen zuständig ist. Diese Anfragen bestehen aus Abhol- und Lieferkunden in Paaren. Um diese Anforderungen zu erfüllen, betreiben wir eine Flotte identischer Fahrzeuge, die eine Vielzahl identischer Produkte von einer Teilmenge von Abholkunden zu ihren jeweiligen Lieferkunden transportieren und dabei einen bestimmten Gewinn erzielen. Die folgenden Einschränkungen

des MVPPDP müssen berücksichtigt werden: • Die Paarungsbeschränkung legt fest, dass jedem Kunden ei

Anfrage (Abholung und Lieferung).

• Vorrangsbeschränkung: Der Abholkunde muss Vorrang vor dem Lieferkunden haben. • Die

Zeitbeschränkung für Fahrten: Jedes Fahrzeug unterliegt einer täglichen Fahrzeitbesch

die bei der Betreuung der Kunden nicht überschritten werden dürfen.

Eine Einschränkung ist die Kapazität jedes Fahrzeugs, die während der Fahrt nicht überschritten werden darf.

Abholung der Ware bei Abholkunden. • Jedes

Fahrzeug sollte leer vom Depot abfahren und dort ankommen. • Pro Kunde ist nur ein Besuch erlaubt.

Der MVPPDP kann formal wie folgt definiert werden [10]: Sei G = (V, A) **ein Graph, in dem** V = **{0, . . . ,** 2n + **1} die Knotenmenge definiert, wobei der Knoten (0,** 2n + 1) einem zentralen Depot **entspricht , während die restlichen Knoten die Abholkunden** P = **{1, . . . , n} und die Lieferkunden** D = {n + **1, . . . , 2n} darstellen.** A = {(i, j) : **i,** j **ÿ V,** i ÿ= **j} definiert die Bogenmenge, wobei jedem Bogen nicht-negative Routenkosten** cij zugeordnet sind. **Jeder Lieferknoten** i **hat einen Umsatz** ri, **der bei seinem Besuch erzielt werden kann. Ebenso hat jeder Knoten** i **ein Angebot** qi **(Abholung,** qi > 0) **oder eine Nachfrage (Lieferung,** qn+i = ÿqi). **Zu Beginn gibt es im Depot weder Angebot noch Nachfrage (q0** = 0). **Außerdem hat jeder Knoten** i **die Betriebsdauer** di . **Es gibt eine Menge von Fahrzeugen** K = **{1, . . . , m}, und jedes Fahrzeug hat eine Ladekapazität** C **und eine maximale Fahrtzeit T.**

Machine Translated by Google

Die Zielfunktion (OF) des MVPPDP zielt darauf ab, den Gesamtgewinn durch Subtraktion der gesamten Reisekosten von den Gesamteinnahmen zu maximieren. Sie kann wie folgt formuliert werden:

$$\text{Maximieren} \ddot{y} \ddot{y} \ddot{y} \quad ri \ddot{y} cij xijk, \tag{1}$$

Das mathematische Modell des MVPPDP kann wie in [4,6] wie folgt formuliert werden:

$$\ddot{y} \ddot{y} \quad xijk \ddot{y} 1 \; j \ddot{y} V \tag{2}$$

$$\ddot{y} \ddot{y} \quad xijk \ddot{y} 1 \; i \ddot{y} V \tag{3}$$

$$xi0k = 0 \; i \ddot{y} V, \; k \ddot{y} K \tag{4}$$

$$x2n+1,jk = 0 \; j \ddot{y} V, \; k \ddot{y} K \tag{5}$$

$$\ddot{y} \quad xijk \ddot{y} xjik = 0 \; j \ddot{y} V \; \{0, 2n + 1\}, \; k \ddot{y} K \tag{6}$$

$$\ddot{y} \quad xijk \ddot{y} xn+i,jk = 0 \; i \ddot{y} P, \; k \ddot{y} K \tag{7}$$

$$\ddot{y} \quad x0jk = \ddot{y} \; xi,2n+1,k = 1 \; k \ddot{y} K \tag{8}$$

$$xijk = 1 \ddot{y} Qjk = Qik + qj \; i \ddot{y} V, \; j \ddot{y} V \; \{0\}, \; k \ddot{y} K \tag{9}$$

$$Qik \ddot{y} C \; i \ddot{y} V, \; k \ddot{y} K \tag{10}$$

$$Q0k = 0 \; k \ddot{y} K \tag{11}$$

$$Bik \ddot{y} Bn+i,k \quad i \ddot{y} P. \tag{12}$$

$$Bjk \ddot{y} Xijk \; Bik + di + tij \quad i \ddot{y} V, \; j \ddot{y} V, \; k \ddot{y} K \tag{13}$$

$$B2n+1,k \ddot{y} T \; k \ddot{y} K \tag{14}$$

$$Bok = 0 \; k \ddot{y} K \tag{15}$$

$$xijk \ddot{y} 0, 1 \; i \ddot{y} V, \; j \ddot{y} V, \; k \ddot{y} K \tag{16}$$

$$Qik, \; Bik \ddot{y} 0 \; i \ddot{y} V, \; k \ddot{y} K \tag{17}$$

Die Beschränkungen (2) und (3) erfordern, dass jeder Knoten höchstens einmal angefahren wird. Die Beschränkungen (4) und (5) verhindern, dass ein Fahrzeug das Ursprungsdepot (Startpunkt) betritt oder das Zieldepot (Endpunkt) verlässt. Beschränkung (6) weist auf Flusserhaltung hin. Beschränkung (7) besagt, dass jedes Kundenpaar (Abholung und Lieferung) vom gleic Einschränkung (8) erfordert, dass alle Fahrzeuge zum Depot starten und dorthin zurückkehren. Einschränkung (9) besagt, dass ein Überschreiten der Fahrzeugkapazität nicht zulässig ist. Einschränkung (10) gibt an, dass die Ladung eines Fahrzeugs durch C begrenzt ist. Einschränkung (11) gibt an, dass alle Fahrzeuge mit einer leeren Ladung starten . Einschränkung (12) gibt an, dass der Lieferkunde nicht vor dem Abholkunden besucht werden kann. Einschränkung (13) gibt an, dass der früheste Zeitpunkt für den Beginn des Dienstes an einem Knoten j bestimmt **wird , indem der Beginn der Dienstzeit am Knoten** i **zur Dienstzeit bei und der Fahrzeit zwischen** i **und j addiert wird. Einschränkung (14) gibt an, dass die maximale Zeit für jedes Fahrzeug durch** T begrenzt ist. **Einschränkung (15) gibt an, dass jedes Fahrzeug zum Zeitpunkt Null am Depot startet . Schließlich werden Er**

4. Vorgeschlagene

Methode In der vorliegenden Studie verwendeten wir den von Karaboga **[13] erstmals vorgeschlagenen Algorithmus „Artificial Bee Colony“ (ABC) . Er ahmt das kollaborative und kollektive Sammelverhalten eines Bienenvolks nach. Im ABC-Algorithmus gibt es drei verschiedene Biene**

Machine Translated by Google

Kundschafterbienen, angestellte Bienen und Zuschauerbienen. Jeder Typ hat eine andere Rolle bei der Erkundung und Nutzung von Nahrungsquellen. Eine Nahrungsquelle stellt die Lösung für ein Problem dar, und Nahrungsquellen mit mehr Nektar gehören zu den besseren Lösungen. Der ABC-Algorithmus funktioniert folgendermaßen: Zunächst werden die Kundschafterbienen nach dem Zufallsprinzip dem Suchraum zugewiesen, um erste Nahrungsquellen zu finden. Dann werden die angestellten Bienen ausgesandt, um die von den Kundschafterbienen entdeckten Orte zu nutzen, wobei jeder angestellten Biene eine Nahrungsquelle zugewiesen wird. Danach beginnen die angestellten Bienen eine Nachbarschaftssuche, um eine bessere Nahrungsquelle in der Nähe zu finden. Nachdem sie eine bessere Nahrungsquelle gefunden haben, verlassen die angestellten Bienen alle vorherigen Nahrungsquellen und nutzen nur die besseren. Diese angestellten Bienen kehren dann zum Stock zurück und teilen ihre Informationen üb Die Entscheidung, ob die Zuschauerbienen den angestellten Bienen folgen, hängt von der Fülle der Nahrungsquellen ab. Wenn die Zuschauerbienen sich entscheiden zu folgen, werden sie zu angestellten Bienen und wiederholen den Vorgang der angestellten Bienen. Nach mehreren Iterationen kann die Nahrungsquelle erschöpft sein. An diesem Punkt werden die angestellten Bienen zu Kundschafterbienen und beginnen wahllos nach einer neuen Nahrungsquelle zu suchen, um die vorherige zu ersetzen. Das Flussdiagramm des ABC-Algorithmus ist in Abbildung 1 dargestellt **[14].**

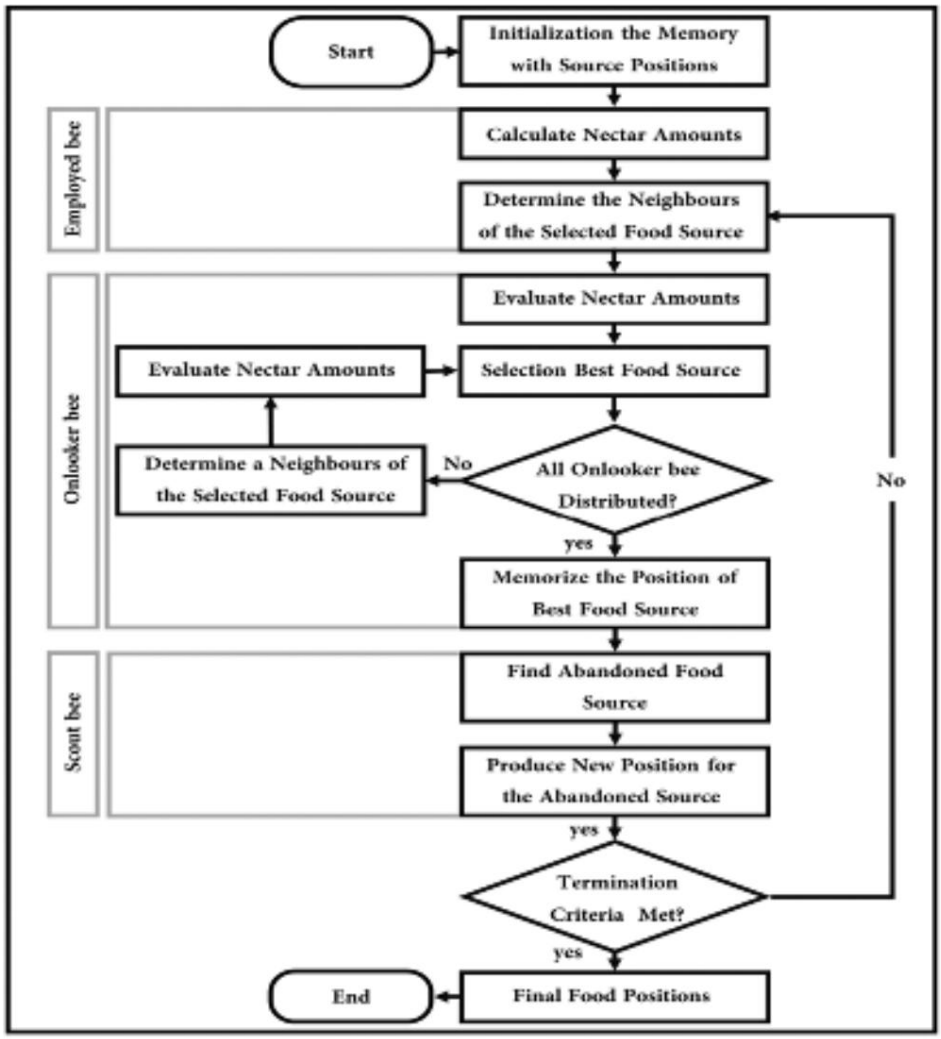


Abbildung 1. Flussdiagramm des Artificial Bee Colony-Algorithmus [14].

Die ABC-Heuristik wurde für die Lösung verschiedener VRP-Varianten verwendet, beispielsweise von Szetoa et al. [15] zur Lösung des Problems der kapazitätsbeschränkten Fahrzeugroutenplanung, von Wu et al. [16] zur Lösung des VRP mit Zeitfenstern und von Yao et al. [17] zur Lösung des periodischen VR

tt

Machine Translated by Google

In diesem Artikel möchten wir die in [7] vorgeschlagenen ersten Lösungen verbessern . Es gibt viele Metaheuristiken, die verwendet wurden, um die Qualität von Lösungen zu verbessern, wie z. B. lokale Suche (LS), simuliertes Abkühlen (SA) und VNS. In unserer Studie verwenden wir auch eine Metaheuristik, die gierige und randomisierte Suchen kombiniert und auf dem Verhalten von Honigbienen basiert, nämlich ABC. Der Hauptgrund für die Auswahl von ABC anstelle anderer Honigbienenalgorithmen ist seine Popularität, wobei die Anzahl der Studien, die ABC im Vergleich zu anderen Varianten verwendet haben, 54 % erreicht, wie in Karabogas Umfrage [18] berichtet wird. Zur Bewältigung des MVPPDP werden zwei Versionen von ABC vorgeschlagen: ABC ohne Clustering (ABC) und ABC mit Clustering (ABC(C)), wie im Folgend

4.1. ABC ohne Clustering (ABC)

In diesem Teil der Forschung wollten wir die anfänglichen Lösungen verbessern, die mit dem verbesserten GRASP in [7] erreicht wurden. Die ABC-Metaheuristik besteht aus drei Phasen: der Phase der beschäftigten Biene, der Phase der Zuschauerbienen und der Phase der Kundschafterbienen. Das Hauptziel der ersten beiden Phasen besteht darin, die Intensivierung des Suchraums zu erhöhen, um gute Lösungen zu erhalten, während die letzte Phase darauf abzielt, die Diversifizierung zu erhöhen und zu verhindern, dass man in lokalen Optima stecken bleibt. Die Details unseres vorgeschlagenen Algorithmu

1. Phase der beschäftigten Bienen:

Nachdem den beschäftigten Bienen die anfänglichen Lösungen zugewiesen wurden, starten sie eine Nachbarschaftssuche, um bessere Lösungen in der Nähe zu finden, indem sie zufällig einen aus einer Reihe von Nachbarschaftsoperatoren auswählen . Die Nachbarschaftsoperatoren zielen darauf ab, eine kleine Änderung an der Lösung vorzunehmen, um sie zu verbessern. Daher wird für jede Lösung zufällig eine Route ausgewählt, um einen ebenfalls zufällig ausgewählten Operator anzuwenden. Die Reihe der Nachbarschaftsoperatoren wird unten beschrieben: • Nur einen Abholkunden verlegen: Wählen Sie zufällig einen Abholkunden aus und verlegen Sie seine Position, sodass die neue Abholposition nicht nach seinem Lieferkunden liegen darf .

• Nur einen Lieferkunden verlegen: Wählen Sie einen Lieferkunden nach dem Zufallsprinzip und verlegen Sie seine Position, sodass die neue Lieferposition nicht vor dem Abholkunden liegen darf. • Verlegen Sie ein Kundenpaar: Wählen Sie ein Kundenpaar nach dem Zufallsprinzip und verlegen Sie die Positionen unter Beibehaltung der Vorrangbeschränkung (d. h. Abholung vor Lieferung). • Vertauschen Sie zwei Kundenpaare: Wählen Sie zwei Kundenpaare nach dem Zufallsprinzip und tauschen Sie ihre Positionen. • Two-Opt-Operator: Wählen Sie zwei aufeinanderfolgende Kundenpaare (ein aufeinanderfolgendes Paar ist ein Paar, bei dem auf einen Abholkunden direkt sein Lieferkunde folgt) und ordnen Sie die Kanten zwischen den Paaren unter Berücksichtigung der Vorrangbeschränkung neu an. • Entfernen Sie ein Kundenpaar nach dem Zufallsprinzip und fügen Sie ein unbesuchtes ein: Entfernen Sie ein Kundenpaar nach dem Zufallsprinzip, wählen Sie ein unbesuchtes Kundenpaar mit einer hohen Einfügur

$$\text{IRKundenpaar} = \frac{\text{Kundenpaar}}{\text{c(Depot,PC)}}, \tag{18}$$

wobei r **der Umsatz des Kundenpaares und** c(Depot,PC) **die Distanz zwischen dem Depot und dem Abholkunden dieses Paars ist. • Entfernen**

Sie ein Kundenpaar mit niedrigem IR und fügen Sie ein nicht besuchtes Paar mit hohem IR ein: Berechnen Sie den IR für alle besuchten Paare mit Gleichung (18) und entfernen Sie ein Kundenpaar mit dem niedrigsten IR-Wert. Berechnen Sie dann den IR für alle nicht besuchten Kundenpaare neu, wählen Sie eines mit dem höchsten IR-Wert aus und fügen Sie es an der besten Position in der Route ein. • Fügen Sie ein Kundenpaar mit hohem IR ein: Berechnen Sie den IR für alle nicht besuchten Kundenpaare, wählen Sie eines mit dem höchsten IR-Wert aus und fügen Sie

Nach Anwendung des ausgewählten Nachbarschaftsoperators wird, wenn die neue Lösung S' durchführbar ist, die Fitness von S' mit Hilfe von Gleichung (1) berechnet und mit der Fitness der alten Lösung S verglichen. Um lokalen Optima zu entgehen, haben wir uns entschieden, den Dämonalgorithmus (DA) zu

Machine Translated by Google

als ein einfacher lokaler Suchalgorithmus, da der DA die Fähigkeit besitzt, eine mögliche, nicht verbesserte neue Lösung zu akzeptieren. Der DA ist eine Variante des SA, der von Creutz [19] vorgestellt w Die Akzeptanzfunktion des DA basiert auf dem Wert des Dämons (Kredit), der zu Beginn mit einem gegebenen Wert D initialisiert wird. Falls die neue Lösung S' keine Verbesserung darstellt, wird die Differenz (ÿE) zwischen der Fitness von S und S' mit dem Dämonwert verglichen und S' wird akzeptiert, wenn diese Differenz kleiner oder gleich D ist. Wenn S' akzeptiert wird, wird es gespeichert und der Wert von D verringert; andernfalls wird die alte Lösung S gespeichert, der D-Wert erhöht und ein Nichtverbesserungszähler für diese Lösung um eins erhöht. Der DA ist einfacher und schneller als der SA, da letzterer mehr Rechenzeit zum Generieren einer Zufallszahl und einer Exponentialfunktion benötigt. Außerdem enthält er nur einen Parameter, der angepasst werden muss. Der DA erwies sich als effektiver Algorithmus, als er zur Lösung mehrerer realer Probleme verwendet wurde und im Vergleich zum SA gute Ergebnisse in Bezug auf die Qualität der Lösung und die Zeit erzielte [19].

2. Zuschauerbienenphase:

Nur Lösungen mit guten OF-Werten haben eine hohe Wahrscheinlichkeit, den Zuschaueragenten zugewiesen zu werden. Es wurden verschiedene Strategien zur Auswahl von Lösungen vorgeschlagen, wie z. B. Roulette-Rad-Auswahl, Turnierauswahl, stochastisches Universal Sampling und rangbasierte Auswahl [20]. Wir haben die beliebteste Strategie übernommen, nämlich die Roulette-Rad-Auswahlstrategie , bei der die Auswahlwahrscheinlichkeit Pi wie fol

$$Pi = \frac{OFi}{\ddot{y}_{j=1}^{N} OFj}, \quad (19)$$

wobei OFi die Zielfunktion für eine Lösung i ist und $\ddot{y}_{j=1}^{N} OFj$ die Summe der Zielfunktionswerte für alle Lösungen in einer Population ist. Nachdem eine Lösung ausgewählt wurde, wird einer der Lösungswege zufällig ausgewählt und alle in der Phase des „Employed Bee“ verwendeten Nachbarschaftsoperatoren werden nacheinander darauf angewendet. Anschließend wird die Eignung der neuen Lösung mithilfe von Gleichung (1) gemessen. Um den Intensivierungsprozess zu erhöhen und sich nur auf vielversprechende Bereiche des Suchraums zu konzentrieren, basiert das Akzeptanzkriterium auf einem einfachen LS-Algorithmus (Local Search). Das heißt, wenn S' besser als S ist, merke dir S' und es besteht keine Notwendigkeit, die verbleibenden Operatoren zu vervollständigen; andernfalls versuche es mit dem nächsten Operator. Wenn die alte Lösung durch keinen Operator verbessert wird, wird die alte Lösung gemerkt und ein Nichtverbesserungszähler für

3. Scout-Bee-Phase: Die

Scout-Bee-Phase zielt darauf ab, die Diversifizierung zu erhöhen und den Lösungsraum zu erweitern, um so zu verhindern, dass man in lokalen Optima stecken bleibt. Die Scout-Bee-Phase beginnt, wenn ein Nichtverbesserungszähler für eine Lösung Si eine vordefinierte Anzahl von Versuchen erreicht, die als Limit bezeichnet wird. Das Limit bezieht sich auf die Anzahl der Versuche, die unternommen wurden, um die Lösung Si zu verbessern , sowohl in der Phase der beschäftigten als auch der Phase der Zuschauerbienen. Wenn ein Nichtverbesserungszähler für Lösung S das Limit erreicht, sollten drei wichtige Aktionen ausgeführt werden: (1) den Nichtverbesserungszähler löschen, (2) die Lösung Si löschen und (3) die Lösung Si rekonstruieren .
Eine geeignete Rekonstruktionsstrategie spielt in der Scout-Bee-Phase eine wichtige Rolle, da sie entweder zur Verbesserung der Lösungsqualität beiträgt oder die Lösung unverändert lässt. Daher haben wir zwei Strategien zur Rekonstruktion der Lösung Si vorgeschlagen . Die Einzelheiten dieser Rekonstruktionsheuristiken werden wie folgt dargestellt: Erste Strategie (ABC(S1)):

- Eine der folgenden drei Heuristiken wird zufällig ausgewählt:

1. Heuristik der zufälligen Einfügung:

Solange die Zeitbeschränkung nicht verletzt wird, werden die Kundenpaare nach dem Zufallsprinzip ausgewählt und direkt in die Route eingefügt. Auf jeden Abholkunden folgt sein Lieferkunde. Daher besteht keine Notwendigkeit, die Kapazitäts- und Vorrangbeschränkungen des Fahrzeug

2. Greedy- und randomisierte Heuristiken:

Machine Translated by Google

Solange die Zeitbeschränkung nicht verletzt wird, werden die Kundenpaare nach dem Zufallsprinzip ausgewählt und an den besten Positionen auf der Route eingefügt. Neben der Zeitbeschränkung dürfen auch die Beschränkungen hinsichtlich Fahrzeugkapazität und Vorrang nicht verletzt werden.

3. Heuristik der gierigen Einfügung:

Das Einfügungsverhältnis wird für alle Kundenpaare nach Gleichung (18) berechnet. Das Kundenpaar mit dem höchsten IR-Wert wird dann an der besten Position in die Route eingefügt. Zeit-, Fahrzeugkapazitäts- und Vorrangbeschränkungen dürfen nicht

verletzt werden. • Zweite Strategie (ABC(S2)): Diese Strategie funktioniert genau wie das GRASP in [

Nachdem die Lösung Si **rekonstruiert wurde, wird die OF mit Hilfe der Gleichung (1) berechnet und Si als neue Lösung gespeichert und verwendet, um den Verbesserungsprozess unter Verwendung der verwendeten und der beobachteten Bienenphasen erneut zu starten.**

Die Gliederung der Routing-Phase unseres ABC wird in Algorithmus 1 dargestellt, wobei jede Notation wie folgt definiert ist: Pop Size stellt die Populationsgröße dar; Max Iter: die maximale Anzahl von Iterationen; Num Vehicles: die Menge der Fahrzeuge; US: Kundenpaare, die nicht bedient wurden; ÿE: **die Differenz zwischen den OF-Werten für eine alte und eine neue Lösung;** D : **Dämonwert; und** Non- **Improvement :** ein Zähler, der **die Anzahl der erfolglosen Versuche zählt, die Lösung zu verbessern.**

Algorithmus 1: Der Pseudocode des vorgeschlagenen ABC.

Pop = {S1 , **S2, S3, .., Sn}; /* Erste Lösungen, die mit GRASP(V2) erzeugt wurden*/ Limit = erster Grenzwert,** D **= erster Dämonwert, Nichtverbesserung** = **0, Global_Best = Beste Lösung in Pop; Für** M = **1 bis** Max_Iter **/*** Verwendete Bienenphase ***/ Für** i = **1 bis Pop_Size Wähle zufällig einen Lösungsweg** Si ; **Wende einen zufällig ausgewählten Nachbarschaftsoperator auf den aktuellen Lösungsweg** Si an , **um eine neue Lösung** S zu erhalten. **Wenn** S **durchführbar ist, berechne die** OF **von** S i. **Ende, wenn**

$\ddot{y}_{ich;}$ $\ddot{y}$ $\ddot{y}_{ich;}$

Wenn von $S_{\ddot{i}}^{\ddot{y}} > VON(S_{-})$)
Merken Sie sich die neue Lösung S $\ddot{y}_{ich;}$

Sonst wenn (ÿE **ÿ** Di)
Merken Sie sich die neue Lösung S $\ddot{y}_{i;}$

aktualisiere den Dämonwert: Di = Di **ÿ** ÿE;
Anders

Merken Sie sich die alte Lösung Si **und**

aktualisieren Sie den Dämonwert: Di = Di + ÿE;

Keine Verbesserung(i) = Keine Verbesserung(i) + **1;**
Ende, wenn

Ende für

/* Zuschauerbienenphase ***/**

Wählen Sie mithilfe der Rouletterad-Auswahlstrategie N Lösungen aus den eingesetzten Bienen aus.
Für i = **1 bis** N

Wählen Sie zufällig einen Lösungsweg Si aus ;

Für j = **1 bis Anzahl der Nachbarschaftsoperatoren**

Anwenden eines ausgewählten Nachbarschaftsoperators (j) **auf den aktuellen Lösungsweg** Si , **um eine neue Lösung** S zu erhalten $\ddot{y}_{ich;}$

Wenn $^{\ddot{y}}$S **durchführbar ist, dann berechnen Sie die** OF **von** S i $\ddot{y}_{ich;}$
Ende, wenn

Wenn von $S_{\ddot{i}}^{\ddot{y}} > VON(S_{-})$
Merken Sie sich die neue Lösung $S^{\ddot{y}}_{ich;}$

Brechen;

Ende, wenn

Ende für

Machine Translated by Google

Wenn sich die Lösung Si **nach Anwendung der Nachbarschaftsoperatoren nicht verbessert hat**
Keine Verbesserung(i) = **Keine Verbesserung(i)** + **1;**
Ende, wenn
Ende für

/* Kundschafterbienenphase ***/**
Für i = **1 bis** N

Wenn Nichtverbesserung(i) > **Grenzwert**
Keine Verbesserung(i) = **0;**
Si = **0;**
Rekonstruieren Sie Si , **indem Sie entweder die S1- oder die S2-Strategie der Scout-Bee-Phase verwenden. Berechnen Sie die** OF **für** Si .
Ende, wenn
Ende für

Weisen Sie die Lösung mit dem höchsten OF der Best_Solution zu.
Wenn (Global_Best < **Beste_Lösung)**
Global_Best = **Beste_Lösung;**
Ende, wenn
Ende für

Ausgabe: die Global_Best-Lösung

4.2. ABC mit Clustering (ABC(C))

In diesem Teil der Studie wurde die in Abschnitt 4.1 beschriebene ABC-Methode **verwendet, um die ursprünglichen Lösungen zu verbessern, die mit den verschiedenen in [7] beschriebenen GRASP(C)-Methoden, nämlich K-means-GRASP(C), modifiziertem K-means-GRASP(C) und ACO-GRASP(C), e Auch hier müssen bei der Verwendung von ABC mit diesen Algorithmen einige Modifikationen berücksichtigt werden. Bei Verwendung der folgenden Operatoren sollte der Einfügungsprozess auf die Kundenpaare innerhalb desselben Clusters beschränkt werden: Entfernen Sie zufällig ein Kundenpaar und fügen Sie ein nicht besuchtes ein, entfernen Sie ein Kundenpaar mit niedriger IR und fügen Sie ein nicht besuchtes Paar mit hoher IR ein und fügen Sie ein Kundenpaar mit hoher IR ein. Außerdem sollte die Einfügungsbeschränkung in der Scout-Bee-Phase in beiden Strategien (S1 und S2) angewendet werden. Wir haben dieselben Parameterwerte verwendet, die in ABC verwendet wurden (ohne Clustering), um die anfänglichen Lösungen in diesem Teil der Studie zu verbessern. Darüber hinaus wurden in den ABC(C))-Algorithmen dieselben Strategien aus der Scout-Bee-Phase (S1 und S2) verwendet. Insgesamt gibt es also sechs Algorithmen, die ABC(C) verwenden: (1) K-means-GRASP(C)-ABC(C)S1, (2) K-means-GRASP(C)-ABC(C)S2, (3) modifiziertes K-means-GRASP(C)-ABC(C)S1, (4) modifiziertes K-means-GRASP(C)-ABC(C)S2, (5) ACO-GRASP(C)-ABC(C)S1 und 6) ACO-GRASP(C)- ABC(C)S2. Der erste Algorithmus bezieht sich beispielsweise auf die Generierung der anfänglichen Lösung unter Verwendung von GRASP, nachdem der Suchraum mithilfe des K-means-Algorithmus geclustert un**

5. Computergestützte Experimente

Alle Algorithmen wurden in MATLAB (R2017b) auf einem Laptop mit einem Intel Core i7-4510U-Prozessor mit 2,00 GHz (2601 MHz, zwei Kerne, vier logische Prozessoren) implementiert. Die folgenden Unterabschnitte enthalten Beschreibungen der verwendeten Datensätze und die Einzelheiten der Parameterabstimmung, bevor die experimentellen Ergebnisse präsentiert werden.

5.1. Testinstanzen

Um die Leistungsfähigkeit unserer vorgeschlagenen Methoden zu testen, haben wir dieselben Instanzen wie in [4] verwendet. Der Datensatz umfasst 36 Instanzen, die in drei unterschiedliche Gruppen eingeteilt wurden : kleine Instanzen (bestehend aus 20 bzw. 50 Kunden, die von zwei bzw. drei Fahrzeugen bedient werden), mittlere Instanzen (bestehend aus 100 bzw. 250 Kunden, die von 4 bzw. 5 Fahrzeugen bedient werden) und große Instanzen (500 bzw. 1000 Kunden, die von 6 bzw. 8 Fahrzeug Darüber hinaus besteht jede Gruppe aus zwölf Instanzen. Diese Vorkommnisse weisen sowohl zeitliche als auch umsatzmäßige Schwankungen auf. Um kurze und lange Routen zu generieren, ist das Gesamtzeitlimit entweder klein oder groß und liegt in einem Bereich von 2500 bis 15.000 s.

Machine Translated by Google

Darüber hinaus werden die Einnahmen auf eine von drei Arten festgelegt: nach dem Zufallsprinzip, proportional zur Nachfrage oder gleich für alle Kunden. Einnahmen werden durch die Lieferung erzielt Kunden nach Versand der Ware durch den Abholkunden.

Die Warenmenge wird als positive Ganzzahl zwischen 1 und 50 angegeben.

5.2. Parameterabstimmung

5.2.1. Parameterabstimmung von ABC ohne Clustering

Der ABC-Algorithmus hat drei Parameter: die Anzahl der Iterationen (Max_Iter), die Größe der Population (Pop_Size) und dem Anfangswert des Limits. Da wir zusätzlich der DA in der eingesetzten Bienenphase, der Anfangswert des Dämons (D) muss angepasst werden.

Die Einzelheiten zum Testen der einzelnen Parameter werden unten dargestellt.

Anzahl der Iterationen

Jede Dateninstanz von 100 Kunden wurde mit unterschiedlichen Iterationszahlen getestet: 10, 50, 100, 500 und 750. Wie aus Tabelle 1 hervorgeht, zeigten die Testergebnisse, dass keine Verbesserung der Zielfunktionswerte nach 500 Iterationen. Somit ist die maximale Die Anzahl der Iterationen wurde mit 500 angenommen.

Tabelle 1. Ergebnisse der Anpassung der Anzahl der Iterationen des ABC.

Datensatz Instanzen	Anzahl der Iterationen				
	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

Bevölkerungsgröße

Die gleichen Datensatzinstanzen wurden erneut getestet, um die entsprechende Anzahl von Lösungen in der Population. Zu Beginn wurde die Populationsgröße auf 30 festgelegt, da es lieferte die besten Ergebnisse, die in [7] angegeben sind. Dann haben wir die Population auf verschiedene Größen erhöht: 40, 50 und 60 Lösungen. Tabelle 2 zeigt die Werte der OF nach der Festlegung der Anzahl der Iterationen auf 500. Die Ergebnisse zeigten, dass eine Erhöhung der Populationsgröße auf mehr als 50 führte nicht zu einer Verbesserung des Zielfunktionswertes. Somit ist die Populationsgröße wurde mit 50 angenommen.

Tabelle 2. Ergebnisse der Anpassung der Populationsgröße des ABC.

Datensatz Instanzen	Bevölkerungsgröße			
	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

Grenzwert

Der Grenzwert des Nichtverbesserungszählers ist ein weiterer wichtiger Parameter von der ABC-Algorithmus, da er die Konvertierung des Suchvorgangs beeinflusst. Somit ist der Wert Die für die Grenze gewählte Anzahl sollte für die Anzahl der Iterationen geeignet sein. Daher haben wir getestet die Grenze mit zwei Werten: 100 und 200. Die Ergebnisse in Tabelle 3 zeigen, dass eine Erhöhung der

Machine Translated by Google

Grenzwertes auf über 100 führte nicht zu einer Verbesserung der Zielfunktion Der Grenzwert wurde daher mit 100 angenommen.

Tabelle 3. Ergebnisse der Feinabstimmung des ABC-Grenzwerts.

Datensatz Instanzen	Limit	
	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

Dämonenwert

Um den besten Anfangswert des Dämons (D) zu ermitteln, wurden mehrere Werte getestet: 1000, 2500, 5000 und 7500. Die Ergebnisse in Tabelle 4 zeigten keine Verbesserung der Zielfunktion wenn der Dämonenwert größer als 5000 war. Daher wurde der Dämonenwert D **mit 5000 angenommen.**

Tabelle 4. Ergebnisse der Anpassung des Dämonwerts des ABC.

Datensatz Instanzen	Dämonenwert			
	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. Parameterabstimmung von ABC(C)

Da wir den gleichen ABC-Algorithmus mit nur geringen Änderungen im ABC mit dem Clustering-Ansatz haben wir die gleichen Parameterwerte übernommen, die in der ABC für unser ABC(C).

5.4. Experimentelle Ergebnisse

Wie bereits erwähnt, haben wir zwei Versionen des ABC vorgeschlagen, um die anfängliche Lösungen: das ABC, das mit GRASP verwendet wird, und das ABC(C), das mit den Clustermethoden verwendet wird, d. h. K-means-GRASP(C), modifiziertes K-means-GRASP(C) und ACO-GRASP(C). Wir haben auch zwei verschiedene Strategien für die Kundschafterphase. S1 konstruiert die Lösung durch zufälliges Auswahl einer von drei Strategien: zufällige Einfügung, gierige und zufällige Einfügung oder gierige Einfügung. Nach dem Testen dieser Strategie stellten wir fest, dass die Heuristik der zufälligen Einfügung die Ergebnisse im Vergleich zu anderen nicht verbessern, also haben wir diese Wahl eliminiert und die andere behalten zwei. Die zweite Strategie, S2, funktioniert wie das GRASP in [7]. Daher haben wir insgesamt acht vorgeschlagene Methoden mit folgenden Namen: (1) ABC(S1), (2) ABC(S2), (3) K-means-GRASP(C)-ABC(C)S1, (4) K-means-GRASP(C)-ABC(C)S2, (5) modifiziertes K-means-GRASP(C)-ABC(C)S1, (6) modifiziertes K-means-GRASP(C)-ABC(C)S2, (7) ACO-GRASP(C)-ABC(C)S1 und (8) ACO-GRASP(C)-ABC(C)S2. Alle vorgeschlagenen Algorithmen wurden fünfmal ausgeführt, wie mit die vorherigen Algorithmen in [6] und [7]. Wir verglichen unsere vorgeschlagenen Algorithmen und die modernste Algorithmen, die in [10,11] vorgestellt werden, nämlich C1-Seq VNS, C2-Seq VNS, C1-Self Adaptive VNS, C2-Self Adaptive VNS und IPPD, zur Lösung des MVPPDP im Hinblick auf die Bearbeitungszeit und erzielter Gewinn.

Erstens, was die Verarbeitungszeit betrifft, wurden alle vorherigen Algorithmen ausgeführt für 5 min für jeden Fall. Wir weisen jedoch darauf hin, dass diese Zeit nicht als akzeptabel angesehen wird für die M-Commerce-Apps (z. B. Hungerstation, MRSOOL usw.), die täglich genutzt werden und auf der

Machine Translated by Google

Reaktionsgeschwindigkeit, daher haben wir in unserem Fall die Abbruchbedingung von 500 Iterationen verwendet, um messen Sie die Verarbeitungszeit. Tabelle 5 zeigt die Vergleichsergebnisse zwischen unseren vorgeschlagenen Algorithmen und modernste Algorithmen in Bezug auf die Zeit (in Sekunden). Auch der Durchschnitt jeder Kategorie (hervorgehobene Zeilen in Tabelle 5) wurde verwendet, um die Leistung zu vergleichen der Algorithmen, wie in Abbildung 2 dargestellt. Aus Tabelle 5 und Abbildung 2 geht hervor, dass unsere Die vorgeschlagenen clusterbasierten Algorithmen übertreffen alle bisherigen Algorithmen in Bezug auf zur Bearbeitungszeit in allen Fällen, außer in Fällen mit 1000 Kunden. Außerdem die anderen vorgeschlagenen Algorithmen, die nicht auf Clustering basieren (ABC(S1) und ABC(S2)) waren schneller als die vorherigen Algorithmen für kleine und mittlere Instanzen. Die Der Grund, warum unsere Algorithmen bei großen Instanzen nicht überzeugen, liegt wahrscheinlich an der Scout Bee-Strategien, die verwendet wurden, nämlich die Greedy-Methode in der ersten Strategie und der GRASP im zweiten. Bei der Greedy-Methode werden die Kundenpaare ausgewählt basierend auf eine bestimmte Maßnahme und dann in die beste Position eingefügt, die zu einer Verringerung führt die Reisekosten. Dieser Ansatz liefert gute Ergebnisse, ist aber zeitaufwendig, insbesondere bei großen Instanzen. Bei der zweiten Strategie werden die nicht besuchten Kundenpaare ausgewählt auf der Grundlage von Berechnung der IR, was bei großen Instanzen ebenfalls zeitaufwendig ist. Insgesamt Allerdings half die Clusterung des Suchraums eindeutig bei der Verarbeitung großer Instanzen und hat die Bearbeitungszeit im Vergleich zu anderen Algorithmen, die alle Kunden bearbeiten, beschleunigt ohne Clusterbildung.

Tabelle 5. Vergleich zwischen den neuesten Algorithmen und den vorgeschlagenen Algorithmen basierend auf Zeit (Sekunden).

Datensatz	Vorherige Algorithmen(5 Min Laufzeit)	K-Mittel-FASSEN-ABC(C)S1	K-Mittel-FASSEN-ABC(C)S2	Geändert K-Mittel-FASSEN-ABC(C)S1	Geändert K-Mittel-FASSEN-ABC(C)S2	ACO-FASSEN-ABC(C)S1	ACO-FASSEN-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	30,65
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	28.45
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	21,56
Durchschnitt-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
Durchschnitt-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
Durchschnitt-100	300	34,68	34,15	1	4,79	6,85	3,30	7,46	83,19
19-250-FS	300	7 3,50	72,85	78,10	73,30	74,19	67,79	176,81	204,37
20-250-FL	300	79,65	73,28	80,50	76,28	77,66	65,10	181,10	195
21-250-PS	300	74,20	72,09	75,23	72,88	74,05	63,51	169,26	195,17
22-250-PL	300	73,82	72,718	79,53	74,40	77,86	65,44	179	193,60
23-250-RS	300	74,67	72,99	77,63	72,38	76,21	66,44	177,25	193,93
24-250-RL	3 00	79,55	73,21.	84,04	73,84	79,62	69,46	178,99	194,47
Durchschnitt-250	300	75,90	72,86	79,17	73,85	76,60	66,29	177,07	196,09
25-500-FS		163,89	155,32	172,74	158,36	162,15	142,52	388,62	474,16
26-500-FL		167,03	155,73	177,18	160,25	170,25	148,98	415,25	466,38
27-500-PS		160,91	146,27	172,04	158,45	162,76	140,62	386,55	465,87
28-500-PL		168,60	144,23	180,26	158,74	169,89	141,34	435,34	464,62

Machine Translated by Google

Tabelle 5. Fortsetzung

Datensatz	Vorherige Algorithmen(5 Min Laufzeit)	K-Mittel-FASSEN-ABC(C)S1	K-Mittel-FASSEN-ABC(C)S2	Geändert K-Mittel-FASSEN-ABC(C)S1	Geändert K-Mittel-FASSEN-ABC(C)S2	ACO-FASSEN-ABC(C)S1	ACO-FASSEN-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
Durchschnitt-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
Durchschnitt-1000	300	437,78	378,73	440,28	420,90	407,16	417,72	1150,64	1373,71

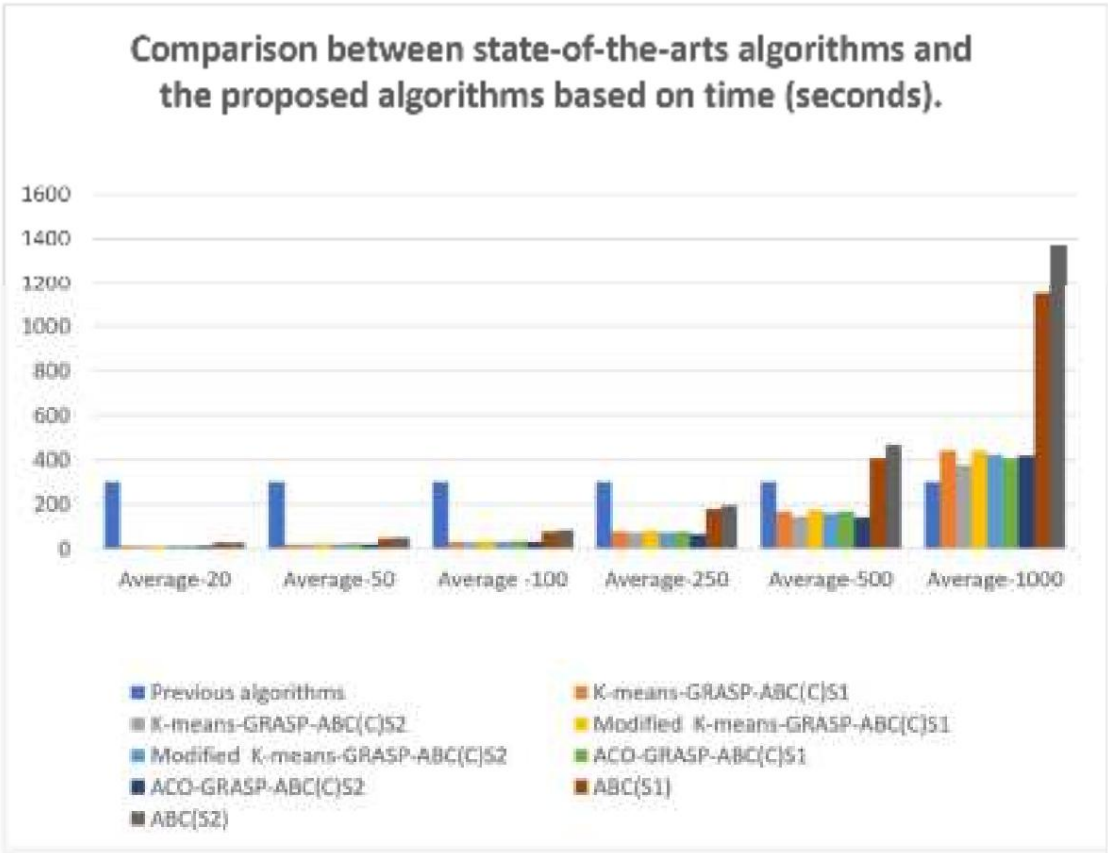


Abbildung 2. Vergleich zwischen den aktuellen Algorithmen und den vorgeschlagenen Algorithmen basierend auf Zeit (Sekunden).

Andererseits verglichen wir unsere vorgeschlagenen Algorithmen mit den vorherigen Algorithmen in Bezug auf den Gewinn, wie er in Gleichung (1) berechnet wird. Tabelle 6 zeigt den durchschnittlichen Gewinn für fünf Läufe, während Tabelle 7 den besten Gewinn aus fünf Läufen und die Lücke zwischen den bester Gewinn und der beste bekannte Gewinn mithilfe der folgenden Gleichung:

$$\text{Lücke} = \frac{\text{Bester Gewinn über fünf Läufe ÿ Bekanntester Gewinn}}{\text{Bekanntester Profi}} \text{ ÿ } 100 \qquad (20)$$

Machine Translated by Google

Tabelle 6. Vergleich unserer vorgeschlagenen Algorithmen mit früheren Methoden basierend auf dem durchschnittlichen Gewinn.

Datensatz	C1 – Folge VNS [9]	C2 – Sequenz VND [9]	C1 – Selbst Anpassungsfähig VND [9]	C2 – Selbst Anpassungsfähig VND [9]	Die IPPD [10]	K-Mittel-FASSEN-ABC(C)S1	K-Mittel-FASSEN-ABC(C)S2	Geändert K-Mittel-FASSEN-ABC(C)S1	Geändert K-Mittel-FASSEN-ABC(C)S2	ACO-FASSEN-ABC(C)S1	ACO-FASSEN-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
Durchschnitt-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 Abweichung	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
Durchschnitt-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 Abweichung	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
Durchschnitt-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 Abweichung	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
Durchschnitt-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 Abweichung	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

Tabelle 6. Fortsetzung

Datensatz	C1 – Folge VNS [9]	C2 – Sequenz VND [9]	C1 – Selbst Anpassungsfähig VND [9]	C2 – Selbst Anpassungsfähig VND [9]	Die IPPD [10]	K-Mittel-FASSEN-ABC(C)S1	K-Mittel-FASSEN-ABC(C)S2	Geändert K-Mittel-FASSEN-ABC(C)S1	Geändert K-Mittel-FASSEN-ABC(C)S2	ACO-FASSEN-ABC(C)S1	ACO-FASSEN-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
Durchschnitt-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 Abweichung	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
Durchschnitt-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 Abweichung	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

Tabelle 7. Vergleich zwischen den vorgeschlagenen Algorithmen und früheren Methoden basierend auf den besten Lösungs- und Lückenwerten.

Datensatz	Am besten Bekannt Profitieren	C1—Seq VNS [4]		C2—Seq VND [4]		C1 – Selbstadaptiver VND [4]		C2 – Selbstadaptiver VND [4]		Die IPPD [10]		K-Mittel-GRASP-ABC(C)S1	
		Am besten Profitieren	Lücke	Am besten Profitieren	Lücke	Am besten Profitieren	Lücke	Am besten Profitieren	Lücke	Am besten Profitieren	Lücke	Am besten Profitieren	Lücke
1-20-FS	30.315,5	30.315,5		30.315,5		30.315,5		30.315,5		30.315,5		30.267,75	ÿ0,15
2-20-FL	39.340,3	38.765,6	0 ÿ1,5	38.765,6	0 ÿ1,5	38.765,6	0 ÿ1,5	38.765,6	0 ÿ1,5	39.179,66	0 ÿ0,41	37.750,35	ÿ4
3-20-PS	52.855,2	51.473,4	2,7	52.561,1	ÿ0,6 0	52.565,1	ÿ0,6 0	52.565,1	ÿ0,6 0	52.566	ÿ0,55	51.515,53	ÿ2,5
4-20-PL	58.365,6	58.365,6	0	58.365,6		58.365,6		58.365,6		58.777	0,7	57.350,35	ÿ1,7
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	ÿ6,5
6-20-RL	43.764,3	43.368,6	ÿ0,9	43.368,6	ÿ0,9	43.368,6	ÿ0,9	43.368,6	ÿ0,9	43.781,81	0,04	42.353,35	ÿ3,2
Durchschnitt-20	43.518,4	43.126,37	ÿ0,85	43.307,65	ÿ0,50	43.308,32	ÿ0,50	43.308,32	ÿ0,50	43.514,91	ÿ0,04	42.221,32	ÿ2,9
7-50-FS	22.441	22.306	ÿ0,6	22.306	ÿ0,6	22.306	ÿ0,6	22.306	ÿ0,6	22.307,15	ÿ0,6	18.662,83	ÿ16,8
8-50-FL	58.050,8	58.050,8	0	58.050,8	0	58.050,8	0	58.050,8	0	58.050,8	0	51.040,14	ÿ12

Machine Translated by Google

Tabelle 7. Fortsetzung

Datensatz	Am besten Bekannt Profitieren	C1—Seq VNS [4]		C2—Seq VND [4]		C1 – Selbstadaptiver VND [4]		C2 – Selbstadaptiver VND [4]		Die IPPD [10]		K-Mittel-GRASP-ABC(C)S1	
		Am besten Profitieren	Lücke	Am besten Profitieren	Lücke	Am besten Profitieren	Lücke	Am besten Profitieren	Lücke	Am besten Profitieren	Lücke	Am besten Profitieren	Lücke
9-50-PS	57.651,7	57.651,7	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	ÿ13,3
11-50-RS	33.340,4	79.350,3	0	33.340,4	0	33.340,4		33.340,4		33.340,4		28.907,33	ÿ13,2
12-50-RL	63.913,7	63.891,20	ÿ0,1	79.350,3	0	79.350,3		79.350,3		77.832,56	0 0 0	76.479,84	ÿ3,6
Durchschnitt-50	44.402,3	42.174,4	ÿ5,3	81.999 .4		63.891,20	ÿ0,10	63.891,20	0 0 0 0	63.891,20	0 0 0 0	63.638,44	ÿ1,95
13-100-FS	81.999,4	0 71.512,4		70.836 ÿ1	125.032	42.174,4	ÿ5,3	81.999,4		42.215,5	ÿ0,10	41.956,25	ÿ0,43
14-100-FL	124.770	80.878,8		80.878,8	131.845	0 71.326,4	ÿ0,3			80.848,1	ÿ5,2	88.891,4	
15-100-PS	131.473	89.278,32		97.991,44	ÿ1,13	124.770	78.818			81.621	ÿ5,3	71.399,2	ÿ0,5 ÿ0,2
16-100-PL	75.309,1	73.610,9	ÿ2,3	ÿ0,2		131.845	ÿ0,2			71.399,2	ÿ0,2 ÿ0,2	71.234,58	ÿ5,83 0 ÿ0,39
17-100-RS	136.594	134.889		113.886	0		ÿ2,6			123.766	ÿ1	124.910	ÿ0,1
18-100-RL	112.427	200.402		200.402	ÿ0,3	88.488,87.	0			80.878,8	0	80.878,8	0
Durchschnitt-100	164.070	162.960		274.563	271.185	ÿ1 .40	73.575,7	ÿ2,4		131.339	ÿ0,4	131.339	ÿ0,4
19-250-FS	160.804	159.245,65	ÿ1,13	167.080		134.186	112.480			88.529,73	ÿ1,23	88.598,43	ÿ1,02
20-250-FL	161.063	ÿ3,7	26-500-FL		ÿ1,3	198.611	ÿ1,8			73.835,5	2	73.544,7	ÿ2,4
21-250-PS	259.319	249.171	27-500-		ÿ1,3	164.070	ÿ1,3			132.870	ÿ2,8	135,6 22	ÿ0,7
22-250-PL	PS 206.788	203.832	ÿ1,5		0	271.193	ÿ0,9			112.428	ÿ1,3	110.062	ÿ3,5
23-250-RS	28-500-PL 331.291				ÿ0,7	159.019,28	0			195.135	ÿ2,7	196.524	ÿ2
24-250-RL	321.471	ÿ3,1	29-500-RS		ÿ1,2	ÿ1,27	ÿ1,2			162.115	ÿ1,2	162.115	ÿ1,2
Durchschnitt-250	281.711	279.105	ÿ0,9	30-500-RL		161.040	ÿ3,8	249.661		272.957	ÿ0,6	269.790	ÿ1,8
25-500-FS	417.282	407.689	ÿ2,4	Durchschnitt-500		205.480	ÿ0,6	ÿ1,8		158.223,42	ÿ1,10	174.822,60	ÿ1,93
277.245,2	270.388,50	ÿ2,61	31-1000-FS	173.826	ÿ4,1	325.522	ÿ3,9			159.064	ÿ5	161.375	ÿ3,5
154.196	ÿ12,7	32-1000-FL	247.385	221.958	ÿ11,5	33-1000-		278.233	ÿ1,3	409.755		247.986	ÿ4,6
PS 464.580	450.349	ÿ3,2	34-1000-PL	712.918	693.407	ÿ2,8		271.615,17	ÿ1,8			200.764	ÿ3
35-1000-RS	492.083	463.625	ÿ6,1	36-1000-RL	705.167			ÿ2,20	151.943,2	ÿ12,2		320.872	ÿ3,2
673.951	ÿ4,6	Durchschnitt-1000	465.993,2	405.958	ÿ6,8			221.322	ÿ11,8	450.337		274.419	ÿ2,7
								ÿ3,2	672.910	ÿ5,9		407.849	ÿ2,3
								472.626	ÿ4,1 . 6	69.926		268.492,33	ÿ3,47
								ÿ5,3	439.844,03	ÿ7,08		154.357	ÿ12,6
												224.999	ÿ9,9
												442.311	ÿ5
												684.121	ÿ4,2
												465.022	ÿ5,8
												676.720	ÿ4,4
												441.255	ÿ6,98

Machine Translated by Google

Tabelle 7. Fortsetzung

Datensatz	K-Mittel-GRASP-ABC(C)S2		Modifizierter K-Means-GRASP-ABC(C)S1		Modifizierter K-Means-GRASP-ABC(C)S2		ACO-GRIFF-ABC(C)S1		ACO-GRIFF-ABC(C)S2		ABC (S1)		ABC (S2)	
	Bester Gewinn	Lücke	Bester Gewinn	Lücke	Beste Gewinnlücke	Beste Gewinnlücke	Bester Gewinn			Lücke	Beste Gewinnlücke	Beste Gewinnlücke		
1-20-FS	30.427,58	0,3	27.007,58	ÿ10,9	27.007,58	ÿ10,9	25.490,76	ÿ15,9	25.490,76	ÿ15,9	30.554,62	0,78	30.315,52	0
2-20-FL	37.556,07	ÿ4,5	40.104,23	1.9	39.240,95	ÿ0,2	36.423,3	ÿ7,4	36.322,33	ÿ7,6	40.104,23	1.9	36.234,48	ÿ7,8
3-20-PS	44.334,2	ÿ16,1	51.515,53	ÿ2,5	44.334,2	ÿ16,1	42.944,78	ÿ18,7	34.107,64	ÿ35,4	53.164,53	0,58	52.565,1	ÿ0,5
4-20-PL	57.350,35	ÿ1,7	57.350,35	ÿ1,7	57.350,35	ÿ1,7	56.603,63	ÿ18,1	ÿ3	47.402,91	ÿ18,7	1.32	57.286,89	ÿ1,8
5-20-RS	34.090,58	ÿ6,5	30.827,18	ÿ15,4	29.843,58	31.741,08	0,04	41.071,62	ÿ12,9	31.741,08	ÿ12,9		36.469,52	0
6-20-RL	42.169,13	ÿ3,6	44.764,07	2,2	43.780,72	ÿ7,4	39.045,86	ÿ3,7	ÿ6,1	40.600,21	ÿ7,2	0	44.707,23	ÿ2
Durchschnitt-20	40.987,99	ÿ5,8	41.928,16	ÿ3,6	40.259,56	21.281,88	ÿ12,3		ÿ10,2	35.944,16	ÿ17,4	2,1	44.023,52	42.624,2 ÿ2
7-50-FS	21.603,83	ÿ3,7	21.603,83	ÿ3,7	21.603,83	49.128,45	ÿ7,9 . 5		ÿ5,1	25.123,15	11,9	1,16	24.975,25	24.998,45 11.3
8-50-FL	51.001,04	ÿ12,1	51.001,04	ÿ12,1	50.896,3	5.643,25	ÿ11,8	100.393,3	ÿ15,3	50.061,24	ÿ13,7	11,2	53.710,99	54.288,08 ÿ6,4
9-50-PS	54.165,39	ÿ6	54.165,39	ÿ6	53.064,14	ÿ13,2	31.251,34	ÿ3,4	ÿ3,4	55.643,25	ÿ3,4		57.651,71	57.651,71 0
10-50-PL	119.099,6	ÿ10,2	119.099,6	ÿ10,2	116.868,6	65.792,21	ÿ9,2	53.915,07	ÿ24,3	106.544,8	ÿ19,6	126.360	ÿ7,4 0 ÿ4,7	133.743,2 0,82
11-50-RS	28.907,33.	ÿ13,2	28.907,33	ÿ13,2	28.907,33	ÿ13,6	33.565,46	ÿ25,1	ÿ6,2	29.969 .16	ÿ10,1	33.058,21	ÿ0,8	33.159,12 ÿ0,5
12-50-RL	76 .697,65	ÿ3,3	7 6.697,65	ÿ3,3	76, 604,84	52.328,66	60.614,78		ÿ17	65.752,8	ÿ17,1	77, 467,34	ÿ2,3	76, 517,36 ÿ3,5
Durchschnitt-50	58.579,14	ÿ8,3	58.579,14	ÿ8,3	57.990,84	ÿ10,6	95.836,25	ÿ23,3	ÿ15,6	55.515,73	ÿ13,1	62.203,92	ÿ2,6	63.392,99 ÿ0,8
13-100-FS	36.188,32	ÿ18,4	35.937,96	ÿ19	38.323,6	52.689,7	ÿ16	95.363,69	ÿ24,4	33.566,06	ÿ24,4	41.812,08	ÿ5,8	41.636,8 ÿ6,2
14-100-FL	63.488,24	ÿ22,5	60.851,27	ÿ25,7	61.351,99				ÿ36,1	49.789,52	ÿ39,2	66.375,92	ÿ19	72.595,41 ÿ11,4
15-100-PS	60.815,37	ÿ14,9	54.190,18	ÿ24,2	60.554,49	ÿ15,3			ÿ15,2	47.498,28	ÿ33,5	69.275,79	ÿ3,1	71.400,63 ÿ0,1
16-100-PL	104.786,4	ÿ16,1	102.121,7	ÿ18,3	111.728,2				ÿ23,3	101.142,5	ÿ19,1	115.579,2	ÿ7,5	128.473,2 2.7
17-100-RS	61.647,98	ÿ23,7	61.962,79	ÿ23,3	61.962,79				ÿ34,8	52.689,7	ÿ34,8	77.484,06	ÿ4,1	77.484,06 ÿ4,1
18-100-RL	101.013,6	ÿ23,3	105.175,3	ÿ20,2	110.683,8				ÿ27,6	97.568,94	ÿ25,9	119.554,3	ÿ9,3	126.871,5 ÿ3,7
Durchschnitt-100	71.323,32	ÿ20,1	70.039,87	ÿ21,5	74.100,81	ÿ17	65.066,42		ÿ27,1	63.709,17	ÿ28,6	81.680,23	ÿ8,5	86.410,27 ÿ3,2
19-250-FS	35.443,33	ÿ52,9	48.079,56	ÿ36,1	39.848,03	ÿ47	36.309,86	ÿ35,8	ÿ51,7	36.612,19	ÿ62,4	ÿ51,3	36.309,86	ÿ51,7 36.612,19 ÿ62,4 ÿ51,3
20-250-FL	67.014,68	ÿ50,9	70.325,74	ÿ48,5	87.575,52	51.239,48	73.654,78		62.831,21	73.143,68	ÿ54	51.239,48	62.831,21 ÿ35,3	ÿ54
21-250-PS	77.941,82	ÿ31,5	81.510,78	ÿ28,4	92.124,73	ÿ19,1	ÿ19,4		ÿ35,3	131.668,8	ÿ35,7	73.654,78	73.143,68 ÿ52,1	ÿ35,7
22-250-PL	135.164,3	ÿ32,5	133.567,4	ÿ33,3	161.377,2	95.906,32	ÿ41,8		ÿ52,1	101.863,1	ÿ34,2	95.906,32	131.668,8 ÿ34,4	ÿ34,2
23-250-RS	97.231,36	ÿ40,7	113.547,4	ÿ30,7	95.471,16	107.628,9	146.270,4		ÿ34,4	ÿ46,7	ÿ37,9	107.628,9	101.863,1 ÿ46,7	ÿ37,9
24-250-RL	187.891,9	ÿ31,5	180.807,7	ÿ34,1	180.803	ÿ34,1			166.523,5		ÿ39,3	146.270,4	166.523,5	ÿ39,3
Durchschnitt-250	100.114,6	ÿ37,7	104.639,8	ÿ34,9	109.533,3	ÿ31,8	85.168,2		ÿ47	95.440,4	ÿ40,6	85.168,2	ÿ47	95.440,4 ÿ40,6
25-500-FS	74.692,46	ÿ55,2	51.581,38	ÿ69,1	76.481,48	ÿ54,2	61.718,84	ÿ49,2	ÿ63	54.247,37	ÿ67,5	78.432,78	ÿ53 86.250,21	ÿ56,7 ÿ48,3
26-500-FL	124.478,4	ÿ51,9	106.505,3	ÿ58,9	131.572,4	72.388,37	ÿ27,1	109.238	ÿ72	93.308,66	ÿ64	112.172	126.203,2 ÿ34,1	ÿ51,3
27-500-PS	152.566,6	ÿ26,2	107.797,6	ÿ47,8	150.645,3				ÿ47,1	112.111,9	ÿ45,7	136.168,2	148.384,5	ÿ28,2

Machine Translated by Google

Tabelle 7. Fortsetzung

Datensatz	K-Mittel-GRASP-ABC(C)S2		Modifizierter K-Means-GRASP-ABC(C)S1		Modifizierter K-Means-GRASP-ABC(C)S2		ACO-GRIFF-ABC(C)S1		ACO-GRIFF-ABC(C)S2		ABC (S1)		ABC (S2)	
	Bester Gewinn	Lücke	Bester Gewinn	Lücke	Beste Gewinnlücke	Beste Gewinnlücke	Bester Gewinn			Lücke	Beste Gewinnlücke	Beste Gewinnlücke		
28-500-PL	249.759,7	ÿ24,6	225.812,4	ÿ31,8	247.180,6	ÿ25,3	152.573,5	ÿ53,9	196.969,2	ÿ40,5	201.973,8	ÿ39	225.040,5	ÿ32
29-500-RS	172.397,2	ÿ38,8	108.998,5	ÿ61,3	175.267,1	ÿ37,7	141.743,7	ÿ29,4	ÿ49,6	128.570,6	ÿ54,3	178.693,9	ÿ36,5	173.003,3	ÿ38,5
30-500-RL	275.488,2	ÿ33,9	241.454	ÿ42,1	294.424,7	167.819,5		ÿ59,7	214.081,3	ÿ48,6	239.675,7	ÿ42,5	261.517,2	ÿ37,3
Durchschnitt-500	174.897,1	ÿ36,9	140.358,2	ÿ49,3	179.261,9	ÿ35,3	117.580,3	ÿ57,5	133.214,8	ÿ51,9	157.852,7	ÿ43	170.066,5	ÿ38,6
31-1000-FS	51.188,29	ÿ70,5	39.530,22	ÿ77,2	51.585,38	ÿ70,3	23.266,5	ÿ60,4	ÿ86,6	25.563,9	ÿ85,2	33.125,19	ÿ80,9	42.197,08 ÿ76,9	ÿ75,7
32-1000-FL	97.093,51	ÿ60,7	92.142,15	ÿ62,7	97.772,35	24.110,85	ÿ38	178.654,3	ÿ90,2	47.620,03	ÿ80,7	57.137,24	73.126,9	ÿ48,1	ÿ70,4
33-1000-PS	293.900,2	ÿ36,7	236.797,1	ÿ49	287.774,9	ÿ24,9	346.244,9	ÿ49,3	ÿ61,5	209.372,1	ÿ54,9	240.944,9	267.814,2	ÿ44,1	ÿ42,3
34-1000-PL	536.199,8	ÿ24,7	492.753,1	ÿ30,8	534.689,9	124.538,1	ÿ34,2		ÿ51,4	381.116,4	ÿ46,5	397.930	430.894,4	ÿ56,7	ÿ39,5
35-1000-RS	264.295,2	ÿ46,2	192.030,8	ÿ60,9	249.078,6	297.212,8			ÿ74,6	181.237	ÿ63,1	212.979,1	232.298,7	ÿ48,2	ÿ52,7
36-1000-RL	458.126,7	ÿ35	407.703,7	ÿ42,1	463.915,7				ÿ57,8	335.935,6	ÿ52,3	364.643,1	380.268,9		ÿ46
Durchschnitt-1000	276.004,15	ÿ40,7	243.492,8	ÿ47,7	280.802,8	ÿ39,7	165.671,2		ÿ64,4	196.807,5	ÿ57,7	217.793,3	ÿ53,2	237.766,7	ÿ48,9

Machine Translated by Google

Der beste bekannte Gewinn wurde in [9] berechnet , **indem jede Instanz 20 Stunden lang ausgeführt wurde.** Dann wurde der beste Gewinn als bester bekannter Gewinn gespeichert. Wie aus den Tabellen **6 und 7 ersichtlich , schneiden unsere vorgeschlagenen Algorithmen in Bezug auf den Gewinn bei kleinen Instanzen gut ab, wobei ABC(S1) sechs neue beste Lösungen und zwei weitere Lösungen lieferte, die den besten bekannten Lösungen entsprachen. ABC(S2) erzielte ebenfalls drei neue beste Lösungen und drei Lösungen, die der besten bekannten Lösung entsprachen. Außerdem erzielte das modifizierte K-means-GRASP(C)-ABC(C)S1 zwei neue beste Lösungen. Zudem lieferten K-means-GRASP(C)-ABC(C)S1, K-means-GRASP(C)-ABC(C)S2, das modifizierte K-means-GRASP(C)-ABC(C)S2 und ACO-GRA- SP(C)-ABC(C)S2 jeweils eine neue beste Lösung. Insgesamt war die allgemeine Leistung aller vorgeschlagenen Algorithmen akzeptabel und lag hinsichtlich des durchschnittlichen Gewinns dieser Instanzen nahe an den vorherigen**

Bei den mittelgroßen und großen Instanzen schnitten unsere vorgeschlagenen Algorithmen im Vergleich zu früheren Methoden jedoch nicht gut ab, wie der große durchschnittliche Lückenprozentsatz im Vergleich zu den bekanntesten Ergebnissen in Tabelle 6 zeigt. Dafür gibt es viele mögliche Gründe :

- **Auch hier scheinen die in der Scout-Phase des ABC verwendeten Methoden Auswirkungen auf die Leistung der vorgeschlagenen Algorithmen zu haben. Die erste ist die gierige randomisierte Strategie, die ein zufälliges Kundenpaar auswählt, was bedeutet, dass das schlechteste Paar (entweder mit geringem Umsatz oder durch eine große Entfernung getrennt) ausgewählt werden kann, was zu Darüber hinaus können die Nachbarschaftsoperatoren in manchen Fällen die Qualität der Lösung aufgrund von Zeitüberschreitungen oder Lastbeschränkungen nicht ausreichend verbessern. Die zweite Methode verwendet GRASP, das Lösungen durch zufällige Auswahl aus der Restricted Candidate List (RCL) erstellt, die halb so groß ist wie das Candidate Solution Set (CSS). Daher ist die Auswahl nicht besuchter Kunden bei kleinen Instanzen kein Problem,** da die Anzahl der Kundenpaare gering ist, was bedeutet, dass die Chance , **das beste Kundenpaar auszuwählen** (d. h. das mit dem höchsten IR2-Wert), hoch ist. Im Gegensatz dazu sinkt bei **mittelgroßen und großen Instanzen, bei denen die Größe der RCL groß ist, die Chance, das beste Kundenpaar auszuwählen, was sich auf die Qualität der Lösungen auswirkt.**
- **Die Anzahl der Lösungen in einer Population, die Anzahl der Iterationen und der Grenzwert spielen eine wichtige Rolle für die Leistung der Algorithmen. In unseren Algorithmen verlassen wir uns auf kleine Zahlen von Lösungen und Iterationen, da eine schnelle Verarbeitungszeit eine kritische Anforderung für tägliche M-Commerce-Apps ist. Diese verringerten Iterationszahlen und Populationsgrößen verringern jedoch die Chancen, eine zufriedenstellende Lösungsqualität zu erreichen .**

- **Alle vorherigen Algorithmen basierten auf VNS, das die Möglichkeit hat, die Größe der Nachbarschaftssuche jedes Mal zu ändern, wenn die Lösungsqualität nicht verbessert wird. Die systematische Erweiterung der Nachbarschaftsstruktur hilft VNS dabei, die Nachbarschaft vollständig zu erkunden und gute Lösungen zu erhalten. Im Gegensatz dazu basiert die Diversifizierungsfunktion in unserem Ansatz entweder auf der Greedy-Funktion (Einfügen des Kundenpaars mit dem höchsten IR in die beste Position) oder auf einer Kombination aus Greedy- und Randomisierungsfunktionen (Einfügen eines zufälligen Kundenpaars in die beste Position oder Einfügen eines zufälligen Kundenpaars aus RCL in die beste Position). Diese Strategien scheinen jedoch bei der Diversifizierung weniger erfolgreich zu sein als VNS, das auf einem organisierten Nachbarschaftserkundungsmechanismus beruht.**

Abbildung 3 vergleicht unsere vorgeschlagenen Algorithmen auf der Grundlage des Durchschnitts der besten Lösungen für jede Kategorie von Instanzen (hervorgehobene Zeilen in Tabelle 6). Es ist klar, dass ABC in beiden Versionen (S1 und S2) bei Instanzen mit 20, 50, 100, 250 und 500 Kunden immer noch allen anderen Algorithmen vorzuziehen ist. Im Gegensatz dazu übertreffen beide Versionen von K-means-GRASP(C)-ABC(C) und modifiziertem K-means-GRASP(C)-ABC(C) GRASP(V2)-ABC in beiden Versionen (S1 und S2) bei Instanzen mit 1000 Kunden. Die Hauptgründe für die schlechte Leistung von Clustering-Algorithmen in kleinen und mittelgroßen Instanzen sind die Verteilung der Kundenpaare in K-means und modifiziertem K-means, wobei der Clustering-Prozess dazu führen kann, dass eine kleine Anzahl von Kunden in einem Cluster platziert wird, was bedeutet, dass das Fahrzeug noch Restzeit zum Bedienen hat, aber keine weiteren Kunden vorhanden sind.

Machine Translated by Google

innerhalb dieses Clusters. Daher kommen wir zu dem Schluss, dass die Clustering-Algorithmen (außer ACO) nur bei großen Instanzen gute Ergebnisse liefern.

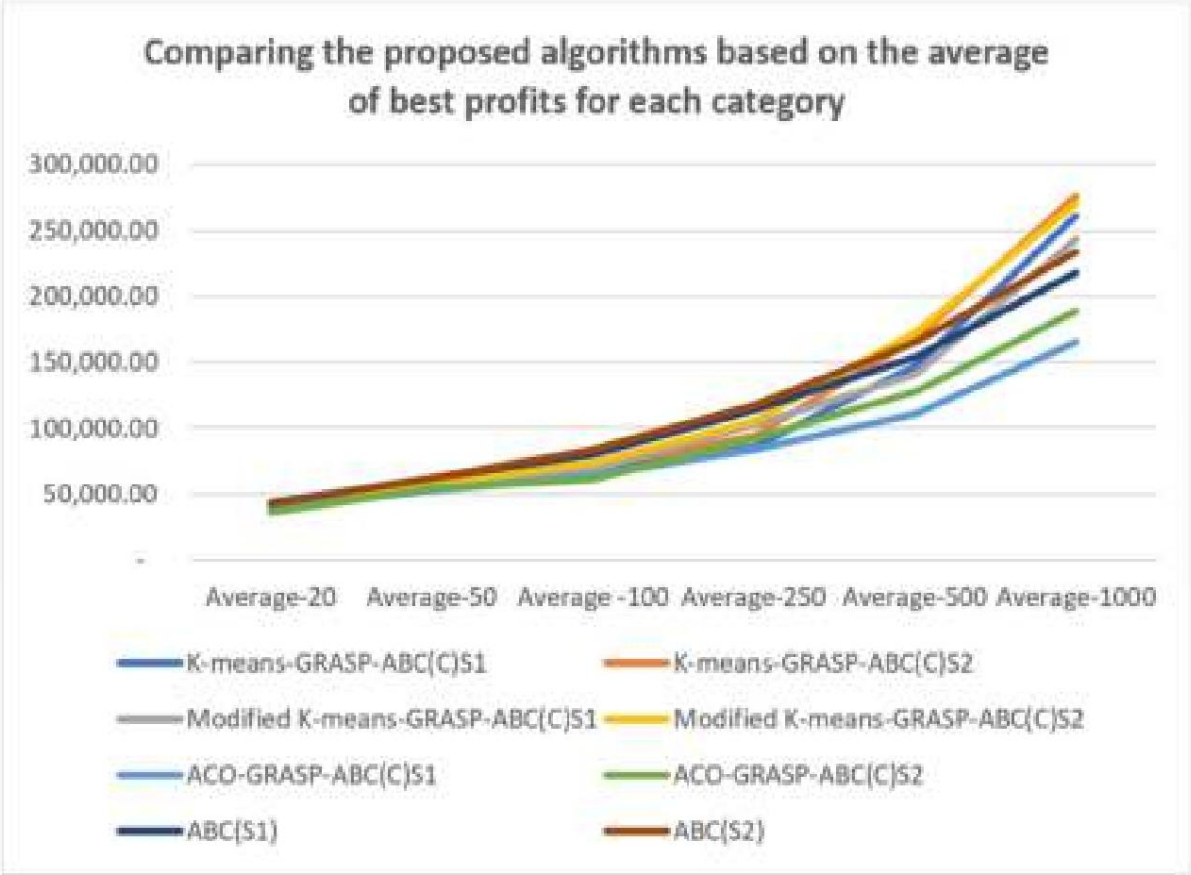


Abbildung 3. Vergleich der vorgeschlagenen Algorithmen basierend auf dem Durchschnitt der besten Gewinne für jede Kategorie.

In Bezug auf ACO-GRASP(C)-ABC(C)(S1 und S2) ist klar, dass es von allen Algorithmen die schlechteste Leistung aufweist. Auch hier hat die Clustermethode von ACO einen erheblichen Einfluss auf die Ergebnisse. Insbesondere trägt die Zufälligkeitsfunktion von ACO zu einer ungenauen Clusterung von Kundenpaaren bei.

Darüber hinaus haben wir, wie aus den hervorgehobenen Zeilen in Tabelle 7 hervorgeht, festgestellt, dass die zweite Strategie für Kundschafterbienen in vielen Fällen besser abschneidet als die erste, da bei der ersten Strategie entweder die besten Kundenpaare oder zufällige Paare ausgewählt werden und die Option der zufälligen Auswahl die Qualität der Lösung beeinträchtigen kann, während bei der zweiten Strategie die Kundenpaare zufällig aus der RCL ausgewählt werden, die die bes

6. Schlussfolgerungen

In diesem Artikel präsentieren wir acht neue Metaheuristiken zur Verbesserung der anfänglichen Lösungen für das Problem der rentablen Abholung und Zustellung mehrerer Fahrzeuge (MVPPDP). Wir haben zwei Versionen des ABC (Artificial Bee Colony) zur Verbesserung der anfänglichen Lösungen vorgeschlagen: das ABC mit GRASP und das ABC (C) mit GRASP und Clustering, d. h. K-means-GRASP (C), modifiziertes K-means-GRASP (C) und ACO-GRASP (C). Wir haben auch zwei separate Strategien für die Kundschafterbienenphase. Die erste Strategie, S1, erstellt die Lösung durch zufällige Auswahl einer von zwei Strategien: gierige und zufällige Einfügung oder nur gierige Einfügung. Die zweite Strategie, S2, basiert auf dem GRASP. Somit haben wir acht verschiedene vorgeschlagene Methoden. Im Vergleich zu den modernsten Methoden in Bezug auf die Verarbeitungszeit sind alle acht Methoden in allen Fällen (mit Ausnahme der Fälle mit 1000 Kunden) vorzuziehen. Darüber hinaus zeigten alle vorgeschlagenen Algorithmen bei kleinen Instanzen gute Ergebnisse

Machine Translated by Google

neue beste Lösungen und fünf gleichwertige Lösungen für die besten bekannten Lösungen. Die Algorithmen hinken jedoch in mittelgroßen und großen Instanzen aufgrund der gierigen Randomisierungsstrategie und des GRASP, die in der Scout-Bee-Phase verwendet wurden, etwas hinterher. Darüber hinaus priorisieren unsere Algorithmen minimale Lösungen und Iterationen für schnelle Verarbeitungszeiten in täglichen M-Commerce-Apps, während die Reduzierung der Iterationsanzahl und Populationsgröße die Wahrscheinlichkeit verringert, eine gute Lösungsqualität zu erzie

Für zukünftige Forschungen werden weitere Nachbarschaftsoperatoren in der Scout-Bee-Phase sorgfältig untersucht, um sie innerhalb der Beschränkungen des MVPPDP anwendbar zu machen und so die Chancen auf die Entwicklung qualitativ hochwertiger Lösungen zu erhöhen. Darüber hinaus werden wir andere populationsbasierte Metaheuristiken ausprobieren, wie GAs, evolutionäre Algorithmen oder andere schwarmbasierte Metaheuristiken, die zu neuen Lösungen für das MVPPDP beitragen können.

Beiträge der Autoren: Konzeptualisierung, AIA; Methodik, AIA; formale Analyse, AIA; Datenkuratierung , AIA; Schreiben – Originalentwurf, AIA; Mittelbeschaffung, AIA; Betreuung, MIH; Schreiben – Überprüfung und Bearbeitung, MIH. Alle Autoren haben die veröffentlichte Version des **Manuskripts gelesen und stimmen ihr zu .**

Finanzierung: Diese Forschung erhielt keine externe Finanzierung.

Datenverfügbarkeitserklärung: http://prolog.univie.ac.at/research/MVPPDP/instances.zip (abgerufen am 17. Juli 2024).

Danksagungen: Die Forscher möchten dem Dekanat für Graduiertenstudien und wissenschaftliche Forschung der Qassim-Universität für die finanzielle Unterstützung (QU-APC-2024-9/1) danken.

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte vorliegen.

Verweise

1. **Curry, D. Einnahmen- und Nutzungsstatistiken für Food Delivery-Apps. 2024. Online verfügbar: https://www.businessofapps.com/data/ Essensliefer-App-Markt/ (abgerufen am 24. Januar 2024).**
2. **Khalid, H. Vorteile einer Essensliefer-App für Restaurants und Kunden. Online verfügbar: https://enatega.com/benefits-of-Essensliefer-App/ (abgerufen am 1. Februar 2024).**
3. **Team, DJ Top Food Delivery Apps in Saudi-Arabien 2023. Online verfügbar: https://www.digitalgravity.ae/blog/top-food- delivery-apps-in-saudi-arabia/ (abgerufen am 2. Januar 2024).**
4. **Gansterer, M.; Küçüktepe, M.; Hartl, RF Das profitable Abhol- und Lieferproblem mit mehreren Fahrzeugen.** OR Spectr. **2017,** 39, **303–319. [Querverweis]**
5. **Talbi, EG** Metaheuristik: Vom Entwurf zur Implementierung; **John Wiley & Sons: Hoboken, NJ, USA, 2009; ISBN 978-0-470-27858-1.**
6. **Alhujaylan, AI; Hosny, MI Ein GRASP-basierter Lösungskonstruktionsansatz für die rentable Abholung und Zustellung mehrerer Fahrzeuge Problem.** Int. J. Adv. Comput. Sci. Appl. **2019,** 10, **111–120. [CrossRef]**
7. **Alhujaylan, AI; Hosny, MI Hybride Clustering-Algorithmen mit GRASP zum Erstellen einer ersten Lösung für den MVPPDP.** Comput. Mater. Contin. **2020,** 62, **1025–1051. [CrossRef]**
8. **Archetti, C.; Speranza, MG; Vigo, D. Vehicle Routing Probleme mit Gewinnen. In** Vehicle Routing: Problems, Methods, and Applications, **2. Auflage; SIAM: Philadelphia, PA, USA, 2014; Kapitel 10; Seiten 273–297. ISBN 978-1-61197-358-7.**
9. **Toth, P.; Vigo, D.** Vehicle Routing: Probleme, Methoden und Anwendungen, **2. Auflage; Gesellschaft für industrielle und angewandte Mathematik: Philadelphia, PA, USA, 2015; ISBN 1611973589.**
10. **Küçüktepe, M. Ein allgemeiner Suchalgorithmus für variable Nachbarschaften für das profitable Abhol- und Lieferproblem mehrerer Fahrzeuge Universität Wien 2014. Online verfügbar: https://scholar.google.com/scholar?hl=en&as_sdt=0,5&q=A+General+Variable+ Nachbarschaftssuchalgorithmus für das profitable Abhol- und Lieferproblem mit mehreren Fahrzeugen.&btnG= (abgerufen am 22. Juni 2024).**
11. **Haddad, MN Eine effiziente Heuristik für Eins-zu-eins-Abhol- und Lieferprobleme. Dissertation, IC/UFF, Niterói, Brasilien, 2017.**
12. **Bruni, ME; Toan, DQ Das Problem der rentablen Abhol- und Lieferroutenplanung für mehrere Fahrzeuge mit unsicheren Reisezeiten.** Transp. Res. Procedia **2021,** 52, **509–516. [CrossRef]**
13. **Karaboga, D. Eine Idee basierend auf einem Honigbienenschwarm zur numerischen Optimierung. Technischer Bericht-tr06, Erciyes University, Fakultät für Ingenieurwissenschaften, Abteilung für Computertechnik. 2005. Online verfügbar: https://scholar.google.com/scholar?hl=en&as_sdt=0,5&q=Eine+auf+Honigbienenschwarm+basierende+Idee+zur+numerischen+Optimierung&btnG= (abgerufen am 22. Juni 2024).**
14. **Alazzawi, AK; Rais, HM; Basri, S.; Alsariera, YA PhABC: Eine hybride künstliche Bienenkolonie-Strategie für die paarweise Generierung von Testsuiten mit Unterstützung von Einschränkungen. In Proceedings der 2019 IEEE Student Conference on Research and Development (SCOReD), Bandar Seri Iskandar, Malaysia, 15.–17. Oktober 2019; S. 106–111, ISBN 9781728126135.**
15. **Szeto, WY; Wu, Y.; Ho, SC Ein künstlicher Bienenkolonie-Algorithmus für das Problem der kapazitätsabhängigen Fahrzeugführung.** Eur. J. Oper. Res. **2011,** 215, **126–135. [CrossRef]**

Machine Translated by Google

16. **Wu, B.; Shi, Z. Ein Clustering-Algorithmus basierend auf Schwarmintelligenz. In Proceedings der 2001 International Conferences on Info- Tech and Info-Net. Proceedings (Cat. No.01EX479), Peking, China, 29. Oktober–1. November 2001; S. 58–66. ISBN 0-7803-7010-4.**
17. **Yao, B.; Hu, P.; Zhang, M.; Wang, S. Künstlicher Bienenkolonie-Algorithmus mit Scan-Strategie für die periodische Fahrzeug-Routingplanung Problem.** Simulation **2013,** 89, **762–770. [CrossRef]**
18. **Karaboga, D.; Akay, B. Eine Untersuchung: Algorithmen zur Simulation der Intelligenz von Bienenschwärmen.** Artif. Intell. Rev. **2009,** 31, **61–85. [CrossRef]**
19. **Creutz, M. Mikrokanonische Monte-Carlo-Simulation.** Phys. Rev. Lett. **1983,** 50, **1411. [CrossRef]**
20. **Bansal, JC; Sharma, H.; Jadon, SS Künstlicher Bienenstock-Algorithmus: Eine Untersuchung.** Int. J. Adv. Intell. Paradig. **2013, 5, 123–159. [Querverweis]**

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultieren.

Machine Translated by Google

Algorithmen

Artikel

Preptimize: Automatisierung der Vorverarbeitung und Prognose von Zeitreihendaten

Mehak Usmani [1], Zulfikar Ali Memon [1], Adil Zulfikar [2] und Rizwan Qureshi 3,*

[1] Fast School of Computing, Nationale Universität für Computer und Neue Wissenschaften, Karachi 65200, Pakistan; k200984@nu.edu.pk (MU); zulfiqar.memon@nu.edu.pk (ZAM)
[2] Fakultät für Elektrotechnik, Nationale Universität für Computer und Neue Wissenschaften, Faisalabad Campus, Faisalabad 38000, Pakistan; adil.zulfiqar@nu.edu.pk
[3] Zentrum für Regenerative Medizin und Gesundheit, Hong Kong Institute of Science and Innovation, Chinesische Akademie der Wissenschaften, Science Park, Hong Kong SAR 999077, China
* Korrespondenz: engr.rizwanqureshi786@gmail.com

Zusammenfassung: Zeitreihenanalysen sind für geschäftliche und finanzielle Entscheidungen von entscheidender Bedeutung, insbesondere angesichts der zunehmenden Integration des Internets der Dinge (IoT). Die Nutzung von Zeitreihendaten für Prognosen erfordert jedoch eine umfassende Vorverarbeitung, um Herausforderungen wie fehlende Werte, Heteroskedastizität, Saisonalität, Ausreißer und Rauschen zu bewältigen. Für univariate und multivariate Zeitreihen, gaußsche und nicht-gaußsche Zeitreihen sowie stationäre und nicht-stationäre Zeitreihen sind unterschiedliche Ansätze erforderlich. Allein die Behandlung fehlender Daten ist komplex und erfordert für jeden Typ einzigartige Lösungen. Das Extrahieren statistischer Merkmale, das Identifizieren von Datenqualitätsproblemen und das Auswählen geeigneter Bereinigungs- und Prognosetechniken erfordert erheblichen Aufwand, Zeit und Fachwissen. Um diesen Prozess zu rationalisieren, schlagen wir eine automatisierte Strategie namens Preptimize vor, die statistische und maschinelle Lerntechniken integriert und Blaupausen für Vorhersagemodelle empfiehlt und die am besten geeigneten Ansätze für einen bestimmten Datensatz als ersten Schritt zur weiteren Analyse vorschlägt. Preptimize liest eine Stichprobe aus einem großen Datensatz und empfiehlt das auf Optimierung basierende Blaupausenmodell, wodurch es auch für Laien einfach zu verwenden ist. Die Ergebnisse verschiedener Experimente zeigten, dass Preptimize in mehreren Sektoren, darunter Aktienkurse, Kryptowährungen und Stromverbrauchsprognosen, entweder besser abschnitt oder eine vergleichbare Leistung wie Benchmarkmodelle aufwies . Dies zeigt die Wirksamkeit des Frameworks bei der Empfeh

Schlüsselwörter: Automatisierung; Optimierung; Zeitreihen; univariat und multivariat; statistische Techniken; maschinelles Lernen; Vorverarbeitung; Prognose; Interpolation; einfache und mehrfache Imputation; feh

check for updates

Zitat: Usmani, M.; Memon, ZA; Zulfiqar, A.; Qureshi, R. Vorbereiten: Automatisierung von Zeitreihendaten Vorverarbeitung und Prognose. Algorithmen **2024**, 17, **332. https://doi.org/10.3390/a17080332**

Wissenschaftlicher Herausgeber: Vittorio Maniezzo

Empfangen: 18. Juni 2024
Überarbeitet: 22. Juli 2024
Akzeptiert: 26. Juli 2024
Veröffentlicht: 1. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz. Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/ 4.0/).

1. Einleitung

Zeitreihendaten, gekennzeichnet durch aufeinanderfolgende Datenpunkte, die zu bestimmten Zeitstempeln aufgezeichnet werden, stellen besondere Herausforderungen in der Datenwissenschaft dar und erfordern spezielle Vorverarbeitungstechniken zur Gewährleistung der Datenqualität. Die Zeitreihenanalyse umfasst das Sammeln, Bereinigen , Vorverarbeiten und Prognostizieren von Daten, zugeschnitten auf verschiedene Arten von Zeitreihen, ob numerisch oder textlich, univariat oder multivariat, saisonal oder nicht saisonal. Bestehende Forschung konzentriert sich überwiegend auf univariate Zeitreihen, während multivariate Zeitreihen zusätzliche Komplexitäten wie Rauschen, fehlende Werte, Ausreißer und Saisonalität mit sich bringen. Zur Bewältigung dieser Herausforderungen wird ein automatisiertes Vorverarbeitungsframework vorgeschlagen, das verschiedene Techniken und maschinelle Lernalgorithmen integriert. Die Zeitreihenanalyse, die in den Bereichen IoT und Finanzen von entscheidender Bedeutung ist, erfordert aufgrund ihrer besonderen Merkmale und Herausforderungen eine umfassende Datenbereinigung und Vorverarbeitung . Das Verständnis der Eigenschaften der Daten, wie Stationarität und Verteilung, ist für die Auswahl geeigneter Vorverarbeitungstechniken von entscheidender Bedeutung. Darüber hinaus zeigen Deep-Learning-Algorithmen wie GANs und Auto-Encoder vielv

Machine Translated by Google

Die Suche nach dem genauesten Prognosemodell mit minimalen Fehlerraten in der numerischen Zeitreihenanalyse erfordert umfangreiche Experimente mit verschiedenen Techniken und deren Kombinationen. Außerdem sollte bekannt sein, welche Technik auf einen bestimmten Datentyp anwendbar ist [1]. Dieser Prozess ist jedoch zeit- und arbeitsaufwendig.

Um fehlende Werte effektiv zu behandeln , **werden unterschiedliche Imputationsmethoden basierend auf den Zeitreiheneigenschaften wie Verteilung und Saisonalität eingesetzt. Mit dem exponentiellen Wachstum von Zeitreihendaten, die von unterschiedlichen Quellen wie Börsen [2], Energieverbrauch [3], Gesundheitsüberwachungssystemen und Wetteraufzeichnungssystemen generiert werden, stehen Analysten vor einer zunehmenden Arbeitsbelastung. Zeitreihen, die durch sequenzielle Daten gekennzeichnet sind, die über Zeitintervalle aufgezeichnet werden, stellen einzigartige Herausforderungen dar, wie z. B. die Behandlung von Rauschen, Saisonalität und fehlenden Werten, die spezielle Vorverarbeitungstechniken erfordern, um die Datenqualität sicherzustellen und die Prognosegenauigkeit zu verbess**

Die Vorverarbeitung, ein entscheidender Schritt bei der Datenanalyse und beim maschinellen Lernen, umfasst eine Reihe von Schritten zur Verbesserung der Datenqualität und Vereinfachung des Modelltrainings. Die Vorverarbeitung ist jedoch zeitintensiv, weist keine vordefinierte Reihenfolge auf und ist in verschiedenen Bereichen unterschiedlich. Im Kontext der Zeitreihenanalyse wird die Vorverarbeitung aufgrund der Echtzeitnatur und der sequentiellen Abhängigkeiten der Daten komplexer. Zu den Herausforderungen gehört der Umgang mit fehlenden, doppelten und falschen Daten, einschließlich Ausreißern. Die vorhandene Literatur konzentriert sich hauptsächlich auf den Umgang mit fehlenden Daten und die

Ausreißer, die erheblich von der Mehrheit der Datenpunkte abweichen, können univariat oder multivariat sein, wobei letztere eine größere Komplexität darstellen. Die Erkennung multivariater Ausreißer erfordert aufgrund des mehrdimensionalen Merkmalsraums häufig maschinelle Lernansätze . Techniken zur Erkennung von Ausreißern umfassen statistische und masc Zu den statistischen Techniken gehören exponentielle Glättung, Z-Score, lineare Regressionsmodelle und Interquartilsabstand, während Techniken des maschinellen Lernens überwachte Methoden wie neuronale Netzwerke und Entscheidungsbäume sowie unüberwachte Methoden wie Clusteranalyse umfassen.

Darüber hinaus kommen **Distanzmessverfahren, informationstheoretische Methoden und Algorithmen wie KNN und COF zum Einsatz.**

Der Umgang mit fehlenden Werten in Zeitreihendaten umfasst entweder das Entfernen unvollständiger Datensätze oder das Imputieren fehlender Werte. Imputationstechniken für Zeitreihendaten, einschließlich Spline-Interpolation und gleitender Durchschnitte, unterscheiden sich aufgrund zeitlicher Eigenschaften von denen, die in Nicht-Zeitreihendaten verwendet werden . Herkömmliche Methoden reichen möglicherweise nicht aus, sodass fortgeschrittene Techniken wie Interpolation, Autoencoder und generative Adversarial Networks (GANs) zur Imputation und Ausreißererkennung erforderlich sind.

Bei der Analyse von Zeitreihendaten werden Eigenschaften wie Saisonalität und Trend berücksichtigt, und Zeitreihen werden anhand ihrer Eigenschaften in Typen eingeteilt. Zu den Herausforderungen bei Zeitreihendaten gehören neben Ausreißern und fehlenden Werten auch Konzeptdrift, Klassenungleichgewicht, Trendbehandlung sowie Probleme mit Saisonalität und Datenkonsistenz.

Laut [4] wird die Zeitreihenanalyse meist mit statistischen Methoden durchgeführt; jedoch gewinnen Techniken des maschinellen Lernens bei der Zeitreihenanalyse zunehmend an Bedeutung , da sie bei entsprechender Vorverarbeitung genauere Ergebnisse versprechen. Die Vorverarbeitung bleibt jedoch von entscheidender Bedeutung, um die Datenqualität zu verbessern und zuverlässige Vorhersag Techniken wie SCREEN [6] bieten einschränkungsbasierte Bereinigungsmethoden für Zeitreihendaten und verbessern so die Genauigkeit und Effizienz der Datenanalyse. Forschungsstudien haben verschiedene Ansätze zur Verbesserung von Vorverarbeitungstechniken für Zeitreihendaten untersucht. Eine Studie [7] nutzte Meta-Learning, um Vorverarbeitungsalternativen auszuwählen, und konzentrierte sich dabei auf die Auswirkungen der Transformation auf die Algorithmusleistung in Datensätzen wie Automobile, Kredit, Iris und Wahl Während Regressionsbäume als Meta-Lerner nur eine begrenzte Wirksamkeit zeigten, verbesserten Transformationen die Ergebnisse deutlich. Eine andere Studie [8] führte eine iterative Methode zur minimalen Reparatur und Parametervorhersage für die Anomaliereparatur ein und erreichte eine erhebliche Verbesserung der Ausführungsgeschwindigkeit über Datensätze hinweg, darunter GPS, Intel Lab Data und China Meteorological Data. In einer separaten Untersuchung [9] wurden Landsat-Zeitreihenbilder in Kombination mit Trajektorienanalysen eingesetzt, um Waldinstabilitäten in Myanmar zu erkennen. Diese Studie hob die Korrelation zwischen instabilen Gebieten und gefällten Bäumen hervor und zeigte

Machine Translated by Google

potenzielle Anwendungen für selektives Logging. Die Änderungserkennung in Zeitreihen wurde umfassend analysiert [10] und umfasste Vorverarbeitungstechniken wie atmosphärische Korrektur und Wolkenerkennung sowie beliebte automatisierte Vorverarbeitungsalgorithmen wie LEDAPS und Fmask. Die Erweiterung vollständig konvolutionaler Netzwerke mit LSTM-RNN wurde für die Klassifizierung von Zeitreihen vorgeschlagen [11] und zeigte eine verbesserte Leistung mit Aufmerksamkeitsmechanismen. Darüber hinaus befassten sich Studien [12] mit Problemen des Datenprofiling und schlugen visuell basierte Softwarelösungen für Datenqualitätsprüfungen und Techniken zur Erkennung von Ausreißern unter Verwendung von Methoden wie der k-Sliding-Window-Technik und neuronalen Netzwerken vor [13].

Eine Übersichtsarbeit [14] zeigte auf, dass weitere Forschung im Bereich der Bereinigung von Zeitreihendaten erforderlich ist. Dabei wurde auch die Bedeutung der Vorverarbeitung für die Maximierung des Datenpotenz Mögliche Bereiche für zukünftige Arbeiten sind die Analyse von Fehlerdatentypen, mehrdimensionale Bereinigungsalgorithmen für Zeitreihen, Anomalieerkennung und spezifische Bereinigungstechniken für Anwendungsdaten. Es wurden mehrere innovative Systeme vorgeschlagen, um die Herausforderungen bei der Bereinigung von Zeitreihendaten in verschiedenen Bereichen anzugehen. Cleanits [15] konzentriert sich beispielsweise auf die Bereinigung industrieller Zeitreihendaten und verwendet Bollinger-Bänder zur Anomalieerkennung und LSTM zur Behebung von Fehlern. Obwohl es in erster Linie für industrielle Daten entwickelt wurde, besteht Potenzial für die Anwendung in anderen Bereichen, für die Flexibilitätsverbesserungen erford

Im Energiesektor wurden Systeme zur Wissensextraktion unter Verwendung von Clustering-Techniken vorgeschlagen [16] , wobei k-Medoids vielversprechende Ergebnisse bei Energiezeitreihen zeigten. Auch die Medizin profitiert von der Analyse von Zeitreihendaten, was zur Entwicklung von Systemen wie Clairvoyance [17] führt, die schnelles Prototyping und die Optimierung klinischer Entscheidungssysteme ermöglichen. Hybridmodelle und -systeme wie AutoMTS wurden zur Bereinigung multivariater Zeitreihendaten eingeführt [18,19], was ihre Bedeutung für die Verbesserung der Datenqualität durch Imputation und Ausreißererkennung zeigt. AutoML und Frameworks wie DAEMON [20,21] konzentrieren sich auf die Automatisierung der Anpassung von maschinellen Lernmodellen und der Ausreißererkennung in Zeitreihendaten, erreichen eine hohe Genauigkeit und demonstrieren das Potenzial für Online-Anomalieerkennungsdienste [22].

Zeitreihendaten mit Saisonalität stellen einzigartige Herausforderungen dar, wie z. B. Multisaisonalität und Heteroskedastizität, die die Analyse und Vorhersage erschweren. Die Forschung [23,24] hat sich diesen Herausforderungen durch Vorverarbeitung der Daten, Erstellung von Modellen und Anpassung an Originalmessungen, gefolgt von Anomalieerkennung, gestellt. Weitere Studien zielen darauf ab, die lokale Trendvorhersage und Ausreißerkorrektur in der Zeitreihenanalyse zu verbessern. Nichtstationarität wurde in Studien behandelt, die sich auf die Bereinigung volatiler und nichtvolatiler Daten konzentrierten [25]. Techniken wie k-Nearest Neighbors und Sliding Windows wurden untersucht [26], wobei Sliding Windows insbesondere bei volatilen Daten wie Aktienmarktzeitreihen bessere Ergebnisse lieferten [27]. Es wurden Anstrengungen unternommen, die Datenqualität mithilfe von Indizes wie dem Attribute Value Quality Index (AVQI) und dem Structured Data Value Quality Index (SDVQI) zu messen , um einen Vergleich und eine Verbesserung von Datenqualitätsmaßen zu ermöglichen.

Optimierungsalgorithmen wie genetische Algorithmen (GAs) werden eingesetzt, um die Merkmalsauswahl in der Vorverarbeitung zu automatisieren und so die Effizienz von Regressionsmodellen zu verbessern [28]. Während bestehende Lösungen wie SCREEN, CLEANITS, DAEMON und Clairvoyance spezifische Herausforderungen der Vorverarbeitung angehen, besteht Bedarf an einem generischen Ansatz, der m Preptimize zeichnet sich durch eine integrierte Lösung für Vorverarbeitung und Modellempfehlung aus , die auch für Laien zugänglich ist und verschiedene Herausforderungen im Bereich Zeitreihen bewältigen kann. SCREEN und DAEMON konzentrieren sich auf Anomalieerkennung und -bereinigung, während CLEANITS auf Imputation spezialisiert ist. Clairvoyance ist in erster Linie ein Prognosetool, das fortgeschrittene Techniken des maschinellen Lernens verwendet. Die breite Anwendbarkeit und der benutzerfreundliche Ansatz von Preptimize unterscheiden es von diesen spezialisierten Tools. Verschiedene Open-Source-Bibliotheken und proprietäre SaaS-Produkte automatisieren die Generierung und Optimierung von Pipelines für maschinelles Lernen und rationalisieren den Vorverarbeitungsprozess [29,30].

Die zu prognostizierenden Zeitreihenwerte können nominal, kategorisch, textuell und numerisch sein. Die numerische Prognose ist ein Regressionsproblem, das mit verschiedenen Techniken wie gleitenden Durchschnitten, Autoregression und exponentieller Glättung sowie fortgeschrittenen Methoden wie ARIMA, SARIMA, VAR und LSTM gelöst werden kann [31]. Hybride Ansätze

Machine Translated by Google

Auch die Kombination statistischer und maschineller Lernmodelle ist üblich [32–34]. Allerdings gibt es aufgrund der unterschiedlichen Eigenschaften von Zeitreihen keine einzige Technik, die für alle Zeitreihen geeignet ist, was den Auswahlprozess zeitaufwändig und komplex macht. Um den Prozess der Analyse einer Zeitreihe zu rationalisieren, sollte ein Standardverfahren eingeführt werden, um Profilierung, Vorverarbeitung, optimierte Modellprognose, Bewertung und Auswahl des leistungsstärksten Modells mit minimaler Fehlerquote zu automatisieren.

Der wichtigste Beitrag dieser Arbeit ist die Einführung von Preptimize, einer automatisierten Strategie zur Analyse von Zeitreihendaten, die statistische und maschinelle Lerntechniken integriert, um optimale Modelle für die Analyse zu empfehlen. Durch das Lesen einer Stichprobe aus einem großen Datensatz schlägt Preptimize auf Optimierung basierende Blaupausenmodelle vor, einschließlich der besten Kombination aus Vorverarbeitungs- und Prognosetechniken, wodurch die Prognosegenauigkeit für verschiedene Arten von Zeitreihendaten verbessert wird. Preptimize rationalisiert den Prozess der Analyse sowohl univariater als auch multivariater Zeitreihen mit fehlenden Werten. Aufgrund der Komplexität der Zeitreihenanalyse unterstreicht diese Forschung die Bedeutung der Auswahl der richtigen Prognosetechnik und Vorverarbeitungsstrategie basierend auf der Art der Zeitreihendaten, um den Prozess der Analyse der Zeitreihendaten zu beschleunigen. Darüber hinaus demonstrierte diese Arbeit das Potenzial des Frameworks für eine breite Anwendbarkeit bei der Zeitreihenprognose und untermauerte seine Wirksamkeit bei der Empfehlung geeigneter Vorhersagemo

2. Materialien und Methoden

Basierend auf der Diskussion im vorherigen Abschnitt kann festgestellt werden, dass ein umfassendes Framework erforderlich ist, das Standardschritte während der Vorverarbeitung und Prognose von Zeitreihen definieren kann. Der folgende Abschnitt enthält eine ausführliche Erläuterung des vorgeschlagenen Frameworks namens Preptimize. Der Schwerpunkt von Preptimize liegt auf der Automatisierung des Prozesses der Zeitreihenanalyse, indem das erste Blueprint-Modell für weitere Analysen bereitgestellt wird. Das vorgeschlagene Framework wurde mit der Programmiersprache Python imple

Preptimize

Preptimize beginnt mit dem Lesen von Zeitreihendaten und dem Generieren eines umfassenden Profils , das Verteilungsanalyse, Saisonalität und Trenderkennung umfasst. Anschließend verarbeitet es die Daten vor, indem es Datenqualitätsprobleme wie fehlende Werte und Ausreißer mithilfe verschiedener Techniken behebt und die vorverarbeiteten Daten mithilfe mehrerer Prognosemodelle auswertet, um das Modell mit der besten Leistung zu ermitteln. Die Schritte sind auch in Abbildung 1 aufgeführt. Da die Gestaltung des Frameworks eine große Aufgabe ist, konzentriert sich dieses Dokument auf numerische Daten und Probleme mit fehlenden Daten. Ziel dieser Arbeit ist es, die meisten verfügbaren Lösungen zum Umgang mit fehlenden Daten in Zeitreihen zu erklären und die bestmögliche Lösung zu bieten. Das vorgeschlagene Framework wird anhand verschiedener Datensätze getestet, darunter Kryptowährungsdaten, Börsendaten und die von einem IoT-Netzwerk generierten

Bei der Implementierung von Preptimize kommen zahlreiche Techniken und Algorithmen zum Einsatz , wie in Tabelle 1 dargestellt. Die Regeln, denen Preptimize folgt, um mögliche Blaupausen für Vorhersagemodelle zu empfehlen, sind aus der vorhandenen Literatur abgeleitet. Daher ist die Organisation und Rationalisierung dieser Regeln und Erkenntnisse ein wesentlicher Beitrag dieser Arbeit. Der grundlegende Pseudocode des vorgeschlagenen Frameworks ist in Abbildung 2 dargestellt. Der folgende Abschnitt hebt nur die gängigsten Techniken hervor, die in verschiedenen Phasen des Prozesses

Tabelle 1. Details der bei der Implementierung von Preptimize verwendeten Techniken.

Bühne	Unterstufe	Details der verwendeten Techniken
Profilerstellung	Statistiken	Varianz, Standardabweichung, Minimum, Maximum, Median, Modus, fehlender Prozentsatz, Quartile, Stationarität, Verteilung, Kurtosis, Schiefe, Kardinalität, Korrelation
	Visualisierung	Verteilungsdiagramm, Streudiagramm, Liniendiagramm, Boxplot, Histogramm, Trend- und Saisonzerlegungsdiagramm, Korrelations-Heatmap, Diagramm mit gleitender Statistik, ACF, PACF

Machine Translated by Google

Tabelle 1. Fortsetzung

Bühne	Unterstufe	Details der verwendeten Techniken
Vorverarbeitungstechniken	Transformation	Transformation für Stationarität mittels Differenz, Log-Differenz, Exponentialfunktion, additiver Zerlegung, multiplikativer Zerlegung, gleitendem Mittelwert und Standardabweichung, ADF-Statistik
	Fehlen	Transformation für die Normalverteilung mit Box–Cox
	Streichung	MCAR, MAR oder MNAR
		Listenweise, paarweise
	Univariat	Gleitende Durchschnitte einschließlich SMA, Exponential, CMA
	Anrechnung	Vorwärtsbefüllung (LOCF), Rückwärtsbefüllung (NOCB)
		GAIN (basierend auf generativen kontradiktorischen Netzwerken)
	Interpolation	Linear, Akima, baryzentrisch, kubisch, kubischer Spline, aus Ableitungen, Index, Krogh, nächster, Pad, Pchip, Polynom, stückweises Polynom, quadratisch, slinear, Spline, Zeit, Null
	Multivariate Imputation	DataWig, Bayesian Ridge, Random Forest, K-Nachbarn, Entscheidungsbaum, Zusätzliche Bäume, KNNImputer, HotDeck
Prognosen	Univariat	Statistische Techniken einschließlich gleitender Durchschnitte (SMA, Exponential, CMA) und andere saisonale und nicht-saisonale Techniken einschließlich AR, ARIMA, SES, SARIMA, HWES
		MLP, SVR-RBF, SVR-POLY, Entscheidungsbaum, Random Forest, RNN, GRU, GRU-bidirektional, Vanilla LSTM, Stacked LSTM, Bidirektionales LSTM, CNN-LSTM
	Multivariat	Statistische Techniken einschließlich VAR, VMA, VARMA
		MLP, Entscheidungsbaum, Random Forest, RNN, GRU, GRU-bidirektional, Gestapeltes LSTM

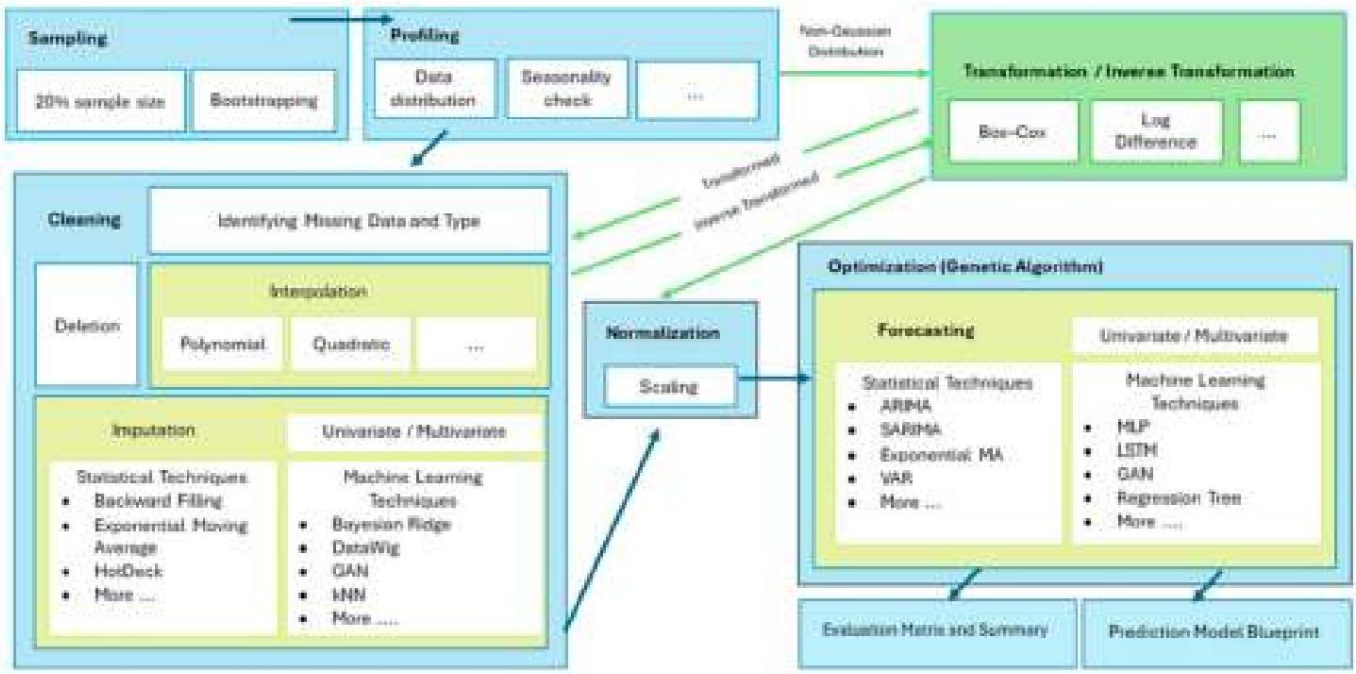


Abbildung 1. Preptimize: Visuelle Darstellung des systematischen Ansatzes, der die Phasen Vorverarbeitung, Modellauswahl und gründliche Auswertung integriert und so die Effizienz der Zeitreihenanalyse verbessert.

Machine Translated by Google


```
Datensatz initialisieren

Sichern Sie den Datensatz, um später Nullwerte verarbeiten zu können
# Profilieren Sie den Datensatz

Profil als Wörterbuch initialisieren

Wenn der Datensatz mehrere Spalten hat
    Profil als multivariat festlegen

    Paarweise Korrelation des Datensatzes drucken

    Heatmap der Korrelationen anzeigen
Anders
    Profil als univariat festlegen

Berechnen Sie die Gesamtzahl der Datensätze und den Typ der Zeitreihe (Univariat/Multivariat).

# Profil jeder Spalte vorbereiten
Für jede Spalte im Datensatz
    Initialisieren Sie dataAttribute als Spaltendaten

    Wenn univariat, legen Sie UnivariateColumnName fest.

    Auf fehlende Daten in der Spalte prüfen
    Berechnen von Nicht-Nullwerten und Nullwerten

    Wenn fehlende Daten vorhanden sind
        Fehlenden Prozentsatz berechnen

        Löschen von Nullwerten zur weiteren Analyse
    Berechnen Sie Schiefe und Kurtosis von dataAttribute

    Verteilungstyp bestimmen (Normal/Nicht Normal)

    Profil der Spalte mit berechneten Statistiken anzeigen

    Erstellen eines saisonalen Zerlegungsdiagramms

    Zeichnen Sie grundlegende Profildiagramme

    Profil für Spalte im Profil speichern
# Originaldatensatz aus Backup wiederherstellen

Initialisieren Sie missingData und missingColCorrelationCompleteCases

Bestimmen Sie den Fehlwerttyp und die Korrelation

Initialisieren Sie verarbeitete Datensätze
# Univariate Zeitreihen verarbeiten

Wenn Univariat

    Profil der univariaten Spalte abrufen

    Wenn fehlende Daten vorhanden sind
        Wenn der fehlende Typ MCAR ist und der fehlende Prozentsatz <= 10 %

            Listenweises Löschen anwenden

        Anders

            Wenn die Daten numerisch und nicht verzerrt sind

                Imputationstechniken anwenden

            Andernfalls, wenn die Daten nicht normal verteilt sind

                Box–Cox-Transformation anwenden
```

Abbildung 2. Fortsetzung

Machine Translated by Google


```
                Wenn die Transformation erfolgreich ist
                    Fügen Sie fehlende Werte mithilfe mehrerer Techniken hinzu und speichern Sie jeden Datensatz in verarbeiteten Datensätzen
                    Inverse Transformation von Daten mit Box–Cox
    Anders
        Verarbeitete Datensätze als Originaldatensatz festlegen
# Prognose univariater Zeitreihen
Initialisieren Sie die Auswertungsmatrix
Wenden Sie einen genetischen Algorithmus an, um das beste Vorhersagemodell basierend auf der höchsten Fitness oder dem niedrigsten RMSE auszuwählen
        Generieren Sie verschiedene potenzielle Lösungen, einschließlich vorverarbeiteter Datensätze und univariater Prognosen
        Techniken und Parametereinstellungen
        Wählen Sie mithilfe der Turnierauswahl die beste Lösung aus der Population aus
    Wenn fehlende Werte im vorverarbeiteten Datensatz verbleiben, Zeilen mit Nullen löschen
    Erstellen Sie Prognosen auf Grundlage ausgewählter Prognosetechniken und -parameter.
    Prognostizierte Daten speichern
    Bewerten Sie das Modell, indem Sie Fehlermetriken berechnen und die Bewertungsmatrix aktualisieren.
Weiter bis n Generationen
Zeigen Sie die Blaupause der fünf leistungsstärksten Vorhersagemodelle
# Multivariate Zeitreihen verarbeiten
Sonst wenn multivariat
    Wenn fehlende Daten vorhanden sind
        MissingDatasetPerc berechnen
        Wenn missingDatasetPerc <= 10 %
            Listenweises Löschen anwenden
        Anders
            Überprüfen Sie die Normalität für alle Spalten
            Wenn alle Spalten normalverteilt sind
                Paarweises Löschen anwenden
                Wenn in den Spalten immer noch Daten fehlen
                    Einfache und multiple Imputation anwenden
            Anders
                Box–Cox-Transformation für nicht-normale Spalten anwenden
                Einfache und multiple Imputation anwenden
                Inverse Transformation von Daten mit Box–Cox
    Anders
        Verarbeitete Datensätze als Originaldatensatz festlegen
    # Prognose multivariater Zeitreihen
    Initialisieren Sie die Auswertungsmatrix
    Wenden Sie einen genetischen Algorithmus an, um das beste Vorhersagemodell basierend auf der höchsten Fitness oder dem niedrigsten RMSE auszuwählen
        Generieren Sie verschiedene potenzielle Lösungen, einschließlich vorverarbeiteter Datensätze und multivariater Prognosen
        Techniken und Parametereinstellungen
        Wählen Sie mithilfe der Turnierauswahl die beste Lösung aus der Population aus
```

Abbildung 2. Fortsetzung

Machine Translated by Google

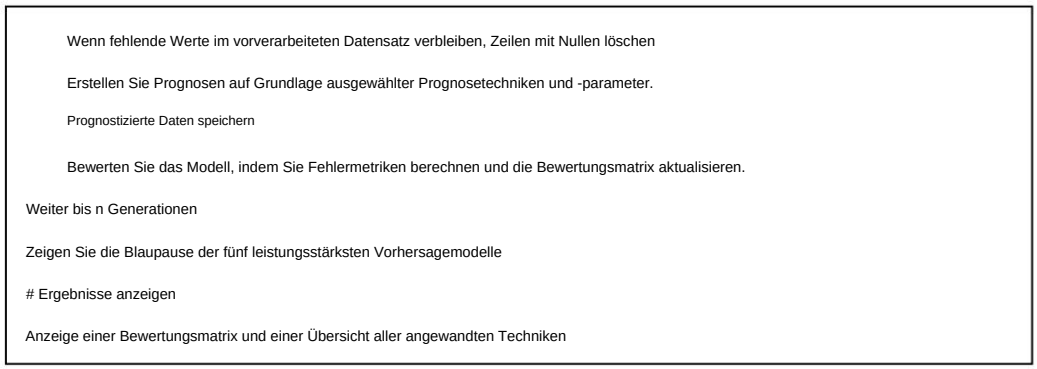


Abbildung 2. Preptimize: Sequentieller Workflow zur Vorverarbeitung von Zeitreihendaten und Modellauswahl. Hier steht # für den Kommentar.

2.1. Datenstichprobe

Im ersten Schritt generiert Preptimize eine Stichprobe aus den Daten, um das ursprüngliche Verhalten darzustellen und dabei die Größe auf 20 % zu reduzieren. Bootstrapping wurde angewendet und der Index wurde beibehalten.

2.2. Datenprofilierung

In dieser Phase werden verschiedene Eigenschaften der Daten berechnet und ein Profil für das Attribut erstellt . Diese Eigenschaften umfassen Mittelwert, Median, Minimal- und Maximalwert, Varianz und Standardabweichung der Daten. Die Daten werden weiter untersucht, indem der Verteilungstyp , die Schiefe und die Kurtosis der Daten ermittelt werden. Die Stationarität der Daten wird ebenfalls berechnet, was Saisonalität und Trend einschließt. Eine kurze Beschreibung dieser Techniken finden Sie weiter unten.

2.2.1. Datenverteilung

Die Datenverteilung gibt alle wahrscheinlichen Werte für eine Variable an und quantifiziert auch die relative Häufigkeit, die angibt, wie häufig ein Wert auftritt. Es sind verschiedene Verteilungen möglich, darunter Normal-, Binomial-, Poisson- oder Bernoulli- Verteilung. Die meisten Situationen folgen der Normalverteilung, die als glockenförmige und symmetrische Kurve um die Linie x = µ ausgedrückt wird. Sie zeigt an , **dass auf beiden Seiten des Zentrums gleich viele** Werte vorhanden sind. Die Gauß-V **von 0 und einer Standardabweichung von 1 ist die Standard-Gauß-Verteilung.**

Die beiden mit der Verteilung verbundenen Eigenschaften sind Schiefe und Kurtosis. Die Schiefe ist ein Maß für die Regelmäßigkeit. Die Datenreihe ist symmetrisch, wenn sie auf beiden Seiten des Mittelpunkts gleich aussieht. Die Kurtosis misst die Häufigkeit von Ausreißern am Ende. Daten mit dicken Enden weisen auf mehr Ausreißer hin und der Kurtosis-Wert ist hoch. Die Kurtosis einer Normalverteilung beträgt 3. Ist sie kleiner als 3, gibt es wenige Ausreißer, ist sie größer als 3, gibt es mehr Ausreißer. Die Schiefe kann negativ sein, was bedeutet, dass das Ende auf der linken Seite ist, andernfalls ist das Ende auf der rechten Seite. Ist die Schiefe null, ist die Verteilung vollkommen symmetrisch. Die Verteilung kann als normal betrachtet werden, wenn die Schiefewerte zwischen ÿ3 und +3 liegen und die Kurtosis bei Verwendung des **Standardmessfehlers in einem Bereich von ÿ10 bis +10** lieg

2.2.2. Stationarität

Eine Zeitreihe, deren Eigenschaften nicht der Zeit unterliegen, ist stationär. Das Vorhandensein eines Trends oder einer Saisonalität zeigt also an, dass die Reihe nicht stationär ist. Die zu verschiedenen Zeiten beobachtete Eigenschaft wird durch Trend und Saisonalität beeinflusst. Eine Zeitreihe hat drei Komponenten: Saison, Trend und Rauschen. Die Saison kann als der sich wiederholende kurzfristige Zyklus in der Reihe definiert werden. Der Trend ist die Steigung in den Daten, die mit der Zeit steigen oder fallen kann, während das Rauschen die zufällige Änderung in der Reihe ist. Der Umgang mit Zeitreihendaten, die Saisons oder Trends enthalten, wird oft als schwierig angesehen . Das Aufzeichnen der Zeitreihe gibt sofortige Informationen über das Vorhandensein einer Saison und eines Trends.

Machine Translated by Google

Es gibt zwei gängige Ansätze zum Umgang mit saisonalen Zeitreihen. Eine Methode besteht darin, die Reihe aufzuteilen, indem die saisonale und die Trendkomponente entfernt werden. Anschließend kann die Reihe zur Analyse verwendet werden oder eine andere Transformationstechnik verwendet werden, um die nicht saisonale Darstellung der Reihe zu erhalten. Die andere Möglichkeit besteht darin, speziell entwickelte Prognosetechniken für saisonale Reihen wie SARIMA zu verwenden.

2.2.3. Identifizieren der Art fehlender Daten

Beim Umgang mit fehlenden Daten werden drei Kategorien unterschieden: völlig zufällig fehlende Daten (MCAR), zufällig fehlende Daten (MAR) und nicht zufällig fehlende Daten (MNAR). Bei der univariaten Analyse führt die Behandlung der Zeit als Variable zu einer höheren Komplexität, da Zeitreihenmerkmale und Kovarianz für die Imputation berücksichtigt werden müssen. Bei MCAR-Daten ist die univariate Imputation relativ unkompliziert. Bei der multivariaten Analyse wird jedoch die Assoziation zwischen Spalten berücksichtigt. MCAR-Daten fehlen zufällig und können oft durch Mittel- oder Medianwerte ersetzt werden, da die fehlenden Daten als wiederherstellbar gelten. MCAR wird durch (1) angegeben, wobei R **fehlende Daten und** Y **die vollständigen Daten sind.**

$$P(Ri|Yi) = P(Ri) \quad \textbf{(1)}$$

MCAR-Daten weisen nach Maßgaben wie Chi-Quadrat, Cramers Regel oder Korrelationskoeffizient keine signifikante Assoziation auf und enthalten typischerweise weniger Nullwerte. Bei großen Datensätzen können Instanzen mit fehlenden Daten weggelassen oder imputiert werden. Zu den Löschmethoden für MCAR gehören das listenweise Löschen, das Entfernen von Datensätzen mit einem beliebigen Nullwert, das geeignet ist, wenn die Anzahl fehlender Daten minimal ist, und das paarweise Löschen, das Beibehalten von Daten für Mengenpaare mit verfügbaren Datenpunkten, wodurch mehr Daten erhalten bleiben. Imputation wird notwendig, wenn Löschmethoden ungeeignet sind. Bei der Imputation werden fehlende Daten durch erwartete Werte ersetzt, wobei je nach Art der Daten verschiedene Techniken angewendet werden. Im Fall von MAR können fehlende Daten einem Muster oder einer Sequenz folgen und gelegentlich außerhalb der Sequenz erscheinen. Wenn eine Beziehung zwischen fehlenden und vorhandenen Daten identifiziert wird, wird dies als MAR betrachtet. Imputationstechniken wie Mittelwert, Median, Modus oder Interpolation können zur Handhabung von MAR eingesetzt werden, wobei mehrere Imputationstechniken für Nicht-Zeitreihendaten und Interpolation für Zeitreihendaten verfügbar sind. MAR ist gegeben durch (2)

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

Bei MNAR haben bestimmte Instanzen fehlende Daten gemeinsam, was die Vorhersage schwierig macht, da sie auf unbeobachteten Informationen beruht. Wenn fehlende Werte in einer Spalte wenig oder keine Korrelation mit anderen Spalten aufweisen und bei der Untersuchung keine starken Assoziationen gefunden werden und die Häufigkeit fehlender Werte hoch ist, wird dies als MNAR betrachtet. MNAR kann nicht ignoriert werden und es gibt keine direkte Methode zur Behandlung außer dem Abrufen der fehlenden Daten. MNAR hat im Vergleich zu MCAR typischerweise mehr Nullwerte und kann nur durch Domänenwissen identifiziert werden. MNAR wird durch

$$P(Ri|Yi) \neq P(Ri|Yi,o) \quad \textbf{(3)}$$

2.3. Datenvisualisierung

Die Verwendung verschiedener Diagramme und Plots ist ein wertvoller Ansatz in der Zeitreihenanalyse zum Verständnis des Datenverhaltens und der Dateneigenschaften. Preptimize generiert eine Reihe von Visualisierungen, wie Streudiagramme, Liniendiagramme und Zerlegungsdiagramme, um einen umfassenden Überblick über die Zeitreihendaten zu bieten. Diese visuellen Tools helfen beim Verständnis des Datenverhaltens, beim Erkennen von Trends und Saisonalität und bei der Auswahl geeigneter Vorverarbeitungstechniken.

Der erste Satz von Diagrammen bietet eine übergreifende Perspektive der Daten und umfasst Streudiagramme , Liniendiagramme, Boxplots und Histogramme. Jedes Attribut in einer multivariaten Zeitreihe wird einzeln dargestellt, um das Wesentliche der Daten zu erfassen.

Ein weiteres wichtiges Diagramm zeigt die Datenverteilung und erleichtert die Beurteilung von Normalität. Diese Auswertung ist für die Bestimmung geeigneter statistischer Techniken von wesentlicher Bedeutung.

Machine Translated by Google

Eine Grafik, die die Saison- und Trendzerlegung veranschaulicht, ist besonders hilfreich, um saisonale Muster innerhalb der Zeitreihe zu erkennen. Diese Erkenntnisse leiten die Anwendung geeigneter Techniken, die auf saisonale Daten zugeschnitten sind, und erfordern möglicherweise Transformationen für eine verbesserte Analyse.

Um die Stationarität der Daten zu beurteilen, generiert Preptimize ein Diagramm, das den gleitenden Mittelwert und die gleitende Standardabweichung gegenüber den Originaldaten aufträgt. Stationarität wird angezeigt, wenn diese Werte im Laufe der Zeit konstant bleiben [36].

Die Heatmap-Visualisierung stellt Korrelationen zwischen mehreren Variablen dar, wobei höhere Korrelationsstufen durch unterschiedliche Farben dargestellt werden. Dies hilft dabei, die Stärke der Beziehungen zwischen Variablen zu verstehen.

Zusätzliche Grafiken konzentrieren sich auf fehlende Daten, einschließlich einer Matrix und eines Balkendiagramms, die Einblicke in die Art der fehlenden Daten und ihre möglichen Beziehungen zu anderen Attributen bieten.

2.4. Datenbereinigung

Die Vorverarbeitungsphase bei der Zeitreihenanalyse umfasst die Datenbereinigung, bei der verschiedene Datenqualitätsprobleme wie fehlende Werte, Ausreißer und Rauschen behoben werden. Im Gegensatz zu anderen Formen der Datenanalyse erfordert die Vorverarbeitung von Zeitreihen spezifische Techniken, die auf ihre einzigartigen Eigenschaften zugeschnitten sind. Preptimize verwendet eine Reihe von Imputationstechniken, um fehlende Daten zu verarbeiten, einschließlich, aber nicht beschränkt auf lineare Interpolation, Spline-Interpolation und maschinelle Lernmethoden wie GANs. Die Wahl der Imputationsmethode basiert auf den spezifischen Eigenschaften des Datensatzes und gewährleistet die bestmögliche Datenqualität für die Pr Mit fast 25 verschiedenen verfügbaren Interpolationstechniken zielt das Framework darauf ab, den am besten geeigneten Ansatz für jeden Datensatz zu finden. Transformationstechniken werden angewendet, wenn Daten von der Normalität abweichen oder Nichtstationarität aufweisen. Die Box-Cox-Transformation wird verwendet, um Nichtnormalität zu beheben, während Methoden wie die saisonale Zerlegung eingesetzt werden, um Trends und Saisonalität aus nichtstationären Daten zu entfernen.

Machine-Learning-Ansätze, darunter Generative Adversarial Networks (GANs), sind in das Framework integriert, um die Vorhersage fehlender Daten zu verbessern. Für multivariate Zeitreihen werden iterative Imputationsmodelle und DataWig eingesetzt, um die Herausforderungen fehlender Daten zu bewältigen.

Um die optimale Imputationstechnik zu bestimmen, wird eine umfassende Bewertung durchgeführt , bei der Prognoseleistung und Bewertungsmetriken berücksichtigt werden. Alternativ kann ein Brute-Force-Ansatz eingesetzt werden, um alle möglichen Imputationstechniken zu testen.

Insgesamt spielt die Vorverarbeitungsphase eine entscheidende Rolle bei der Vorbereitung von Zeitreihendaten für die Analyse und Prognose und erfordert spezielle Techniken, die auf die individuellen Merkmale und Herausforderungen dieser Daten zugeschnitten sind.

2.4.1. Vorwärts- und Rückwärtsbefüllung

Last Observation Carried Forward (LOCF) ist eine weit verbreitete Einzelimputationstechnik , die den letzten gemessenen Wert durch Vorwärtsauffüllen imputiert. In LOCF muss mindestens ein Post-Baseline-Messwert vorhanden sein. Next Observation Carried Backward (NOCB) ist eine analoge Methode zu LOCF, funktioniert aber in umgekehrter Richtung durch Rückwärtsauffüllen; das heißt, es verwendet den ersten aufgezeichneten Wert nach dem nicht verfügbaren Wert und überträgt ihn rückwärts.

2.4.2. Interpolation

Interpolationstechniken bieten eine Möglichkeit, fehlende Datenpunkte vorherzusagen, indem sie Werte zwischen bekannten Datenpunkten schätzen. Die lineare Interpolation, die einfachste Methode, verbindet benachbarte Punkte mit geraden Linien, wobei angenommen wird, dass die Punkte auf der Linie die fehlenden Werte darstellen. Die vielseitigere polynomische Interpolation verwendet Polynomfunktionen unterschiedlicher Ordnung, um Beziehungen zwischen Attributen darzustellen. Beispiele für polynomische Interpolationsmethoden sind die Lagrange-Polynom-Interpolation und die Newton-Polynom-Interpolation . Die Lagrange-Interpolation ist zwar vielseitig, kann aber bei größeren Datensätzen Probleme bereiten.

Machine Translated by Google

Die Newton-Polynominterpolation verwendet einen rekursiven Divisionsprozess, um komplexe Polynome zu konstruieren. Dies kann für die Vorhersage effektiv sein, kann aber zu numerischen Komplexitäten und einer Empfindlichkeit gegenüber Ausreißern führen. Die Spline-Interpolation hingegen verwendet stückweise Polynome, um glattere Kurven an kleinere Gruppen von Datenpunkten anzupassen. Die kubische Spline-Interpolation wird aufgrund ihrer Glätte und Stabilität häufig bevorzugt.

Zu den weiteren im Framework verwendeten Interpolationsmethoden gehören stückweise polynomische, Akima-, baryzentrische, Krogh- und Pchip-Interpolationen, die jeweils unterschiedliche Ansätze zur Schätzung fehlender Werte in Zeitreihendaten bieten.

2.4.3. Gleitender

Durchschnitt Gleitende Durchschnitte wie einfache gleitende Durchschnitte (SMAs), kumulative gleitende Durchschnitte (CMAs) und exponentielle gleitende Durchschnitte (EMAs) werden häufig zur Imputation in der Zeitreihenanalyse verwendet. SMA berechnet den Durchschnitt einer bestimmten Anzahl von Datenpunkten mithilfe einer einfachen Summationsformel. CMA berechnet den Durchschnitt für alle Einträge bis zum aktuellen Zeitpunkt, wobei jeder Wert gleich gewichtet wird. EMA verwendet exponentiell abnehmende Gewichtungsfaktoren zur Berechnung des Durchschnitts, wobei der Wert zu jedem Zeitpunkt durch einen Koeffizienten und den vorherigen EMA-Wert bestimmt wird. Diese gleitenden Durchschnitte können bei der Datenimputation hilfreich sein, sofern die Zeitreihe stationär ist. Stationarität, die ein konsistentes Verhalten über die Zeit anzeigt, kann mithilfe statistischer Tests wie dem Augmented Dickey–Fuller (ADF)-Test **[37] oder durch Beobachtung gleitender Statistike**

2.4.4. Deep-Learning-Algorithmen Die

Datenimputation mit Deep-Learning-Techniken, insbesondere rekurrierenden neuronalen Netzwerken (RNNs) und generativen kontradiktorischen Netzwerken (GANs), hat sich in der aktuellen Literatur als vielversprechender Ansatz für die Zeitreihenanalyse herausgestellt. RNNs und ihre Variationen eignen sich hervorragend für die Modellierung von Sequenzdaten, während GANs gut darin sind, fehle Preptimize nutzt ein GAN für Einzel- und Mehrfachimputationen. Das GAN arbeitet mit zwei Modellen, dem Generator (G) und dem Diskriminator (D), die zusammenarbeiten, um realistische Datenproben zu generieren und zwischen echten und gefälschten Proben zu unterscheiden . Das Training wird so lange fortgesetzt, bis der Diskriminator nicht mehr zwischen echten und generierten Proben unterscheiden kann , was die Fähigkeit des Generators anzeigt, glaubwürdige Beispiele zu erzeugen.

In GANs versucht D , **den Stichproben mit maximaler Wahrscheinlichkeit das richtige Label zuzuordnen, während** G **versucht,** log (1 ÿ D(G(z))) zu minimieren. **Nach [38]** spielen D **und** G **das Zwei-Spieler-Minimax-Spiel mit der Wertfunktion** V (G, D), **wie in (4) gezeigt.**

$$\text{minGmaxD } V(D, G) = \text{Exÿpdaten (x) } \textbf{[log } D(x)] + \text{Ezÿpz (z) } \textbf{[log(1 ÿ D(G(Z)))]} \quad \textbf{(4)}$$

In Preptimize wird ein in **[39] vorgestelltes Modell zur Imputation fehlender Daten mithilfe eines GAN verwendet. Die Methode heißt Generative Adversarial Imputation Nets (GAIN). Bei der Methode** beobachtet G **reale Daten, imputiert das fehlende Element und gibt finalisierte Daten aus.**
D **nimmt dann ein finales Datum und versucht herauszufinden, welche Elemente beobachtet und welche zugeschrieben werden.**

2.4.5. Mehrfachimputation

Mehrfachimputationen (MIs) bieten einen umfassenden Ansatz zum Umgang mit fehlenden Werten, indem sie den vollständigen Satz von Merkmalen nutzen, um die fehlenden Werte zu approximieren. MI generiert mehrere Datensätze durch Imputation fehlender Werte und ermöglicht so den Ersatz fehlender Werte durch mehrere mögliche Werte. Diese imputierten Werte werden aus jedem Datensatz geschätzt und dann zu einem endgültigen Datensatz zusammengefasst. MI ist in Fällen anwendbar, in denen die Daten völlig zufällig fehlen (MCAR), wenn sie zufällig fehlen (MAR) und sogar, wenn die Daten nicht zufällig fehlen (MNAR).

Bei der iterativen Imputation hingegen wird jedes Merkmal als Funktion der anderen Merkmale in einem sequentiellen Prozess modelliert. Frühere imputierte Werte werden verwendet, um nachfolgende Merkmale vorherzusagen, bis alle fehlenden Werte ausgefüllt sind. Preptimize verwendet die Itera

Machine Translated by Google

Python, das verschiedene Schätzer wie den Bayesian Ridge-Regressor, den Random-Forest-Regressor, den K-Neighbors-Regressor, den Decision Tree-Regressor und den Extra Trees-Regressor bietet.

Darüber hinaus verwendet die KNNImputer-Technik, die in derselben Python-Bibliothek bereitgestellt wird, den k-Nearest-Neighbors-Ansatz, um fehlende Werte zu imputieren, indem Beobachtungswerte für benachbarte Datenpunkte berücksichtigt werden.

DataWig, eine weitere Imputationsmethode, hat sich in der Literatur als effektiv bei der Imputation fehlender Werte erwiesen. DataWig [40] ist eine Deep-Learning-Bibliothek, die MXNet **als Backend nutzt, um Vorhersagen für fehlende Daten zu generieren.**

Darüber hinaus **wird zur Imputation die für ihre Recheneffizienz bekannte Hot-Deck-** Methode **[41]** eingesetzt. Bei dieser Methode werden fehlende Werte imputiert, indem **innerhalb einer Datenmatrix ein zufällig ähnlicher Datensatz ausgewählt und vorhandene Werte aus dem Spenderobjekt verwendet werden. Das Hot-Deck operiert jeweils mit einem Attribut und dient als einzelne Imputationstechnik.**

2.5. Prognosen **In**

der Literatur gibt es verschiedene Ansätze zur Vorhersage von Zeitreihen.
Beispiele hierfür sind die exponentielle Glättung, der autoregressive integrierte gleitende Durchschnitt (ARIMA) mit saisonalen und nicht-saisonalen Varianten, mehrere Deep-Learning-Techniken und viele andere. Es gibt unterschiedliche Techniken für die Prognose univariater und multivariater Zeitreihen. Preptimize verwendet einen genetischen Algorithmus [42], um den riesigen Raum möglicher Modelle und Konfigurationen zu erkunden und das am besten geeignete für einen gegebenen Datensatz zu finden. Jedes Individuum in der Population stellt eine potenzielle Lösung dar, die aus einer Vorverarbeitungsstrategie, einem Prognosemodell und Parametereinstellungen besteht. Jedes Individuum wird anhand einer Fitnessfunktion bewertet , in d
Auf der Grundlage ihrer Fitnesswerte werden einzelne Personen als Eltern für die nächste Generation ausgewählt. Die Mutation erfolgt auf der Grundlage des Bit-Flip-Schemas; 3 Individuen werden per Turnierauswahl zufällig aus der Population ausgewählt. Dieser Ansatz hilft bei der Automatisierung des Modellauswahlprozesses und spart potenziell viel Zeit und Aufwand bei gleichzeitiger Erzielung einer optimalen oder nahezu optimalen Prognosegenauigkeit.

2.5.1. Univariate Prognose

Bei der univariaten Prognose werden zukünftige Werte einer einzelnen Variablen auf Grundlage historischer Daten aus aufeinanderfolgenden Zeitintervallen vorhergesagt. Die Wahl der Prognosetechnik hängt davon ab, ob die Zeitreihe Saisonalität und Trendkomponenten aufweist.

Es gibt zwei Haupttypen univariater Prognosemethoden: statistische Methoden und fortgeschrittene Methoden auf Basis maschinellen Lernens. Zu den statistischen Methoden gehören Autoregression (AR), gleitender Durchschnitt (MA), autoregressiver gleitender Durchschnitt (ARMA), autoregressiver integrierter gleitender Durchschnitt (ARIMA) und einfache exponentielle Glättung. Gleitende Durchschnittstechniken wie einfacher gleitender Durchschnitt, exponentieller gleitender Durchschnitt und kumulativer gleitender Durchschnitt sind üblich.

Zu den Optionen für Zeitreihen mit Saisonalität und Trend zählen der saisonale autoregressive integrierte gleitende Durchschnitt (SARIMA), der saisonale autoregressive integrierte gleitende Durchschnitt mit exogenen Regressoren und die exponentielle Glättung nach Holt-Winters (HWES), wobei ARIMA und SARIMA am erfolgreichsten sind.

ARIMA kombiniert Autoregression, gleitenden Durchschnitt und Differenzbildung, um Stationarität zu erreichen . Die korrekte Identifizierung der ARIMA-Parameter (p, d, q) ist entscheidend. Dies erfolgt normalerweise durch Anpassen von Modellen mit unterschiedlichen Ordnungen anhand von Kriterien wie dem Akaike-Informationskriterium (AIC), um die beste Ordnung auszuwählen. Das AIC misst die Modellleistung im Verhältnis zur Anzahl der Parameter. Die Autokorrelationsfunktion (ACF) und die partielle Autokorrelationsfunktion (PACF) können nicht verwendet werden, um zuverlässige Werte für p und q zu ermitteln [43]. Die Residuenanalyse ist bei der ARIMA-Modellierung unverzichtbar, um sicherzustellen, dass die Residuen weißem Rauschen ähneln und damit keine Autokorrelation aufweisen. Der Ljung-Box-Test wird häufig verwendet, um die Autokorrelation der Residuen zu bewerten, wobei p-Werte über e

Die Prognose von Zeitreihen hat im Bereich des maschinellen Lernens und Deep Learning große Aufmerksamkeit erregt. Preptimize berücksichtigt auch diese ML-basierten Prognosemethoden. Di

Machine Translated by Google

In der Literatur werden zu diesem Zweck verschiedene Techniken vorgestellt, darunter Long Short-Term Memory (LSTM), Support Vector Regression (SVR), das Multi-Layer Perceptron (MLP), Entscheidungsbäume, Random Forest, das Recurrent Neural Network (RNN) und seine Variante, die Gated Recurrent Unit (GRU).

Unter diesen Techniken sticht LSTM hervor, das für seine Effektivität bei der Verarbeitung sequentieller Daten bekannt ist. Traditionelle RNNs stehen jedoch vor Herausforderungen wie verschwindenden und explodierenden Gradienten, was zu längeren Trainingszeiten und geringerer Präzision führt. Diese Probleme werden von LSTM und GRU angegangen, die bei der Bewältigung von Gradientenproblemen effizienter sind.

GRU unterscheidet sich von RNNs durch die Verwendung von Update- und Reset-Gates zur Steuerung des Informationsflusses und zur Lösung von Gradientenproblemen. Diese Gates bestimmen, wie viele vorherige Informationen beibehalten oder verworfen werden, wodurch die Effizienz des Modells verbessert wird. Im vorgeschlagenen Framework werden Varianten von GRU verwendet, darunter auch bidirektionale Versionen.

LSTM, eine weitere Variante von RNN, besteht aus mehreren LSTM-Einheiten, die jeweils mit Zell-, Eingabe-, Ausgabe- und Forget-Gates ausgestattet sind. Diese Gates spielen im LSTM-Modell eine entscheidende Rolle, da sie die sequentielle Verarbeitung von Daten und die Generierung von Vorhersagen erleichtern

Dieses Framework verwendet verschiedene Konfigurationen neuronaler Netzwerke zur Prognose, darunter Vanilla LSTM, Stacked LSTM, Bidirectional LSTM und CNN-LSTM. Vanilla LSTM besteht aus einer verborgenen Schicht und einer Ausgabeschicht, während Stacked LSTM mehrere verborgene Schichten umfasst. Bidirectional LSTM verarbeitet Daten sowohl vorwärts als auch rückwärts und kombiniert dann die Interpretationen. CNN-LSTM integriert ein CNN, um Sequenzteile zu verstehen, und übergibt sie dann zur Interpretation an LSTM, wodurch ein Hybridmodell entsteht.

Beim Vergleich von GRU und LSTM lässt sich hinsichtlich aller Leistungsparameter kein klarer Gewinner erkennen . GRU wird mit weniger Parametern und geringerem Speicherverbrauch schneller ausgeführt, während LSTM eine höhere Präzision aufweist, insbesondere bei größeren Datensätzen.

MLP dient als grundlegendes künstliches neuronales Netzwerk mit vollständig verbundenen Feedforward-Schichten, die für Klassifizierungs- und Regressionsaufgaben geeignet sind. Es besteht aus mindestens einer Eingabeschicht, einer verborgenen Schicht und einer Ausgabeschicht und verwendet nichtlineare Aktivierungsfunktionen, um linear nicht trennbare Daten zu verarbeiten.

Für Regressionsaufgaben bietet die Support Vector Regression (SVR) Flexibilität bei der Definition der Fehlertoleranz. Sie versucht, die L2-Norm des Koeffizientenvektors zu minimieren und gleichzeitig Abweichungen mit Schlupfvariablen zuzulassen. Entscheidungsbäume, einschließlich Regressionsbäume, werden zur Prognose von Zeitreihen eingesetzt, während Random Forest mehrere Regressionsbäume kombiniert, um die Vorhersagegenauigkeit durch Bagging und Merkmalszufälligkeit zu verb

2.5.2. Multivariate Prognose

Im vorgeschlagenen Rahmen werden für die Prognose multivariater Zeitreihen eine Reihe statistischer und maschineller Lernmethoden verwendet. Während statistische Ansätze für multivariate Zeitreihen begrenzt sind, umfasst der Rahmen wichtige Techniken wie Vektorautoregression (VAR), Vektorgleitender Durchschnitt (VMA) und Vektorautoregressionsgleitender Durchschnitt (VARMA).

Im Gegensatz zu univariaten Techniken wie AR und ARIMA sind vektorbasierte Methoden wie VAR, VMA und VARMA bidirektional, sodass Variablen sich gegenseitig beeinflussen können. Diese Techniken erfordern jedoch Stationarität, die häufig mithilfe des ADF-Tests bestätigt wird. VAR stellt Beziehungen zwischen Variablen über ein Gleichungssystem her, wobei die Auswahl der Verzögerungen von Kriterien wie dem Akaike-Informationskriterium (AIC) geleitet wird. VMA arbeitet mit einer gleitenden Durchschnittsmatrix und einem k-dimensionalen Vektor, während VARMA sowohl autoregressive als auch gleitende Durchschnittsordnungen integriert. Für maschinell lernende Prognosen verwendet das Framework Techniken wie MLP, Regressionsbaum, Random Forest, RNN, GRU und LSTM, die jeweils im vorhergehenden Abschnitt ausführlich beschrieben werden .

3. Experimentelle Ergebnisse und Diskussion

Diese Studie evaluiert Preptimize, ein Framework für die Analyse von Zeitreihendaten, in verschiedenen Bereichen und Szenarien, darunter Börsendaten, Kryptowährungsdaten und IoT-generierte Stromverbrauchsdaten. Jedes Experiment beinhalte

Machine Translated by Google

Daten, Anwendung verschiedener Prognosetechniken und Bewertung der Leistung anhand von Metriken wie MAE, MSE, RMSE und MAPE. Die Details der verwendeten Datensätze sind in Tabelle 2 aufgeführt. Das Framework verwendet Vorverarbeitungs- und Prognosetechniken, die auf jeden Datensatztyp zugeschnitten sind, wobei die Auswahl durch einen genetischen Algorithmus (GA) optimiert wird. Preptimize ist darauf ausgelegt, den GA zur Optimierung der Modellauswahl zu nutzen. Um ein umfassendes Verständnis der Funktionalität des Frameworks zu gewährleisten, wurden alle möglichen Modelle bewertet. Der GA wurde angepasst, um alle potenziellen Modelle einzubeziehen und sowohl die besten als auch die schlechtesten Performer zu identifizieren. Dies wurde erreicht, indem den M In praktischen Anwendungen empfiehlt der GA jedoch in der Regel die fünf leistungsstärksten Modelle auf der Grundlage des niedrigsten RMSE, wobei Effizienz und Effektivität im Gleichgewicht stehen. Zu den Bewertungsmetriken gehören MAE, MSE, RMSE und MAPE, wobei eine detaillierte Analyse in einem umfassenden Bericht bereitgestellt wird, der von Preptimize generiert wird. Die Experimente werden separat für univariate und multivariate Datensätze durchgeführt und decken den Aktienmarkt, Kryptowährungen und Daten zum Stromverbrauch im Haushalt ab. Jedes Experiment wird sorgfältig beschrieben, einschließlich Datensatzdetails, Trainings-/Testverfahren und Bewertungsergebnissen. Die Details und Ergebnisse dieser Experimente sin

Tabelle 2. Details der in den Experimenten verwendeten Datensätze.

Name	Beschreibung
Experiment 1 – Univariate Zeitreihen	
Börse	**Symbol: MSFT (Microsoft) Spalten: Datum, Schluss Anzahl Datensätze: 503 Termine: 03.08.2020 bis 01.08.2022 Intervall: Täglich**
Börse	**Wie 1-A, jedoch mit fehlenden Daten**
Kryptowährung	**Kryptowährung: BTC-USD (Bitcoin) Spalten: Datum, Schluss Anzahl Datensätze: 748 Termine: 01.08.2020 bis 17.08.2022 Intervall: Täglich**
Experiment 2 – Multivariate Zeitreihen	
IoT – Stromverbrauch	**Domäne: Home Spalten: Datum, Global_active_power, Globale Blindleistung, Spannung, Globale Intensität, Untermessung_1, Untermessung_2, Untermessung_3 Anzahl Datensätze: 10.081 Termine: 24.04.2007 bis 30.04.2007 Intervall: Minute**
Börse	**Symbol: MSFT (Microsoft) Spalten: Datum, Eröffnung, Höchststand, Tiefststand, Schlussstand Anzahl Datensätze: 503 Termine: 03.08.2020 bis 01.08.2022 Intervall: Täglich**

3.1. Experiment 1: Univariate Reihen

Preptimize wurde zunächst mit Marktdaten getestet, darunter Aktien und Kryptowährungen , ein gängiger Typ von Zeitreihendaten, der durch die schwankenden Preise von Angeboten im Laufe der Zeit gekennzeichnet ist. Experiment 1 ist in drei Teile gegliedert. Teil 1 nutzte vollständige Aktiendaten, um die Leistung von Preptimize bei qualitativ hochwertigen Daten zu bewerten. Teil 2 führte absichtlich fehlende Daten ein, um den Umgang des Frameworks mit unvollständigen Datensätzen zu bewerten. Teil 3 nutzte Kryptowährungsdaten mit fehlenden Werten, um die Fähigkeiten von Preptimize weiter zu testen. Diese Experimente zielten darauf ab, die Wirksamkeit des Frameworks in verschiedenen Szenarien und Datenqualitäten im Bereich der Fina

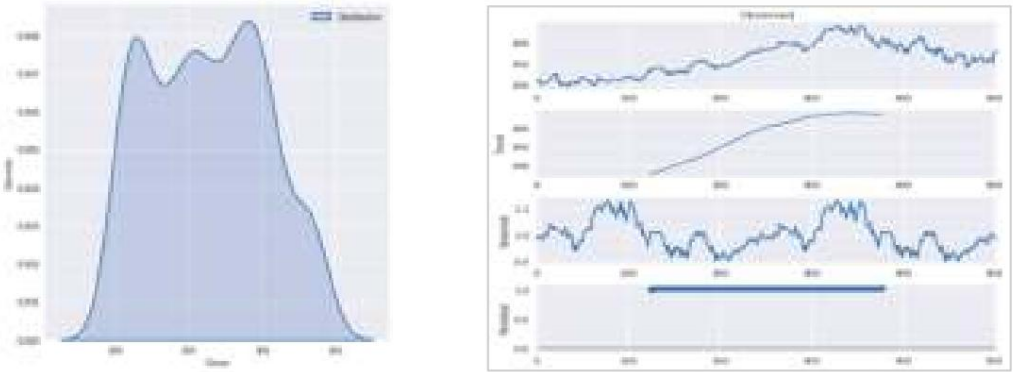
Im ersten Teil des Experiments untersucht diese Studie Börsendaten von Yahoo Finance und konzentriert sich dabei auf die an der NASDAQ-Börse notierte Microsoft Corporation

Machine Translated by Google

Mithilfe der Python-Bibliothek von yfinance wurden Marktdaten von August 2020 bis August 2022 gesammelt, wobei insbesondere die Schlusskurse an jedem Handelstag (ausgenommen Wochenenden) untersuc Dabei handelt es sich um einen univariaten Zeitreihendatensatz mit rund 500 Datensätzen, die in Trainings- und Testdatensätze unterteilt wurden. Der Datensatz war insbesondere vollständig und wies keine Qualitätsprobleme auf, sodass in der Anfangsphase des Experiments keine Vorverarbeitung erforderlich w

Insgesamt wurden 139 Modelle zur Prognose der Zeitreihe verwendet. Der erstellte Bericht bietet einen umfassenden Überblick über den Datenanalyseprozess. Trotz Platzbeschränkungen enthält der Bericht wichtige Funktionen wie grundlegende Profilmerkmale, Diagramme mit gleitendem Mittelwert und Standardabweichung, Streudiagramme, Boxplots, Verteilungsdiagramme und Trendzerlegungsdiagramme . Darüber hinaus enthält der Bericht Trainingsprotokolle, Fehlerberechnungen und Leistungsbewertungen für alle Modelle. Abbildung 3a,b veranschaulicht die Verteilung der Daten, weist auf Normalität hin und zeigt die Zerlegung von Trend und Saisonalität innerhalb der Zeitreihe. Umfassende Analysen, einschließlich p-Werten und Augmented Dickey–Fuller (ADF)-Statistiken, deu

Die Datei mit der Zusammenfassung der Modellleistung enthielt Ergebnisse von etwa 139 Modellen. Einige davon sind in Tabelle 3 aufgeführt . Dabei werden nur die 5 Modelle mit der besten und die 5 Modelle mit der schlechtesten Leistung angezeigt.



(A) **(B)**

Abbildung 3. (a) Die Grafik zeigt die normale Datenverteilung der MSFT-Schlusskurse für 2 Jahre. (b) Zerlegungsgrafik der Schlusskurse zur Trennung der Trend- und Saisonalitätskomponente. Die Grafik zeigt einen klaren Aufwärtstrend und keine Saisonalität.

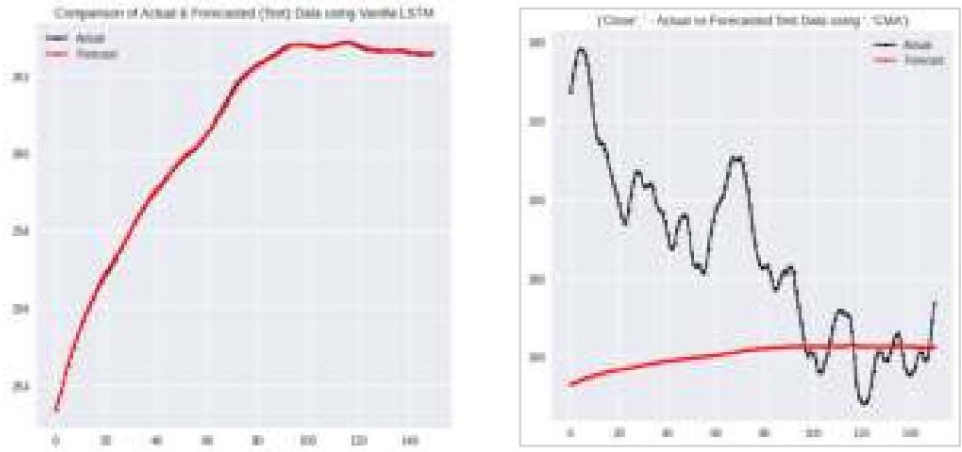
Verschiedene Prognosetechniken wurden auf einen univariaten Aktienmarktdatensatz angewendet, wobei sowohl maschinelles Lernen als auch statistische Methoden zum Einsatz kamen. Maschinelle Lerntechniken wie LSTM, GRU und MLP sowie statistische Methoden wie einfacher gleitender Durchschnitt, exponentieller gleitender Durchschnitt und ARIMA wurden mit verschiedenen Hyperparametern getestet, darunter verborgene Schichten, Neuronen und Epochen. Ebenso wurden verschiedene Varianten statistischer Techniken evaluiert, wie etwa verschiedene Ordnungen für ARIMA. Diese Modelle wurden anhand verschiedener Metriken am Testdatensatz evaluiert.

Die Ergebnisse zeigten, dass das Vanilla LSTM Deep Learning-Modell im Datensatz die beste Leistung erbrachte und einen RMSE von 0,0294 mit 50 Epochen und 40 Einheiten erreichte. Die von diesem Modell generierten Vorhersagewerte wurden mit den erwarteten Werten im Testsatz verglichen, wie in Abbildung 4a dargestellt. Unter den statistischen Techniken schnitt der exponentielle gleitende Durchschnitt (EMA) mit einem MAE von 2,0925 und einem RMSE von 2,5814 am besten ab. EMA belegte den 127. Platz von 139 Modellen. Im Gegensatz dazu zeigte das auf dem kumulativen gleitenden Durchschnitt (CMA) basierende Modell die schlechteste Leistung mit einem MAE von 27,2085 und einem RMSE von 35,1176. Abbildung 4b zeigt einen Vergleich zwischen den tatsächlichen Werten und den vom CMA-Modell erzeugten Vorhersagewerten.

Machine Translated by Google

Tabelle 3. Zusammenfassung der Modellleistung des Experiments basierend auf univariaten Zeitreihen.

Rang	Vorverarbeitung	Modell	Einstellungen	MAE	RMSE
Teil 1: Aktienmarkt-Zeitreihen ohne Probleme mit der Datenqualität					
1	N / A	Vanille-LSTM	Epoche = 50, Einheiten = 40	0,02	0,03
2	N / A	GRU	Epoche = 75, Einheiten = 50	0,02	0,03
3	N / A	Vanille-LSTM	Epoche = 25, Einheiten = 50	0,02	0,03
4	N / A	GRU	Epoche = 50, Einheiten = 40	0,03	0,04
127	N / A	Exponentiell	-	2.09	2,58
135	N / A	Entscheidungsbaum-Regressor	Aufteilung =	7,57	8.03
136	N / A	Zufällig Wald	MSE 100 Bäume, Aufteilung = MSE	7,69	8.15
138	N / A	AR	-	21.11	26,88
139	N / A	CMA	-	27.21	35.12
Teil 2: Börsenzeitreihen mit fehlenden Daten					
1	CMA	ARIMA	(7, 1, 2)	0,57	0,61
2	Spline 3	ARIMA	(7, 1, 7)	0,65	0,84
	Exponentiell	Exponentiell	-	1,92	2.43
6 59	Zeit	SVR	Kernel = Poly	3,64	4,73
64	Derivate	SVR	Kernel = RBF	3,68	4,75
69	Stückweises Polynom	Gestapeltes LSTM	Epoche = 100, Einheiten = 50	3,66	4,76
70	Derivate	RNN	Epoche = 25, Einheiten = 50	3,67	4,76
1269	Polynom 7	CMA	-	27,82	35,76
1270	Polynom 9	CMA	-	28,46	36,52
Teil 3: Kryptowährungszeitreihen mit fehlenden Daten					
	GEWINNEN	Bidirektionales LSTM	Epoche = 100, Einheiten = 50	9,09	10.03
1	Index	Gestapeltes LSTM	Epoche = 100, Einheiten = 50	8,86	10,64
	pchip	Vanille-LSTM	Epoche = 100, Einheiten = 50	10,82	12.11
2	Exponentiell	GRU-bidirektional	Epoche = 25, Einheiten = 50	12,95	14,67
3 10	Kubisch	AR	-	243,97	291,00
693	GAIN-	DT-Regressor	Aufteilung =	260,87	324,39
703	Chip	Zufälliger Wald	MSE 100 Bäume, Aufteilung = MSE	292,03	362,15
729 824	CMA	SVR	Kernel = Polynom	2575,77	2576,76
892	Spline 5	CMA	-	7431.54	9344.14



(A) (B)

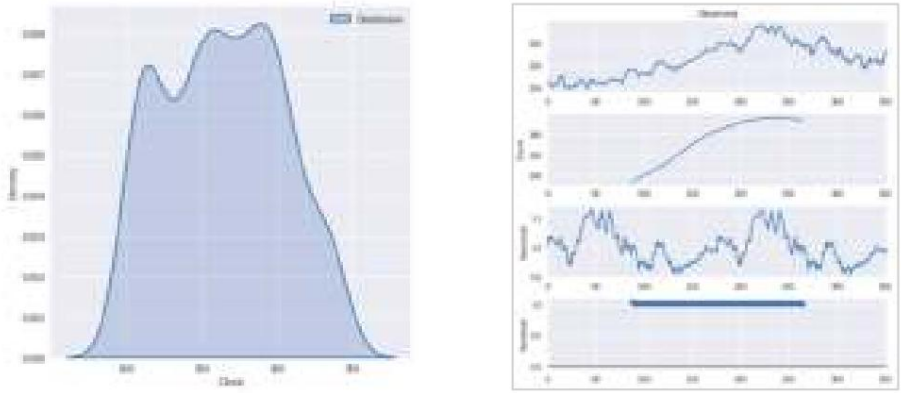
Abbildung 4. (a) Vergleich der tatsächlichen und prognostizierten Daten, die von Vanilla LSTM unter Verwendung von 50 Epochen generiert wurden und 40 Einheiten. (b) Vergleich der tatsächlichen und prognostizierten Daten, generiert durch den kumulativen gleitenden Durchschnitt (CMA).

ARIMA ist eine der effektivsten und effizientesten Techniken für Zeitreihen Prognosen. In Preptimize wurde ARIMA mit einer optimierten Reihenfolge von (2, 1, 2) implementiert, validiert durch den Ljung–Box-Test. Trotz dieser Optimierungen ist das resultierende Der RMSE von 3,6001 war nicht besonders beeindruckend.

Machine Translated by Google

Der Aktienmarktdatensatz wurde im zweiten Teil des Experiments mit fehlenden Datenproblemen verwendet. Wie bereits erwähnt, wurden die Daten absichtlich entfernt, um die Leistung des Frameworks durch Vergleich mit den Originaldaten zu testen.

Preptimize wurde auf diesen Datensatz angewendet, der die fehlenden Tupel enthielt. Das Framework generierte einen über 3500 Seiten umfassenden Bericht mit einer detaillierten Beschreibung der Schritte, die auf den gegebenen Datensatz angewendet wurden, einschließlich Vorverarbeitung und abschließender Analyse. Neben anderen Einzelheiten enthielt der Bericht Details zum fehlenden Typ der Daten. Das Framework fasste die Daten als völlig zufällig fehlende Daten (MCAR), normalverteilt und stationär zusammen, ohne Saison und Trend. Abbildung 5a,b zeigt die Trend- und Saisonaufteilung der Zeitreihe bzw. die Verteilung der Daten.



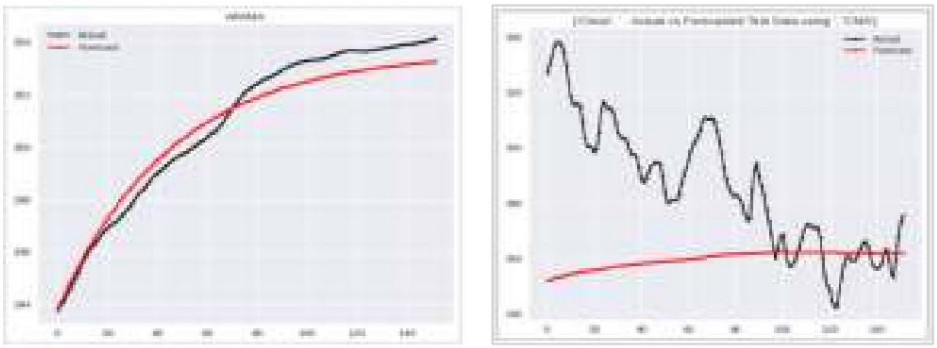
(A) **(B)**

Abbildung 5. (a) Zerlegungsdiagramm des Schlusskurses zur Trennung der Trend- und Saisonalitätskomponente. Das Diagramm zeigt einen leichten Aufwärtstrend und keine Saisonalität. (b) Das Diagramm zeigt die Normalverteilung der Schlusskursdaten.

Die Leistungsübersicht aller auf die Daten angewandten Modelle wurde berechnet und auf einem Blatt gespeichert. Das Übersichtsblatt zeigte, dass etwa 1270 Modelle angewandt wurden. Diese 1270 Modelle wurden als Ergebnis verschiedener Kombinationen von Vorverarbeitungs- und Prognosetechniken generiert. Beispielsweise wurde zuerst eine Vorverarbeitungsstrategie wie Forward Fill angewandt, um die fehlenden Daten zu ergänzen, dann wurden alle anderen Prognosetechniken für univariate Prognosen angewandt, darunter statistische und maschinelle Lerntechniken, und dann wurde jedes Modell anhand von vier verschied

Im Experiment wurden mehrere Vorverarbeitungstechniken und -modelle ausprobiert, darunter Interpolation, Vorwärts- und Rückwärtsfüllen, gleitende Durchschnitte und Deep-Learning -Techniken wie GAN zur Imputation. Jedes Modell wurde anhand verschiedener Maßstäbe anhand des Testdatensatzes bewertet . Die Ergebnisse zeigten, dass ARIMA (p, d, q) mit Ordnung (7, 1, 2) die beste Leistung erbrachte, wenn fehlende Daten mithilfe von CMA imputiert wurden. Der Ljung-Box-Test bestätigte das Fehlen von Autokorrelation im weißen Rauschen und im Modell. Dieses ARIMA-Modell ergab einen MAE von 0,56997 und einen RMSE von 0,6099. Die fünf besten Modelle gehörten überwiegend ARIMA an und verwendeten unterschiedliche Imputationstechniken wie Spline-Ordnung 3, Polynomordnung 5, Rückwärtsfüllen und Auffüllen. Nach ARIMA zeigte der exponentielle gleitende Durchschnitt mit mehreren Imputationstechniken eine gute Leistung, darunter lineare, slineare, stückweise polynomische und Ableitungstechniken. Die ersten 85 Modelle, die eine gute Leistung erbrachten, waren alle statistische Modelle. Umgekehrt schnitt das CMA-basierte Modell, das mit einem Polynom vom Grad 9 verarbeitet wurde, mit einem MAE von 28,4643 und einem RMSE von 36,5159 am schlechtesten ab. Abbildung 6 veranschaulicht die Leistung der besten und schlechtesten Modelle im Experiment, wobei Abbildung 6a die vom ARIMA-Modell (7,1,2) für den durch CMA imputierten Datensatz vorhergesagten Werte im Vergleich zu den ursprünglichen Werten des Testsatzes zeigt, während Abbildung 6b die Leistung des CMA-Modells demonstriert . Die Support Vector Regression mit einem **mithilfe von Zeitinterpolation imputierten polynomischen Kernelmodell** erwies sich in der Kategorie Maschinelles Lernen als die beste **und ergab einen MAE von 3,6430 und einen RMSE von 4,7306. Die für die genannten Modelle berechneten Fehler sind in Tabelle 2 aufgeführt.**

Machine Translated by Google

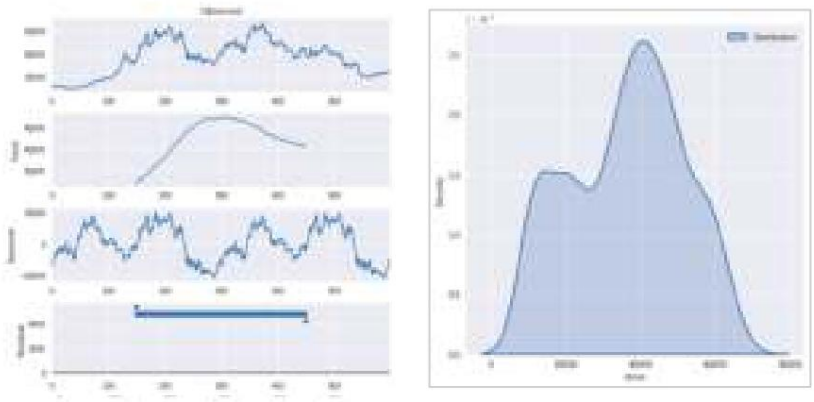


(A) **(B)**

Abbildung 6. (a) Vergleich der tatsächlichen und prognostizierten Daten, generiert durch ARIMA (7,1,2), imputiert durch CMA. (b) Vergleich der tatsächlichen und prognostizierten Daten, die von CMA generiert und durch Polynome imputiert werden Interpolation vom Grad 9.

Der dritte Teil des Experiments konzentrierte sich auf die Analyse von Kryptowährungsdaten, insbesondere Bitcoin (BTC-USD) von der Coinbase-Börse. Kryptowährungen funktionieren unabhängig von traditionellen Banksystemen und nutzen die Blockchain-Technologie für Transaktionen.

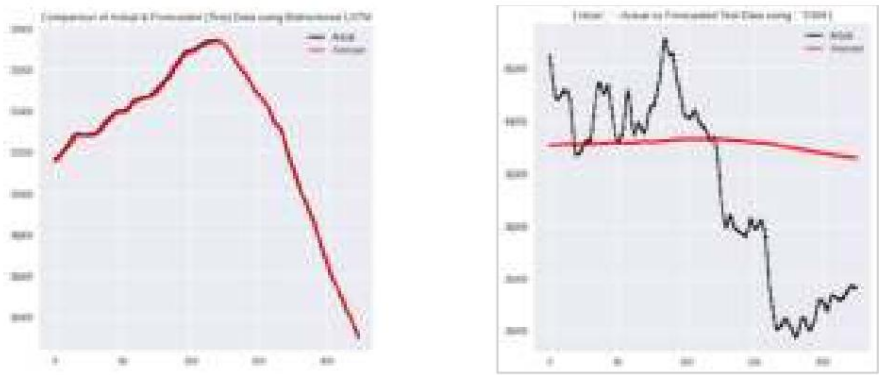
Authentifizierung und Datenhaltung. Der Datensatz, der etwa zwei Jahre umfasst, von August 2020 bis August 2022, bestand aus 747 Aufzeichnungen von Bitcoin-Preisen in USD. Zur Bewertung der Die Leistung des Frameworks bei der Verarbeitung fehlender Daten wurde durch zufällig ausgewählte 30 % des Datensatzes belegt. entfernt. Preptimize verarbeitete die Daten und identifizierte sie als stationär, ohne erkennbare Saisonalität oder Trend. Trotz der numerischen Natur des Datensatzes und der fehlenden Schiefe wies er eine beträchtliche Varianz auf, was auf eine erhebliche Preisvolatilität über den Zeitraum hindeutet. Abbildung Zeigen Sie die Trend- und Saisonaufteilung der Zeitreihe bzw. die Verteilung der Daten. Verschiedene Imputationstechniken, darunter Deep-Learning-Methoden, Interpolation und gleitende Durchschnitte, wurden eingesetzt, um die fehlenden Werte zu ermitteln. Anschließend wurde de aufgeteilt in Trainings- und Test-Sets für die Prognose. Insgesamt wurden 892 Modelle trainiert und evaluiert, wobei sich Bidirektionales LSTM in Kombination mit GAIN-Imputation als die beste Performer, der einen RMSE von 10,028 und einen MAE von 9,0879 erreichte. Unter den statistischen Techniken Die Autoregression (AR) zeigte in Kombination mit der kubischen Interpolation vielversprechende Ergebnisse und erreichte den Rang 693. Umgekehrt das CMA-basierte Modell mit Spline-Interpolation und einem Polynom fünften Grades zeigte die schlechteste Leistung mit einem RMSE von 9344,1381 und einem MAE von 7431,54. Bemerkenswert ist , dass baumbasierte maschinelle Lernalgorithmen suboptimal abschnitten, wobei die Ents Tree-Regressor- und Random-Forest-Modelle rangieren auf Platz 703 bzw. 729, wie in Tabelle 2. Leistungsdiagramme der besten und schlechtesten Modelle sind in Abbildung 8 dargestellt.



(A) **(B)**

Abbildung 7. (a) Zerlegungsgraph der Schlusskurse von Bitcoin (b) Datenverteilung: Der Graph zeigt normale Datenverteilung des Schlusskurses von Bitcoin in USD.

Machine Translated by Google



(A) **(B)**

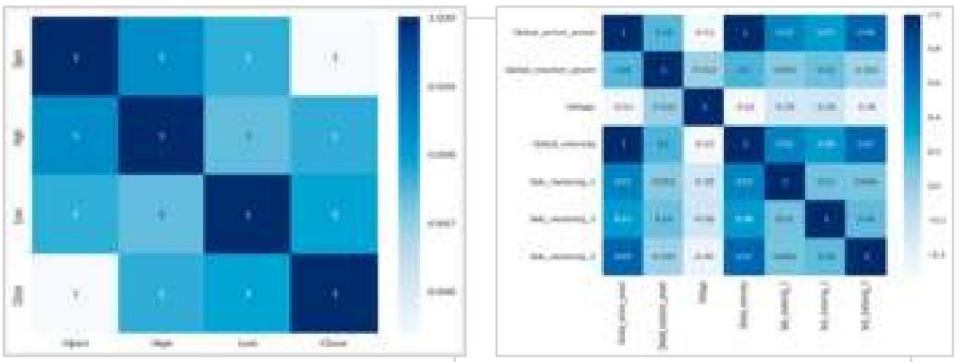
Abbildung 8. (a) Vergleich der tatsächlichen und prognostizierten Daten, die durch Bidirectional LSTM imputed generiert wurden durch GAIN. (b) Vergleich der tatsächlichen und prognostizierten Reihen, die durch CMA generiert wurden, unterstellt durch Spline von Polynom 5. Grades.

Die Ergebnisse zeigen, dass Preptimize bei verschiedenen Arten von Zeitreihen gut funktioniert Daten. Zum Beispiel zeigte das Vanilla LSTM-Modell die beste Performance an der Börse Daten mit einem RMSE von 0,0294, während ARIMA-Modelle mit verschiedenen Imputationstechniken Die beste Leistung wurde bei Datensätzen mit fehlenden Werten erzielt, was die Vielseitigkeit des Frameworks unterstreicht. und Wirksamkeit.

3.2. Experiment 2: Multivariate Reihen

Das zweite Experiment konzentrierte sich auf die Bewertung der Leistung von Preptimize bei multivariaten Zeitreihendaten mit und ohne fehlende Werte. Die vollständigen Daten aus der Aktie Markt wurden im ersten Teil verwendet, während der zweite Teil fehlende Instanzen einführte unter Verwendung eines Datensatzes zum Stromverbrauch einzelner Haushalte.

Im ersten Teil erfasste der Aktienmarktdatensatz die Leistung von Microsoft (MSFT) basierend auf Eröffnungs-, Höchst-, Tiefst- und Schlusswerten vom 3. August 2020 bis zum 1. August 2022, insgesamt 503 Datensätze. Preptimize begann mit der Datenprofilierung und enthüllte eine hohe paarweise Korrelation zwischen den Spalten, was auf Abhängigkeit hinweist. Die Spalten 'offen' und 'hoch' waren jedoch als nicht stationär erwiesen und möglicherweise Trend- und Saisonalitätskomponenten enthalten, dargestellt in Abbildung 9; die Heatmap veranschaulicht die paarweise Korrelation.



(A) **(B)**

Abbildung 9. (a) Korrelations-Heatmap, die die Korrelation zwischen Paaren im multivarianten Datensatz zum Aktienmarkt zeigt. (b) Korrelations-Heatmap, die die Korrelation zwischen Paaren im Stromverbrauch zeigt multivariater Datensatz.

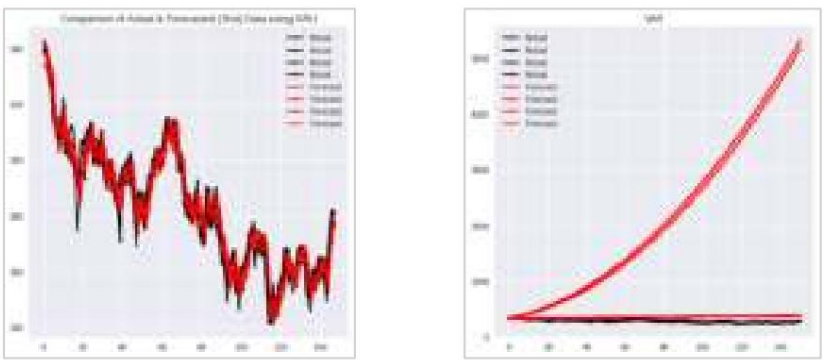
Beim Übergang zur Prognose, da keine Vorverarbeitung erforderlich war, wendete Preptimize verschiedene Techniken an, darunter VAR, VMA, VARMA, MLP, Entscheidungsbaum-Regressor, Zufalls forest, RNN, GRU, GRU-bidirektional und Stacked LSTM. Die Transformation wurde angewendet auf 'offene' und 'hohe' Spalten, um sie vor der Prognose stationär zu machen. Unter den Modellen

Machine Translated by Google

Wie in Tabelle 4 erwähnt, erwies sich GRU mit 50 Einheiten und 100 Epochen als der beste Performer, mit 3,9242 MAE und 5,1312 RMSE. Abbildung 10a zeigt die Leistung des GRU Modell auf dem multivariaten Testsatz. Statistische Modelle wie VAR, VMA und VARMA zeigte eine schlechte Leistung, da das grundlegende Muster in der multivariaten Reihe nicht verstanden wurde . Abbildung 10b zeigt die Leistung des VAR-Modells bei den multivariaten Daten. Darüber hinaus ergab die Performance-Analyse der einzelnen Spalten, dass die 'offenen' und 'hohe' Säulen, die zunächst als nicht stationär angesehen wurden, stellten eine Herausforderung für die Vorhersage dar. Trotz Bei Anwendung der Differenztransformation zeigten diese Spalten vorhergesagte Werte, die voneinander abwichen erheblich von den Originaldaten ab, was zu hohen RMSE-Werten führt.

Tabelle 4. Zusammenfassung der Modellleistung des Experiments basierend auf multivariaten Reihen.

Rang	Vorverarbeitung	Modell	Einstellungen	MAE	RMSE
Teil 1: Multivariate Zeitreihen am Aktienmarkt ohne Probleme mit der Datenqualität					
	N / A	GRU	Epoche = 100, Einheiten = 50	3,92	5.13
1	N / A	RNN	Epoche = 100, Einheiten = 50	3,93	5.13
2	N / A	RNN	Epoche = 25, Einheiten = 50	3,95	5.19
3	N / A	MLP	Epoche = 100, Einheiten = 20	4,04	5.19
4	N / A	Zufälliger Wald	100 Bäume, geteilter	5,47	7.08
17	N / A	VARMA	MSE (2,	407,98	657,08
23	N / A	VMA	1, 3) (0, 1, 1)	365,41	752,85
24 25	N / A	VAR	Bestellung = 5	963,64	1678.42
Teil 2: Multivariate Zeitreihe zum Stromverbrauch einzelner Haushalte mit fehlenden Daten					
	Exponentiell	GRU	Epoche = 100, Einheiten = 20	0,12	0,64
1	CMA	GRU	Epoche = 100, Einheiten = 20	0,13	0,64
	Bayesianischer Grat	MLP	Epoche = 100, Einheiten = 50	0,13	0,65
2	Zufälliger Wald	Gestapeltes LSTM	Epoche = 100, Einheiten = 50	0,13	0,65
3 4	KNNImputer 5 (0,1,1)	VMA		0,43	1,88
330	KNNImputer 5 (2,1,3) polynomial_3	VARMA		0,43	1,88
331		RNN	Epoche = 25, Einheiten =	9,35	29.31
449 450	20 polynomial_3	DT-Regressor	MSE aufteilen	14,53	58,72



(A) **(B)**

Abbildung 10. (a) GRU-Leistungsvergleich für multivariate Zeitreihendaten während des Tests. (b) VAR Vergleich der Modellleistung für multivariate Zeitreihendaten während des Tests.

Der zweite Teil des multivariaten Reihenexperiments testete Preptimize mit dem Datensatz zum Stromverbrauch einzelner Haushalte, mit dem Ziel, dessen Verhalten anhand unvollständiger multivariater Zeitreihendaten zu verstehen [44]. Dieser Datensatz umfasst sieben Spalten, ohne Datum und Uhrzeit, Aufzeichnung des Stromverbrauchs über vier Jahre mit einem einminütiger Sampling-Rate. Mit rund 2.075.259 Instanzen wies der Datensatz fehlende Werte überwiegend im April 2007. Einzelheiten sind in Tabelle 5 aufgeführt.

Aufgrund der Größe des Datensatzes wurden nur Daten einer Woche mit fehlenden Instanzen verarbeitet . Mit 40 Vorverarbeitungsstrategien und 25 Prognosemodellen wurden insgesamt 10 wurden ausgewertet. Aufgrund von Rechenkapazitätsbeschränkungen wurde das Experiment jedoch aufgeteilt in Teile, wobei es mehrere Tage dauerte, bis Ergebnisse vorlag. Nach der Durchführung früherer Experimente wurden 18 erfolgreiche Vorverarbeitungstechniken mit 25 Prognosemodellen zur Auswertung kombiniert.

Machine Translated by Google

Tabelle 5. Datensatz zum Stromverbrauch einzelner Haushalte, der im Experiment 2 verwendet wurde.

NEIN.	Spaltenname	Einheiten	Beschreibung
1	Globale Wirkleistung	Kilowatt	Die gesamte Wirkleistung, die der Haushalt verbraucht
2	Globale Blindleistung	Kilowatt	Die gesamte vom Haushalt verbrauchte Blindleistung
3	Spannung	Volt	Durchschnittliche Spannung
4	Globale Intensität	Verstärker	Durchschnittlicher Strom
5	Submetering_1	Wattstunden	Aktive Energie für die Küche
6	Submetering_2	Wattstunden	Aktive Energie für die Wäsche
7	Submetering_3	Wattstunden	Aktivenergie für Klimaanlagen

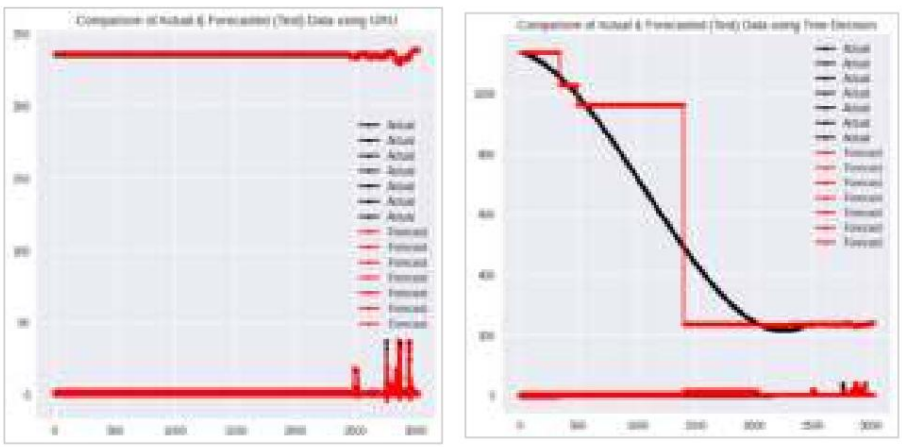
Der umfassende Bericht, der von Preptimize erstellt wurde, begann mit der Profilierung der Daten, wobei jedes Attribut einzeln analysiert wurde. Eine Korrelations-Heatmap, die zuvor in Abbildung 8b dargestellt wurde, veranschaulicht die paarweisen Beziehungen zwischen den Spalten. Insbesondere die globale Intensität und die globale Die Wirkleistung wies eine starke positive Korrelation auf, während die Untermessung 3, verbunden mit Klimakontrollsysteme korrelieren stark mit der globalen Wirkleistung und der globalen Intensität. Im Gegensatz dazu zeigte die globale Blindleistung keine Korrelation mit anderen Eigenschaften, insbesondere Spannung und Untermessung 3.

Darüber hinaus wurde im Bericht der Prozentsatz fehlender Daten (ca. 37 %) ermittelt. Identifizieren von Spalten wie Untermessung 1 und Untermessung 2 als nicht normalverteilt mit hohen Kurtosis-Werten. Preptimize klassifizierte alle vier Spalten mit fehlenden Daten als MNAR-Typ. Versuche, die Verteilungen für Submetering 1 und Submetering 2 zu normalisieren, wurden behindert durch die Möglichkeit von Null- oder negativen Werten, was eine Box-Cox-Transformation ausschließt.

Nach der Trend- und Saisonzerlegung wurden verschiedene Imputationstechniken, darunter iterative Methoden, angewendet. Jeder imputierte Datensatz wurde anhand von 18 verschiedenen Prognosemodelle, die zu einer umfassenden Leistungsbewertung führen.

Preptimize erstellte einen Bewertungsbogen mit 450 Modellen, der die Vorverarbeitungstechniken, die während der Imputation eingesetzt werden. Wie bereits gezeigt , präsentiert die Leistungsmetriken ausgewählter Modelle und hebt das Modell mit der besten Leistung hervor als GRU mit einem exponentiellen gleitenden Durchschnitt imputiert, was einen RMSE von 0,6393 ergibt und MAE von 0,1225. Umgekehrt zeigte der Entscheidungsbaum-Regressor, der mit polynomischer Interpolation dritten Grades imputiert wurde, die schlechteste Leistung mit einem RMSE von 14,5308 und MAE von 58,7152.

Bemerkenswerterweise zeigten statistische Modelle eine gleichbleibende Leistung, wenn die Daten imputiert mit KNNImputer mit zwei Nachbarn, mit VMA-, VARMA- und VAR-Ranking genau. Einige Modelle des maschinellen Lernens schnitten jedoch schlechter ab. Leistungsdiagramme für GRU- und Entscheidungsbaum-Regressormodelle sind in Abbildung 11a bzw. b dargestellt.



(A) (B)

Abbildung 11. (a) GRU imputiert mit exponentiellem gleitendem Durchschnitt: Leistungsvergleich für Leistung Verbrauchsreihendaten während des Tests. (b) Entscheidungsbaum imputiert mit polynomischer Interpolation von Grad 3: Leistungsvergleich für Stromverbrauchsreihen im Test.

Machine Translated by Google

Zusammenfassend zeigte das Experiment, dass sich der exponentielle gleitende Durchschnitt, Bayesian Ridge, KNN-basierte Techniken und DataWig als effektive Imputationsmethoden erwiesen, während die Polynom-Interpolation dritten Grades, die Spline-Interpolation dritten Grades und HotDeck suboptimale Ergebnisse lieferten.

4. Schlussfolgerungen und Diskussion

Die Zeitreihenanalyse ist aufgrund der Vielzahl verfügbarer Techniken immer komplexer geworden. Die Wahl der geeigneten Prognosemethode und Vorverarbeitungsstrategie auf der Grundlage der Eigenschaften der Zeitreihe ist entscheidend für das Erzielen zuverlässiger Ergebnisse. Bei der Zeitreihenanalyse müssen die richtigen Schritte befolgt werden, um die besten in praktischen Umgebungen anwendbaren Modelle zu ermitteln. Statistische und maschinelle Lernmodelle sowie verschiedene Imputationstechniken zum Umgang mit fehlenden Daten bieten vielfältige Analyseoptionen. Angesichts der umfangreichen Forschungsarbeiten, die die Leistung jeder Technik bei bestimmten Datensätzen detailliert beschreiben, besteht Bedarf an einem System, das diese Techniken effi

Durch Experimente mit verschiedenen Zeitreihendatensätzen wird die Wirksamkeit des vorgeschlagenen Frameworks Preptimize deutlich. Preptimize empfiehlt mehrere Blaupausen für Vorhersagemodelle , die als erster Schritt zur Analyse der gegebenen Zeitreihen dienen und viel Zeit und Aufwand sparen. Forscher können diese Modelle weiter verbessern, indem sie Hybridkombinationen erstellen, Parameter anpassen und weitere Funktionen hinzufügen, um die Prognosegenauigkeit zu verbessern . Preptimize generiert detaillierte Berichte, die alle Schritte der Datenverarbeitung abdecken, zusammen mit Diagrammen, Grafiken, Statistiken und einem umfassenden Bewertungsblatt, das die Leistung aller anhand verschiedener Fehlermetriken getesteten Modelle anhand des Datensatzes zeigt.

Um die Leistung dieses Frameworks zu bewerten, wurden verschiedene Experimente mit unterschiedlichen Datensätzen durchgeführt . Die Wirksamkeit des Frameworks kann daran gemessen werden, wie genau das Blueprint-Modell mit den in der Literatur verwendeten Modellen übereinstimmt, und indem die Genauigkeit auf Basis des RMSE verglichen wird. Benchmark-Leistungen für Modelle zur Vorhersage des Aktienmarkts basieren auf LSTM, SVR und Hybridmodellen, die statistische Methoden mit Techniken des maschinellen Lernens kombinieren und oft RMSE-Werte von nur 0,02 für Aktienkurse melden. Preptimize schlug auch LSTM und ARIMA zur Prognose von Aktienkursen vor. Benchmark-Leistungen für die Aktienprognose von Microsoft (MSFT) meldeten einen RMSE von 0,05 mit einem Hybridmodell, das LSTM und RNN kombiniert **[45]. Das mit Preptimize vorgeschlagene Blueprint-Modell erreichte einen RMSE von 0,029 unter Verwendung von Vanilla LSTM. Die hohe Volatilität von Bitcoin mit erheblichen Preisschwankungen über kurze Zeiträume stellt für traditionelle Prognosemodelle eine Herausforderung da Preptimize empfahl LSTM (RMSE 10,02) zur Bitcoin-Prognose, was mit den Erkenntnissen aus der Literatur übereinstimmt. Eine Studie [46] überprüfte verschiedene Prognosemodelle, darunter Feedforward -Neuralnetze, LSTM, GRU, ARIMA und KI-Techniken, anhand eines Datensatzes zum Stromverbrauch einzelner Haushalte, um ihre Leistung zu vergleichen. Der durchschnittliche RMSE über 12 verschiedene Modelle hinweg betrug 0,83, wobei der niedrigste RMSE von 0,75 sowohl von LSTM als auch von GRU erreicht wurde. Im Gegensatz dazu empfahl Preptimize die Verwendung von GRU, das durch exponentielle Glättung vorverarbeitet wurde , was zu einem RMSE von 0,639 führte. Basierend auf der Diskussion können wir den Schluss ziehen, dass Preptimize effektiv geeignete Vorhersagemodelle für einen gegebenen Datensatz empfiehlt und so eine schnelle Entscheidungsfindung mit minimalem Aufwand ermöglicht.**

Preptimize weist bestimmte Einschränkungen auf, die berücksichtigt werden sollten. Erstens ist es ressourcenintensiv und erfordert aufgrund seines umfassenden Ansatzes, bei dem zahlreiche Techniken auf eine Datenprobe angewendet werden, um die optimalen Vorverarbeitungs- und Prognosemethoden zu ermitteln, erhebliche Rechenleistung und Zeit. Zweitens unterstützt Preptimize in erster Linie Datensätze mit einer Normalverteilung. Bei der Verarbeitung nicht normaler Verteilungen transformiert es die Reihe in eine annähernde Normalverteilung, was zu zusätzlicher Komplexität und potenziellen Ungenauigkeiten führen kann, insbesondere wenn die Transformation keine vollständige Normalverteilung erreicht. Wir möchten diese Probleme in zukünftigen Arbeiten angehen.

Während sich das vorgeschlagene Framework auf die Lösung des Problems fehlender Daten in der Zeitreihenanalyse konzentriert, bleiben zahlreiche andere Herausforderungen wie Ausreißererkennung, Konzeptdrift und Heteroskedastizität bestehen. Die Bewältigung dieser Herausforderungen erfordert eine weitere A

Machine Translated by Google

der Zeitreihenanalyse, eine Aufgabe, die in zukünftigen Forschungsphasen verfolgt wird. Zukünftige Arbeiten könnten die Erweiterung von Preptimize zur Verarbeitung komplexerer Arten von Zeitreihendaten untersuchen, beispielsweise solcher mit mehreren Jahreszeiten oder unregelmäßigen Intervallen. Darüber hinaus könnte die Integration weiterer Hybridmodelle auf der Grundlage maschineller Lerntechniken und die Ausweitung der Anwendbarkeit des Frameworks auf andere Bereiche wie das Gesundheitswesen und die Umweltüberwachung seinen Nutzen weiter steigern.

Obwohl wir die Robustheit und Vollständigkeit des Rahmens anhand mehrerer Szenarien getestet haben, erkennen wir an, dass zusätzliche Szenarien, wie saisonale univariate Daten, hohe Prozentsätze fehlender Werte, nicht normale Verteilungen, multisaisonale Daten und Abwärtstrends, weitere wertvolle Erkenntnisse liefern könnten. Es war jedoch nicht möglich, alle möglichen Szenarien in die anfängliche Arbeit einzubeziehen. Wir ermutigen andere Forscher, diesen Rahmen mit unterschiedlichen Datensätzen zu untersuchen, um diese Forschungslücke zu schließen.

Beiträge der Autoren: Konzeptualisierung, Methodik, Validierung, formale Analyse, Visualisierung, Schreiben – Vorbereitung des Originalentwurfs, MU; Überwachung, ZAM; Überprüfung und Bearbeitung, RQ und AZ. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu.

Finanzierung: Diese Forschung erhielt keine externe Finanzierung.

Erklärung des Institutional Review Board: Nicht zutreffend.

Einverständniserklärung: Nicht zutreffend.

Datenverfügbarkeitserklärung: Die in dieser Studie verwendeten Datensätze sind online verfügbar. Der Datensatz zum Stromverbrauch ist verfügbar unter https://archive.ics.uci.edu/ml/datasets/individual+household+ Strom+Verbrauch (abgerufen am 1. Juni 2024) Der Finanzdatensatz des Aktienmarktes ist verfügbar unter https://aroussi.com/ (abgerufen am 1. Juni 2024). Der Kryptowährungsdatensatz ist verfügbar unter https://pypi.org/project/Historic-Crypto/ (abgerufen am 1. Juni 2024).

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Verweise

1. Liu, Z.; Zhu, Z.; Gao, J.; Xu, C. Prognosemethoden für Zeitreihendaten: Eine Untersuchung. IEEE Access **2021, 9, 91896–91912.** [CrossRef]
2. Asadi, S.; Esmaeil, H.; Farhad, M.; Masoud, NM Hybridisierung von evolutionären Levenberg-Marquardt-Neuralnetzen und Daten Vorverarbeitung für Börsenprognosen. Knowl. Based Syst. **2012,** 35, **245–258.** [CrossRef]
3. Di Persio, L.; Fraccarolo, N. Energieverbrauchsprognosen durch Gradient-Boosting-Regressionsbäume. Mathematics **2023,** 11, **1068.** [Querverweis]
4. Cryer, JD; Kellet, N. Time Series Analysis: With Applications in **R, 2. Auflage; Springer: New York, NY, USA, 2008; Band 2, S. 31.**
5. Zhou, Q.; Ooka, R. Einfluss der Datenvorverarbeitung auf die Leistung neuronaler Netzwerke zur Reproduktion von CFD-Simulationen von nicht-isotherme Luftstromverteilung in Innenräumen. Energy Build. **2021,** 230, **110525.** [CrossRef]
6. Song, S.; Zhang, A.; Wang, J.; Yu, PS SCREEN: Stream-Datenbereinigung bei Geschwindigkeitseinschränkungen. In Proceedings of the 2015 ACM SIGMOD Internationale Konferenz zum Datenmanagement, Melbourne, Australien, 31. Mai–4. Juni 2015.
7. Bilalli, B.; Abelló, A.; Aluja-Banet, T.; Wrembel, R. Automatisierte Datenvorverarbeitung mittels Meta-Learning. In Proceedings der International Conference on Model and Data Engineering, Almería, Spanien, 21.–23. September 2016; Springer: Cha
8. Zhang, A.; Song, S.; Wang, J.; Yu, PS Zeitreihendatenbereinigung: Von der Anomalieerkennung bis zur Anomaliebehebung. Proc. VLDB Endow. **2017,** 10, **1046–1057.** [CrossRef]
9. Shimizu, K.; Ponce-Hernandez, R.; Ahmed, OS; Ota, T.; Win, ZC; Mizoue, N.; Yoshida, S. Verwendung von Landsat-Zeitreihenbildern zur Erkennung von Waldstörungen in selektiv abgeholzten tropischen Wäldern in Myanmar. Can. J. For. Res. **2017,** 47, **289–296.** [CrossRef]
10. Zhu, Z. Veränderungserkennung mithilfe von Landsat-Zeitreihen: Ein Überblick über Frequenzen, Vorverarbeitung, Algorithmen und Anwendungen. ISPRS J. Photogramm. Remote Sens. **2017,** 130, **370–384.** [CrossRef]
11. Karim, F.; Majumdar, S.; Darabi, H.; Chen, S. LSTM vollkonvolutionelle Netzwerke zur Klassifizierung von Zeitreihen. IEEE Access **2017, 6, 1662–1669.** [CrossRef]
12. Gschwandtner, T.; Erhart, O. Kenne deinen Feind: Qualitätsprobleme von Zeitreihendaten erkennen. In Proceedings of the IEEE Pacific Visualization Symposium (PacificVis), Kobe, Japan, 10.–13. April 2018.
13. Jeenanunta, C.; Abeyrathna, KD; Dilhani, MS; Hnin, SW; Phyo, PP Zeitreihen-Ausreißererkennung für kurzfristige Stromlast Nachfrageprognose. Int. Sci. J. Eng. Technol. (ISJET) **2018, 2, 37–50.**
14. Wang, X.; Wang, C. Zeitreihendatenbereinigung: Eine Übersicht. IEEE Access **2019, 8, 1866–1881.** [CrossRef]
15. Ding, X.; Wang, H.; Su, J.; Li, Z.; Li, J.; Gao, H. Cleanits: Ein Datenbereinigungssystem für industrielle Zeitreihen. Proc. VLDB Endow. **2019,** 12, **1786–1789.** [CrossRef]

Machine Translated by Google

16. **Ruiz, L.; Pegalajar, M.; Arcucci, R.; Molina-Solana, M. Eine Zeitreihen-Clustering-Methodik zur Wissensextraktion im Energiebereich Verbrauchsdaten.** Expert Syst. Appl. **2020,** 160, **113731.** [CrossRef]
17. **Jarrett, D.; Yoon, J.; Bica, I.; Qian, Z.; Ercole, A.; Schaar, MVD Hellsehen: Ein Pipeline-Toolkit für medizinische Zeitreihen. In Proceedings der International Conference on Learning Representations, virtuell, 3.–7. Mai 2021.**
18. **Desai, V.; Dinesha, HA Ein hybrider Ansatz für Datenvorverarbeitungsmethoden. In Proceedings der IEEE International Conference for Innovation in Technology (INOCON), Bangalore, Indien, 6.–8. November 2020.**
19. **Sousa, R.; Amado, C.; Henriques, R. AutoMTS: Vollständig autonome Verarbeitung multivariater Zeitreihendaten aus heterogenen Sensornetzwerken. In Proceedings der International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, virtuell, 29.–30. November 2020; Springer: Cham, Schweiz, 2020.**
20. **Chen, X.; Deng, L.; Huang, F.; Zhang, C.; Zhang, Z.; Zhao, Y.; Zheng, K. Daemon: Unüberwachte Anomalieerkennung und -interpretation für multivariate Zeitreihen. In Proceedings der 2021 IEEE 37th International Conference on Data Engineering (ICDE), Chania, Griechenland, 19.–22. April 2021; S. 2225–2230.**
21. **Chauhan, K.; Jani, S.; Thakkar, D.; Dave, R.; Bhatia, J.; Tanwar, S.; Obaidat, MS Automatisiertes maschinelles Lernen: Die neue Welle des maschinellen Lernens. In Proceedings der 2. Internationalen Konferenz über innovative Mechanismen für Industrieanwendungen (ICIMIA), Bangalore, Indien, 5.–7. März 2020.**
22. **Sarafanov, M. AutoML für Zeitreihen: Auf jeden Fall eine gute Idee. Online verfügbar: https://towardsdatascience.com/automl-for- time-series-definitelya-good-idea-c51d39b2b3f (abgerufen am 1. September 2023).**
23. **Sun, B.; Ma, L.; Shen, T.; Geng, R.; Zhou, Y.; Tian, Y. Eine robuste datenbasierte Methode für Multisaisonalität und Heteroskedastizität in Vorverarbeitung von Zeitreihen.** Wirel. Commun. Mob. Comput. **2021,** 2021, **6692390.** [CrossRef]
24. **Zhang, GP; Min, Q. Neuronale Netzprognosen für saisonale und Trend-Zeitreihen.** Eur. J. Oper. Res. **2005,** 160, **501–514.** [Querverweis]
25. **Ranjan, KG; Prusty, BR; Jena, D. Vergleich zweier Datenbereinigungsmethoden, angewendet auf volatile Zeitreihen. In Proceedings der International Conference on Power Electronics Applications and Technology in Present Energy Scenario (PETPES), Mangalore, Indien, 29.–31. August 2019.**
26. **Ranjan, KG; Tripathy, DS; Prusty, BR; Jena, D. Eine verbesserte, auf gleitender Fenstervorhersage basierende Ausreißererkennung und -korrektur für volatile Zeitreihen.** Int. J. Numer. Model. Electron. Netw. Devices Fields **2021,** 34, **e2816.** [CrossRef]
27. **Lv, P.; Wu, Q.; Xu, J.; Shu, Y. Aktienindexvorhersage basierend auf Zeitreihenzerlegung und Hybridmodell.** Entropy **2022,** 24, **146.** [Querverweis]
28. **Brunel, B.; Alsamad, F.; Piot, O. Auf dem Weg zum automatisierten maschinellen Lernen in der Schwingungsspektroskopie: Einsatz und Einstellungen genetischer Algorithmen zur Vorverarbeitung und Regressionsoptimierung.** Chemom. Intell. Lab. Syst. **2021,** 219, **104444.** [CrossRef]
29. **Kumar, S. 8 AutoML-Bibliotheken zur Automatisierung der Machine-Learning-Pipeline. 2020. Online verfügbar: https://medium.com/swlh/8-AutoML-Bibliotheken zum Automatisieren von Machine-Learning-Pipelines – 3da0af08f636 (abgerufen am 1. August 2023).**
30. **Jang, W.-J.; Lee, S.-T.; Kim, J.-B.; Gim, G.-Y. Eine Studie zum Datenprofiling: Fokussierung auf den Attributwert-Qualitätsindex.** Appl. Sci. **2019, 9, 5054.** [CrossRef]
31. **Ghaderpour, E.; Pagiatakis, SD; Hassan, QK Eine Untersuchung zur Änderungserkennung und Zeitreihenanalyse mit Anwendungen.** Appl. Sci. **2021,** 11, **6141.** [CrossRef]
32. **Zou, H.; Yang, Y. Kombination von Zeitreihenmodellen für Prognosen.** Int. J. Forecast. **2004,** 20, **69–84.** [CrossRef]
33. **Lim, B.; Zohren, S. Zeitreihenprognose mit Deep Learning: Eine Untersuchung.** Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **2021,** 379, **20200209.** [CrossRef]
34. **Abbasimehr, H.; Paki, R.; Bahrini, A. Ein neuartiger Ansatz basierend auf der Kombination von Deep Learning-Modellen mit statistischen Methoden für COVID-19-Zeitreihenprognose.** Neural Comput. Appl. **2022,** 34, **3135–3149.** [CrossRef] [PubMed]
35. **Brown, TA** Konfirmatorische Faktorenanalyse für angewandte Forschung; **The Guilford Press: New York, NY, USA, 2006.**
36. **Haroon, D. Zeitreihen-Differenzierung. In** Python Machine Learning Case Studies: Fünf Fallstudien für den Datenwissenschaftler; **Apress: New York, NY, USA, 2017; S. 110.**
37. **Agiakloglou, C.; Newbold, P. Empirische Beweise für Dickey-Fuller-artige Tests.** J. Time Ser. Anal. **1992,** 13, **471–483.** [CrossRef]
38. **Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks.** Commun. ACM **2020,** 63, **139–144.** [CrossRef]
39. **Yoon, J.; Jordon, J.; Schaar, M. Gain: Fehlende Datenimputation mit generativen kontradiktorischen Netzen. In Proceedings of the International Konferenz zum maschinellen Lernen, Stockholm, Schweden, 10.–15. Juli 2018.**
40. **Biessmann, F.; Rukat, T.; Schmidt, P.; Naidu, P.; Schelter, S.; Taptunov, A.; Lange, D.; Salinas, D. DataWig: Imputation fehlender Werte für Tabellen.** J. Mach. Learn. Res. **2019,** 20, **1–6.**
41. **Joenssen, DW; Bankhofer, U. Hot-Deck-Methoden zur Imputation fehlender Daten. In Proceedings des International Workshop on Machine Learning and Data Mining in Pattern Recognition, Berlin, Deutschland, 13.–20. Juli 2012.**
42. **Holland, JH** Anpassung in natürlichen und künstlichen Systemen: Eine einführende Analyse mit Anwendungen in Biologie, Kontrolle und Künstlicher Intelligence; **MIT Press: Cambridge, MA, USA, 1992.**
43. **McDowall, D.; McCleary, R.; Bartos, BJ** Interrupted Time Series Analysis, **21. Auflage; SAGE: New York, NY, USA, 1980; S. 41.**
44. **Dua, D.; Graff, C.** UCI Machine Learning Repository; **University of California, School of Information and Computer Science: Irvine, CA, USA, 2019. Online verfügbar: http://archive.ics.uci.edu/ml (abgerufen am 1. Juni 2024).**

Machine Translated by Google

45. Sourav, D.; Apan, P.; Sayan, S.; Sayan, G.; Udatya, D.; Chandra, D.; Shilpi, B. Ein neuartiges Hybridmodell mit Lstm und Rnn für Aktienmarktprognose. Int. J. Eng. Res. Technol. **2024,** 13. **[CrossRef]**

46. Gasparin, A.; Slobodan, L.; Cesare, A. Deep Learning für Zeitreihenprognosen: Der elektrische Lastfall. CAAI Trans. Intell. Technol. **2022, 7, 1–25. [CrossRef]**

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Artikel

Boolesche Operationen auf Basis von Deep Learning, Zeitreihen, Fehlererkennung und
Prädiktive Analyse im Containerkranbetrieb

Amruta Awasthi, Lenka Krpalkova und Joseph Walsh *

Fakultät für Naturwissenschaften, Technologie, Ingenieurwesen und Mathematik (STEM), Munster Technological University (MTU), Kerry Campus, V92 HD4V Kerry, Irland;
amruta.awasthi@research.ittralee.ie (AA); lenka.krpalkova@mtu.ie (LK)
*** Korrespondenz: joseph.walsh@mtu.ie**

Zusammenfassung: Deep Learning ist in der Schiffslogistik und bei der Fehlererkennung, -diagnose und -vorhersage von Containerkränen von entscheidender Bedeutung. Eine neuartige Deep-Learning-Technik unter Verwendung von Long Short-Term Memory (LSTM) erkannte und antizipierte Fehler in einem System mit unausgewogenen Daten. Das LSTM-Modell wurde anhand realer Betriebsfehlerdaten von Containerkränen trainiert. Der benutzerdefinierte Algorithmus verwendet die Synthetic Minority Oversampling TECHnique (SMOTE), um die unausgewogenen Daten für Betriebsdatenfehler (d. h. zu wenige Stichproben der Minderheitsklasse) auszugleichen. Zum Programmieren wurde Python verwendet. Es werden Korrelationsmatrizen und Kovarianzmatrizen von Pearson, Spearman und Kendall präsentiert. Der Trainings- und Validierungsverlust des Modells wird angezeigt und die verbleibenden Daten werden vorhergesagt. Der Testsatz (30 % der tatsächlichen Daten) und die prognostizierten Daten hatten RMSEs von 0,065. Eine Zusätzlich wurden die Fehlerausgaben für die Zeitreihe für die nächsten n Sekunden projiziert, wobei die n Sekunden vom Benutzer eingegeben wurden. Die Genauigkeit betrug 0,996, die Präzision 1,00, der Rückruf 0,500 und der F1-Score 0,667, gemäß den erstellten Bewertungskriterien. Experimente haben gezeigt, dass die Technik in der Lage ist, kritische Elemente zu identifizieren. Zukünftige Versuche werden daher die Struktur des Modells verbessern, um Fehler bei industriellen Big Data vorherzusagen. Der Vorteil besteht jedoch darin, dass es mit unausgewogenen Daten umgehen kann, was in den meisten Branchen üblich ist. Mit zusätzlichen Daten kann das Modell weiter verbessert werden.

Schlüsselwörter: Predictive Analytics; Deep Learning; Datenanalyse; Boolesche Zeitreihendaten; Fehlererkennung; tiefe neuronale Netzwerke; Rekurrente neuronale Netzwerke (RNNs); Long Short-Term Memory (LSTM)-Netzwerke; Synthetic Minority Oversampling TECHNique (SMOTE); Containerkräne; maritime Industrie

check for updates

Zitat: Awasthi, A.; Krpalkova, L.; Walsh, J. Deep Learning-basierte Boolesche Werte, Zeitreihen, Fehlererkennung, und prädiktive Analyse im Containerkranbetrieb . Algorithmen **2024,** 17, **333. https://doi.org/10.3390/a17080333**

Wissenschaftliche Redakteure: Mateus Mendes und Balduíno Mateus

Empfangen: 16. Juni 2024
Überarbeitet: 20. Juli 2024
Akzeptiert: 24. Juli 2024
Veröffentlicht: 1. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.
Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/4.0/).

1. Einleitung

1.1. Hintergrund

Der Containerschifffahrtsmarkt wird im Jahr 2024 ein Volumen von 116,04 Milliarden US-Dollar haben und soll bis 2029 auf 134,03 Milliarden US-Dollar anwachsen [1]. Mit dem Anstieg des globalen Handelsvolumens steigen gleichzeitig **die Anforderungen an die Produktivität und Betriebseffizienz der Hafenanlagen . Neben diesen Anforderungen ist die Gewährleistung der Sicherheit und Zuverlässigkeit des Containerkranbetriebs von größter Bedeutung [2]. Containerkräne spielen eine zentrale Rolle in der modernen maritimen Logistik, da sie den effizienten Warentransport zwischen Schiffen und Ufern in Containerterminals weltweit** ermöglichen **[3]. Die hoch aufragenden Portalkräne sind unverzichtbar, um Container von Schiffen zu lad**

Containerkräne, die entlang der Kaianlage arbeiten, nutzen komplexe Technologien wie Spreizbalken, um Container zwischen Schiffen und Terminals zu transportieren und tragen so zur allgemeinen Effizienz des Hafens bei. Bei diesem Vorgang muss der Kran präzise so angeordnet werden, dass er zu den verschiedenen Buchten des Schiffs passt, um Zugang zu den Containern entlang der Länge des Schiffs zu erhalten. Darüber hinaus können Containerkräne neben dem Schiffsbetrieb auch zum Stapeln von Containern innerhalb des Terminalgeländes eingesetzt werden, um eine Z

Machine Translated by Google

Transport [4]. Die Kräne erzeugen enorme Datenmengen, die analysiert werden können, und es können wertvolle Erkenntnisse gewonnen werden, die eine effizientere, bessere und schnellere Entscheidungsfindung in der Industrie ermöglichen.

In jüngster Zeit konzentriert sich die Forschung zunehmend auf datengesteuerte Fehlerdiagnosemethoden, insbesondere solche auf Basis neuronaler Netze. Dies ist auf Fortschritte in der Sensortechnologie und eine erhöhte Datenspeicherkapazität zurückzuführen, die die Datenverarbeitungsfähigkeiten neuronaler netzbasierter Methoden verbessert haben. Das Versäumnis, Prozessfehler zu erkennen, kann sich nachteilig auf Produktivität, Sicherheit, Ausgabequalität und Wirtschaftlichke
Effiziente und fortschrittliche automatisierte Diagnosegeräte sind in modernen Unternehmen von entscheidender Bedeutung, um abnormales Prozessverhalten zu erkennen, zu diagnostizieren und zu beheben. Das Hauptziel der Fehlererkennung und -isolierung besteht darin, die Bediener umgehend auf potenzielle Probleme aufmerksam zu machen, damit sie die erforderlichen Maßnahmen ergreifen können , um nach dem Auftreten von Problemen einen Systemausfall zu verhindern

Eine rechtzeitige Fehlererkennung und -vorhersage ist von entscheidender Bedeutung für die Optimierung der Kranleistung und die Gewährleistung der Sicherheit beim Containerkranbetrieb.

Die Bedeutung einer rechtzeitigen Fehlererkennung kann aus mehreren wichtigen Perspektiven verstanden werden:

1. **Optimierung von Produktivität und Effizienz: Eine rechtzeitige Fehlererkennung und -vermeidung optimiert die Leistung und Betriebseffizienz von Kränen. Durch die umgehende Identifizierung und Behebung von Fehlern können Betreiber die Auswirkungen auf Produktivität und Durchsatz minimieren und sicherstellen, dass Kräne mit maximaler Kapazität und Effizienz arbeiten [6]. Dies kann durch eine frühzeitige Vorhersage von Fehlern erreicht werden, sodass sie sich nicht zu Störungen entwickeln und Ausfälle verursachen .** Diese Leistungsoptimierung trägt zur **Gesamtproduktivität und Rentabilität des Containerterminalbetriebs bei .**
2. **Vermeidung von Betriebsstörungen: Durch rechtzeitige Fehlererkennung können potenzielle Probleme oder Störungen im Kranbetrieb frühzeitig erkannt werden [7]. Durch die frühzeitige Identifizierung von Fehlern können die Betreiber Korrekturmaßnahmen ergreifen, um Betriebsstörungen wie Kranausfälle oder Verzögerungen beim Güterumschlag zu verhindern. Dieser proaktive Ansatz trägt dazu bei, den reibungslosen Betriebsablauf in Containerterminals aufrechtzuerhalten, Ausfallzeiten zu minimieren und eine effiziente Nutzung der Ressourcen sicherzustellen [8].**
3. **Verbesserung der Sicherheitsprotokolle: Durch die frühzeitige Erkennung von Fehlern können die Betreiber Sicherheitsprotokolle und -verfahren implementieren, um die mit dem Kranbetrieb verbundenen Risiken und Gefahren zu mindern. Durch die umgehende Behebung von Fehlern können die Betreiber Unfälle, Verletzungen und Schäden an Ladung oder Infrastruktur verhindern, das Wohlbefinden des Personals schützen und die Integrität der Terminaleinrichtungen bewahren [9].**
4. **Reduzierung der Wartungskosten: Das frühzeitige Erkennen, Diagnostizieren und Vorhersagen von Fehlern kann dazu beitragen, die mit dem Kranbetrieb verbundenen Wartungskosten zu senken. Durch das Identifizieren potenzieller Fehler, bevor diese zu erheblichen Ausfällen eskalieren, können die Betreiber vorbeugende und präskriptive Wartungsmaßnahmen ergreifen, um zugrunde liegende Probleme zu beheben und die Lebensdauer der Ausrüstung zu verlängern [10]. Dieser proaktive Wartungsansatz trägt dazu bei, Ausfallzeiten, Reparaturkosten und Betriebsstörungen zu minimieren, was letztlich zu Kosteneinsparungen für die Terminalbetreiber führt [11].**
5. **Einhaltung gesetzlicher Normen: Eine rechtzeitige Fehlererkennung gewährleistet die Einhaltung gesetzlicher Normen und Branchenvorschriften für den Kranbetrieb. Durch die umgehende Identifizierung und Behebung von Fehlern können die Betreiber die Einhaltung der von den Aufsichtsbehörden vorgeschriebenen Sicherheitsrichtlinien und Betriebsprotokolle gewährleisten. Diese** Einhaltung gewährleistet nicht nur die Sicherheit von Personal und Anlagen, sondern hilft auch , **potenzielle Bußgelder und rechtliche Haftungen im Zusammenhang mit der Nichteinhaltung zu vermeiden [12,13].**

Zusammenfassend lässt sich sagen, dass eine rechtzeitige Fehlererkennung und -vorhersage von entscheidender Bedeutung sind, um die Kranleistung zu optimieren , die Sicherheit im Containerkranbetrieb zu gewährleisten und den Beteiligten in der gesamten maritimen Lieferkette Vorteile zu verschaffen.

1.2. Forschungsziele, Motivation und Beiträge **Ein entscheidender Aspekt**

zur Gewährleistung der Sicherheit, Effizienz und Zuverlässigkeit im Containerterminalbetrieb ist das Erkennen und Eindämmen von Fehlern, die beim Kranbetrieb auftreten können.

Machine Translated by Google

Fehler beim Betrieb von Containerkränen am Kai können erhebliche Folgen haben, die von möglichen Schäden an Ladung und Infrastruktur bis hin zu Störungen in der Logistik der Lieferkette reichen. Darüber hinaus tragen menschliches Versagen, Gerätefehler und Umwelteinflüsse zur Komplexität der Fehlererkennung bei und stellen eine Herausforderung für die Gewährleistung eines unterbrechungsfreien Betriebs dar [14]. Systemausfälle können durch die Identifizierung und Vorwegnahme von Fehlern verhindert werden, was zu einem robusteren und langlebigeren System führt. Bei der Fehlervorhersage im Betrieb von Containerkränen werden Datenanalyse, maschinelles Lernen und Deep-Learning-Methoden verwendet, um mögliche Fehler oder Anomalien vorherzusagen, die während des Betriebs von Containerkränen auftreten können [6].

Ziel dieser Forschung ist die Entwicklung proaktiver Strategien zur Fehlererkennung und -vorhersage durch Nutzung der Leistungsfähigkeit von Deep Learning, Datenanalyse und prädiktiver Modellierung, um so die Sicherheit, Effizienz und Zuverlässigkeit des Containerterminalbetriebs zu verbessern. Die Lösung dieses Problems erfordert ein umfassendes Verständnis der Betriebsdynamik von Containerterminals und die Anwendung fortschrittlicher KI-Technologien und -Methoden zur Fehlererkennung , -diagnose und -vorhersage [15].

Durch die Analyse verschiedener Datenquellen, darunter Kransensordaten, Betriebsprotokolle vergangener Tage, Wetterbedingungen und andere relevante Aspekte, können prädiktive Modelle erstellt werden, um Probleme im Voraus vorherzusehen und zu minimieren. Die Umsetzung dieser proaktiven Strategie verbessert die Sicherheit, Effektivität und Gesamteffizienz des Containerterminalbetriebs.

Forschungsergebnisse in diesem Bereich haben erhebliche Auswirkungen auf die Schifffahrtsindustrie , Hafenbehörden, Terminalbetreiber und andere am Containerterminalbetrieb beteiligte Interessengruppen . Ziel ist es, die Betriebssicherheit zu erhöhen, Ausfallzeiten zu minimieren, die Ressourcennutzung zu optimieren und die Gesamtleistung des Terminals zu verbessern, indem Fehler in Betriebsdaten effektiv erkannt und behoben werden. In diesem Artikel wird untersucht, wie Deep Learning die Fehlererkennung, -vorhersage und Zeitreihenanalyse im Containerkranbetrieb verbessern kann, was derzeit ein dringendes Bedürfnis der Schifffahrtsindustrie ist und zur Verbesserung der B

Das Potenzial dieser Forschung geht über Containerkräne hinaus und erstreckt sich auch auf andere Branchen mit unausgewogenen Datenproblemen, in denen das Auftreten seltener Ereignisse oder Anomalien erheblich von Nutzen sein k

Auch andere Industriezweige können von dieser Forschung erheblich profitieren, wie etwa die Fertigung, wo sie Geräteausfälle oder -defekte frühzeitig in Produktionslinien erkennt; der Energiesektor, wo sie Anomalien in Stromerzeugungs- oder -verteilungssystemen vorhersagen kann; und der Transportsektor, wo sie Unregelmäßigkeiten in der Fahrzeugleistung oder in Sicherheitssystemen erkennt. Darüber hinaus können Unternehmen gezielt eingreifen und vorbeugende Maßnahmen ergreifen, indem sie kritische Elemente identifizieren, die mit Fehlern in Zusammenhang stehen.

Zum Beispiel:

Wartung: Priorisieren Sie Wartungspläne für Komponenten, die am wahrscheinlichsten Fehler verursachen;

Prozessoptimierung: Passen Sie Betriebsabläufe anhand von Fehlermustern an.

Ressourcenzuweisung: Um Störungen zu minimieren, weisen Sie Ressourcen (Personal, Materialien) strategisch zu.

Diese Forschung ist motiviert durch die Dringlichkeit, die Fehlererkennung und -vorhersage bei Containerkränen in der milliardenschweren Containerschifffahrtsindustrie zu verbessern, die das Rückgrat der gegenwärtigen Weltwirtschaft bildet. Herkömmliche regelbasierte Techniken versagen häufig, weil sie komplizierte, dynamische Einstellungen und ungleichmäßige Datenverteilungen nicht bewältigen können. Deep Learning, insbesondere LSTM-Netzwerke, bietet eine praktikable Lösung für diese Mängel. Durch die Erfassung zeitlicher Zusammenhänge und das Lernen aus früheren Daten können LSTMs komplexe Muster modellieren und die Genauigkeit erhöhen.

1.3. Übersicht über Deep Learning

Deep Learning ist ein Zweig der künstlichen Intelligenz (KI), bei dem Computersysteme trainiert werden , Aufgaben wie das Erkennen von Mustern, Treffen von Entscheidungen und Lösen von Problemen auszuführen, ohne dass sie explizit dafür programmiert wurden. Dabei kommen Algorithmen zum Einsatz , die als neuronale Netzwerke bekannt sind und von der Struktur und Funktion des menschlichen Gehirns inspiriert sin

Datengesteuerte Maschinenzustandsüberwachung wird in modernen

Machine Translated by Google

Produktionssysteme, da kostengünstige Sensoren weit verbreitet und mit dem Internet verbunden sind. Deep Learning bietet wertvolle Methoden zur Verarbeitung und Auswertung großer Betriebsdatenmengen [17]. Deep Learning ermöglicht es Computern, Muster zu erkennen, Informationen zu klassifizieren und hochpräzise Vorhersagen zu treffen, indem sie aus großen Datenmengen lernen.

Beim Deep Learning werden neuronale Netzwerke eingesetzt, die aus zahlreichen Schichten bestehen, sogenannten tiefen Architekturen, um aus Eingaben Wissen zu gewinnen und Vorhersagen zu erstellen. Diese Technologie findet Anwendung in verschiedenen Bereichen, darunter Bild- und Spracherkennung, Verarbeitung natürlicher Sprache, autonome Fahrzeuge und in diesem Fall die Fertigungsindustrie.

Mehrere Schlüsselfaktoren machen den Einsatz von Deep Learning bei der Fehlererkennung, -diagnose und -vorhersage in verschiedenen Branchen, einschließlich des Betriebs von Containerkränen an Kais, von größter Bedeutung: 1. Deep-Learning-Algorithmen wie Convolutional Neural Networks (CNNs) und Recurrent Neural Networks (RNNs) haben außergewöhnliche Fähigkeiten im Umgang mit komplexen und hochdimensionalen Daten bewiesen **[18]. Beim Betrieb von Containerkränen, wo Datenströme von verschiedenen Sensoren, Steuerungssystemen und Betriebsprotokollen in Hülle und Fülle vorhanden sind ,** können Deep-Learning-Modelle diese Daten effektiv verarbeiten, um subtile **Muster und Anomalien zu erkennen,** die auf Fehler hinweisen.

2. **Deep-Learning-Modelle zeichnen sich durch die Erfassung zeitlicher Abhängigkeiten und dynamischer Beziehungen innerhalb sequenzieller Daten aus [19]. Beim Betrieb von Containerkranen, bei dem sich Kranbewegungen, Umschlagprozesse und Umgebungsbedingungen entwickeln, können Deep-Learning-Algorithmen historische Daten analysieren, um zukünftige Zustände vorherzusagen und so eine proaktive Fehlererkennung und -vermeidung** zu ermöglichen **[20].**
3. **Aufgrund ihrer Skalierbarkeit und Anpassungsfähigkeit eignen sich Deep-Learning-Modelle gut für die Echtzeit- und Offline-Überwachung und Entscheidungsfindung in dynamischen Betriebsumgebungen. Durch kontinuierliches Lernen aus neuen Daten und Anpassung an veränderte Bedingungen können Deep-Learning-basierte Fehlererkennungssysteme die Betriebssicherheit erhöhen, die Ressourcennutzung optimieren und potenzielle Risiken im Containerkranbetrieb mindern [21,22].**

Insgesamt liegt die Bedeutung des Einsatzes von Deep Learning zur Fehlererkennung, Diagnose und Prognose im Containerkranbetrieb in seiner Fähigkeit, die Leistungsfähigkeit fortschrittlicher Datenanalysen zu nutzen, um die Betriebssicherheit, Effizienz und Belastbarkeit in der mariti Mit dem technologischen Fortschritt wird der Einsatz von Deep-Learning-Algorithmen immer unverzichtbarer, um einen reibungslosen und sicheren Betrieb von Containerterminals im Ökosystem der globalen Lieferkette zu gewährleisten.

Durch den Einsatz fortschrittlicher Deep-Learning-Techniken zielt die Forschung darauf ab, die inhärenten Komplexitäten der Fehlererkennung und -vorhersage im Containerkranbetrieb zu bewältigen, bei dem Betriebsdaten als diskrete Boolesche Werte im Zeitverlauf dargestellt werden. Der vorgeschlagene Algorithmus versucht, zeitliche Abhängigkeiten und subtile Muster in Booleschen Zeitreihendaten effektiv zu erfassen, um eine genaue Vorhersage potenzieller Fehler vor ihrem Auftreten während des Kranbetriebs zu ermöglichen. Durch empirische Validierung und Leistungsbewertung versucht dieses Papier, die Wirksamkeit und Anwendbarkeit des vorgeschlagenen Deep-Learning-Algorithmus bei der Verbesserung der Betriebssicherheit, Effizienz und Zuverlässigkeit im Containerterminalmanagement zu demonstrieren.

1.4. Arbeitsablauf

In den folgenden Abschnitten der Workflow-Analyse wird der aktuelle Stand der Technik eingehend untersucht, indem die unzähligen Facetten der Fehlererkennung im Containerkranbetrieb untersucht werden . Darüber hinaus werden die methodischen Ansätze, die zur Bewältigung dieser Herausforderung eingesetzt werden, zusammen mit empirischen Erkenntnissen aus relevanten Studien untersucht. Diese umfassende Analyse wird ein differenziertes Verständnis der Fehlervorhersage im Zusammenhang mit dem Containerkranbetrieb ermöglichen und ihre Auswirkungen und potenziellen Weiterentwicklungsmöglichkeiten im akademischen und wissenschaftlichen Diskurs beleuchten.

Der folgende Abschnitt informiert über die Methodik und die Daten, die in dieser Forschung verwendet wurden . Er präsentiert einen neuen, maßgeschneiderten, Python-basierten Algorithmus, der SMOTE integ

Machine Translated by Google

und LSTM-Techniken, um unausgewogene Daten durch benutzerdefinierte Vorverarbeitung, Analyse und Einbindung von Bewertungsmetriken zu beheben. Die Workflow-Struktur ist in Abbildung 1 unten dargestellt:

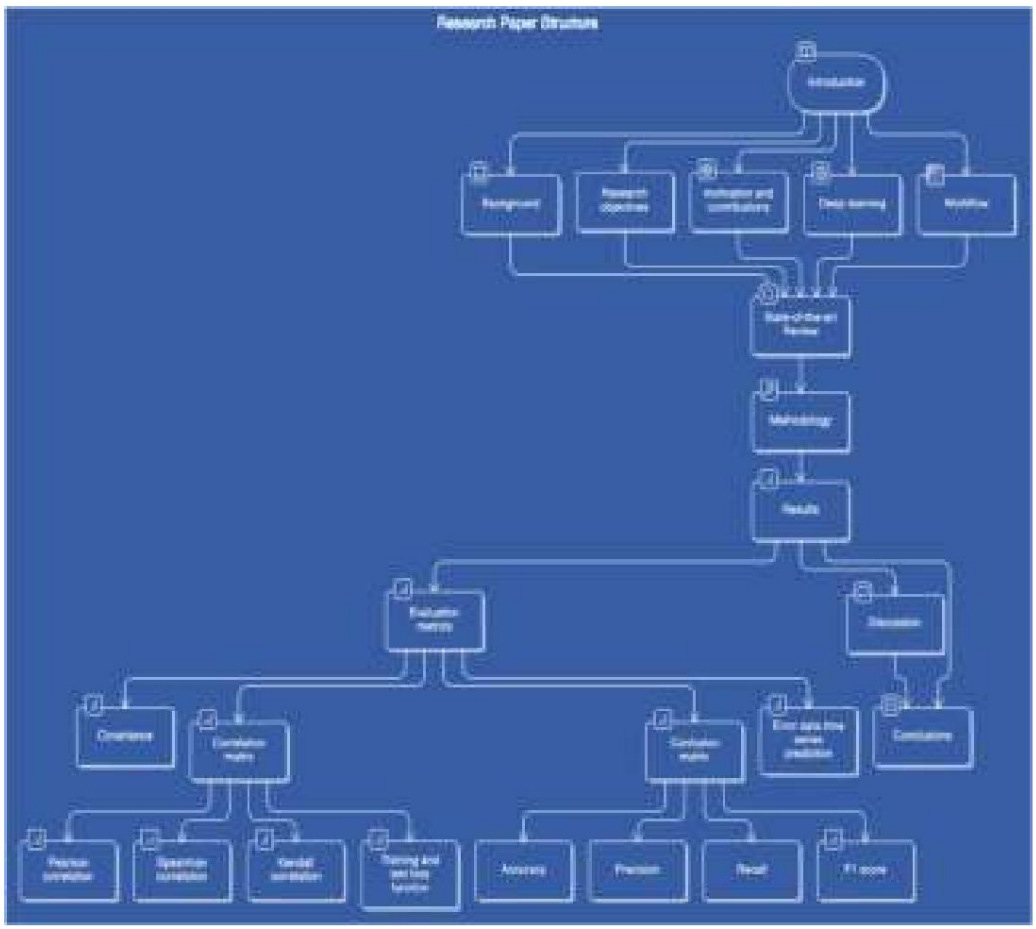


Abbildung 1. Arbeitsablauf bei der Erstellung von Forschungsarbeiten.

2. Überprüfung des aktuellen Stands der Technik

Wie bereits erwähnt, ist die Gewährleistung der Sicherheit und Zuverlässigkeit des Kranbetriebs von größter Bedeutung. Zeitreihendaten aus dem Betrieb solcher Schwermaschinen stellen aufgrund ihrer Größe und komplexen Muster besondere Herausforderungen dar. Die rechtzeitige Identifizierung von Betriebsfehlern kann Unfälle und Ausfallzeiten reduzieren [23]. Modelle des maschinellen Lernens und des Deep Learning haben sich als leistungsstarke Werkzeuge zur Erkennung von Anomalien in solchen Daten erwiesen. Es gibt umfangreiche Literatur zur Fehler- und Ausfallerkennung in ML [24]. Zur Fehlererkennung und -vorhersage **gibt es jedoch sehr wenig . Es ist entscheidend, einen Fehler zu erkennen und vorherzusagen, da ein Fehler ein Vorbote von Fehlern und Ausfällen ist. Abbildung 2 unten veranschaulicht diesen Zusammenhang zwischen Fehler, Defekt und Ausfall.**

Die Erkennung von Betriebsfehlersignalen einer Maschine bezieht sich auf den Prozess der Identifizierung abnormaler oder unerwarteter Signale im Betrieb einer Maschine. Diese Signale können auf Probleme, Fehlfunktionen oder Abweichungen vom normalen Verhalten des Systems oder der Maschine hinweisen. Bei dieser Untersuchung wurden tatsächliche STS-Krandatenbanken an Bord und außerhalb von Häfen und Kränen verwendet, wie z. B. Garantie-, Alarm- und Betriebsdaten, bei denen von der Kran-SPS generierte Fehlersignale verfügbar waren.

Die Bedeutung des Einsatzes von Deep Learning zur Fehlererkennung im Containerkranbetrieb liegt in seiner Fähigkeit, komplexe Muster, zeitliche Abhängigkeiten und nichtlineare Beziehungen in Betriebsdaten effektiv zu erfassen. Die wichtigsten Punkte, die seine Bedeutung unterstreichen, sind die folgenden:

Machine Translated by Google



Abbildung 2. Fehler als Vorbote von Ausfällen [25].

1. **Komplexität der Betriebsdaten: Der Betrieb von Containerkränen erzeugt riesige Datenmengen, darunter Sensorwerte, Kranbewegungen, Frachtumschlag und Umweltbedingungen . Deep-Learning-Algorithmen wie RNNs und CNNs eignen sich hervorragend für die Verarbeitung hochdimensionaler und sequenzieller Daten [18] und** sind daher gut für die Analyse der **komplexen Dynamik von Kranbetrieben geeignet .**
2. **Zeitliche Abhängigkeiten: Kranoperationen weisen zeitliche Abhängigkeiten auf, bei denen Aktionen und Ereignisse zu einem Zeitpunkt nachfolgende Zustände und Ergebnisse beeinflussen. RNNs, insbesondere LSTM-Netzwerke, sind darauf ausgelegt, zeitliche Abhängigkeiten in sequenziellen Daten zu erfassen und daraus zu lernen , wodurch eine genaue Modellierung der zeitlichen Entwicklung von Kranoperationen und die Erkennung abnormaler Muster oder Abweichungen** ermöglicht wird **[2**
3. **Feature-Learning: Deep-Learning-Modelle können relevante Features automatisch aus Rohdaten lernen, sodass kein manuelles Feature-Engineering mehr erforderlich ist. Im Rahmen der Fehlererkennung im Containerkranbetrieb können Deep-Learning-Algorithmen informative Features aus Sensordaten, Kranbahnen und Betriebsprotokollen** extrahieren und **so Anomalien und Abweichungen, die auf Fehler hinweisen, effektiver erkennen.**
4. **Anpassungsfähigkeit an komplexe Umgebungen: Der Betrieb von Containerkränen unterliegt verschiedenen Umweltfaktoren wie Wind, Wetterbedingungen und Seegang, die die Leistung der Kräne beeinflussen und Unsicherheiten mit sich bringen können [28,29]. Deep-Learning-Modelle mit ihrer Fähigkeit, aus vielfältigen und verrauschten Datenquellen zu lernen, weisen Robustheit und Anpassungsfähigkeit an komplexe Betriebsumgebungen auf und verbessern die Zuverlässigkeit und Genauigkeit von Fehlererkennungssystemen [22].**
5. **Skalierbarkeit und Generalisierung: Deep-Learning-Algorithmen sind von Natur aus skalierbar und können große Datenmengen verarbeiten, wodurch sie sich für die Analyse der umfangreichen Datensätze eignen, die durch den Betrieb von Containerterminals generiert werden. Darüber hinaus können Deep-Learning-Modelle unbekannte Daten gut verallgemeinern, was eine effektive Fehlererkennung über verschiedene Terminalkonfigurationen, Krantypen und Betriebsbedingungen hinweg** ern
6. **Kontinuierliches Lernen und Verbesserung: Deep-Learning-Modelle unterstützen kontinuierliches Lernen und Verbesserung durch iteratives Training mit neuen Daten. Wenn zusätzliche Betriebsdaten verfügbar werden, können Deep-Learning-basierte Fehlererkennungssysteme aktualisiert und verfeinert werden, um sich an veränderte Betriebsbedingungen anzupassen und so ihre Wirksamkeit im Laufe der Zeit** zu verbessern **[16,31].**

Zur Fehlererkennung beim Betrieb von STS-Containerkränen wurden verschiedene Methoden vorgeschlagen. Die Autoren von [32] führten eine neue Methode zur Erkennung von Kranschienenverschleiß ein, die die Genauigkeit deutlich verbesserte, während [33] menschliche Fehler bei Turmdrehkranführern identifizierte und bewertete und sich dabei auf die Bedeutung der Überwachung des Antikollisionssystems konzentri Die Autoren von [34] beschäftigten sich mit der Frage der Erkennung von physischen Stößen auf die Ecken von Frachtcontainern während Handhabungsvorgängen unter Verwendung der Aufprallerkennungsmethode. Dabei wurde die Methode der Aufprallerkennung verwendet, um wiederkehrende Aufpralle von Kaikränen auf dieselben Containerstandorte während des Anhängens zu erkennen. Diese gleichzeitigen Aufpralle deuten auf fa

Machine Translated by Google

Verhaken, was innerhalb kurzer Zeit zu erheblichen Metallverformungen führen kann. Die Studie unterstreicht die Notwendigkeit der Erkennung von Schäden an Schiffscontainern für die Sicherheit und Effizienz von Häfen . In [35] entwickelten die Autoren einen schnellen, farbbasierten Objekterkennungsalgorithmus für Assistenzsysteme für Kaikranführer, der die Erkennungsrate bei gleichzeitig hoher Genauigkeit deutlich erhöhte. Zusammengefasst unterstreichen diese Studien die Bedeutung kontinuierlicher Verbesserung von Fehlererkennungsmethoden zur Erhöhung der Sicherheit und Effizienz im Betrieb von STS-Containerkränen. Mehrere Studien haben die Wirksamkeit von Deep Learning für die Fehlervorhersage im Betrieb von Containerkränen nachgewiesen. Eine rechtzeitige Fehlererkennung ist unabdingbar, um die Kranleistung zu optimieren und die Sicherheit zu gewährleisten. Unbemerkte Fehler oder Probleme können die Produktion und Sicherheit gefährden, indem sie Ineffizienzen, Geräteschäden und mögliche Unfälle verursachen **[36].**

Menschliches Versagen ist ein wesentlicher Faktor für Unfälle beim Einsatz mobiler Krane. Eine effektive Kransteuerung und präzise Positionierung hängen von voreingestellten Geschwindigkeiten für jede Richtung ab, die gemäß den Spezifikationen des Kransystems bestimmt werden. Das rechtzeitige Erkennen von Problemen bei Geschwindigkeit, Beschleunigung oder Verzögerung kann die Leistung verbessern und Überlastungen oder Unfälle vermeiden. Die Entwicklung intelligenter Assistenzsysteme, die Fehler in Echtzeit erkennen, kann die Sicherheit des Kranbetriebs erhöhen, indem sie die Um Brückenkräne sicher zu bedienen und Fehlfunktionen der Anlage zu minimieren, ist umfassendes Fachwissen erforderlich. Dennoch ist der Mangel an qualifizierten Bedienern in letzter Zeit Anlass zur Sorge. Eine frühzeitige Erkennung von Fehlern, wie z. B. Probleme mit der Seilspannung oder Unterschiede im Lastgewicht, kann dazu beitragen, Unfälle und Anlagenschäden zu vermeiden und so die Sicherheit zu erhöhen, ohne allein von den Fähigkeiten des Bedieners abhängig zu sein [38].

Um effiziente automatisierte Krantrajektorien erfolgreich zu implementieren, ist es unerlässlich, Fehler umgehend zu identifizieren **und eine benutzerfreundliche Schnittstelle bereitzustellen, die den Bedienern bei der Auswahl und Ausführung von Bewegungen hilft. Die rechtzeitige Identifizierung von Fehlern bei der Trajektorienplanung oder -ausführung kann die Leistung verbessern und Unfälle verhindern [36]. Der derzeit für die Planung von Kranhubwegen eingesetzte manuelle Planungsprozess ist mühsam und fehleranfällig, was vor allem an den dynamischen und überfüllten Bedingungen auf Baustellen liegt. Die Implementierung automatisierter Systeme, die Fehler oder Kollisionen in Echtzeit erkennen können, kann die betriebliche Effizienz und Präzision** verbessern **[39].**

Die Forschung zur Fehlervorhersage im STS-Containerkranbetrieb ist vielfältig und deckt verschiedene Aspekte des Kranbetriebs ab. Die Autoren von [40] konzentrierten sich auf die Analyse von Ermüdungsschäden alternder Stahlstrukturen, ein kritischer Faktor bei der Fehlervorhersage. Die Autoren von [41] identifizierten spezifische Risikofaktoren bei Stauereivorgängen, die als Grundlage für Fehlervorhersagestrategien dienen können . In [42] befassten sich die Autoren mit der Notwendigkeit robuster Steuerungssysteme für Offshore- Containerkräne, die zur Fehlerminimierung beitragen können. Die Autoren von [43] betonten die Bedeutung der Überwachung der strukturellen Integrität bei der Fehlervorhersage, da sie eine potenzielle Methode zur Früherkennung darstellt. Schließlich stellte **[44] eine Verringerung der mittleren Betriebsdauer zwischen Ausfällen aufgrund schwerer Operationen, erhöhter Wartungsvorfälle und Arbeitsstunden fest. Es wurden wichtige Leistungsindikatoren entwickelt, um die Leistung von Schiff-Land-Kranen zu bewerten und Erkenntnisse für zukünftige Verbesserungen beim Betrieb von Containerkrananlagen zu liefern. Die Ergebnisse legen nahe, dass die Leistungsindikatoren als Leitfaden für zukünftige Inspektionen und Verbesserungen im Containerkranbetrieb**

Es werden vielversprechende Fortschritte bei der Verwendung von Deep Learning zur Fehlererkennung und **-vorhersage in Zeitreihendaten erzielt .** Um Probleme wie verrauschte Beschriftung, Ensemblemethoden zur **Fehlererkennung [45] und innovative Modelle wie GRU-erweitertes DCNN zur besseren Fehlerdiagnose in chemischen Prozessen [46] anzugehen, wurden mehrere Strategien vorgeschlagen. Darüber hinaus wurden Deep-Learning-Modelle zur Anomalieerkennung in Telemetriekanälen von Raumfahrzeugen untersucht, was die Notwendigkeit automatisierter Methoden zur effektiven Überwachung von Systemanomalien** hervorhob **[47]. Diese Studien zeigen, wie Deep-Learning-Methoden wie rekurrierende Einheiten und Convolutional Neural Networks die Effizienz, Generalisierbarkeit und Genauigkeit der Fehlererkennung in verschiedenen Industriebereichen verbessern könn**

Deep Learning hat bemerkenswerte Fähigkeiten bei der Analyse komplexer Datenmuster und der Erstellung von Vorhersagen gezeigt. Tiefe neuronale Netzwerke (DNNs) zeichnen sich durch das Lernen von Darstellu

Machine Translated by Google

aus großen Datenmengen, wodurch sie subtile Signale erkennen können, die auf drohende Fehler im Kranbetrieb hinweisen. Techniken wie RNNs und CNNs wurden angepasst, um sequentielle und räumliche Abhängigkeiten in Kransensordaten, Betriebsprotokollen und Umgebungsfaktoren zu modellieren . Zu den wichtigsten Deep-Learning-Architekturen, die für die booleschen Zeitreihendaten der STS-Containerkräne verwendet werden können, gehören die folgenden [18,48–50]: • Rekurrente neuronale Netzwerke (RNNs): RNNs sind gut im Umgang mit sequentiellen Daten und daher für Zeitreihenanalysen geeignet. RNNs können zeitliche Abhängigkeiten in Kranbetriebsdaten effektiv erfassen [51].

• Long Short-Term Memory (LSTM): LSTM-Netzwerke, eine Variante von RNNs, sind darauf ausgelegt, das Problem des verschwindenden Gradienten zu lösen. LSTMs eignen sich besonders gut für die Modellierung langfristiger Abhängigkeiten und wurden erfolgreich zur Vorhersage des Kranverhaltens eingesetzt. • Gated Recurrent Unit (GRU): GRUs sind eine weitere RNN-Variante, die den Speicher beibehält und gleichzeitig rechnerisch effizient ist. GRUs haben sich bei der Vorhersage von Eingangskraftwerten für die Kransteuerung als vielversprechend erwiesen.

RNNs sind eine robuste maschinelle Lerntechnik, die sich hervorragend für die Verarbeitung sequentieller Daten wie Zeitreihen und die Verarbeitung natürlicher Sprache eignet. RNNs enthalten eine Rückkopplungsschleife, die es ihnen ermöglicht, komplizierte Korrelationen in Eingabesequenzen zu lernen und diese Strukturen auf nichtlineare Weise zu nutzen [52]. RNNs werden in verschiedenen Bereichen wie der Stimmungsanalyse, der Prognose des Energiebedarfs und der monatlichen Durchflussvorhersage umfassend eingesetzt [53]. Ein wesentlicher Vorteil von RNNs ist ihre Fähigkeit, Eingabesequenzen unterschiedlicher Länge zu verarbeiten und Ausgabesequenzen unterschiedlicher Länge zu erzeugen. Diese Eigenschaften machen sie für Aufgaben wie Sprachmodellierung, maschinelle Übersetzung und Spracherkennung geeignet [51]. Tiefe neuronale Netzwerke, die in der Lage sind, noch komplexere Darstellungen der Eingabedaten zu

Um das Problem langfristiger Abhängigkeiten zu lösen, das für reguläre RNNs problematisch sein kann, wurden anspruchsvollere RNN-Architekturen wie LSTMs und GRUs entwickelt . Diese Designs erfassen langfristige Abhängigkeiten in den Eingabedaten besser, indem sie Gating-Techniken einsetzen, um Informationen selektiv zu behalten und zu vergessen.

In der verfügbaren Literatur sind die in der Fertigungsindustrie zur Erkennung, Diagnose und Vorhersage boolescher Zeitreihenfehler auf der Grundlage von maschinellem Lernen verwendeten Methoden in Tabelle 1 und die Methoden für Deep Learning in Tabelle 2 aufgeführt :

Tabelle 1. Methoden des maschinellen Lernens.

Methodik	Beschreibung	Referenz
Fehlererkennung und -diagnose (FDD)	Diese Methoden erkennen neu auftretende Unregelmäßigkeiten und prognostizieren deren Folgen für das Systemverhalten.	[55]
Datenbasierte Fehlerdiagnose	Dieser Ansatz verwendet Algorithmen des maschinellen Lernens zur mechanischen Fehlererkennung oder Fehlerprognose in Fertigungsanlagen.	[24]
Eine Reihe von ML-, DL- und DHL Algorithmen	Diese Algorithmen können möglicherweise eine frühzeitige Fehlererkennung durchführen, die zu künftigen Maschinenausfällen führen würden.	[56]
Automatisierter maschineller Lernansatz	Dieser Ansatz wird zur Echtzeit-Fehlererkennung und -diagnose in diskreten Fertigungsmaschinen verwendet.	[57]

Containerkräne sind für Containerterminals unverzichtbar und für die Weltwirtschaft von grundlegender Bedeutung, da der Großteil der weltweiten Waren durch sie hindurchgeht. Ihr Automatisierungsgrad ist jedoch niedriger als der von Lagerkränen und fahrerlosen Transportfahrzeugen (AGVs) [21]. Hier wurden verschiedene Methoden untersucht, aber die Lücke besteht in der weiteren Verfeinerung und Ausweitung der Anwendung von Deep-Learning-Techniken, um spezifische Herausforderungen bei der Fehlererkennung, -diagnose und -vorhersage in der Datenanalyse von Kaicontainerkränen anzugehen. Dies würde letztendlich die Sicherheit, Effizienz und Automatisierung des Hafenbetriebs verbessern. Der wesentliche Teil ist die frühzeitige Fehlerdiagnose, die mit dem ersten Schritt der Fehleridentifizierung und -vorhersage einhergeht. Die vorgeschlagene Methode zielt genau dara

Machine Translated by Google

Tabelle 2. Deep-Learning-Methoden.

Methodik	Beschreibung	Referenz
Deep Learning für Zeitreihen Anomalieerkennung	Dieser Ansatz verwendet spezialisierte Deep-Learning-Modelle zum Erkennen anomaler Muster in Zeitreihendaten.	[58]
Tiefe generative Modelle	Zur Verarbeitung unausgeglichener Zeitreihendaten aus Fertigungsanwendungen werden die Variational Autoencoder-Rekonstruktion entlang des Projektionspfads (VAE-RaPP) und ein Fence Generative Adversarial Network (Fence GAN) eingesetzt.	[59]
Deep Metric-Learning-Modell	Dieses Modell berücksichtigt unausgeglichene Fehlerdaten und eine Quadrupel-Datenpaar-Entwurfsweise.	[60]
ATT-1D CNN-GRU-Modell	Dieses Modell kombiniert ein eindimensionales CNN, GRU, Aufmerksamkeitsmechanismus und KG (Wissensgraph) zur Fehlerdiagnose.	[61]
Adaptives Lernen auf Basis von Deep Learning Fehlervorhersage	Dieser Ansatz integriert ein umfassendes Modul zur Extraktion von Faltungsmerkmalen, ein angepasstes Gating-Modul und ein mehrschichtiges progressives Extraktionsmodul.	[62]

3. Methodik Der hier

für das Training und die Auswertung des Deep-Learning-Modells verwendete Datensatz enthielt die tatsächlichen Daten von Containerkränen an Containerterminals. Er lag in einem proprietären Format vor und konnte nicht direkt gelesen werden. Er wurde mithilfe der Python-Programmierung in ein lesbares Format konvertiert. Er bestand aus den Betriebsdaten der SPS, Sensoren und anderen Berechnungen. Die extrahierten Fehlerdaten waren Boolesch, was bedeutet, dass die Ausgabe 0 oder 1 war.

Softwaresysteme werden zur Diagnose verwendet und speichern SPS-Daten 30 Tage lang, danach werden sie verworfen. Es ist ein wichtiges Werkzeug für den Kranbetrieb und die tägliche Leistungsbewertung. Es bietet eine zuverlässige Datenquelle in Zeitreihen, Rohdaten direkt von den Sensoren und deren Berechnungen. Es erkennt und lokalisiert periodische Fehler im Kran. Es erstellt ungefähr 680 bis 730 monatliche Dateien und speichert Daten pro Stunde. Es speichert ungefähr 23.000 SPS- Variablen in einem durchschnittlichen Kran, kann aber in einigen Kränen viel mehr enthalten (z. B. 31.521 SPS-Variablen in STS CC2124). Das sind 518,4 Milliarden Proben pro Monat (20.000 Daten bei 100 ms). Die Datenmerkmale wurden ausgewählt (in diesem Fall Fehler und einige andere Identifikationssignale) und die Daten aus diesen Variablen wurden mit Python-Code in CSV-Dateien in Excel konvertiert. Die Fehlerdaten enthielten ungefähr 3000 Variablen. Nur zwei Spalten waren nicht boolesch, einschließlich der letzten für Zeitstempel/Zeitreihen (mit falschem Format). Es gab ungefähr 3000 Spalten und mehr als 16.100.000 (16 Millionen) Zeilen.

Bei der Analyse des Betriebsdatensatzes wurden acht Kräne aus verschiedenen Häfen untersucht. Die detaillierte Analyse von drei Kränen ergab schließlich, dass die meisten 2852–2878 Fehlersignale aufwiesen. Allerdings schalteten sich nur 3 bis 6 Signale ein, was darauf hindeutet, dass bei den Kränen tatsächlich ein Fehler aufgetreten ist. Dies deutet auf die Seltenheit von Fehlerereignissen und den Mangel an relevanten Daten hin, die weiter analysiert werden könnten, was eine erhebliche Herausforderung für das Projekt darstellte.

Um die Daten aus den von den Kränen erfassten Proletariatsdaten zu extrahieren, wurde ein eigener Python-Code entwickelt. Die daraus resultierenden Excel-Dateien wurden dann zur weiteren Datenverarbeitung verwendet

Der maßgeschneiderte Algorithmus wurde unter Verwendung von SMOTE- und LSTM-Algorithmen in der Python-Programmierung entwickelt, zusammen mit angepassten Vorverarbeitungsverfahren, die unausgewogene Daten berücksichtigen und Bewertungsmetriken integrieren.

Vorverarbeitung der Daten: Zum Extrahieren boolescher Daten wurde eine benutzerdefinierte Vorverarbeitung eingesetzt. Zunächst wurden die unverarbeiteten Daten geladen, die aus Zeitreihen und Sensorwerten bestanden . Die Sicherstellung der Datenqualität war von größter Bedeutung, sodass alle fehlenden Werte und Ausreißer berücksichtigt wurden. Anschließend wurden Boolesche Merkmale extrahiert, die aus binären Flags oder Wahr/ Falsch-Werten bestanden.

Umgang mit unausgewogenen Daten: Im Falle eines unausgewogenen Datensatzes, wenn eine Klasse erheblich unterrepräsentiert ist, kann SMOTE eingesetzt werden. Diese Methode erzeugt künstliche Stichproben für die unterrepräsentierte Kategorie, indem durch Interpolation vorhandener Datenpunkte neue Instanzen erstellt werden. Ziel ist es, eine ausgewogene Klassenverteilung zu erreichen, die eine gerechte Vertretung beider Klassen garantiert. Als nächstes wird

Machine Translated by Google

künstlich erzeugte Proben. Anschließend wird der Datensatz zu Trainings- und Validierungszwecken in separate Teilmengen aufgeteilt. Die Daten wurden zum Testen des Modells im Verhältnis 7:3 aufgeteilt, wobei 70 % die Trainingsdaten und die restlichen 30 % die Testdaten waren. Diese Daten wurden in ein dreidimensionales Format umgeformt, wobei die Dimensionen die Proben, Zeitschritte und Merkmale waren.

Aufbau des LSTM-Modells: Die Entwicklung der LSTM-Struktur war entscheidend. Ein LSTM-Neuralnetz wird mithilfe von Eingabeschichten aufgebaut, die sequentielle Daten darstellen, LSTM-Schichten, die zeitliche Abhängigkeiten erfassen, und einer Ausgabeschicht, die eine binäre Klassifizierung dur Das Modell wurde durch Auswahl einer Verlustfunktion, beispielsweise binäre Kreuzentropie, eines Optimierers, beispielsweise Adam, und einer Bewertungsmetrik, beispielsweise Genauigkeit, kompiliert. Das Modell hatte eine Batch-Größe von 50 und wurde 40 Epochen lang trainiert, bestehend aus 2 LSTM-Schichten und einer dichten Schicht für die endgültige Ausgabe. Zu Beginn wurde ein Startwert bereitgestellt, der dabei half, reproduzierbare Ergebnisse für das Modell zu generieren, sodass nicht bei jedem Durchlauf des Modells zufällig unterschiedliche Ergebnisse generiert wurden . Nachdem das Modell trainiert war, wurde seine Leistung anhand des **Validierungssatzes bewertet** und bei Bedarf Anpa

Modellbewertung: Das Modell wurde anhand einer Reihe von Metriken wie Genauigkeit, Präzision, Rückruf und F1-Score bewertet. Die Konfusionsmatrix bietet eine umfassende Aufschlüsselung der Vorhersagen, einschließlich wahrer Positiv-, wahrer Negativ-, falscher Positiv- und falscher Negativ-Vorhersagen. Die Bewertungsmetriken analysierten das Verhalten des Modells und nahm

Das Flussdiagramm des benutzerdefinierten Algorithmus ist unten in Abbildung 3 dargestellt.

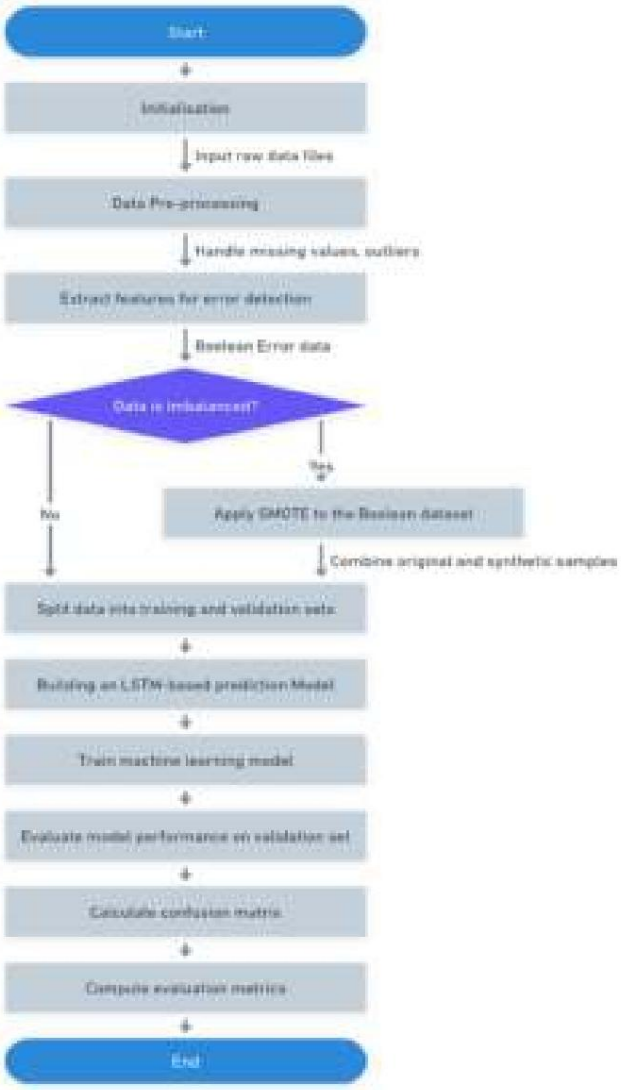


Abbildung 3. Benutzerdefiniertes Algorithmus-Flussdiagramm basierend auf SMOTE und LSTM.

Machine Translated by Google

4. Ergebnisse

Dieser Abschnitt stellt die experimentellen Ergebnisse der Anwendung des vorgeschlagenen Deep Learning dar. ing-Algorithmus für die Fehlererkennungs- und -vorhersageaufgabe.

4.1. Bewertungsmaßstäbe

Bewertungsmetriken sind quantitative Maße, die zur Bewertung der Leistung eines Modells oder Algorithmus verwendet werden, insbesondere im maschinellen Lernen, in der Statistik und bei der Datenanalyse. Diese Metriken helfen dabei, die Leistung eines Modells zu bestimmen und Verbesserungen und Optimierungen vorzunehmen. Die in den folgenden Metriken als var_xxxx angezeigten Variablen stellen spezifische Fehlersignale dar, die von Containerkränen generiert werden.

4.1.1. Kovarianz

Kovarianz ist eine durchgängig additive Beziehung zwischen den beiden Datenproben. Diese Beziehung kann zwischen zwei Variablen zusammengefasst werden, die als Kovarianz bezeichnet werden. Sie wird als Durchschnitt des Produkts zwischen den Werten aus jeder Probe berechnet, wobei die Werte zentriert wurden (der Mittelwert abgezogen wurde) [63]. Kovarianz bedeutet keine Kausalität – sie misst nur die statistische Beziehung. Das Verständnis der Kovarianz kann beim Analysieren von Daten helfen, Muster und Abhängigkeiten zu erkennen. Eine Kovarianzmatrix stellt die Beziehungen zwischen Variablenpaaren in einem Datensatz dar. Sie quantifiziert das Ausmaß, in dem zwei Variablen gemeinsam variieren.

$$\text{Cov}(X,Y) = (\text{Summe}(x - \text{Mittelwert}(X)) \times (y - \text{Mittelwert}(Y))) \times 1/(n - 1) \quad (1)$$

In dieser Heatmap entspricht jede Zelle der Kovarianz zwischen zwei Variablen. Die diagonalen Zellen (von oben links nach unten rechts) stellen die Kovarianz einer Variablen mit sich selbst dar. Diese Werte sind immer positiv. Die nichtdiagonalen Zellen (oben rechts und unten links) stellen die Kovarianz zwischen verschiedenen Variablen dar. Diese Werte können positiv oder negativ sein. Die Heatmap verwendet Farbintensität, um Kovarianzwerte darzustellen. Dunkelrot zeigt eine starke positive Kovarianz (nahe 1) an, was bedeutet, dass die Variablen dazu neigen, gemeinsam zuzunehm Dunkelblau zeigt eine negative Kovarianz (nahe −1) an, was auf eine inverse Beziehung schließen lässt. Hellere Schattierungen stellen eine schwächere Korrelation dar (näher an 0).

Die Heatmap oben in Abbildung 4 zeigt, dass die Beschriftungen der x- und y-Achse bestimmten Variablen entsprechen (z. B. „var_2663“ und „var_831“). var_2663 steht für .tINTF_General_Visu.tLaneProtection.tCover3.tMotor4.tMotor.tPT100.tAlarm.bStatus und var_831 für .tINTF_ General_Visu.tLaneProtection.tCover1.tMotor4.tCB.tSignalWithError.tError.bHasError.

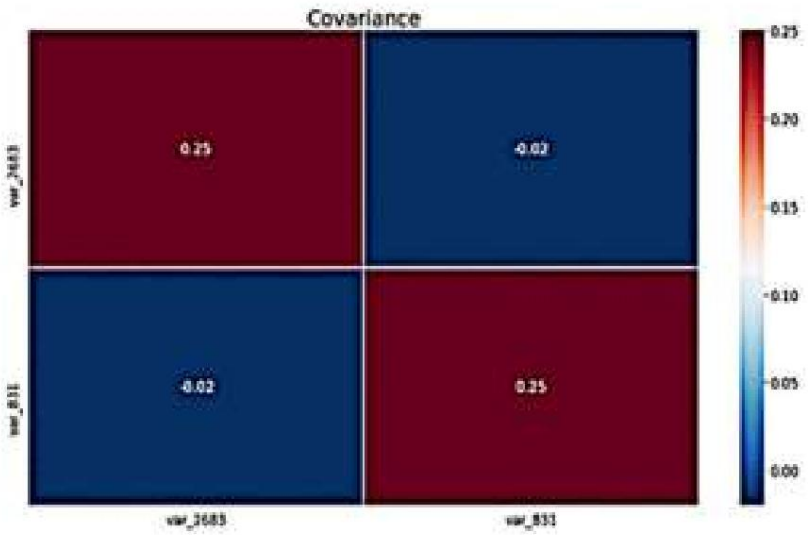


Abbildung 4. Kovarianz.

Machine Translated by Google

Der obere linke Quadrant ist dunkelrot gefärbt, was auf eine robuste positive Kovarianz (etwa 0,25) zwischen var_2663 und sich selbst hinweist. Der untere rechte Quadrant zeigt ebenfalls eine solide positive Kovarianz für var_831. Die hellblauen Quadranten außerhalb der Diagonale deuten auf eine schwache inverse Beziehung (etwa ÿ0,02) zwischen var_2663 und var_831 hin, was darauf hinweist, dass Motor 4 eine positive Kovarianz mit dem Temperatursensorfehler im Steuerblock für Motor 4 aufweist.

4.1.2. Korrelationsmatrix

Die Korrelationsmatrix ist eine häufig verwendete multivariante Technik zum Bestimmen stark verbundener Variablen. Dieser Aspekt des Projekts konzentrierte sich auf das Korrelieren aller Variablen in den Fehlersignalen und einigen Betriebssignalen, die als abhängige Variablen betrachtet werden, wie z. B. Kran an, Kran aus und Drehverschluss verriegelt. Fehlervariablen mit nur einem eindeutigen Wert, 0, wurden entfernt, da homogene Variablen nur einen eindeutigen Wert enthalten und keine zusätzlichen Informationen zum Modell beitragen. Die Korrelation zwischen den Variablen wurde analysiert und visualisiert, um mehr Einblick in die Daten zu geben. Die Korrelation wurde verwendet, um die Beziehung zwischen Variablen zu bestimmen. Sie kann positiv sein, was bedeutet, dass sich beide Variablen in die gleiche Richtung bewegen, oder negativ, was bedeutet, dass, wenn der Wert einer Variablen steigt, die Werte der anderen Variablen sinken. Die Korrelation kann auch neutral oder null sein, was bedeutet, dass die Variablen keine Beziehung haben oder nicht miteinander verbunden sind.

Pearson-Korrelation

Der Korrelationskoeffizient nach Pearson wird berechnet, indem die Kovarianz zweier Variablen durch das Produkt ihrer Standardabweichungen geteilt wird. Es handelt sich um die Normalisierung der Kovarianz zwischen den beiden kontinuierlichen Variablen, um einen interpretierbaren Wert zu erhalten. Er quantifiziert den Grad, in dem eine Änderung einer Variablen mit einer Änderung einer anderen Varia

$$\text{Pearson-Korrelation (r)} = \text{Kovarianz(X,Y)}/(\text{Standardabweichung(X)} \times \text{Standardabweichung(Y)} \quad (2)$$

Diese Heatmap, die in Abbildung 5 oben gezeigt ist, vergleicht die beiden Variablen var_2683, d. h. .tINTF_General_Visu.tLaneProtection.tCover3.tMotor4.tMotor.tPT100_Group.tPT100_ 2.tError.bHasError und „var_831“, d. h. .tINTF_General_Visu.tLaneProtection.tCover1. tMotor4.tCB.tSignalWithError.tError.bHasError. Die Heatmap zeigt den Korrelationskoeffizienten zwischen diesen Variablen. Die diagonalen Quadranten (wo eine Variable mit sich selbst verglichen wird) zeigen einen Korrelationskoeffizienten von 1 (wie erwartet). Die nicht diagonalen Quadranten zeigen eine schwache negative Korrelation von ungefähr ÿ0,079. Die Farbskala reicht von ÿ1 bis 1. Dunkelrot **steht für stärkere negative Korrelationen, während Dunkelblau stärkere positive Korrelationen anzeigt.**

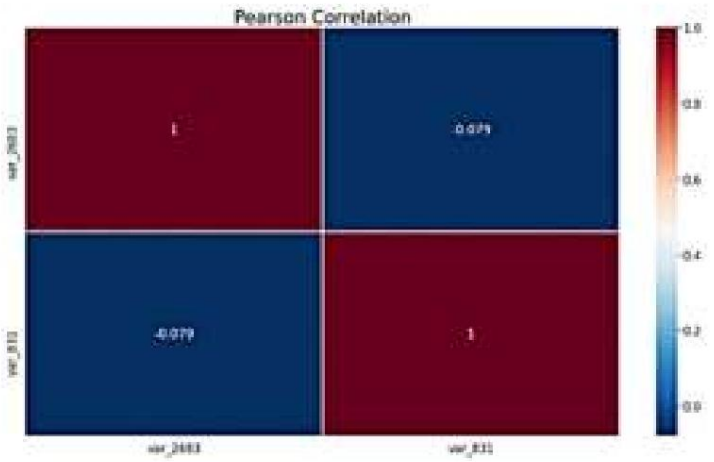


Abbildung 5. Pearson-Korrelation.

Machine Translated by Google

Spearman-Korrelation Die

Spearman-Korrelation wurde aus der relativen Rangfolge der Werte jeder Stichprobe berechnet. Dieser Ansatz wird häufig in der nichtparametrischen Statistik verwendet, beispielsweise in statistischen Methoden, bei denen keine bestimmte Verteilung der Daten, wie etwa die Gauß-Verteilung, angenommen wird [64]. Dieser Koeffizient misst die Stärke und Richtung der monotonen Beziehung zwischen zwei rangierten Variablen. Er ist hilfreich bei der Identifizierung und Bewertung nichtlinearer Beziehungen und ordinaler Daten.

SpearmanKorrelation = Kovarianz(Rang(X),Rang(Y))/std(Rang(X)) × std(Rang(Y)) (3)

Diese Heatmap, die oben in Abbildung 6 gezeigt wird, ist in vier Quadranten unterteilt, um den Spearman- Korrelationskoeffizienten zwischen zwei Variablen anzuzeigen, die auf der Y-Achse als var_2683 —.tINTF_ **General_Visu.tLaneProtection.tCover3.tMotor4.tMotor.tPT100_Group.tPT100_2.tError.bHasError** und auf der **X-Achse als var_831—.tINTF_General_Visu.tLaneProtection.tCover1.tMotor4.tCB. tSignalWithError.tError.bHasError bezeichnet sind. Die beiden diagonalen Quadranten zeigen eine perfekte positive Korrelation (einen Wert von 1) für jede Variable mit sich selbst (angezeigt durch die rote Farbe und die ff- Zahl 1). Die nicht diagonalen Quadranten zeigen eine sehr niedrige negative Korrelation von ÿ0,079 zwischen den beiden Variablen (angezeigt durch Dunkelblau).**

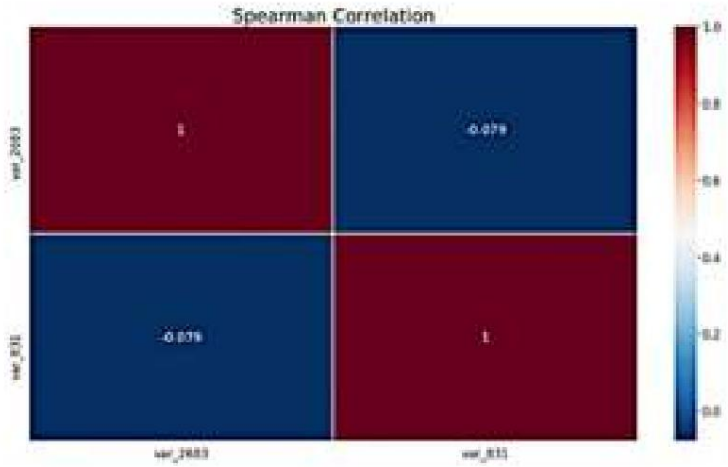


Abbildung 6. Spearman-Korrelation.

Kendall-Korrelation

Der Kendall-Tau-Rangkorrelationskoeffizient misst den Grad der Ähnlichkeit zwischen zwei **Ranggruppen , die derselben Gruppe von Objekten zugewiesen wurden. Im Gegensatz zum Spearman-Koeffizienten misst der Kendall-ffTau -Koeffizient nur die Richtungsübereinstimmung, nicht die Rangunterschiede.**

Kendall-Korrelation = (Anzahl übereinstimmender Paare ÿ **Anzahl nicht übereinstimmender Paare)/n(n** ÿ **1)/2** **(4)**

Die in Abbildung 7 **dargestellte Kendall-Korrelations-Heatmap bestätigt die Ergebnisse der anderen Korrelationen. Sie konzentriert sich auf dieselben beiden Variablen wie die anderen Korrelations-Heatmaps, d. h. var_2683 —.tINTF_General_Visu.tLaneProtection.tCover3.tMotor4.tMotor.tPT100_ Group.tPT100_2.tError.bHasError auf der Y-Achse und var_831—.tINTF_General_ Visu.tLaneProtec Die Quadranten oben links und unten rechts sind dunkelrot, was auf eine perfekte Korrelation von 1 hinweist. Das ist zu erwarten, da jede Variable mit sich selbst korreliert. Die Quadranten oben rechts und unten links sind blau, was auf eine schwache negative Korrelation (ungefähr ÿ0,079) zwischen var_2683 und var_831 hinweist.**

Machine Translated by Google

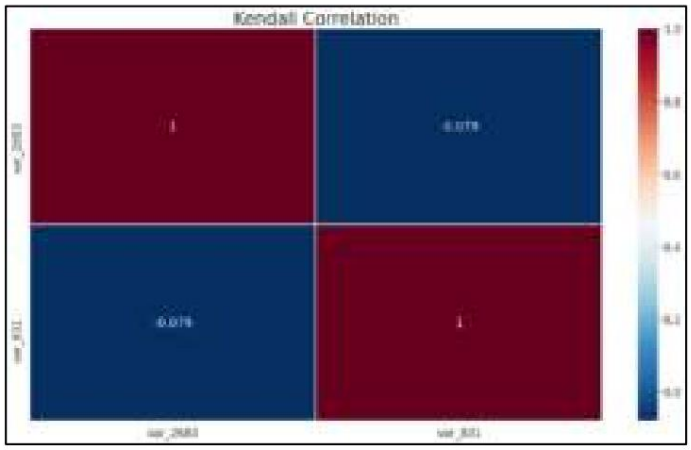


Abbildung 7. Kendall-Korrelation.

4.1.3. Trainings- und Testverlustfunktion

Beim maschinellen Lernen sind Trainingsverlust und Testverlust (oder Validierungsverlust) wichtige Kennzahlen, die zur Bewertung der Leistung eines Modells während des Trainingsvorgangs verwendet werden. Diese Verlustfunktionen helfen dabei zu messen, wie gut das Modell mit den Trainings- und unbekannten Daten (Test- oder Validierungsdaten) abschneidet. Der Trainingsverlust ist der Fehler oder die Diskrepanz zwischen den vorhergesagten und tatsächlichen Ausgaben des Trainingsdatensatzes. Er bietet ein Maß dafür, wie gut das Modell zu den Trainingsdaten passt. Der Testverlust (oder Validierungsverlust) ist der Fehler zwischen den vorhergesagten und tatsächlichen Ausgaben des Test- oder Validierungsdatensatzes. Dieser Datensatz wird während des Trainings nicht verwendet, sondern beiseite gelegt, um die Leistung des Modells mit unbekannten Daten zu

Der Wert des mittleren quadratischen Fehlers (RMSE) wird berechnet. Die Quadratwurzel des Mittelwerts Der quadrierte Fehler liefert ein Maß in denselben Einheiten wie die Zielvariable.

Das Diagramm in Abbildung 8 zeigt zwei Linien: Training (blau) und Test (orange). Die horizontale Achse stellt die Anzahl der Epochen dar, die von 0 bis 40 reicht, während die vertikale Achse den Verlust darstellt, der von 0,0 bis 0,7 reicht. Beide Linien zeigen zunächst einen starken Rückgang des Verlusts während der frühen Epochen, was auf effektives Lernen aus den Trainingsdaten hinweist. Im Laufe der Epochen flachen beide Linien allmählich ab, was auf abnehmende Erträge hindeutet.
Lernen und das Risiko einer Überanpassung, wenn der Testverlust zunimmt. Die Trainingslinie stellt den Verlust der Trainingsdaten dar und die Testlinie den Verlust unsichtbarer Testdaten.
Die Überwachung beider Linien hilft bei der Bewertung der Modellleistung und der Generalisierungsfähigkeit und unterstützt bei der Diagnose von Problemen wie Überanpassung, Unteranpassung und Konvergenz. Diese Grafik ist ist bei der Bewertung von ML-Modellen akademisch relevant und wird zur Optimierung von Trainingsstrategien und Verbesserung der Modellleistung verwendet.

4.1.4. Verwirrungsmatrix

Eine Konfusionsmatrix ist ein grundlegendes Tool zur Bewertung der Leistung von Klassifizierungsmodellen . Sie bietet einen umfassenden Überblick darüber, wie gut ein Modell verschiedene Klassen vorhersagt. Die Konfusionsmatrix basiert auf einem binären Klassifizierungsproblem mit zwei Klassen: 0 und 1. Diese Klassen können verschiedene Ergebnisse darstellen, wie z. B. bestanden/nicht bestanden, ja/nein oder wahr/falsch. Die Matrix ist in vier Quadranten unterteilt: • Wahre Negative (TN):

Instanzen, die korrekt als Klasse 0 vorhergesagt wurden. Dies sind erwünschte Vorhersagen – Instanzen werden korrekt als negativ (Klasse 0) identifiziert;

- **Falsch-Positive (FP): Instanzen, die fälschlicherweise als Klasse 1 vorhergesagt werden, obwohl sie tatsächlich der Klasse 0 angehören. Dabei handelt es sich um Fehler – Instanzen, die fälschlicherweise als positiv (Klasse 1) klassifiziert werden, obwohl sie es nicht sind.**
- **Falsch-Negative (FN): Instanzen, die fälschlicherweise als Klasse 0 vorhergesagt werden, obwohl sie tatsächlich der Klasse 1 angehören. Auch diese sind Fehler – Instanzen, die fälschlicherweise als negativ klassifiziert werden, obwohl sie positiv sind.**

Machine Translated by Google

• **True Positives (TP): Instanzen, die korrekt als Klasse 1 vorhergesagt wurden. Dies sind erwünschte Vorhersagen, bei denen Instanzen korrekt als positiv identifiziert werden.**

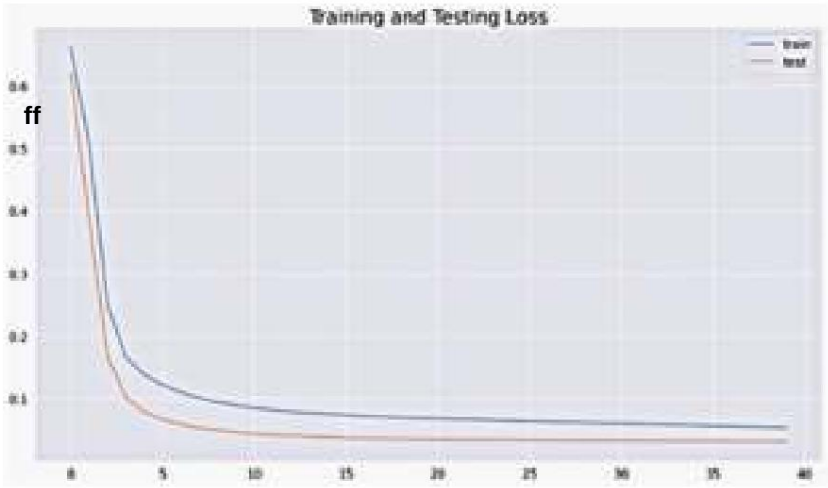


Abbildung 8. Die Trainings- und Testverlustfunktion.

Die in Abbildung **9** **dargestellte Konfusionsmatrix fasst die Leistung des Klassifizierungsmodells zusammen. Sie bietet Einblicke, wie gut die Vorhersagen des Modells mit den tatsächlichen Ergebnissen übereinstimmen.**

1. **True Negative (TN): Das Modell hat negative Instanzen korrekt vorhergesagt (das tatsächliche Ergebnis war negativ). Es gibt 997 True Negatives; 2. False**

Positive (FP): Das Modell hat positive Instanzen falsch vorhergesagt (das tatsächliche negative Ergebnis). Hier gibt es neun False Positives; 3. False

Negative (FN): Das Modell hat negative Instanzen falsch vorhergesagt (das tatsächliche positives Ergebnis). Die Matrix zeigt vier falsch-negative Ergebnisse;

4. **True Positive (TP): Das Modell hat positive Fälle richtig vorhergesagt (das tatsächliche Ergebnis war positiv). In dieser Matrix gibt es vier True Positives.**

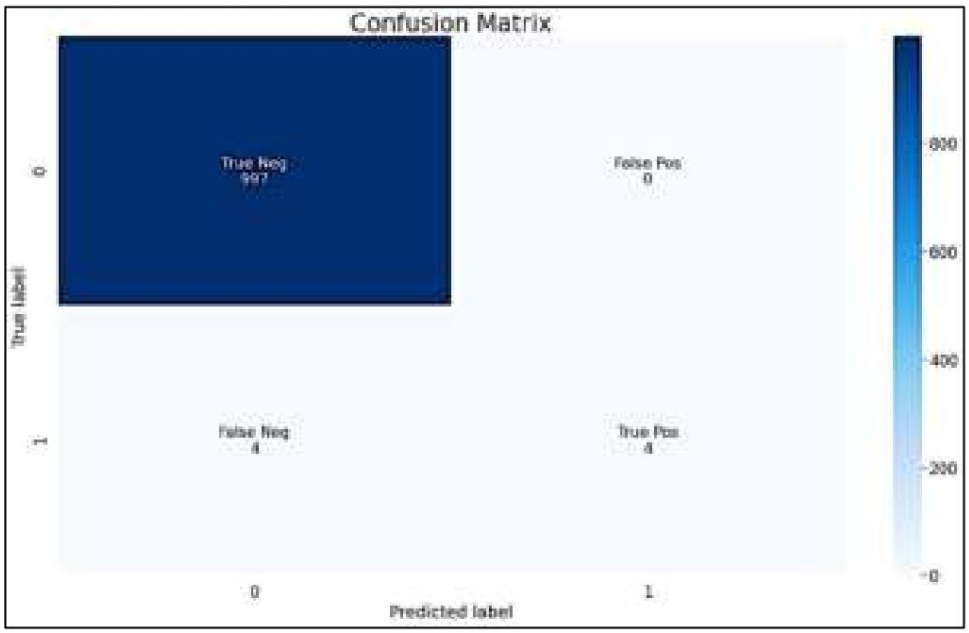


Abbildung 9. Die Konfusionsmatrix.

Machine Translated by Google

4.1.5. Genauigkeit, Präzision, Rückruf, F1-Score

Die in Abbildung **10** unten **dargestellten Genauigkeit, Präzision, Rückruf und F1-Werte waren ebenfalls berechnet und gedruckt für die prognostizierten Datenauswertungsmetriken. Diese Modellleistungen wurden dann in einer Textdatei gespeichert.**

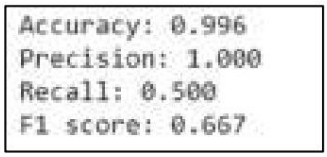


Abbildung 10. Ergebnisse der Auswertungsmetriken.

Genauigkeit: Die Anzahl der richtigen Vorhersagen wird durch die Gesamtzahl geteilt von Vorhersagen.

$$\text{Genauigkeit} = (\text{Wahre Positive} + \text{Wahre Negative})/\text{Gesamtzahl der Instanzen} \quad (5)$$

Präzision: Hiermit wird die Genauigkeit positiver Klassenvorhersagen gemessen, indem wie viele von ihnen tatsächlich zur positiven Klasse gehören.

$$\text{Präzision} = \text{Wahre Positive}/(\text{Wahre Positive} + \text{Falsche Positive}) \quad (6)$$

Recall (Sensitivität): Hiermit wird der Anteil korrekter positiver Vorhersagen berechnet aller positiven Beispiele im Datensatz.

$$\text{Rückruf} = \text{Wahre Positive}/(\text{Wahre Positive} + \text{Falsche Negative}) \quad (7)$$

F1-Score: Dies bietet einen einzelnen Score, der Präzision und Rückruf in einer Zahl vereint. Es ist ein harmonisches Mittel aus Präzision und Rückruf.

$$\text{F1-Score} = 2 \times ((\text{Präzision} \times \text{Rückruf})/(\text{Präzision} + \text{Rückruf})) \quad (8)$$

Genauigkeit: Das Erreichen einer Genauigkeit von 0,996 weist auf eine insgesamt solide Leistung hin.

Präzision: Ein Präzisionswert von 1,00 deutet darauf hin, dass das Modell selten wahre Fehler als Nicht-Fehler.

Rückruf: Der Rückrufwert von 0,500 zeigt an, dass das Modell nur die Hälfte der tatsächlichen Fehler. Dieser Kompromiss zwischen Präzision und Trefferquote kommt bei unausgewogenen Datensätzen häufig vor.

F1-Score: Der F1-Score (0,667) gleicht Präzision und Rückruf aus und bietet eine umfassende Sicht auf die Wirksamkeit des Modells.

4.2. Vorhersage von Zeitreihen mit Fehlerdaten

Das Endergebnis ist die Vorhersage der folgenden n Sekunden an Daten, in denen der Benutzer liefert die n Sekunden als Eingabe am Anfang und die entsprechenden Daten werden gespeichert in eine CSV-Datei. Dazu werden die Daten der nächsten 1 s für die nächsten 60 s vorhergesagt mit Hilfe von historische Daten für die Analyse verfügbar. Die Ausgabe aus der vorhergesagten CSV-Datei für einen der Fehlervariablen sind unten in Tabelle 3 aufgeführt:

Tabelle 3. Vorhergesagte Fehlerdaten.

Zeitstempel	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

Tabelle 3. Fortsetzung

Zeitstempel	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

Tabelle 3. Fortsetzung

Zeitstempel	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

5. Diskussion

Für eine Optimierung der Kranleistung sind eine rechtzeitige Fehlererkennung und -vorhersage von entscheidender Bedeutung. Sie sind von entscheidender Bedeutung für die Sicherheit, indem sie Unfälle verhindern, die Ausrüstung vor Schäden und die Reduzierung menschlicher Fehler. Diese Praktiken steigern die Produktivität und geringere Betriebskosten.

Es ist wichtig, die Grenzen der prädiktiven Analytik zu verstehen. In erster Linie Fortschritte ist ohne einen umfangreichen und qualitativ hochwertigen Trainingsdatensatz unmöglich. Zweitens ist ein Eine klare Beschreibung des zu prognostizierenden Konzepts und der vergangenen Ereignisse ist entscheidend. Herausforderungen in der prädiktiven Analytik sind die Identifizierung einer Nadel im Heuhaufen und die Vorhersage ihrer Vorkommen. Es gibt jedoch gewisse Einschränkungen bei traditionellen Methoden. Diese Einschränkungen im Containerkranbetrieb sind wie folgt:

1. Vom Menschen verursachte Fehler: Menschliche Bediener spielen bei Kaikranen immer noch eine bedeutende Rolle Operationen trotz Automatisierungsfortschritten. Ihre Entscheidungen und Handlungen können dazu führen, Fehler, die die Produktivität und Sicherheit beeinträchtigen [64].
2. Komplexe Dynamik: Containerkräne arbeiten in einer dynamischen Umgebung mit unterschiedlichen Containergewichte, Windverhältnisse und Schiffsbewegungen. Die Vorhersage von Fehlern in solchen ein komplexes System ist eine Herausforderung [65–67].
3. Effizienz und Ausfallzeiten: Die Minimierung von Ausfallzeiten aufgrund von Fehlern ist von entscheidender Bedeutung. Durch Prognosen können kostspielige Verzögerungen vermieden und die Kranauslastung optimiert werden [68,69].
4. Unausgewogene Daten: Ein Datensatz wird als unausgewogen bezeichnet, wenn es erhebliche Unterschiede gibt in der Anzahl der Beispiele jeder Klasse, was dazu führt, dass eine oder mehrere Klassen im Datensatz unterrepräsentiert [70]. Dieses Phänomen wurde in mehrere reale Datensätze, die in industriellen Umgebungen gesammelt wurden.

Daher wurde im Rahmen dieser Forschung ein neuer, maßgeschneiderter Algorithmus entwickelt, um diese Einschränkungen und Herausforderungen, indem wir uns auf die Erkennung und Vorhersage von Fehlern konzentrieren, bevor diese in Störungen und Ausfälle umwandeln.

Die folgenden Punkte befassen sich mit allen Herausforderungen und Einschränkungen, die während die Umsetzung:

1. Da es einen erheblichen Zeitunterschied zwischen den Datenarraydateien gab, Array-Dateien konnten nicht zur Analyse zusammengeführt werden; stattdessen wurde die Analyse für jede einzelne durchgeführt Datenarraydatei separat.

Machine Translated by Google

2. **Für die Zeitreihenprognose wurden die Daten von Millisekunden auf Sekunden herunterskaliert, da die Daten in Millisekunden nicht kontinuierlich waren, d.ÿh. (1, 2, 3, ..., 999).**
3. **Es besteht möglicherweise eine hohe Fehlerwahrscheinlichkeit, da die Datensatzgröße klein war, sodass das LSTM-Modell neue Daten nicht richtig und korrekt vorhersagen konnte. Durch das Hinzufügen weiterer Daten kann dieses Problem behoben werden.**

Trotz der vielversprechenden Ergebnisse bleibt die Implementierung von Deep-Learning-Modellen zur Fehlervorhersage im realen Kranbetrieb eine Herausforderung. Dazu gehören Probleme mit der Datenqualität, der Interpretierbarkeit komplexer Modelle und der Integration in vorhandene Systeme. Zukünftige Forschungsrichtungen können diese Herausforderungen durch fortschrittliche Modellarchitekturen, Datenfusionstechniken und gemeinsame Anstrengungen zwischen akademischen und industriellen Interessenvertretern angehen. Die Wege für zukünftige Forschung können mit verbesserter Datenqualität und -quantität vorangetrieben werden, die derzeit hinterherhinken.

6. Schlussfolgerungen

Die Forschung zu Containerkränen umfasst verschiedene Aspekte im Zusammenhang mit ihrer Konstruktion, ihrem Betrieb und ihren Auswirkungen auf die Hafenlogistik. Es gibt jedoch einen eklatanten Mangel an Studien, die sich auf die frühzeitige Fehlererkennung zur Vermeidung von Störungen und Ausfällen konzentrieren. Die rechtzeitige Fehlervorhersage durch Deep-Learning-Algorithmen hat das Potenzial, die Kranleistung zu revolutionieren und optimale Effizienz im Containerkranbetrieb sicherzustellen.

Es wurde ein maßgeschneiderter Deep-Learning-Algorithmus auf LSTM-Basis entwickelt, der eine Genauigkeit von 99,6 % bei der Vorhersage von Kranfehlersignalen unter Verwendung von LSTM und SMOTE zur Klassifizierungsvorhersage erreichte, mit einer Präzision von 1,00, einem Recall von 0,500 und einem F1-Score von 0,667. Es sind jedoch weitere Daten erforderlich, um die Ergebnisse des Modells zu verbessern.

Die aktuelle Datenverfügbarkeit ist auf einen Monat mit ungefähr 30.000 Variablen beschränkt, darunter weniger als 3.000 Fehlervariablen. Die verwendbaren Daten waren ein kleiner Stichprobensatz mit nur 3 bis 6 Fehlervariablen, die gelegentlich aktiviert wurden. Die Studie liefert auch andere Ergebnisse, wie Korrelation, Kovarianz und Konfusionsmatrizen. Die Experimente zeigten die Fähigkeit der Technik, entscheidende Elemente zu identifizieren, was zu Bemühungen führte, die Struktur des Modells zu verbessern, um eine verbesserte Fehlervorhersage in industriellen Big Data zu ermöglichen.

Das Ausgabemodell basiert zwar auf bereits vorhandenen Daten, doch seine Fähigkeit, mit unausgewogenen Daten umzugehen, was in den meisten Branchen ein häufiges Problem ist, stellt einen erheblichen Vorteil dar. Mit weiteren Daten kann das vorhandene Modell verbessert werden. Kontinuierliche Daten über mehrere Monate hinweg liefern einen umfassenderen Fehlerdaten-Beispielsatz und trag
Sobald in Zukunft ein größerer Datensatz vorliegt, **können andere Algorithmen wie Random Forest, Naïve Bayes und XG Boost untersucht** werden.

Um die **Vorteile der prädiktiven Analytik im Hafenbetrieb voll ausschöpfen zu können, sind kontinuierliche Forschung und Innovationen auf diesem Gebiet unabdingbar .**

Beiträge der Autoren: Konzeptualisierung, AA; Datenkuratierung, AA; formale Analyse, AA; Methodik , AA; Software, AA; Überwachung, LK und JW; Validierung, AA; Visualisierung, AA; Schreiben – Originalentwurf, AA; Schreiben – Überprüfung und Bearbeitung, AA, LK und JW. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu.

Finanzierung: Dieses Projekt wurde teilweise durch das Stipendium 13/RC/2094_P2 der Science Foundation Ireland unterstützt und im Rahmen des Europäischen Fonds für regionale Entwicklung durch das Regionale Operationelle Programm Süd & Ost an Lero – das Forschungszentrum für Software der Science Foundation Ireland **(www.lero.ie – Forschungszentrum für Software) – kofinanziert . abgerufen am 20. Juli 2024) während der ersten zwei Jahre dieses Projekts.**

Datenverfügbarkeitserklärung: Die in diesem Artikel präsentierten Datensätze sind aufgrund von Datenschutz- oder Datensicherheitsproblemen des Containerhafens nicht ohne weiteres verfügbar. Anfragen zum Zugriff auf die Datensätze sollten an den entsprechenden Autor dieses Dokuments gerichtet werden.

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Machine Translated by Google

Verweise

1. **Mordor Intelligence. Containerschifffahrtsmarkt – Größen-, Anteils- und Wachstumsanalyse.** Mordor Intelligence-Branchenberichte. [Online]. **Online verfügbar: https://www.mordorintelligence.com/industry-reports/global-container-shipping-market (abgerufen am 17. Juli 2024).**
2. **Lindeberg, E.; Verbesserung der Betriebsleistung von STS-Kranen. Västerås, Schweden, Februar 2011. Online verfügbar: www.porttechnology.org (abgerufen am 18. Juni 2021).**
3. **Ship-to-Shore (STS)-Containerkräne – Marktgröße, Anteil, Branchenprognose 2028. Online verfügbar: https://www. Fortunebusinessinsights.com/ship-to-shore-sts-container-cranes-market-102880 (abgerufen am 15. Juni 2021).**
4. **Zrniü, N.; Hoffmann, K. Entwicklung des Designs von Ship-To-Shore-Containerkränen: 1959–2004. Im** Internationalen Symposium über Geschichte der Maschinen und Mechanismen; **Springer: Dordrecht, Niederlande, 2007. [CrossRef]**
5. **Khoukhi, A.; Khalid, MH Hybride Computertechniken zur Fehlererkennung und -isolierung, eine Übersicht.** Comput. Electr. Eng. **2015,** 43, **17–32. [CrossRef]**
6. **Saltzer, JH; Frans Kaashoek, M. 2.2: Fehler, Ausfälle und fehlertolerantes Design 2.2.1: Fehler, Ausfälle und Module. In** Principles of Computer System Design; **Massachusetts Institute of Technology LibreTexts: Davis, CA, USA, 2024; Kapitel 2.2, S. 1–461. Online verfügbar: https://eng.libretexts.org/@go/ page/58498 (abgerufen am 19. Juli 2024).**
7. **Safaei, MM; Hejazian, S.; Pedrammehr, S.; Pakzad, M.; Ettefagh, M.; Fotouhi, M. Schadenserkennung an Portalkranen mit bewegter Masse mittels künstlichem neuronalen Netzwerk.** Buildings **2024,** 14, **458. [CrossRef]**
8. **Wang, Z.; Cheng, J.; Hu, H. Ein proaktiv-reaktiver Ansatz für die kontinuierliche Liegeplatzzuweisung und Kaikranzuweisung Probleme mit hybrider Unsicherheit.** J. Mar. Sci. Eng. **2024,** 12, **182. [CrossRef]**
9. **Gattuso, D.; Pratico, FG; Longo, R.; Cassone, G.; Vigna, M.; Sceni, R. Schienenverkehr in Güterterminals: Sicherheitsaspekte und vorgeschlagene Methodik. In** 5. IEEE International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2017—Proceedings, Neapel, Italien, 26.–28. Juni 2017; **Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; S. 164–169. [CrossRef]**
10. **Huang, M.; Liu, Z.; Tao, Y. Mechanische Fehlerdiagnose und -vorhersage im IoT basierend auf der Fusion von Sensordaten aus mehreren Quellen.** Simul. Modell. Praxis. Theorie **2020,** 102, **101981. [CrossRef]**
11. **Meddaoui, A.; Hain, M.; Hachmoud, A. Die Vorteile der vorausschauenden Wartung für eine exzellente Fertigung: Eine Fallstudie zur Entwicklung zuverlässiger Methoden zur Vorhersage von Ausfällen.** Int. J. Adv. Manuf. Technol. **2023,** 128, **3685–3690. [CrossRef]**
12. **Fernandes, M.; Canito, A.; Bolón-Canedo, V.; Conceição, L.; Praça, I.; Marreiros, G. Datenanalyse und Funktionsauswahl für vorausschauende Wartung: Eine Fallstudie in der metallurgischen Industrie.** Int. J. Inf. Geschäftsführer **2019,** 46, **252–262. [CrossRef]**
13. **Yeoh, JKW; Wong, JH; Peng, L. Integration von Kraninformationsmodellen in BIM zur Überprüfung der Einhaltung von Hebeplananforderungen . In Proceedings des 33. Internationalen Symposiums für Automatisierung und Robotik im Bauwesen (ISARC 2016), Auburn, AL, USA, 18.–21. Juli 2016; Online verfügbar: https://www.iaarc.org/publications/fulltext/ISARC2016-Paper192.pdf (abgerufen am 20. Juli 2024).**
14. **Chu, F.; Gailus, S.; Liu, L.; Ni, L. Die Zukunft der Hafenautomatisierung|McKinsey. 2018. Online verfügbar: https://www.mckinsey.com/Branchen/Reiselogistik und Infrastruktur/Unsere Erkenntnisse/Die Zukunft automatisierter Häfen (abgerufen am 18. Mai 2024).**
15. **Lepenioti, K.; Pertselakis, M.; Bousdekis, A.; Louca, A.; Lampathaki, F.; Apostolou, D.; Mentzas, G.; Anastasiou, S. Maschinelles Lernen für prädiktive und präskriptive Analysen von Betriebsdaten im Smart Manufacturing. In** Advanced Information Systems Engineering Workshops; **Springer: Cham, Schweiz, 2020. [CrossRef]**
16. **Was ist Deep Learning? | Ein Leitfaden für Anfänger. Online verfügbar: https://www.scribbr.com/ai-tools/deep-learning/ (abgerufen am 19. Mai 2024).**
17. **Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, RX Deep Learning und seine Anwendungen zur Überwachung des Maschinenzustands.** Mech. Syst. Signalprozess. **2019,** 115, **213–237. [CrossRef]**
18. **Kumar, A.; Gaur, N.; Chakravarty, S.; Alsharif, MH; Uthansakul, P.; Uthansakul, M. Analyse der Spektrumerfassung mit Deep Lernalgorithmen: CNNs und RNNs.** Ain Shams Eng. J. **2024,** 15, **102505. [CrossRef]**
19. **Ubal, C.; Di-Giorgi, G.; Contreras-Reyes, JE; Salas, R. Vorhersage langfristiger Abhängigkeiten in Zeitreihen mit rekurrenten Künstliche Neuronale Netze.** Mach. Learn. Knowl. Extr. **2023, 5, 1340–1358. [CrossRef]**
20. **Heilig, L.; Stahlbock, R.; Voß, S. Von der Digitalisierung zur datengesteuerten Entscheidungsfindung in Containerterminals.** arXiv **2019, arXiv:1904.13251.**
21. **Shin, H.-S.; Lee, S.-P.; Ha, Y.-S.; Kim, H.-S. Entwurf eines Lernmodells zur Containerkransteuerung mittels Deep Learning.** J. Adv. Mar. Eng. Technol. **2023,** 47, **2023–2234. [CrossRef]**
22. **Chalapathy, R.; Chawla, S. Deep Learning zur Anomalieerkennung: Eine Untersuchung.** arXiv **2019, arXiv:1901.03407.**
23. **Shagluf, A.; Longstaff, AP; Fletcher, S. (PDF) Wartungsstrategien zur Reduzierung von Ausfallzeiten aufgrund von Positionsfehlern von Maschinen, Maintenance Performance Measurement and Management (MPMM). Online verfügbar : https://www.researchgate.net/publication/267864966_Maintenance_Strategies_to_Reduce_Downtime_Due_to_Machine_Positional_Errors (abgerufen am 18. Juli 2024).**
24. **Fernandes, M.; Corchado, JM; Marreiros, G. Maschinelle Lerntechniken zur mechanischen Fehlerdiagnose und Fehlerprognose im Kontext realer Anwendungsfälle in der industriellen Fertigung: Eine systematische Literaturübersicht.** Appl. Intell. **2022,** 52, **14246–14280. [CrossRef] [PubMed]**

Machine Translated by Google

25. Softwaretests – Fehler vs. Defekt vs. Fehler vs. Störung vs. Ausfall – GeeksforGeeks. Online verfügbar: https://www.geeksforgeeks. org/software-testing-bug-vs-defect-vs-error-vs-fault-vs-failure/ (abgerufen am 22. April 2024).

26. Jalayer, M.; Orsenigo, C.; Vercellis, C. Fehlererkennung und Diagnose für rotierende Maschinen: Ein Modell basierend auf Faltungs LSTM, Fast Fourier und kontinuierliche Wavelet-Transformationen. Comput. Ind. **2021,** 125, **103378. [CrossRef]**

27. Zhao, R.; Yan, R.; Wang, J.; Mao, K. Lernen, den Maschinenzustand mit faltenden bidirektionalen LSTM-Netzwerken zu überwachen. Sensors **2017,** 17, **273. [CrossRef] [PubMed]**

28. Zrnic, N.; Petkovi´c, Z.; Bošnjak, SM Automatisierung von Containerkränen zwischen Schiff und Land: Ein Überblick über den aktuellen Stand der Technik. FME Transactions **2005,** 33, **111–121. Online verfügbar: https://www.researchgate.net/publication/228936324_Automation_of_ship-to-shore_ Containerkrane - Ein Überblick über den aktuellen Stand der Technik (abgerufen am 24. März 2024).**

29. Tran, QH; Huh, J.; Nguyen, VB; Kang, C.; Ahn, JH; Park, IJ Sensitivitätsanalyse für die Konstruktion von Containerkränen zwischen Schiff und Land. Appl. Sci. **2018, 8, 1667. [CrossRef]**

30. Li, Z. Deep Learning-basierte Ansätze für die vorausschauende Wartung. Ein Rahmenwerk für intelligente Fehlerdiagnose und -prognose im Zeitalter von Industrie 4.0. Dissertation, Norwegische Universität für Wissenschaft und Technologie, Trondheim, Norwegen, 2018.

31. Goodfellow, I.; Bengio, Y.; Courville, A. 2015. Online verfügbar: http://www.deeplearningbook.org/ (abgerufen am 15. Juni 2024).

32. Liu, W.; Xiao, J. Eine neue Methode zur Erkennung des Verschleißes von Kranschienen und deren Fehleranalyse. J. Mech. Strength **2014,** 36, **878–883. Online verfügbar: https://www.researchgate.net/publication/293095272_A_new_method_of_detection_the_wear_of_crane_ Schienen und ihre Fehleranalyse (abgerufen am 19. März 2024).**

33. Mohammadfam, I.; Borgheipour, H.; Tehrani, G.; Madadi, S. Identifizierung und Bewertung menschlicher Fehler bei Turmdrehkranführern mithilfe von SHERPA- und CREAM-Techniken. J. Health Saf. Work. **2020,** 10, **5–8.**

34. Jakovlev, S.; Eglynas, T.; Jusis, M.; Voznak, M.; Partila, P.; Tovarek, J. Erkennen von physischen Stößen an den Ecken von Versandcontainern während der Handhabung durch Kaikräne. J. Mar. Sci. Eng. **2023,** 11, **794. [CrossRef]**

35. Gao, X.; Yeh, HG; Marayong, P. Ein schneller farbbasierter Objekterkennungsalgorithmus für ein Assistenzsystem für Kaikranführer . In Proceedings der 11. jährlichen IEEE International Systems Conference, SysCon 2017—Proceedings, Montreal, QC, Kanada, 24.–27. April 2017. [CrossRef]

36. Westerberg, S.; Shiriaev, A. Teleoperation von Forstmaschinen auf Basis virtueller Umgebungen: Entwicklung zukünftiger Interaktionsmethoden. J. Hum.-Robot. Interact. **2013, 2, 84–110. [CrossRef]**

37. Ren, W.; Wu, Z.; Zhang, L. Echtzeitplanung eines Hebeschemas in mobilen Kransteuerungen. Can. J. Civ. Eng. **2016,** 43, **542–552. [CrossRef]**

38. Watanabe, M.; Momoi, Y.; Odai, M.; Ieshige, K. Gleichzeitige Schätzung der Seilspannung und Last von Kränen mittels Motorantrieb Signale. Trans. JSME **2023,** 89. **[CrossRef]**

39. Lei, Z.; Behzadipour, S.; Al-Hussein, M.; Hermann, U. Anwendung der robotischen Hindernisvermeidung bei der Kranbahnplanung. Proceedings der 28. ISARC, Seoul, Republik Korea, 29. Juni–2. Juli 2011.

40. Lehner, P.; Kÿrivý, V.; Krejsa, M.; Paÿrenica, P.; Kozák, **J. Stochastische Lebensdauervorhersage für bestehende Stahlkonstruktionen unter Belastung durch Brückenkräne.** Procedia Struct. Integr. **2018,** 13, **1539–1544. [CrossRef]**

41. Shang, KC; Tseng, WJ Eine Risikoanalyse der Stauereiarbeiten in Containerterminals in Seehäfen. J. Mar. Sci. Technol. **2010,** 18, **201–210. [CrossRef]**

42. Ismail, RMTR; That, ND; Ha, QP Modellierung und robuste Trajektorienverfolgung für Offshore-Containerkransysteme. Autom. Constr. **2015,** 59, **179–187. [CrossRef]**

43. Liu, J.; Qin, X.; Sun, Y.; Zhang, Q. Reaktionsvorhersagemodell für Strukturen von Kaicontainerkränen basierend auf Überwachung Daten. J. Perform. Constr. Facil. **2021,** 35. **[CrossRef]**

44. Jo, JH; Kim, S. Entwicklung von Leistungskennzahlen zur Leistungsbewertung von Ship-to-Shore-Kranen in Containerterminals Operationen. J. Mar. Sci. Eng. **2020, 8, 6. [CrossRef]**

45. Cheng, C.; Liu, X.; Zhou, B.; Yuan, Y. Intelligente Fehlerdiagnose mit verrauschten Etiketten über halbüberwachtes Lernen auf industriellen Zeitreihen. IEEE Trans. Ind. Inform. **2023,** 19, **7724–7732. [CrossRef]**

46. Lee, XY; Kumar, A.; Vidyaratne, L.; Rao, AR; Farahat, A.; Gupta, C. Ein Ensemble von Faltungsmethoden zur Fehlererkennung mithilfe von Vibrationssignalen. In Proceedings der 2023 IEEE International Conference on Prognostics and Health Management (ICPHM 2023), Montreal, QC, Kanada, 5.–7. Juni 2023; S. 172–179. [CrossRef]

47. Zhang, J.; Zhang, M.; Feng, Z.; Ruifang, LV; Lu, C.; Dai, Y.; Dong, L. Gated recurrent unit-enhanced deep convolutional neural network for real-time industrial process fault diagnostic. Prozesssicherheit. Umwelt. Prot. **2023,** 175, **129–149. [CrossRef]**

48. So codieren Sie RNN- und LSTM-Neuralnetze in Python. Online verfügbar: https://www.nbshare.io/notebook/249468051/ So programmieren Sie RNN- und LSTM-Neuralnetze in Python/ (abgerufen am 28. April 2022).

49. Quiñones-Grueiro, M.; Llanes-Santiago, O.; Silva Neto, AJ Fehlerdiagnose in industriellen Systemen. Zucht. Syst. Entscheidung. Kontrolle **2021,** 309, **1–14. [CrossRef]**

50. Illustrierter Leitfaden zu LSTMs und GRUs: Eine Schritt-für-Schritt-Erklärung|von Michael Phi|towards Data Science. Online verfügbar: https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-sa-step-by-step-explanation-44e9eb85bf21 (abgerufen am 13. April 2021).

51. Maheswaranathan, N.; Williams, AH; Golub, MD; Ganguli, S.; Sussillo, D. Reverse Engineering rekurrierender Netzwerke zur Sentimentklassifizierung enthüllt Linienattraktordynamik. In Proceedings der 33. Konferenz über neuronale Informationsverarbeitungssysteme (NeurIPS 2019), Vancouver, BC, Kanada, 8.–14. Dezember 2019; S. 15696–15705.

Machine Translated by Google

52. Alizadeh, Z.; Yazdi, J.; Kim, JH; Al-Shamiri, AK Bewertung von Techniken des maschinellen Lernens zur monatlichen Flussvorhersage. Wasser **2018,** 10, **1676. [CrossRef]**

53. Ahmad, J.; Tahir, A.; Larijani, H.; Ahmed, F.; Aziz Shah, S.; Hall, AJ; Buchanan, WJ Energiebedarfsprognose von Gebäuden mithilfe zufälliger neuronaler Netzwerke. J. Intell. Fuzzy Syst. **2020,** 38, **4753–4765. [CrossRef]**

54. Li, S.; Li, W.; Member, S.; Cook, C.; Gao, Y. Tiefes, unabhängig rekurrierendes neuronales Netzwerk (IndRNN). arXiv **2019, arXiv:1910.06251v3.**

55. Abid, A.; Khan, MT; Iqbal, J. Ein Überblick über Fehlererkennungs- und Diagnosetechniken: Grundlagen und mehr. Artif. Intell. Rev. **2021,** 54, **3639–3664. [CrossRef]**

56. Shahin, M.; Chen, FF; Hosseinzadeh, A.; Zand, N. Einsatz von maschinellem Lernen und Deep-Learning-Algorithmen zur Minimierung von Ausfallzeiten in Fertigungssystemen: Ein Diagnosedienst zur frühzeitigen Fehlererkennung. Int. J. Adv. Manuf. Technol. **2023,** 128, **3857–3883. [CrossRef]**

57. Leite, D.; Martins, A.; Rativa, D.; De Oliveira, JFL; Maciel, AMA Ein automatisierter maschineller Lernansatz für Echtzeit Fehlererkennung und Diagnose. Sensoren **2022,** 22, **6138. [CrossRef]**

58. Darban, ZZ; Webb, GI; Pan, S.; Aggarwal, CC; Salehi, M. Deep Learning zur Erkennung von Zeitreihenanomalien: Eine Untersuchung. arXiv **2022, arXiv:2211.05244v2.**

59. Song, J.; Lee, YC; Lee, J. Tiefes generatives Modell mit Zeitreihen-Bildkodierung zur Erkennung von Fertigungsfehlern im Druckgussverfahren . J. Intell. Manuf. **2023,** 34, **3001–3014. [CrossRef]**

60. Gui, X.; Zhang, J. Deep Metric Learning-Modell zur Diagnose unausgewogener Fehler. arXiv **2021, arXiv:2107.03786.**

61. Liu, H.; Ma, R.; Li, D.; Yan, L.; Ma, Z. Maschinenfehlerdiagnose basierend auf Deep Learning für Zeitreihenanalyse und Wissensgraphen . J. Signal Process. Syst. **2021,** 93, **1433–1455. [CrossRef]**

62. Pang, JL Adaptive Fehlervorhersage und Wartung in Produktionslinien mittels Deep Learning. Int. J. Simul. Modell. **2023,** 22, **734–745. [CrossRef]**

63. Brownlee, J.; So berechnen Sie die Korrelation zwischen Variablen in Python. Machine Learning Mastery. Online verfügbar: https: // machinelearningmastery.com/how-to-use-correlation-to-understand-the-relationship-between-variables/ (abgerufen am 14. April 2022).

64. Li, A. Priorisierung des Risikos menschlicher Fehler im Kranbetrieb basierend auf CPT und ICWGT. PLoS ONE **2024,** 19, **e0297120. [CrossRef] [PubMed]**

65. Eglynas, T.; Jusis, M.; Jakovlev, S.; Senulis, A.; Andziulis, A.; Gudas, S. Analyse der Effizienz von Schiffscontainern Handhabungs-/Ladekontrollmethoden und -verfahren. Adv. Mech. Eng. **2019,** 11. **[CrossRef]**

66. Maboudi, M.; Alamouri, A.; De Arriba López, V.; Bajauri, MS; Berger, C.; Gerke, M. Drohnengestützte Containerkran-Inspektion: Konzept, Herausforderungen und vorläufige Ergebnisse. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. **2021,** V-1-2021, **121–12 [Querverweis]**

67. Shan, J.; Zhang, M.; Zhang, D.; Zhou, X.; Li, H.; Wang, B. Containerkrane. In Handbook of Port Machinery; **Springer: Singapur, 2024; S. 451–678. [CrossRef]**

68. Viellechner, A.; Spinler, S. Neuartige Datenanalyse trifft auf konventionelle Containerschifffahrt: Vorhersage von Verzögerungen durch Vergleich verschiedener Algorithmen des maschinellen Lernens. Online verfügbar: https://hdl.handle.net/10125/63897 (abgerufen am 12. Juni 2024).

69. Tian, X.; Yan, R.; Liu, Y.; Wang, S. Eine intelligente Methode zum Vorhersagen und Optimieren für gezielten und kostengünstigen Seetransport. Transp. Res. Teil B Methodol. **2023,** 172, **32–52. [CrossRef]**

70. Umgang mit unausgewogenen Daten zur Klassifizierung – GeeksforGeeks. Online verfügbar: https://www.geeksforgeeks.org/handling-unausgewogene Daten für die Klassifizierung/ (abgerufen am 12. Juni 2024).

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultieren.

Machine Translated by Google

Algorithmen

MDPI

Artikel

Eine effiziente Optimierung der Monte Carlo Baumsuche Algorithmus für Amazonen

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

[1] Hochschule für Informatik und Technologie, Jilin-Universität, Changchun 130012, China; zhanglj2121@mails.jlu.edu.cn (LZ); zouhan1221@mails.jlu.edu.cn (HZ)
[2] Zentrallabor für symbolische Berechnung und Wissenstechnik des Bildungsministeriums, Jilin-Universität, Changchun 130012, China * Korrespondenz: zhuyungang@jlu.edu.cn

Zusammenfassung: Amazons ist ein computergestütztes Brettspiel mit komplexen Positionen, die für Menschen eine große Herausforderung darstellen. In diesem Artikel schlagen wir eine effiziente Optimierung des Monte-Carlo-Baumsuchalgorithmus (MCTS) für Amazons vor, indem wir die Strategie „Move Groups“ und die Optimierungsstrategie „Parallel Evaluation“ (MG-PEO) kombinieren. Insbesondere erklären wir die hohe Effizienz der Strategie „Move Groups“, indem wir ein neues Kriterium definieren: die gewinnende Konvergenzdistanz. Wir heben auch das potenzielle Problem der Strategie hervor, in ein lokales Optimum zu fallen, und schlagen vor, dass der Mechanismus der Parallel Evaluation diesen Mangel ausgleichen kann. Darüber hinaus haben wir strenge Leistungsanalysen und Experimente durchgeführt. Die Ergebnisse der Leistungsanalyse zeigen, dass der MCTS-Algorithmus mit der Strategie „Move Groups“ die Spielbarkeit des Amazons-Spiels im Vergleich zum herkömmlichen MCTS-Algorithmus um das 20- bis 30-fache verbessern kann. Die Optimierung der Parallel Evaluation verbessert die Spielbarkeit des Amazons-Spiels um das 2- bis 3-fache. Experimentelle Ergebnisse zeigen, dass der MCTS- Algorithmus mit der Strategie „MG-PEO“ im Durchschnitt eine um 23 % höhere Spielgewinnrate erzielt als der herkömmliche MCTS-Algorithmus. Darüber hinaus gewann das in diesem Artikel vorgeschlagene MG-PEO Amazons-Programm den ersten Preis im Amazons-Wettbewerb bei der China Co

Schlüsselwörter: Amazonen; Algorithmusoptimierung; Bewegungsgruppen; Parallele Auswertung

check for updates

Zitat: Zhang, L.; Zou, H.; Zhu, Y. Eine effiziente Optimierung des Monte-Carlo-Baumsuchalgorithmus für Amazons. Algorithmen **2024,** 17, **334. https://doi.org/10.3390/a17080334**

Wissenschaftliche Herausgeber: Wenxin Li und Haifeng Zhang

Empfangen: 22. April 2024
Überarbeitet: 16. Juli 2024
Akzeptiert: 27. Juli 2024
Veröffentlicht: 1. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz. Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/ 4.0/).

1. Einleitung

Amazons ist ein Brettspiel für zwei Spieler, das 1988 von dem Argentinier Walter Zamkauskas eingeführt wurde [1,2]. Es ist das designierte Spiel für den Olympia- Wettbewerb für Computerspielprogrammierer. Aufgrund seines komplexen Gameplays ist Amazons für Menschen eine große Herausforderung **[3], was es zu einem beliebten Thema für Computerspielwettbewerbe und Forschung macht. Die Spielregeln lauten wie folgt: Das Spiel wird auf einem 10 × 10 Quadrate großen Brett gespielt. Jeder Spieler kontrolliert vier identische Figuren, die „Amazonen“ genannt werden. Diese Figuren und Pfeile folgen denselben Regeln wie eine Dame im Schach: Sie können eine beliebige Anzahl Felder in jede Richtung – horizontal, vertikal oder diagonal – ziehen, dürfen aber keine anderen Figuren oder Pfeile überspringen. Ziel ist es, die Bewegung des Gegners durch das Abschießen von Pfeilen einzuschränken, bis keine seiner vier Figuren mehr gezogen werden kann. Wenn in der Runde eines Spielers alle vier seiner Figuren bewegungsunfähig sind, verliert dieser Spieler und der andere Spieler ge**

Die größte Herausforderung des Amazonen-Spiels liegt in seiner Komplexität: Jeder Zug bietet eine große Anzahl von möglichen Entscheidungen, typischerweise mehr als 500 [4,5], und das Spiel kann sich über mehr als 80 Züge erstrecken **[6,7]. Diese hohe Komplexität macht Amazonen zu einem hervorragenden experimentellen Beispiel für das Testen von Suchstrategiealgorithmen [8,9].** Seit seiner Erfindung hat es **erhebliche Aufmerksamkeit von Informatikern**

Die Monte-Carlo-Methode [12] ist eine Rechenmethode, die durch wiederholte Zufallsstichproben numerische Ergebnisse erzeugt. In Kombination mit Baumsuchalgorith

Machine Translated by Google

und geeigneten Bewertungsfunktionen hat der Monte-Carlo-Baumsuchalgorithmus (MCTS) beeindruckende Ergebnisse in Brettspielen mit hohem Verzweigungsfaktor (wie Go) [13–15] und in Echtzeitstrategiespielen (wie StarCraft [16]) geliefert. So besiegte beispielsweise DeepMinds AlphaGo-Programm [17] 2016 den koreanischen Go-Weltmeister Lee Sedol mit 4:1. 2018 besiegte OpenAIs OpenAI Five-Programm [18] das Weltmeisterteam in Dota 2 mit überwältigendem Vorsprung; beide Erfolge basieren auf dem Monte- Carlo-Baumsuchalgorithmus.

Die Optimierungsstrategie „Move Groups-Parallel Evaluation“ (MG-PEO) berücksichtigt die Regeln des Amazonen-Brettspiels, indem sie einen vollständigen Zug in zwei Teile unterteilt: Amazonen-Bewegung und Pfeilschuss. Sowohl die Amazonen-Bewegung als auch der Pfeilschuss werden als Baumknoten behandelt, anstatt einen vollständigen Zug als einen einzigen Dieser Ansatz ist in der Literatur als Move Groups-Strategie bekannt [19]. Obwohl diese Strategie auf das Amazons-Spiel angewendet wurde [20], fehlt noch eine objektive Analyse ihrer Wirksamkeit. Darüber hinaus ist die Meinung weit verbreitet, dass diese Methode nur Vorteile ohne Nachteile bietet. In diesem Artikel erklären wir die hohe Effizienz der Move Groups-Strategie in Amazons, indem wir ein neues Kriterium definieren: die gewinnende Konvergenzdistanz . Wir identifizieren auch bestimmte Mängel dieser Strategie. Um diese zu beheben, schlagen wir vor, einen parallelen Bewertungsmechanismus zu verwenden, um die Einschränkungen der Move Groups-Strategie auszugleichen. Unsere experimentellen Ergebnisse zeigen, dass MG-PEO eine effektive Fusionsoptimierungsstrategie ist.

Die wichtigsten Beiträge dieses Papiers sind wie folgt:

- Wir schlagen eine effektive Optimierung des Monte-Carlo-Baumsuchalgorithmus für das Amazons-Spiel vor: die MG-PEO-Algorithmusoptimierung, die die Move Groups-Strategie mit der Parallel Evaluation-Strategie kombiniert. • Wir bieten eine Leistungsanalyse der Move Groups-Strategie in Amazons. Durch die Definition eines neuen Kriteriums, der gewinnenden Konvergenzdistanz, haben wir die hohe Effizienz der Strategie erklärt. Darüber hinaus heben wir auch hervor, dass diese Strategie bestimmte Mängel aufweist. Die Move Groups-Strategie beinhaltet die Verfeinerung der Regeln des Amazons- Spiels und die Verwendung von Simulationen, um zusätzlichen Layer-Knoten entsprechende Wertungen zuzuweisen . Durch die Verwendung dieser Wertungen (z. B. die Upper Confidence Bound Apply to Tree (UCT)-Strategie [15]) können wir eine effektive Beschneidung erreichen, um Zeit zu sparen. Das Hinzufügen zusätzlicher Layer-Knoten bedeutet jedoch, dass im schlimmsten Fall, in dem keine Beschneidung stattgefunden hat, mehr Knoten durchlaufen und ausgewertet werden müssen, was möglicherweise zu einem erhöhten zeitaufwändigen Overhead führt. Dieser Overhead entsteht, da der Algorithmus versucht, die lokal optimale Lösung zu finden. • Wir schlagen vor, einen Mechanismus zur parallelen Bewertung zu verwenden, um die potenziellen Mängel der Move Groups-Strategie zu beheben. Durch die Parallelisierung der Bewertungsfunktion für das Amazons-Spiel möchten wir den Erweiterungsprozess der Monte-Carlo-Baumsuche beschleunigen und die Wahrscheinlichkeit min

Darüber hinaus haben wir strenge Leistungsanalysen und Spielexperimente durchgeführt. Die Leistungsanalyse zeigt, dass der MCTS-Algorithmus mit der Move Groups-Strategie die Spielfähigkeit von Amazonen im Vergleich zum herkömmlichen MCTS-Algorithmus um das 20- bis 30-fache steigern kann, während die Optimierungskomponente Parallel Evaluation die Spielfähigkeit von Amazonen um das 2- bis 3-fache steigern kann. Versuchsergebnisse zeigen, dass der MCTS-Algorithmus mit der MG-PEO-Strategie im Vergleich zum herkömmlichen MCTS-Algorithmus eine durchschnittliche Steigerung der Spielgewinnrate um 23 % erreicht und eine stärkere Spielfähigkeit aufweist.

Der Rest dieses Dokuments ist wie folgt strukturiert: In Abschnitt 2 präsentieren wir vorläufige Informationen zum Amazons-Spiel und diskutieren dessen Relevanz für den MCTS-Algorithmus. In Abschnitt 3 liefern wir eine detaillierte Beschreibung des MCTS-Algorithmus mit der MG-PEO-Strategie. In Abschnitt 4 bieten wir Leistungsanalysen der Move Groups-Strategie bzw. der Parallel Evaluation-Strategie. Abschnitt 5 beschreibt eine Reihe von Experimenten

Machine Translated by Google

durchgeführt, um die Effizienz des MCTS-Algorithmus mit der MG-PEO-Strategie zu überprüfen. Abschließend werden in Abschnitt **6 Schlussfolgerungen und Ansätze für zukünftige Forschung zusammengefasst .**

2. Verwandte Arbeiten

In diesem Abschnitt geben wir einen Überblick über die relevante Forschung, die die Grundlage unserer Arbeit bildet. Insbesondere werden wir drei Schlüsselmodule der Literatur diskutieren, die für Amazons relevant sind: historische Forschung zu Amazons-Algorithmen, die Anwendung des MCTS- Algorithmus auf Amazons und Wissen über die Bewertung von Spielsituationen.

2.1. Amazons Spielalgorithmus

Das Amazons-Spiel wurde umfassend untersucht. Die Entwicklung seines Algorithmus kann grob in drei Phasen unterteilt werden . Zunächst **dominierten traditionelle Methoden mit Wurzeln in der mathematischen Theorie, wie der Minimax-Suchalgorithmus [21] .** In der zweiten Phase **entstanden Monte-Carlo-Algorithmen, darunter verschiedene Varianten der Monte-Carlo-Baumsuche [19, 20, 22–24]. In der dritten Phase traten Deep-Learning-Techniken unter Nutzung der Amazons-Spieldatenbanken in den Vordergrund [25, 26]. Insbesondere das repräsentativste Amazons-Programm, Invader [27–29], verlässt sich auf MCTS zur prospektiven Exploration, ergänzt durch neuronales Netzwerktraining anhand von Schachdatenbanken, was zu beeindruckenden Spielfähigkeiten im Amazons-Spiel führt. Man kann sagen, dass die weitere Optimierung der Monte-Carlo -Algorithmen für die Weiterentwicklung der Amazons-Forschung von entscheidender Bedeutung bleibt.**

2.2. Monte-Carlo-Baumsuchalgorithmus **Die**

Entwicklung des Monte-Carlo-Baumsuchalgorithmus in Amazons kann grob in zwei Forschungsrichtungen unterteilt werden: Algorithmusoptimierung in Bezug auf die Baumsuche und Algorithmusoptimierung in Bezug auf Bewertungsfunktionen. In Bezug auf die Optimierung des Baumsuchalgorithmus haben Forscher an der Verbesserung der Suchalgorithmen gearbeitet, um die Leistung von Amazons zu verbessern. Beispielsweise haben Guo et al. das Spiel in frühe, mittlere und späte Phasen unterteilt und in jeder Phase unterschiedliche Strategien angewendet [30]. Li et al. haben den UCT- Algorithmus auf Amazons angewendet **[31]; Quan et al. haben die Überlegenheit des UCT-Algorithmus in Amazons** nachgewiesen **[32]. Andererseits ist auch die Optimierung der Bewertungsfunktion einer der Forschungsschwerpunkte. Durch die Optimierung der Bewertungsfunktion kann das aktuelle Spiel genauer bewertet und der Suchalgorithmus zu fundierteren Entscheidungen geführt werden. Lieberum et al. widmen ihre Forschung diesem Aspekt [33,34] und konzentrieren sich hauptsächlich auf die Verbesserung der Hyperparameterabstimmung und die Entwicklung neuer Bewertungsfunktionen. Insgesamt kann der Monte-Carlo-Baumsuchalgorithmus nicht ohne einen einzigen Aspekt verbessert werden. In diesem Artikel stellt unsere vorgeschlagene Fusion der Move-Groups-Strategie und der Parallel-Evaluation- Strategie eine umfassende Betrachtung der beiden historischen Forschungsschwerpunkte dar.**
Diese Fusion des Monte-Carlo-Baumsuchalgorithmus ist hochmodern und effektiv.

2.3. Bewertungsfunktion

Derzeit stehen fast alle Suchalgorithmen vor einem Problem: der Situationsbewertung.
Eine gute Situationsbewertung ist beim Amazonenspiel sehr wichtig und kann sogar über die Qualität des Spielprogramms entscheiden **[33].**

Ähnlich wie viele Brettspiele kann man Amazons grob in drei Phasen unterteilen: Eröffnung, Mitte und Ende, wobei jede Phase unterschiedliche Strategien hat. Die Hauptindikatoren der Bewertungsfunktion von Amazons sind Territorium, Position und Mobilität, und die Bedeutung dieser drei Merkmale variiert in den verschiedenen Phasen. Basierend auf diesen Ideen schlugen **Guo et al. [30] eine phasenweise Bewertungsfunktion vor, wie in Gleichung (1) dargestellt:**

$$\text{Wert} = \mathbf{k1} \ddot{y} \mathbf{t1} + \mathbf{k2} \ddot{y} \mathbf{t2} + \mathbf{k3} \ddot{y} \mathbf{p1} + \mathbf{k4} \ddot{y} \mathbf{p2} + \mathbf{k5} \ddot{y} \text{Mobilität} \quad \textbf{(1)}$$

wobei t1 und t2 Gebietsmetrikwerte sind, die jeweils die Bewertung der Kontrolle jeder Seite über den Raum darstellen, basierend auf der QueenMove-Methode und der KingMove-Methode. p1 und p2 sind Positionsmetrikwerte, die die strategische Position der Figuren jeder Seite widerspiegeln, ebenfalls basierend auf der QueenMove- und der KingMove-Methode rela

Machine Translated by Google

ist der Flexibilitätsbewertungswert der beiden Amazon-Teile. k1, k2, k3, k4 und k5 sind die entsprechenden Gewichtswerte. Wie diese Werte berechnet werden, wird weiter unten beschrieben.

Laut der Literatur [2] werden zur Berechnung von Gebiets- und Positionswerten zwei Arten von Methoden verwendet: QueenMove und KingMove. QueenMove bedeutet, dass ein Zug gemacht werden kann, solange in keiner der acht Richtungen Hindernisse zu überwinden sind (die Hindernisse umfassen Figuren und Pfeile), während KingMove bedeutet, dass die Distanz eines einzelnen Zuges eins beträgt, d. h., es kann nur ein Zug auf ein benachbartes Feld gemacht werden. Der QueenMove- Wert stellt die Mindestanzahl von Zügen dar, die die vier Figuren eines Spielers benötigen, um ein bestimmtes Feld zu erreichen (d. h. keine Figuren oder Pfeile); der KingMove-Wert stellt die Mindestanzahl von Zügen dar, die die vier Figuren eines Spielers benötigen, um ein bestimmtes Feld zu erreichen. Wie in Abbildung 1 auf der linken Seite dargestellt, zeigt der Wert 2 in der oberen linken Ecke des oberen linken (A, 1) Felds an, dass die weiße Figur mindestens zwei Züge benötigt, um dieses Feld mit der QueenMove-Methode zu erreichen, und der Wert 4 in der unteren rechten Ecke zeigt an, dass die schwarze Figur vier Züge benötigt. In ähnlicher Weise zeigen **die Werte 2 und 7 im selben Quadrat in Abbildungÿ1 auf der rechten Seite die Situation an, wenn die KingMove-Methode verwendet wird.**

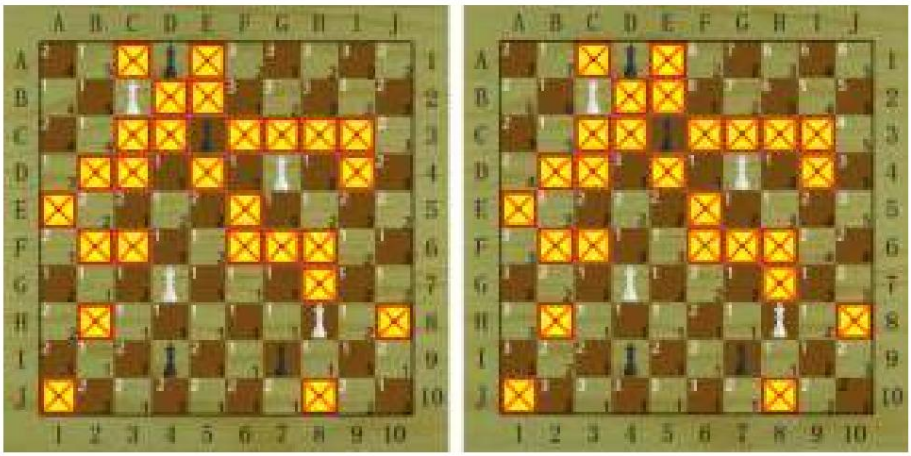


Abbildung 1. Berechnungsergebnisse der Methoden QueenMove und KingMove. Links ist das Ergebnis der KingMove-Berechnung und rechts das Ergebnis der QueenMove-Berechnung.

Dementsprechend können wir für alle Indikatoren die folgende Berechnungsformel verwenden: t1 stellt die Differenz in der Kontrolle des Gebiets mit der QueenMove-Methode dar. t2 stellt die Differenz in der Kontrolle des Gebiets mit der KingMove-Methode dar. Wie in den Gleichungen (2) und (3) gezeigt:

$$ti = ÿÿ(Bi , Wi) \quad (2)$$

$$ÿ(Bi , Wi) = \begin{cases} ÿ\mathbf{0}, & Bi(A) = Wi(A) = \mathbf{ÿ} \\ \mathbf{k}, & Bi(A) = Wi(A)\ ÿ = \mathbf{ÿ} \\ \mathbf{1}, & Bi(A) < Wi(A) \\ \mathbf{ÿ1}, & Bi(A) > Wi(A) \end{cases} \quad (3)$$

wobei B **schwarze Figuren und** W **weiße Figuren darstellt, wobei der Index 1 die Methode QueenMove und 2 die Methode KingMove angibt.** A **steht für leere Felder.** k **steht für das Übergewicht des führenden Feldes und der Wertebereich ist größer als ÿ1 und kleiner als 1. Wenn der Spieler mit dem ersten Zug weiß ist,** ist k **größer als 0, andernfalls kleiner als 0. p1 stellt die Positionen der beiden Figuren auf dem Kontrollfeld A bei Verwendung der Methode**

QueenMove dar. p2 stellt die Positionen der beiden Figuren auf dem Kontrollfeld A bei Verwendung der Methode KingMove dar. Wie in den Gleichungen (4) und (5) gezeigt:

Machine Translated by Google

$$\mathbf{p1} = \mathbf{2ÿ\ 2}\ ÿB1(A)\ \mathbf{ÿ\ 2}\ ÿW1(A) \quad \textbf{(4)}$$

$$\mathbf{p2} = ÿ_{A}\ \min(1, \max(ÿ1, B2(A)\ \mathbf{ÿ}\ W2(A)/6)) \quad \textbf{(5)}$$

Mobilität **bedeutet die Flexibilität der Teile. Wie in den Gleichungen (6) und (7) gezeigt:**

$$\text{Mobilität} = \mathbf{ÿ}_{\mathbf{wÿweißes\ Stück}} F(w)\ \mathbf{ÿ\ ÿ}_{\mathbf{býschwarze\ Figur}} F(b) \quad \textbf{(6)}$$

$$F(a) = \begin{cases} \mathbf{ÿ\ umgebende\ Räume,\ a\ ist\ der\ Raum} \\ \mathbf{0,} & \mathbf{a\ ist\ das\ Hindernis} \\ \mathbf{ÿsÿSa}\ \frac{\text{F-Raum(e)}}{\mathbf{D(ein,s)}}\mathbf{,} & \mathbf{a\ ist\ ein\ Amazon\text{-}Stück} \end{cases} \quad \textbf{(7)}$$

wobei F **die Flexibilität bezeichnet,** Sa **die Menge der Räume, die von einem Spielstein erreicht werden können, indem die QueenMove-Methode, und D(a,s) ist die minimale Anzahl von Zügen von** a **nach** s **unter Verwendung der KingMove-Methode.**

Wir haben die Hyperparametrisierungseinstellungen gemäß der Literatur [2] beibehalten; detaillierte Informationen zu jedem Bewertungsindikator finden Sie in Tabelle 1.

Tabelle 1. Tabelle der Gewichtskoeffizienten.

	Bewertungsmetriken Gebiet		**Position**			**Mobilität**
	Bewertungsfaktor t1		**t2**	**p1**	**p2**	**Mobilität**
	Gewichtskoeffizient k1		**k2**	**k3**	**k4**	**k5**
Etappen	**Öffnung**	**0,14**	**0,37**	**0,13**	**0,13**	**0,20**
	Mitte	**0,30**	**0,25**	**0,20**	**0,20**	**0,05**
	Ende	**0,80**	**0,10**	**0,05**	**0,05**	**0,00**

Tabelle 1 zeigt, dass t1 im Verlauf des Spiels allmählich an Bedeutung gewinnt. beginnend kleiner in der Öffnung, in der Mitte größer werdend und seinen Höhepunkt erreichend im Endspiel. Dies liegt daran, dass im Verlauf des Spiels die Kontrolle über das Brett immer entscheidend für Sieg oder Niederlage. t2 ist in der Eröffnungsphase äußerst wichtig, um sicherzustellen, dass Sicherheit benachbarter Regionen und territorialer Grenzen, aber seine Rolle nimmt ab, wenn die Spielverlauf und kann im Endspiel fast ignoriert werden. Es spiegelt die Dringlichkeit von die anfängliche Layoutverteidigung und der territoriale Wettbewerb. p1 und p2 bezeichnen Figurenpositionen, Diese spielen vor allem in der Eröffnungsphase eine Rolle, wenn es um eine ausgewogene Figurenverteilung geht entscheidend für die nachfolgende strategische Aufstellung; im Endspiel ist das Muster des Spiels steht bereits fest, die Spielsteinverteilung ist nicht mehr der entscheidende Punkt;

Mobilität **bezeichnet Flexibilität, die in der frühen Phase des Spiels am wichtigsten ist. da das Blockieren der Spielsteine ernsthafte Nachteile mit sich bringen kann; im weiteren Spielverlauf und Wenn das erste Layout festgelegt ist, wird die Verteilung der Teile zunehmend kritischer. Spiel fortschreitet und das Layout fertiggestellt ist, der Einfluss der Figurenmobilität auf die Spiel Situation wird geschwächt, sodass sein Gewicht im mittleren und späten Spiel vernachlässigbar wird. Dies spiegelt die Notwendigkeit strategischer Flexibilität in den frühen Phasen der Freizügigkeit wider.**

3. Forschungsmethodik

In diesem Abschnitt beschreiben wir zunächst den Gesamtrahmen des MCTS-Algorithmus, der mit der MG-PEO-Strategie verbessert wurde. Anschließend gehen wir auf die Optimierungsdetails ein und konzentrieren uns dab auf zwei Aspekten: Bewegungsgruppen und parallele Auswertung.

Machine Translated by Google

3.1. Gesamtrahmen des Algorithmus **Wir**

schlagen eine Optimierungsstrategie für den MG-PEO-Algorithmus vor, die auf der Monte-Carlo- Baumsuche basiert [22]. **Wie in Abbildung 2 dargestellt, beinhaltet der optimierte Monte-Carlo-Baumsuchalgorithmus die iterative Konstruktion eines Suchbaums, wobei in jeder Runde ein neuer Baum erstellt wird, bis die Zeitbeschränkung erreicht ist oder das Spiel endet. An diesem Punkt stoppt die Suche und gibt die am besten funktionierende Root-Operation zurück (eine Kombination von Aktionen von bestimmten Knoten in der zweiten und dritten Schicht).**

Im Einzelnen werden für jede Sucheiteration die folgenden vier Schritte ausgeführt:

(1) Auswahl: Ausgehend vom Stammknoten wird eine bestimmte Strategie angewendet, um rekursiv durch den Baum zu gehen und erweiterbare Knoten auszuwählen. Ein Knoten ist erweiterbar, wenn er einen Nicht-Endzustand darstellt und nicht besuchte untergeordnete Knoten hat.

(2) Erweiterung: Erweitern Sie den Baum, indem Sie je nach den verfügbaren Operationen einen (oder mehrere) untergeordnete Knoten hinzufügen .

(3) Simulation: Führen Sie eine Simulation von einem neuen Knoten aus, um Ergebnisse gemäß der Standardstrategie zu erzeugen.

(4) Backpropagation: Sichern Sie die Simulationsergebnisse über die ausgewählten Knoten, um sie zu aktualisieren ihre Statistiken.

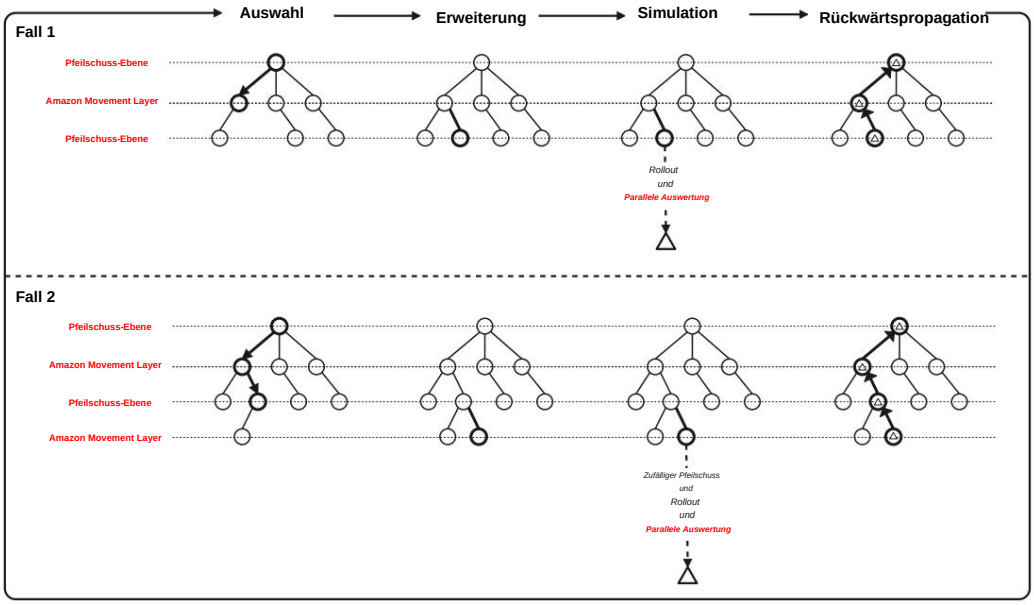


Abbildung 2. Allgemeines MCTS plus MG-PEO-Framework.

Wie in Abbildung 2 zu sehen ist, arbeitet der Algorithmus aufgrund der Verwendung der Move Groups-Strategie während der Simulationsphase in zwei Szenarien: Wenn der aktuelle Knoten ein Knoten der Arrow Shot-Ebene ist, führt er Rollout und Parallel Evaluation aus. Wenn der aktuelle Knoten ein Knoten der Amazon Movement-Ebene ist, führt er Random Arrow Shot, Rollout und Parallel Evaluation aus.

Der Pseudocode für das Gesamtframework wird in Algorithmus 1 beschrieben.

Algorithmus 1: Allgemeines Algorithmus-Framework

Eingabe: s0 (Der aktuelle Spielstand)
Ausgabe: Optimale Entscheidung
1: Funktion MCTSSEARCH (s0) 2: 3: 4:
Erstelle den Stammknoten v0 mit dem
Zustand s0 , solange du im Rechenbudget bleibst.
Mache folgende
5: Schritte: v1 ÿ Auswahl (v0)
6: v2 ÿ Erweiterung (v1, s0) ÿ ÿ Simulation (v2, s0)

Machine Translated by Google

Backup (v2,ÿ) gibt
7: 8: **BESTCHILD(v0) zurück**

3.2. Strategie zum Verschieben von Gruppen im Algorithmus

Die Kernidee der Move Groups-Strategie ist wie folgt: Gemäß den Regeln wird ein Knoten im ursprünglichen Suchbaum (Amazon Movement und Arrow Shot kombiniert) in zwei Knoten zerlegt (die Amazon Movement bzw. Arrow Shot darstellen). Indem wir die Simulationsphase von MCTS nutzen, um einer zusätzlichen Ebene (der neu hinzugefügten Ebene) Wertebewertungen zuzuweisen , können wir die Ziele des Beschneidens und der Zeitersparnis erreichen, basierend auf den Wertebewertungen der zusätzlichen Ebene (z. B. mithilfe der UCT-Strategie).

Insbesondere wenn ein Knoten der Extraschicht seine untergeordneten Knoten noch nicht erweitert hat, ist sein Wert anfänglich unendlich. Daher erweitert jeder Knoten der Extraschicht seine untergeordneten Knoten zunächst einmal und verwendet eine Shallow-Rollout-Strategie, um einen anfänglichen Wert zu erhalten. Anschließend können wir den Knoten der Extraschicht mit dem höchsten Wert für weitere Erweiterung und Rollout auswählen, bis die untergeordneten Knoten eines bestimmten Knotens der Extraschicht vollständig generiert sind und der Knoten erneut ausgewählt wird. An diesem Punkt können wir dann die optimale Entscheidung aus den untergeordneten Knoten ableiten, was den Kern des frühen Beschneidens darstellt. Es ist zu beachten, dass sich aufgrund des dynamischen MCTS- Backpropagation-Mechanismus die Werte der Knoten der Extraschicht dynamisch ändern, wenn ihre untergeordneten Knoten erweitert werden. Daher kann ein frühes Beschneiden auch am zweiten, dritten oder anderen Knoten der Extraschicht erfolgen. Einzelheiten zur Berechnung und Aktualisierung der Werte finden Sie in Abschnitt 3.3. Da der gesamte Schritt eine Kombination aus zwei Schritten ist, nämlich „Amazon Movement“ und „Arrow Shot“, und die zusätzliche Ebene nur einen Schritt darstellt, muss die Gesamtzahl der Knoten in der zusätzlichen Ebene kleiner sein als die Gesamtz

Die Move Groups-Strategie erhöht jedoch durch die Einführung der zusätzlichen Ebene auch die Gesamtzahl der Baumknoten (für eine vollständige einstufige Operation). Wenn alle Knoten ohne vorheriges Beschneiden erweitert werden, wirkt sich der zusätzliche Aufwand (Erweiterung und Auswertung), der durch die Knoten in der zusätzlichen Ebene entsteht, negativ auf den Algorithmus aus. Dies führt außerdem dazu, dass – im schlimmsten Fall – die Hinzufügung der Move Groups-Strategie zum MCTS eine etwas schlechtere Leistung erbringt als der MCTS-Algorithmus ohne die Move Groups-Richtlinie. Eine vollständige theoretische Analyse der Move Groups-Richtlinie finden Sie ausführlich in Abschnitt 4.1.

3.3. Parallele Auswertungsstrategie im Algorithmus

Die Kernidee der Strategie der parallelen Auswertung besteht darin, dass der zeitaufwändige Auswertungsteil der MCTS-Simulation in paralleler Optimierung verarbeitet wird. In der Literatur [35] wurde gezeigt , **dass die Beschleunigung dieser arbeitsintensiven Auswertungsfunktion es ermöglicht, den Monte- Carlo-Baumsuchalgorithmus häufiger zu iterieren. Diese Metrik ist sehr nützlich, um die Tatsache auszugleichen, dass die Strategie der Bewegungsgruppen in ein lokales Optimum (Worst Case) fallen könnte; insbesondere ermöglicht die Beschleunigung der Auswertungsfunktion, dass die Monte-Carlo-Baumsuche so schnell wie möglich skaliert wird, wodurch die Auswirkungen des Worst-Case-Szenarios im Zusammenhang mit der Strategie der Bewegungsgruppen minimiert werden. Der Pseudocode für den Teil der parallelen Auswertung ist in Algorithmus 2 dargestellt.**

Algorithmus 2: Situationsbewertung

Eingabe: s0 (Der aktuelle Spielstatus)
Ausgabe: Situationsbewertungswert 1:
Funktion SituationEvaluation(s0) 2: 3:
t1, t2, p1, p2, Mobilität **ÿ ParallelEvaluationMetrics (s0)**
Wert = **k1 ÿ t1** + **k2 ÿ t2** + **k3 ÿ p1** + **k4 ÿ p2** + **k5 ÿ** Mobilitätsrendite **Wert**
4:

Knoten werden bewertet und die daraus resultierenden Werte bestimmen die Knotenauswahl. Traditionelle Monte-Carlo-Baumsuchalgorithmen wählen normalerweise den untergeordneten Knoten mit der hö

Machine Translated by Google

Knotenpunktzahl (Gewinnrate) basierend auf dem aktuellen Spielstand [36]. Diese Auswahlmethode ist jedoch nicht ideal, da sie dazu neigt, unmittelbare Gewinne zu priorisieren und zu lokalen Optima führen kann. Um dieses Problem zu lösen, verwendet dieses Dokument die „Shallow Rollout-Strategie“ [37] und den „baumbasierten oberen Konfidenzintervall“-Algorithmus [19], um die Knotenauswahl zu verfeinern. Insbesondere der Nutzungswert, der ausschließlich aus Knoten auf Entscheidungsebene abgeleitet wird, kann ungenau sein. Um die Genauigkeit zu verbessern, wird eine Shallow Rollout-Strategie eingesetzt, bei der Szenarien simuliert werden, die näher am Endspiel liegen. Darüber hinaus berücksichtigt der Algorithmus die Größenordnung des explorativen Werts eines Knotens , der für Knoten mit hoher Unsicherheit und zahlreichen zukünftigen Möglichkeiten erhöht ist . Zusammenfassend lässt sich sagen, dass der Ansatz sowohl den partiellen Gleichgewichtsnutzungswert als auch den Explorationswert bewertet, um eine fundiertere Entscheidungsfindung zu ermöglichen.

1. **„Shallow Rollout Strategy“: Dies bedeutet, dass der Nutzenwert nicht nur auf Grundlage des nächsten Zuges, sondern auch auf Grundlage der Spielposition berechnet wird, nachdem der Zug mehrmals zufällig simuliert wurde. In diesem Algorithmus legen wir die Anzahl der Rollouts auf zwei und die Tiefen auf 4 und 5 fest. Konkret führen wir nach der Erweiterung des vollständigen Entscheidungsknotens (Amazon-Bewegung plus Pfeilschuss) zwei Rollouts für die aktuelle Situation durch. In jedem Rollout simulieren wir zwei Schritte: die nächste Entscheidung des Gegners (Amazon-Bewegung plus Pfeilschuss) und unsere nachfolgende Entscheidung (Amazon- Bewegung plus Pfeilschuss), um einen eher zukunftsorientierten Nutzenwert zu erhalten. Einzelheiten finden Sie** im **Pseudocode-Algorithmus 3. Die Analyse der relevanten Parameter finden Sie in Anhang A.**

2. **„Baumbasiertes oberes Konfidenzintervall“: Wie in Gleichung (8) gezeigt, besteht die Berechnungsformel von UCT aus zwei Komponenten. Der erste Term,** Xi , **stellt den Nutzwert der Informationen im erstellten Suchbaum dar, und der zweite Term stellt den Explorationswert der nicht besuchten Knoten dar. Die Konstante** C **wird verwendet, um das Verhältnis des Kompromisses des Algorithmus zwischen den beiden Werten anzupassen. Beachten Sie, dass der Wert der Wurzelformel aufgrund des Vorhandenseins nicht erreichbarer untergeordneter Knoten unendlich ist, was sicherstellt, dass nicht erreichbare untergeordnete Knoten unter dem aktuellen Knoten mindestens einmal besucht werden .**

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

wobei Xi **die Gewinnrate des Kindknotens ist.** N **ist die Anzahl der Zugriffe auf den Elternknoten .** ni **ist die Anzahl der Zugriffe auf den Kindknoten.** C **ist eine Konstante (ÿ 2 und Szepesvári [36]** zeigte, dass der Wert die Hoeffding-Ungleichung mit Belohnungen **im Bereich von [0,** 1] **erfüllt .** $\frac{1}{=}$**).**

Hier können wir den vollständigen Pseudocode bereitstellen, wie in Algorithmus 3 gezeigt.

Algorithmus 3: MCTS mit der MG-PEO-Strategie 1: Funktion

```
Selection(v) 2: while v is
nonTerminal do
3:          wenn v nicht vollständig expandiert ist, dann
4:              zurückgeben v
5:          sonst
6:              v ÿ BESTCHILD(v)
7:      zurückgeben v

8: Funktion Expansion(v,&s0) 9: wenn
v ein Amazon Movement Node ist, dann
10:             ch ÿ Arrow Shot Node nach dem Zufallsprinzip
11:         erstellen, sonst wenn v ein Arrow Shot
12:             Node ist. ch ÿ Amazon Movement Node nach dem Zufallsprinzip
13:         erstellen. s0 ÿ Status des Boards aktualisieren.
14:         return ch

15: Funktion Simulation(v, s0).
```

Machine Translated by Google


```
S ÿ s0
//* Tiefe = 4 und 5 (zweimal unabhängige Reihenfolge)
S ÿ Rollout(S,Tiefe)
ÿ ÿ Situationsbewertung(en)
Rückkehr ÿ
Funktion Rollout(S,Tiefe)
wiederholen
wenn Terminal (S) ist , dann
Pass
        wenn v der Amazon Movement Node ist, dann
            Pfeil ÿ Zufälliger Pfeil (S)
            S ÿ doAction(v,Pfeil)
            Tiefe ÿ Tiefe - 1
        sonst, wenn v der Arrow Shot Node ist, dann
            Verschiebe ÿ übergeordnetes Element von v
            S ÿ führe Aktion aus (Bewegen, v)
            Tiefe ÿ Tiefe - 1
    bis Tiefe = 0 oder Gameover
    Rückkehr S
Funktion Backup(v,ÿ)
    N(v) ÿ N(v) + 1
    Q(v) ÿ Q(v) + ÿ(v, Elternteil von v)
    v ÿ übergeordnetes Element von v
Funktion BESTCHILD(v)
    returniere argmaxv ÿÿKinder von v  Q(v ÿ) / N(v ÿ) + c  sqrt(2 ln N(v) / N(v ÿ))
```

4. MG-PEO-Leistungsanalyse

In diesem Abschnitt entwerfen wir zwei Module zur Analyse der MG-PEO-Strategie. Zunächst analysieren wir die Leistung der Move Groups-Strategie bei Amazon. Zweitens führen wir eine ausführliche Diskussion über die Leistungsverbesserung des Parallel Evaluation-Mechanismus bei Amazonen.

4.1. Leistung der Move Groups-Strategie bei Amazon

Die Wirksamkeit eines Monte Carlo-Baumsuchalgorithmus basierend auf dem Gesetz der großen Zahlen hängen weitgehend von der Anzahl der Situationen ab, die durchsucht werden können. Je mehr Situationen der Algorithmus suchen kann, desto einfacher wird es sein, „Tricks“ zu finden, die normalerweise schwer zu finden sind finden und somit den Gegner schlagen. Mit anderen Worten, je mehr Rechenleistung, desto effektiver ist der Monte-Carlo-Algorithmus. Um also objektiv zu zeigen, Einfluss der Move Groups-Strategie auf die tatsächliche Leistung des MCTS-Algorithmus, Wir bieten eine neue Definition (gewinnende Konvergenzdistanz) zur Bewertung der Leistung des Algorithmus.

Definition 1. Gewinnende Konvergenzdistanz. Dies berechnet die Anzahl der Baumknoten, die erweitert werden durch den Suchbaum, sowohl in der Tiefe als auch in der Breite, vom aktuellen Zustand (Wurzelknoten) bis zum Gewinnerzustand (Spiel vorbei). Im Allgemeinen gilt: Je weniger neue Knoten der Monte-Carlo-Baum erweitern muss, um den Sieg zu erringen so schnell wie möglich, desto kürzer ist die gewinnende Konvergenzdistanz.

Um die gewinnende Konvergenzdistanz zu visualisieren, verwenden wir den ersten Eröffnungsschritt als Beispiel. Da die Anzahl der Knoten, die vom ersten Schritt bis zum Gewinn generiert werden können, Zustand unmessbar ist, verwenden wir die lokale Optimallösung des ersten Schrittes anstelle der Gewinnzustand, wie in Abbildung 3 dargestellt.

Auf dieser Grundlage können wir den Monte Carlo-Suchbaum mit der hinzugefügten Strategie „Gruppen verschieben“ und den Monte Carlo-Suchbaum ohne die hinzugefügte Strategie „Gruppen verschieben“ darstellen.

Machine Translated by Google

und berechnen Sie die besten und schlechtesten erweiterten Knotenzahlen, wie in Abbildung 4 dargestellt. und Tabelle 2.



Abbildung 3. Die im ersten Schritt der Öffnung angenommene lokale Optimallösung. Von links nach rechts: anfänglich Staat, nach der Amazon-Bewegung und nach dem Pfeilschuss.

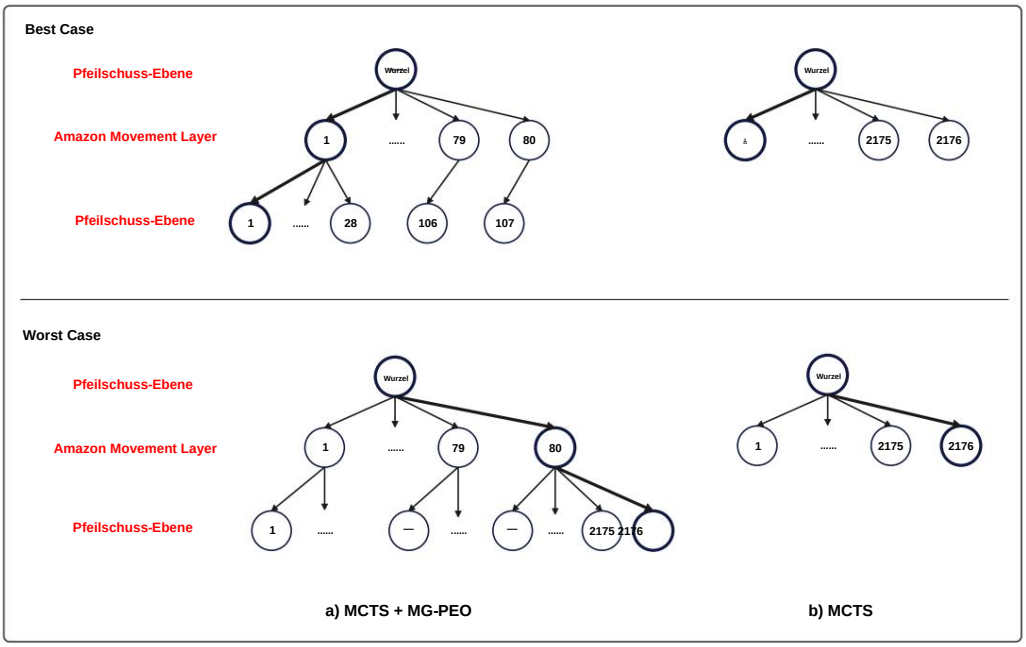


Abbildung 4. Diagramm der Berechnung der siegreichen Konvergenzdistanz. Die Zahl in der Abbildung bezieht sich auf die Reihenfolge, in der der Knoten auf einer bestimmten Ebene erweitert wird.

Tabelle 2. Gewinnende Konvergenzdistanz zweier Algorithmen (Einheit: Knoten).

Algorithmus	Best Case	Worst Case
MCTS + MG-PEO	187	2256
MCTS	2176	2176

Rote Ergebnisse zeigen das beste Ergebnis in der gleichen Spalte an.

In Abbildung 4 und Tabelle 2 wird im ersten Schritt der Monte-Carlo-Baumsuchalgorithmus mit Die hinzugefügte Strategie „Move Groups“ muss zuerst 80 Knoten der Amazon Movement-Ebene erweitern. Da der Wert eines Knotens auf der Amazon Movement-Ebene unendlich ist, wenn er noch nicht seine untergeordneten Knoten erweitert hat, muss jeder Knoten der Amazon Movement-Ebene weiter erweitert werden einen Arrow Shot Layer-Knoten, um einen Anfangswert zu erhalten. Anschließend können wir auswählen der Amazon Movement-Layer-Knoten mit der höchsten Wertung für den nächsten untergeordneten Knoten Erweiterung. Im besten Fall hat der erste Amazon Movement Layer-Knoten den höchsten Wert Punktzahl. Während der Erweiterung der nachfolgenden Arrow Shot-Layer-Knoten wird der Wert Punktzahl von Der übergeordnete Knoten bleibt auf dem in allen Amazon Movement-Ebenen beobachteten Maximum Knoten; die anderen Knoten in der Amazon Movement-Schicht führen die Erweiterung von

Machine Translated by Google

die Knoten der nächsten Schicht, was ein Spiegelbild des frühen Beschneidens ist. Schließlich die Anzahl der Knoten erweitert ist 187. Im schlimmsten Fall sind alle möglichen Knoten vollständig erweitert und durchlaufen, Dadurch wurden insgesamt 2256 Knoten erweitert.

Der Monte Carlo-Baumsuchalgorithmus ohne Einbeziehung der Move Groups Strategie muss den Knoten mit dem höchsten Wert als endgültige Entscheidung auswählen nach Generierung aller möglichen Knoten, unabhängig vom besten oder schlechtesten Szenario. Das Ergebnis ist ein Insgesamt werden 2176 Knoten erweitert.

Zusammenfassend können wir folgendes feststellen: Im besten Fall ist die siegreiche Konvergenz Die Distanz des MCTS mit der Move Groups-Strategie ist deutlich kürzer als die von das traditionelle MCTS, während im schlimmsten Fall die gewinnende Konvergenzdistanz des MCTS mit der Move Groups-Strategie ist etwas schwächer als das traditionelle MCTS. Dies deutet darauf hin, dass die Move Groups-Strategie starke Beschneidungs- und Zeitersparnisfähigkeiten besitzt, Gleichzeitig gibt es aber auch einige Mängel.

4.2. Durchführung der parallelen Evaluierungsstrategie

In diesem Unterabschnitt befassen wir uns mit der Anzahl der Threads, die für den parallelen Evaluierungsmechanismus erforderlich sind, sowie mit der Beschleunigungsrate, die Multithreading bringen kann. Der Verbrauch der Bewertungsfunktion resultiert hauptsächlich aus der Zeit für die Berechnung jedes Bewertungsmetrik, der Teil der parallelen Bewertung ist im Wesentlichen eine mehrfädige Berechnung von Auswertungsmetriken. Der Pseudocode für die Metrikfunktion der parallelen Auswertung wird angezeigt in Algorithmus 4.

Algorithmus 4: Parallele Bewertungsmetriken

```
Eingabe: s0 (Der aktuelle Spielstatus)
Ausgabe: t1,t2,p1,p2,mobility
1: Funktion ParallelEvaluationMetrics(s0)
2:      //* Die Berechnung der Metriken Territory(t1, t2) und Position(p1, p2) erfolgt mit dem gleichen
          unabhängige Funktion viermal, basierend auf dem King Move und QueenMove
          Methoden, die einmal für Schwarz und einmal für Weiß berechnet werden.
3: //* Wir optimieren seine externen Aufrufe parallel.
4: Modul1 = Lambda (Bereich)
5: für den Index im Bereich:
6:          Farbe ÿ Index/2
7:          Typ ÿ Index%2
            CalTP (Farbe, Typ)
8: 9:   parallel_für(Blockbereich(0,2),Modul1)

10:     //* Die Mobilitätsmetrik (Mobilität) wird Funktion für Funktion berechnet, und die
          Die interne Logik besteht darin, jede Brettposition (GridSize = 100) unabhängig zu verarbeiten.
11:     //* Wir optimieren die innere Schleifenverarbeitungslogik parallel.
12: Modul2 = Lambda (Bereich)
13:         für den Index im Bereich:
14: Farbe ÿ Index/2
15:           Typ ÿ Index%2
16:           CalEveryBoardPosition()
17:     parallel_für(Blockbereich(0,Rastergröße),Modul2)
```

A. Diskussion über die Anzahl der Threads: Da die Hauptmetriken der Bewertung Funktion umfassen Gebiet, Position und Manövrierfähigkeit Metriken, Werte des Gebiets und Positionsmetriken können mit einer einzigen Funktion berechnet werden. Da die Berechnungen jedoch Diese beiden Metriken beziehen sich auf die Farben der Stücke und die Berechnungsmethode (KingMove-Methode und QueenMove-Methode) sind vier unabhängige Berechnungen erforderlich für die Werte der Gebiets- und Positionsmetriken. Der Mobilitätsindikator wird berechnet durch eine Funktion, die eine unabhängige Berechnung erfordert. Der Algorithmus ist optimiert in

Machine Translated by Google

parallel für die Gebiets- und Positionsindikatoren, die theoretisch insgesamt vier Fäden. Für den Manövrierbarkeitsindikator werden theoretisch zehn Fäden für interne Berechnungen im Zusammenhang mit der Größe des Boards.

B. Untersuchung der Beschleunigungsrate, die Multithreading bringen kann: Um Um das Beschleunigungsverhältnis genau zu untersuchen, haben wir die für die Berechnung benötigte Zeit gezählt Funktion jeder Bewertungsmetrik. Die Daten sind in Tabelle 3 dargestellt.

Tabelle 3. Modulausführungszeit und Beschleunigungsverhältnis.

Modul	Serielle Ausführungszeit	Parallele Ausführungszeit	Beschleunigungsverhältnis
Modul 1	69 uns	29 uns	2.379
Modul2	45 uns	23 uns	1.956
Ganze Auswertungsfunktion	114 us	54 uns	2.111

Wie aus Tabelle 3 hervorgeht, beträgt die Berechnungszeit für das optimierte Gebiet und die optimierte Position Metriken erreichten eine 2,379-fache Beschleunigung, während die Berechnungszeit für die optimierte Mobilität Metrik erreichte eine 1,956-fache Beschleunigung. Das Gesamtbeschleunigungsverhältnis für den Evaluierungsteil war 2,111. In Abschnitt A haben wir die Anzahl der Threads besprochen und festgestellt, dass die theoretische Die Beschleunigung der Parallelisierung sollte der Anzahl der Threads entsprechen. Die Zeit Der Verbrauch jedes Moduls vor und nach der Optimierung unterschied sich etwas von dem Diskussion in Abschnitt A. Wir berücksichtigen, dass der Overhead von parallelen Bibliotheken von Drittanbietern[38] hat einige Rechenleistung verbraucht. Siehe Anhang B für Details zu den Analyse der CPU-Auslastung. Dennoch reduzierte der parallele Optimierungsmechanismus die Zeit die zur Verarbeitung derselben Aufgaben erforderlich sind, um das 2,111-fache, was zu einer erheblichen Leistungssteigerung führt Verbesserung der praktischen Anwendung.

5. Experiment und Ergebnisse

Um die Wirksamkeit der in diesem Artikel vorgeschlagenen Optimierungsstrategie für den MG-PEO-Algorithmus zu überprüfen, wurden in diesem Abschnitt drei Experimente entworfen: Das erste ve die relevanten Leistungsmetriken des MCTS-Algorithmus mit der MG-PEO-Strategie und der traditioneller MCTS-Algorithmus, und es wird ein Signifikanztest der Vergleichsergebnisse durchgeführt. Das zweite Experiment konzentriert sich auf die Bewertung des Ausmaßes der Wirkung jedes Optimierungsmoduls von MG-PEO auf MCTS, d. h. das Ablationsexperiment. Das dritte Experiment ist allgemeine experimentelle Analyse des MG-PEO.

Darüber hinaus ist es erwähnenswert, dass das MG-PEO Amazons-Programm vorgeschlagen hat in diesem Papier gewann den ersten Preis im Amazons-Wettbewerb bei den China Collegiate 2023 Computerspiele-Meisterschaft und Nationales Computerspiele-Turnier.

5.1. Informationen zur experimentellen Umgebung

Um die Zuverlässigkeit der experimentellen Ergebnisse zu gewährleisten, führten wir alle Experimente in einer einheitlichen Umgebung und detaillierte Informationen über die experimentellen Umgebungseinstellungen sind in Tabelle 4 aufgeführt.

Tabelle 4. Informationen zur experimentellen Umgebung.

Hardware und Einschränkungen	Details
Grafikkarte	GeForce RTX 3070
CPU	Intel Core i7-11800H
Entscheidungszeit	5 Sekunden
Anzahl der vom gleichen Gegner gespielten Spiele	2
Anzahl der Sätze pro Spiel	3
Experimentelle Matching-Plattform	SAU GAME-Plattform
Visual Studio-Laufzeitmodus	[Version x64]

Gegen denselben Gegner werden insgesamt sechs Sätze gespielt, davon drei Sätze als erster Spieler und drei als letzter zweiter Spieler.

Machine Translated by Google

5.2. Quantitative und qualitative Experimente

5.2.1. Gewinnprozentsatzstatistik und -analyse

Um die Wirksamkeit der Optimierungsstrategie des MG-PEO-Algorithmus objektiv und genau zu bewerten, haben wir zum Vergleich den MCTS-Algorithmus ohne Move-Groups-Optimierung und ohne Parallel Evaluation-Optimierung ausgewählt und insgesamt 100 Spiele unter denselben Bedingungen gespielt, wobei jedes Spiel in jeweils 3 Spiele (erste Hand und zweite Hand) unterteilt war. Konkret haben wir 33 Spieler ausgewählt, die an einem Computerspielwettbewerb teilgenommen haben, und 17 Spieler, die ihre Ranglisten (Rangliste) auf der Botzone- Plattform (https://www.botzone.org.cn/) als Open Source veröffentlicht hab

Aufgrund der Notwendigkeit von Drittanbieter-Bibliotheken haben wir die EXE-Programme für alle Spieler lokal exportiert und die Spiele dann gegeneinander auf der SAU-Plattform durchgeführt (http://computergames.caai.cn/platform/SAU_Game_Platform_2.1.0_r3.rar, abgerufen am 20. März 2024), und die Ergebnisse der Spiele sind in Tabelle 5 dargestellt. Detaillierte Informationen zu allen Spielen finden Sie in Anhang C.

Tabelle 5. Die Gewinnrate der beiden Algorithmen.

Algorithmus	Gewinnprozentsatz
MCTS	74 %
MCTS + MG-PEO	97 %

Rote Ergebnisse zeigen das beste Ergebnis in der gleichen Spalte an.

In Tabelle 5 hat der MCTS-Algorithmus eine Gewinnrate von 74 % im Amazons Duel auf der SAU-Matchmaking-Plattform. In derselben Testumgebung zeigt der MCTS-Algorithmus mit der MG-PEO-Strategie eine höhere Stabilität und seine Gewinnrate im Amazons Match auf der SAU-Matchmaking-Plattform erreicht 97 %. Zusammenfassend lässt sich sagen, dass der MCTS-Algorithmus mit der MG-PEO-Strategie bei der Berücksichtigung und Nutzung von Strategien eine umfassendere und tiefergehende Optimierung aufweist als der MCTS-Algorithmus und daher in **diesem Testszenario eine höhere Gewinnrate aufweist .**

Es ist erwähnenswert, dass wir durch die Beobachtung der Spiele der beiden Algorithmen festgestellt haben, dass der optimierte MCTS-Algorithmus in einigen kritischen Runden oft genauere Entscheidungen trifft als der nicht optimierte MCTS-Algorithmus. In diesen Schlüsselrunden geht es oft um entscheidende Positionen für Strategien wie „Gebiet einkreisen“ oder „Gegner blockieren“. Darüber hinaus haben wir während der Spiele festgestellt, dass der optimierte MCTS-Algorithmus gegenüber einigen traditionellen Algorithmen (wie Minimax und traditionellem MCTS) einen klaren Vorteil hat. Im Vergleich zu neuronalen Netzwerkprogrammen, die mit einer großen Menge an Spieldaten trainiert wurden, lässt seine Leistung jedoch leicht zu wünschen übrig. Dies bietet jedoch eine Richtung, in der die Leistung des Algorithmus im nächsten Schrit

5.2.2. Visualisierung und Analyse der erfolgreichen Konvergenzdistanz

In diesem Unterabschnitt zählen wir die Anzahl der Matchballrunden, in denen die Gewinnprozentsatzbewertung des Algorithmus 90 % oder mehr erreicht (die Schwelle für einen soliden Sieg) in 100 Spielen, die von den beiden Algorithmen auf der SAU-Plattform gespielt wurden, um die Definition der Konvergenzdistanz eines Gewinns zu visualisieren und die Stärken und Schwächen der Spielfähigkeiten der beiden Algorithmen des Amazons-Spiels zu vergleichen. Hier betrachten wir den Ausgang eines Spiels im Allgemeinen als klar, wenn der Algorithmus eine Gewinnquote von 90 % oder mehr auswertet (abgesehen von geringfügigen Fehlern der dominanten Seite); solche Momente gelten als Matchbälle, und wir glauben, dass die Algorithmen umso stärker sind, je früher dies geschieht. Die durchschnittlichen resultierenden Daten für die Anzahl der Matchballrunden für die beiden Algorithm

Machine Translated by Google

Tabelle 6. Die durchschnittliche Anzahl von Matchballrunden in 100 Spielen, bei denen die Gewinnratenbewertungswerte zweier Algorithmen 90 % überschreiten.

Algorithmus	Mittelwert
MCTS	42,57
MCTS + MG-PEO	39,68

Wenn der Spieler verliert, werden die Matchballrunden als maximale Rundenanzahl gewertet, also 47. Darüber hinaus zeigen rote Ergebnisse das beste Ergebnis in der gleichen Spalte an.

Wie aus Tabelle 6 ersichtlich , beträgt die durchschnittliche Anzahl der Matchballrunden, die der MCTS-Algorithmus benötigt, um die stabile Gewinnschwelle (Bewertungswert über 90 %) zu erreichen, 42,57 Runden. Andererseits übertrifft der MCTS-Algorithmus mit der MG-PEO-Strategie MCTS hinsichtlich des Erreichens derselben stabilen Gewinnschwelle mit durchschnittlich 39,68 Runden.

Dies zeigt, dass der MCTS-Algorithmus mit der MG-PEO-Strategie schneller konvergiert als MCTS. Er benötigt weniger Durchschnittsrunden als der herkömmliche MCTS-Algorithmus. Der MCTS-Algorithmus mit der MG-PEO-Strategie verwendet die Move Groups-Strategie und die Parallel Evaluation-Strategie, um den Strategieauswahlprozess des MCTS-Algorithmus zu optimieren, sodass er in kürzerer Zeit die besten Suchpfade finden und im Voraus einen stabilen Gewinnzustand mit einer Gewinnrate von über 90 % erreichen kann.

Zusammenfassend lässt sich sagen, dass der MCTS-Algorithmus mit der MG-PEO-Strategie eine kürzere Gewinnkonvergenzdistanz aufweist als der herkömmliche MCTS-Algorithmus. Er benötigt weniger Runden, um eine höhere Gewinnrate zu erreichen, und weist somit eine überlegene Spielfähig

5.2.3. Signifikanztest In

diesem Unterabschnitt führen wir einen Signifikanztest für die Daten zur Matchpoint-Rundenzählung durch, wie in Abschnitt 5.2.2 gezählt, um zu bestimmen, ob es einen signifikanten Unterschied zwischen ihnen gibt . Wir führen die Tests mit einem Signifikanzniveau von 0,05 durch. Die vollständigen Daten zur Matchpoint-Rundenzählung sind in Anhang C detailliert aufgeführt. Zuerst verwenden wir den Shapiro-Wilk-Test, um anzunehmen, dass beide Datensätze einer Normalverteilung gehorchen. Allerdings entspricht keiner der Datensätze einer Normalverteilung . Basierend auf diesem Ergebnis verwenden wir den Mann-Whitney-U-Test, um anzunehmen, dass es keinen signifikanten Unterschied zwischen den beiden Datengruppen gibt. Dann vergleichen wir die Verteilungseigenschaften der Matchpoint-Rundendaten der beiden Algorithmen. Die detaillierten Ergebnisse sind in Tabelle 7 aufgeführt.

Tabelle 7. Bedeutung der Testergebnisse.

Daten zu Matchballrunden	p-Wert für Shapiro–Wilk-Test (P1) 0,0002	p-Wert für Mann–Whitney U Prüfung (P2)
MCTS	0,0013	1,16 × 10ÿ8
MCTS + MG-PEO		

Als Ergebnis des Shapiro-Wilk-Tests zeigten beide getesteten Algorithmen einen P1- Wert von weniger als 0,05 für die Daten zur Anzahl der Matchpunktrunden, was uns dazu veranlasste, die Nullhypothese abzulehnen, dass die Daten für MCTS+MG-PEO und die Daten für MCTS beide einer Normalverteilung entsprechen. Die Ergebnisse des Mann-Whitney-U-Tests zeigen, dass der P2- Wert für die Daten zur Anzahl der Matchpunktrunden zwischen den beiden Algorithmen deutlich kleiner als 0,05 ist, sodass wir die ursprüngliche Hypothese ablehnen. Das heißt, es gibt einen signifikanten Unterschied in der durchschnittlichen Anzahl der Gewinnschritte zwischen den beiden Algorithmen. Insbesondere ist die durchschnittliche Anzahl an Gewinnzügen, die MCTS mit der MG-PEO- Strategie benötigt, erheblich geringer als die, die für herkömmliche MCTS gegen 100 Spieler erforderlich ist. Dies legt die Vermutung nahe, dass MCTS mit der MG-PEO-Strategie herkömmliche MCTS in Bezug **auf die Spielfähigkeit bei Amazonen übertrifft .**

Machine Translated by Google

5.3. Ablationsexperiment

In diesem Unterabschnitt entwerfen wir zur Bewertung der Wirksamkeit der Move Groups-Strategie und der Parallel Evaluation-Strategie in Amazons zwei Experimente: eines, das nur die Move Groups-Strategie ohne Verwendung der Parallel Evaluation-Strategie verwendet, und ein anderes, das nur die Parallel Evaluation-Strategie ohne Verwendung der Move Groups-Strategie verwendet. Die experimentellen Ergebnisse sind in Tabelle 8 dargestellt. Einerseits konzentriert sich die Move Groups-Strategie auf MCTS-Iterationen innerhalb einer begrenzten Zeit auf die Unterknoten wertvollerer Knoten der zusätzlichen Ebene, indem sie ihre Wertbewertungen nutzt und so die Erweiterung unwichtiger Knoten vermeidet. Während die Move Groups-Strategie starke Beschneidungsfähigkeiten aufweist, ändern sich die Wertbewertungen der Knoten der zusätzlichen Ebene aufgrund des dynamischen Backpropagation-Aktualisierungsmechanismus von MCTS dynamisch mit der Erweiterung ihrer Unterknoten. Folglich können alle Knoten in der zusätzlichen Ebene vollständig erweitert und mehrmals durchlaufen werden (es hat keine Beschneidung stattgefunden), was möglicherweise zu einem höheren Zeitaufwand im Vergleich zu MCTS-Algorithmen ohne die Move Groups-Strategie führt. Da andererseits der Umfang der Situationsbewertung begrenzt ist, bietet die Verwendung einer Optimierung für parallele Bewertungen ohne Übernahme der Strategie „Gruppen verschieben“ nur eine begrenzte Leistungsverbesserung . Daher ist die kombiniert

Tabelle 8. Analyse des Ablationsexperiments der Move Groups-Strategie und der Parallel Evaluation-Strategie.

Algorithmus	Gewinnprozentsatz	Anzahl der MCTS-Iterationen innerhalb von 1 s
MCTS	74 %	2176
MCTS + PEO	89 %	10.026
MCTS + MG	83 %	5663
MCTS + MG-PEO	97 %	45.536

Die „Anzahl der MCTS-Iterationen innerhalb von 1s“ bezieht sich auf die Anzahl der durchgeführten Monte-Carlo-Baumsuchiterationen innerhalb einer Sekunde. Bei der Move Groups-Strategie resultiert jede MCTS-Iteration entweder in einem Amazon Movement- Knoten oder einem Arrow Shot-Knoten. Ohne die Move Groups-Strategie resultiert jede MCTS-Iteration in einem vollständigen Entscheidungsknoten (Amazon Movement plus Arrow Shot). Darüber hinaus zeigen rote Ergebnisse das beste Ergebnis in derselben Spalte an.

5.4. Generalisierungsanalyse **In**

diesem Unterabschnitt haben wir eine allgemeine Analyse der Strategie „Gruppen bewegen“ und der Strategie „Parallele Auswertung“ durchgeführt und ihre Übertragbarkeit auf andere Brettspiele eingehend diskutiert.

1. **Strategie der Zuggruppen: In manchen Spielen kann der Verzweigungsfaktor sehr groß sein, aber viele Züge sind ähnlich. Die Strategie der Zuggruppen schafft eine zusätzliche Entscheidungsebene, in der alle möglichen Aktionen gruppiert werden, sodass wir den großen Verzweigungsfaktor des Spiels effizienter handhaben können. Diese Strategie ist für andere Spiele von Vorteil, wie etwa Go [39] und das mehrarmige Banditenspiel mit Zuggruppen [40]. Wir glauben, dass es sich immer noch lohnt, diese Strategie zu erforschen und zu untersuchen.**
2. **Parallele Evaluierungsstrategie: Die parallele Evaluierungsstrategie kann eine effektive Beschleunigung bieten, egal ob bei einfachen regelbasierten Spielen wie Gomoku, komplexeren Spielen wie Schach und Go oder sogar modernen innovativen Brettspielen. Ihr Hauptvorteil liegt in ihrer Fähigkeit, den Entscheidungsbaum effizient zu erkunden, was eine häufige Anforderung in verschiedenen Brettspielen ist.**

6. Schlussfolgerungen und zukünftige Arbeiten

In diesem Artikel schlagen wir eine effiziente Optimierung des Monte Carlo Tree Search-Algorithmus für das Amazons-Spiel vor, die die Monte Carlo Tree Search-Fähigkeit durch die Move Groups-Strategie verfeinert. Der Parallel Evaluation-Mechanismus befasst sich mit dem Risiko, dass die Move Groups-Strategie in ein lokales Optimum fällt (diese Komponenten sind

Machine Translated by Google

komplementär zueinander). Die experimentellen Ergebnisse bestätigen, dass die vorgeschlagene Methode in diesem Papier verbessert die Spielbarkeit des Amazons-Spiels deutlich.

Auch künftige Forschungsanstrengungen sind in unseren Plänen enthalten. Erstens, um die Zeitersparnis zu maximieren Potenzial der Move Groups Strategie Beschneidung, hoffen wir, den Einsatz von intelligenteren Ansätze (wie heuristische Funktionen) zur Steuerung des Bereinigungsverhaltens der Move Groups Strategie. Zweitens demonstrieren die Experimente in diesem Artikel die mächtige Leistungssteigerungspotenzial der parallelen Optimierung, daher planen wir, mehr zu berücksichtigen parallele Möglichkeiten und effizientere parallele Strategien in der Zukunft. Schließlich haben wir auch hoffe, dass dieser Algorithmus erweitert werden kann, um weitere Spielprobleme zu lösen.

Autorenbeiträge: Alle Autoren haben zur Konzeption und Gestaltung der Studie beigetragen. Methodik, Software, Validierung, Untersuchung, Datenanalyse, Modellverifizierung, Ressourcen, Datenkuratierung, Visualisierung , Schreiben – Bearbeiten und Überprüfen: LZ- und HZ-Betreuung, Projektverwaltung, Finanzierun Erwerb: YZ. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu.

Finanzierung: Diese Forschung erhielt keine externe Finanzierung.

Datenverfügbarkeitserklärung: Ein Teil des in dieser Studie verwendeten Wettbewerbskodex ist der öffentlich zugängliche Open-Source-Code, der auf der folgenden Website zu finden ist: https://www.botzone.org.cn/, Zugriff am 2. März 2024.

Interessenkonflikte: Die Autoren erklären, dass sie keine finanziellen Interessen oder persönlichen Beziehungen haben relevant für den Inhalt dieses Artikels.

Anhang A

In Anhang A diskutieren und analysieren wir die Auswirkungen der Rollout-Tiefe in der Simulationsphase auf die Fähigkeit des Algorithmus, Amazons zu spielen, und nutzen die gewinnende Ko Distanzmetrik (siehe Abschnitt 4.2 für Details), um die Spielfähigkeit zu bewerten. Shallow Rollout Strategien können simulierte Positionen näher am Endspiel und weiter vorausschauende Spiele berücksichtigen, um genauere Positionsbewertungswerte zu berechnen, die verwendet werden können, um besser die Knotenauswahl für MCTS steuern. Daher ist es wichtig, die Tiefe der Rollout; eine zu geringe Tiefe wird den Nutzungswert (Gewinnrate) in der Position Bewertungsformel weniger glaubwürdig, während eine zu große Tiefe zu einer einzigen Simulation führt zu langer Iterationsprozess und die Anzahl der Knoten, die in einem begrenzten Zeitraum erweitert werden der Zeit wird kürzer, wodurch die Genauigkeit der Monte-Carlo-Baumsuchentscheidung reduziert wird. Wir haben den Rollout-Tiefe-Wert von 2 auf 10 angepasst und festgestellt, dass der optimale Wert von Die Rollouttiefe beträgt 4 oder 5, wie in Tabelle A1 gezeigt.

Tabelle A1. Parametrische Analyse der Rollout-Tiefe.

Rollouttiefe	1	2	3	4	5	6	7–10	4 und 5
Matchballrunden 40,25		40.14	40,0	39,9	39,84	40,04	40.1	39,68

Rote Ergebnisse zeigen das beste Ergebnis in der gleichen Spalte an.

Darüber hinaus haben wir in unseren Experimenten festgestellt, dass Rollout zweimal (Auswahl von 4 und 5 in der Tiefe) besser funktioniert als Rollout einmal (Auswahl von nur 4 oder 5), d. h. die Anzahl der R bei Matchbällen ist kleiner, die Distanz der Konvergenz zum Gewinnen ist kürzer, und die Algorithmus ist besser in der Lage, Amazonen zu spielen.

Anhang B

Anhang B zeigt die detaillierte CPU-Auslastung des Algorithmus mit und ohne Strategie der parallelen Auswertung. Die Überwachungsumgebung ist der Performance-Profiler in Microsoft Visual Studio 2019, wobei der Algorithmus nur eine Entscheidung trifft, die 20 s. Wie in Abbildung A1 dargestellt, stellt der obere Teil des Bildes die CPU-Auslastung dar, und der schwankende Teil der CPU-Auslastungswerte entspricht dem Algorithmus Entscheidungsphase, während der verbleibende Teil der Input- und Outputphase entspricht. Leerlaufphasen.

Machine Translated by Google

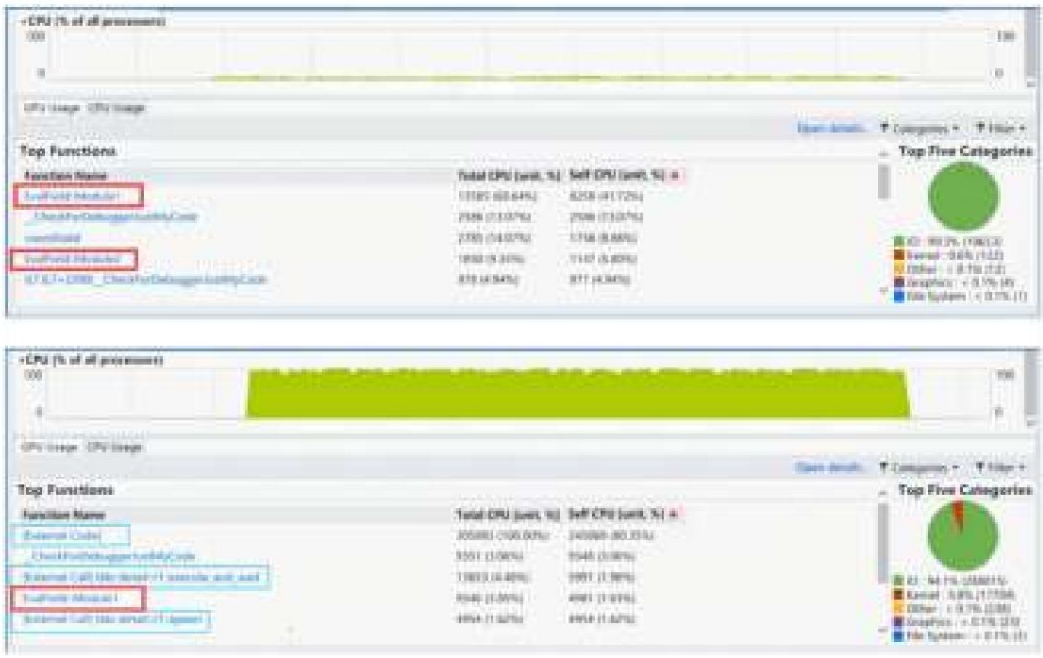


Abbildung A1. CPU-Auslastung und die 5 Module, die unter der Bedingung [Debug ×64] die meisten CPU-Ressourcen belegen. Der obere Teil zeigt die Überwachungsergebnisse des Algorithmus, der sequenziell ausgeführt wird, während der untere Teil zeigt die Überwachungsergebnisse des parallel ausgeführten Algorithmus.

Aus Abbildung A1 ist ersichtlich, dass die CPU-Auslastung des Algorithmusprogramms in Die parallele Ausführung ist deutlich höher als bei der sequentiellen Ausführung. Wenn der Algorithmus Das Programm wird sequentiell ausgeführt, die beiden rot umrahmten Module der Auswertungsfunktion werden unter den Top 5 Modulen, die die meisten CPU-Ressourcen beanspruchen. Dies deutet darauf hin, dass parallele Optimierung des Auswertungsteils ist sehr notwendig. Bei paralleler Ausführung des Algorithmus-Programm, die blau umrahmte Drittanbieter-Parallelbibliothek nimmt einen erheblichen Menge an CPU-Ressourcen. Dies deutet indirekt darauf hin, dass die Leistungsverbesserung nach der Parallelisierung wird die theoretische Beschleunigung aufgrund erheblicher Ressourcen Verbrauch durch externe Codes und externe Anrufe.

Anhang C

Anhang C zeigt die kompletten Spielinformationen, sowie die Matchballrunde Anzahl, wie in den Tabellen **A2 und A3 gezeigt .**

Tabelle A2. Vollständige Spielinformationen (MCTS plus MG-PEO).

ID	Erster Spieler	Zweiter Spieler	Ergebnis	Übereinstimmen Punkt Runden	Gewinner
1	Amazonen basierend auf Qt's QMainWindow-Framework	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
2	Göttliche Bewegung (unsere)	Amazonen basierend auf Qt's QMainWindow-Framework	3:0	39	Göttliche Bewegung (unsere)
3	Gomoku	Göttliche Bewegung (unsere)	0:3	39	Göttliche Bewegung (unsere)
4	Göttliche Bewegung (unsere)	Gomoku	3:0	36	Göttliche Bewegung (unsere)
5	Othello	Göttliche Bewegung (unsere)	0:3	35	Göttliche Bewegung (unsere)
6	Göttliche Bewegung (unsere)	Othello	3:0	33	Göttliche Bewegung (unsere)
7	Gottes Algorithmus	Göttliche Bewegung (unsere)	0:3	43	Göttliche Bewegung (unsere)

Machine Translated by Google

Tabelle A2. Forts.

ID	Erster Spieler	Zweiter Spieler	Ergebnis	Übereinstimmen Punkt Runden	Gewinner
8	Göttliche Bewegung (unsere)	Gottes Algorithmus	3:0	42	Göttliche Bewegung (unsere)
9	Amazons Gaming-System	Göttliche Bewegung (unsere)	0:3	39	Göttliche Bewegung (unsere)
10	Göttliche Bewegung (unsere)	Amazons Gaming-System	3:0	37	Göttliche Bewegung (unsere)
11	Vorab Kleine Königinnen	Göttliche Bewegung (unsere)	0:3	40	Göttliche Bewegung (unsere)
12	Göttliche Bewegung (unsere)	Vorab Kleine Königinnen	3:0	37	Göttliche Bewegung (unsere)
13	Schachmatt	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
14	Göttliche Bewegung (unsere)	Schachmatt	3:0	38	Göttliche Bewegung (unsere)
15	Cliffhanger-Amazonen	Göttliche Bewegung (unsere)	0:3	40	Göttliche Bewegung (unsere)
16	Göttliche Bewegung (unsere)	Cliffhanger-Amazonen	3:0	37	Göttliche Bewegung (unsere)
17	Qi Kaides Sieg	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
18	Göttliche Bewegung (unsere)	Qi Kaides Sieg	3:0	36	Göttliche Bewegung (unsere)
19	Super Amazon	Göttliche Bewegung (unsere)	0:3	39	Göttliche Bewegung (unsere)
20	Göttliche Bewegung (unsere)	Super Amazon	3:0	36	Göttliche Bewegung (unsere)
21	Gottes Algorithmus	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
22	Göttliche Bewegung (unsere)	Gottes Algorithmus	3:0	37	Göttliche Bewegung (unsere)
23	Reisende Amazonen	Göttliche Bewegung (unsere)	0:3	40	Göttliche Bewegung (unsere)
24	Göttliche Bewegung (unsere)	Reisende Amazonen	3:0	35	Göttliche Bewegung (unsere)
25	Amazon Supreme Schach	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
(unsere) 26	Göttliche Bewegung	Amazon Supreme Chess	3:0	39	Göttliche Bewegung (unsere)
27	Schachtruppen fallen aus Himmel	Göttliche Bewegung (unsere)	0:3	43	Göttliche Bewegung (unsere)
28	Göttliche Bewegung (unsere)	Schachtruppen fallen aus Himmel	3:0	41	Göttliche Bewegung (unsere)
29	Nelkenamazonen	Göttliche Bewegung (unsere)	0:3	40	Göttliche Bewegung (unsere)
30	Göttliche Bewegung (unsere)	Nelkenamazonen	3:0	37	Göttliche Bewegung (unsere)
31	Shao Guangs Amazonen	Göttliche Bewegung (unsere)	0:3	43	Göttliche Bewegung (unsere)
32	Göttliche Bewegung (unsere)	Shao Guangs Amazonen	3:0	39	Göttliche Bewegung (unsere)
33	KI spielt Schach	Göttliche Bewegung (unsere)	0:3	43	Göttliche Bewegung (unsere)
34	Göttliche Bewegung (unsere)	KI spielt Schach	3:0	39	Göttliche Bewegung (unsere)
35	Canopus One	Göttliche Bewegung (unsere)	0:3	39	Göttliche Bewegung (unsere)
36	Göttliche Bewegung (unsere)	Canopus One	3:0	35	Göttliche Bewegung (unsere)
37	Gewinnen Sie das Eröffnungsteam	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
38	Göttliche Bewegung (unsere)	Gewinnen Sie das Eröffnungsteam	3:0	39	Göttliche Bewegung (unsere)
39	Wukong-Amazonen 40	Göttliche Bewegung (unsere)	0:3	34	Göttliche Bewegung (unsere)
	Göttliche Bewegung (unsere)	Wukong-Amazonen	3:0	30	Göttliche Bewegung (unsere)
41	Reisende Amazonen	Göttliche Bewegung (unsere)	0:3	44	Göttliche Bewegung (unsere)
42	Göttliche Bewegung (unsere)	Reisende Amazonen	3:0	40	Göttliche Bewegung (unsere)
43	Yue	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
44	Göttliche Bewegung (unsere)	Yue	3:0	40	Göttliche Bewegung (unsere)
45	Final Amazon	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
46	Göttliche Bewegung (unsere)	Final Amazon	3:0	38	Göttliche Bewegung (unsere)
47	Informationen zur Universität - Monte Carlo	Göttliche Bewegung (unsere)	1:2	42	Göttliche Bewegung (unsere)
48	Göttliche Bewegung (unsere)	Informationen zur Universität - Monte Carlo	3:0	40	Göttliche Bewegung (unsere)
49	Haha Hallo 50	Göttliche Bewegung (unsere)	1:2	42	Göttliche Bewegung (unsere)
	Göttliche Bewegung	Haha, hallo	3:0	39	Göttliche Bewegung (unsere)
(unsere) 51	Drachensieg 52	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
	Göttliche Bewegung	Drachensieg	3:0	37	Göttliche Bewegung (unsere)
(unsere) 53	Super Amazon 54	Göttliche Bewegung (unsere)	0:3	43	Göttliche Bewegung (unsere)
	Göttliche Bewegung	Super Amazon	3:0	38	Göttliche Bewegung (unsere)
(unsere) 55	Basispaarung Team 56	Göttliche Bewegung (unsere)	0:3	44	Göttliche Bewegung (unsere)
	Göttliche Bewegung (unsere)	Basispaarungsteam	3:0	41	Göttliche Bewegung (unsere)
57	Bestehen Sie das Level	Göttliche Bewegung (unsere)	3:0	47	Bestehen Sie das Level

Machine Translated by Google

Tabelle A2. Forts.

ID	Erster Spieler	Zweiter Spieler	Ergebnis	Übereinstimmen Punkt Runden	Gewinner
58	Göttliche Bewegung (unsere)	Bestehen Sie das Level	0:3	47	Bestehen Sie das Level
59	Jiaotong-Universität Dalian Amazonen Team 1	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
60	Göttliche Bewegung (unsere)	Jiaotong-Universität Dalian Amazonen Team 1	3:0	39	Göttliche Bewegung (unsere)
61	Gehen Sie an Land	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
62	Göttliche Bewegung (unsere)	Gehen Sie an Land	3:0	40	Göttliche Bewegung (unsere)
63	Leer	Göttliche Bewegung (unsere)	0:3	44	Göttliche Bewegung (unsere)
64	Göttliche Bewegung (unsere)	Leer	3:0	41	Göttliche Bewegung (unsere)
65	Stier	Göttliche Bewegung (unsere)	0:3	42	Göttliche Bewegung (unsere)
66	Göttliche Bewegung (unsere)	Stier	3:0	40	Göttliche Bewegung (unsere)
67	Wukong-Amazonen	Göttliche Bewegung (unsere)	1:2	34	Göttliche Bewegung (unsere)
68	Göttliche Bewegung (unsere)	Wukong-Amazonen	3:0	31	Göttliche Bewegung (unsere)
69	Gangzi Fans Support Team	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
70	Göttliche Bewegung (unsere)	Gangzi Fans Support Team	3:0	39	Göttliche Bewegung (unsere)
71	Thai Pants Scharf	Göttliche Bewegung (unsere)	0:3	44	Göttliche Bewegung (unsere)
72	Göttliche Bewegung (unsere)	Thai Pants Scharf	3:0	41	Göttliche Bewegung (unsere)
73	Grüner Graskuchen	Göttliche Bewegung (unsere)	0:3	34	Göttliche Bewegung (unsere)
74	Göttliche Bewegung (unsere)	Grüner Graskuchen	3:0	30	Göttliche Bewegung (unsere)
75	Konnte die Luft nicht greifen und Schnapp dir die Pflanzenhirn-Hypoxie Team	Göttliche Bewegung (unsere)	0:3	40	Göttliche Bewegung (unsere)
76	Göttliche Bewegung (unsere)	Konnte die Luft nicht greifen und Schnapp dir die Pflanzenhirn-Hypoxie Team	3:0	37	Göttliche Bewegung (unsere)
77	DG 78	Göttliche Bewegung (unsere)	0:3	44	Göttliche Bewegung (unsere)
	Göttliche Bewegung (unsere)	DG	3:0	39	Göttliche Bewegung (unsere)
79	Warum ist Tang Yang ein Gott? 80	Göttliche Bewegung (unsere)	0:3	42	Göttliche Bewegung (unsere)
	Göttliche Bewegung (unsere)	Warum ist Tang Yang ein Gott?	3:0	39	Göttliche Bewegung (unsere)
81	Dream Team	Göttliche Bewegung (unsere)	0:3	43	Göttliche Bewegung (unsere)
82	Göttliche Bewegung (unsere)	Dream Team	3:0	39	Göttliche Bewegung (unsere)
83	Pferdegesichtsrock Täglich und Einfach zu tragen	Göttliche Bewegung (unsere)	0:3	43	Göttliche Bewegung (unsere)
84	Göttliche Bewegung (unsere)	Pferdegesichtsrock Täglich und Einfach zu tragen	3:0	40	Göttliche Bewegung (unsere)
85	Genshin Impact Expert Divine	Göttliche Bewegung (unsere)	0:3	41	Göttliche Bewegung (unsere)
86	Move (unserer)	Genshin Impact Expert Divine	3:0	37	Göttliche Bewegung (unsere)
87	Code Apprentice Divine	Move (unserer)	0:3	42	Göttliche Bewegung (unsere)
	Move (unser) 88 89	Code Apprentice	3:0	41	Göttliche Bewegung (unsere)
	Amazon Drift Notes Divine Move	Divine Move (unserer)	0:3	39	Göttliche Bewegung (unsere)
90	(unser)	Amazon Drift Notes Divine	3:0	37	Göttliche Bewegung (unsere)
91	Die Liebe wird verschwinden, oder?	Move (unsere)	0:3	38	Göttliche Bewegung (unsere)
92	Göttlicher Zug (unserer)	Die Liebe wird verschwinden, oder?	3:0	33	Göttliche Bewegung (unsere)
93	Die kleine Su und die Katze 94	Göttliche Bewegung (unsere)	0:3	43	Göttliche Bewegung (unsere)
	Göttlicher Zug (unserer) 95 Ich	Die kleine Su und die	3:0	42	Göttliche Bewegung (unsere)
	will Marla nicht 96 Göttlicher	Katze Divine Move (unsere)	0:3	44	Göttliche Bewegung (unsere)
	Zug (unserer) 97 Parameter-	Will keinen Marla	3:0	43	Göttliche Bewegung (unsere)
	Anpassungsteam 98 Göttlicher Zug (unserer)	Divine Move (unseren)	0:3	44	Göttliche Bewegung (unsere)
		Parameteranpassungsteam Divine	3:0	41	Göttliche Bewegung (unsere)
99	AlphaAmazon 100 Divine	Move (unser)	0:3	43	Göttliche Bewegung (unsere)
	Move (unser)	AlphaAmazon	3:0	42	Göttliche Bewegung (unsere)

Machine Translated by Google

Tabelle A3. Vollständige Spielinformationen (MCTS).

ID	Erster Spieler	Zweiter Spieler	Ergebnis	Übereinstimmen Punkt Runden	Gewinner
1	Amazonen basierend auf Qt's QMainWindow-Framework	MCTS (Basislinie)	1:2	43	MCTS (Basislinie)
2	MCTS (Basislinie)	Amazonen basierend auf Qt's QMainWindow-Framework	3:0	41	MCTS (Basislinie)
3	Gomoku	MCTS (Basislinie)	3:0	47	Gomoku
4	MCTS (Basislinie)	Gomoku	3:0	45	MCTS (Basislinie)
5	Othello	MCTS (Basislinie)	2:1	47	Othello
6	MCTS (Basislinie)	Othello	2:1	44	MCTS (Basislinie)
7	Gottes Algorithmus	MCTS (Basislinie)	0:3	43	MCTS (Basislinie)
8	MCTS (Basislinie)	Gottes Algorithmus	3:0	40	MCTS (Basislinie)
9	Amazons Gaming-System	MCTS (Basislinie)	0:3	42	MCTS (Basislinie)
10	MCTS (Basislinie)	Amazons Gaming-System	3:0	40	MCTS (Basislinie)
	Vorab Kleine Königinnen 11	MCTS (Basislinie)	3:0	47	Vorab Kleine Königinnen
12	MCTS (Basislinie)	Vorab Kleine Königinnen	3:0	43	MCTS (Basislinie)
13	Schachmatt	MCTS (Basislinie)	2:1	47	Schachmatt
14	MCTS (Basislinie)	Schachmatt	3:0	40	MCTS (Basislinie)
	Cliffhanger Amazonen 15	MCTS (Basislinie)	3:0	47	Schachmatt
16	MCTS (Basislinie) 17	Cliffhanger-Amazonen	3:0	44	MCTS (Basislinie)
	Qi Kaides Sieg	MCTS (Basislinie)	0:3	44	MCTS (Basislinie)
18	MCTS (Basislinie) 19	Qi Kaides Sieg	3:0	41	MCTS (Basislinie)
	Super Amazon	MCTS (Basislinie)	2:1	47	Super Amazon
20	MCTS (Basislinie) 21	Super Amazon	2:1	43	MCTS (Basislinie)
	Gottes Algorithmus	MCTS (Basislinie)	3:0	47	Gottes Algorithmus
22	MCTS (Basislinie) 23	Gottes Algorithmus	2:1	40	MCTS (Basislinie)
	Reisende Amazonen	MCTS (Basislinie)	2:1	47	Reisende Amazonen
24	MCTS (Basislinie)	Reisende Amazonen	3:0	42	MCTS (Basislinie)
25	Amazon Supreme Schach	MCTS (Basislinie)	0:3	40	MCTS (Basislinie)
26	MCTS (Basislinie)	Amazon Supreme Chess	3:0	41	MCTS (Basislinie)
27	Schachtruppen fallen aus Himmel	MCTS (Basislinie)	0:3	43	MCTS (Basislinie)
28	MCTS (Basislinie)	Schachtruppen fallen aus Himmel	3:0	39	MCTS (Basislinie)
29	Nelkenamazonen	MCTS (Basislinie)	0:3	44	MCTS (Basislinie)
30	MCTS (Basislinie) 31	Nelkenamazonen	3:0	40	MCTS (Basislinie)
	Shao Guangs Amazonen	MCTS (Basislinie)	1:2	43	MCTS (Basislinie)
32	MCTS (Basislinie)	Shao Guangs Amazonen	3:0	40	MCTS (Basislinie)
33	KI spielt Schach	MCTS (Basislinie)	2:1	47	KI spielt Schach
34	MCTS (Basislinie) 35	KI spielt Schach	3:0	42	MCTS (Basislinie)
	Canopus One	MCTS (Basislinie)	0:3	43	MCTS (Basislinie)
36	MCTS (Basislinie)	Canopus One	2:1	40	MCTS (Basislinie)
37	Gewinnen Sie das Eröffnungsteam	MCTS (Basislinie)	3:0	47	Gewinnen Sie das Eröffnungsteam
38	MCTS (Basislinie)	Gewinnen Sie das Eröffnungsteam	3:0	44	MCTS (Basislinie)
39	Wukong-Amazonen	MCTS (Basislinie)	0:3	42	MCTS (Basislinie)
40	MCTS (Basislinie) 41	Wukong-Amazonen	3:0	39	MCTS (Basislinie)
	Reisende Amazonen	MCTS (Basislinie)	0:3	41	MCTS (Basislinie)
42	MCTS (Basislinie) 43	Reisende Amazonen	3:0	36	MCTS (Basislinie)
	Yue	MCTS (Basislinie)	0:3	44	MCTS (Basislinie)
44	MCTS (Basislinie) 45	Yue	3:0	41	MCTS (Basislinie)
	Final Amazon	MCTS (Basislinie)	0:3	42	MCTS (Basislinie)
46	MCTS (Basislinie)	Final Amazon	3:0	37	MCTS (Basislinie)
47	Informationen zur Universität - Monte Carlo	MCTS (Basislinie)	3:0	47	Information Universität-Monte-Carlo
48	MCTS (Basislinie)	Informationen zur Universität Monte Carlo	0:3	47	Information Universität-Monte-Carlo

Machine Translated by Google

Tabelle A3. Forts.

ID	Erster Spieler	Zweiter Spieler	Ergebnis	Übereinstimmen Punkt Runden	Gewinner
49	Haha Hallo	MCTS (Basislinie)	3:0	47	Haha, hallo
50	MCTS (Basislinie) 51	Haha, hallo	3:0	41	MCTS (Basislinie)
	Drachensieg	MCTS (Basislinie)	0:3	43	MCTS (Basislinie)
52	MCTS (Basislinie) 53	Drachensieg	3:0	38	MCTS (Basislinie)
	Super Amazon	MCTS (Basislinie)	0:3	41	MCTS (Basislinie)
54	MCTS (Basislinie) 55	Super Amazon	3:0	38	MCTS (Basislinie)
	Basispaarungsteam	MCTS (Basislinie)	3:0	45	Basispaarungsteam
56	MCTS (Basislinie) 57	Basispaarungsteam	0:3	42	Basispaarungsteam
	Bestehen Sie das Level	MCTS (Basislinie)	3:0	47	Bestehen Sie das Level
58	MCTS (Basislinie)	Bestehen Sie das Level	0:3	47	Bestehen Sie das Level
59	Jiaotong-Universität Dalian Amazonen Team 1	MCTS (Basislinie)	0:3	44	MCTS (Basislinie)
60	MCTS (Basislinie)	Jiaotong-Universität Dalian Amazonen Team 1	3:0	42	MCTS (Basislinie)
61	Gehen Sie an Land	MCTS (Basislinie)	3:0	47	Gehen Sie an Land
62	MCTS (Basislinie) 63	Gehen Sie an Land	0:3	47	Gehen Sie an Land
	Leer	MCTS (Basislinie)	0:3	45	MCTS (Basislinie)
64	MCTS (Basislinie)	Leer	3:0	41	MCTS (Basislinie)
65	Stier	MCTS (Basislinie)	0:3	45	MCTS (Basislinie)
66	MCTS (Basislinie)	Stier	3:0	40	MCTS (Basislinie)
67	Wukong-Amazonen	MCTS (Basislinie)	2:1	47	Wukong-Amazonen
68	MCTS (Basislinie)	Wukong-Amazonen	3:0	39	MCTS (Basislinie)
69	Gangzi Fans Support Team	MCTS (Basislinie)	3:0	44	Gangzi-Fans Support-Team
70	MCTS (Basislinie)	Gangzi Fans Support Team	1:2	41	Gangzi-Fans Support-Team
71	Thai Pants Scharf	MCTS (Basislinie)	0:3	42	MCTS (Basislinie)
72	MCTS (Basislinie) 73	Thai Pants Scharf	3:0	40	MCTS (Basislinie)
	Grüner Graskuchen	MCTS (Basislinie)	0:3	42	MCTS (Basislinie)
74	MCTS (Basislinie)	Grüner Graskuchen	3:0	37	MCTS (Basislinie)
75	Konnte die Luft nicht greifen und Schnapp dir die Pflanzenhirn-Hypoxie Team	MCTS (Basislinie)	0:3	41	MCTS (Basislinie)
76	MCTS (Basislinie)	Konnte die Luft nicht greifen und Schnapp dir die Pflanzenhirn-Hypoxie Team	3:0	40	MCTS (Basislinie)
77	DG	MCTS (Basislinie)	0:3	43	MCTS (Basislinie)
78	MCTS (Basislinie)	DG	3:0	39	MCTS (Basislinie)
79	Warum ist Tang Yang ein Gott	MCTS (Basislinie)	2:1	47	Warum ist Tang Yang ein Gott
80	MCTS (Basislinie) 81	Warum ist Tang Yang ein Gott?	3:0	44	MCTS (Basislinie)
	Dream Team	MCTS (Basislinie)	0:3	43	MCTS (Basislinie)
82	MCTS (Basislinie)	Dream Team	3:0	44	MCTS (Basislinie)
83	Pferdegesichtsrock Täglich und Einfach zu tragen	MCTS (Basislinie)	0:3	44	MCTS (Basislinie)
84	MCTS (Basislinie)	Pferdegesichtsrock Täglich und Einfach zu tragen	3:0	40	MCTS (Basislinie)
85	Genshin Impact-Experte	MCTS (Basislinie)	0:3	44	MCTS (Basislinie)
86	MCTS (Basislinie)	Genshin Impact-Experte	3:0	39	MCTS (Basislinie)
87	Code-Lehrling	MCTS (Basislinie)	0:3	42	MCTS (Basislinie)
88	MCTS (Basislinie)	Code-Lehrling	3:0	40	MCTS (Basislinie)
89	Amazon Drift Notizen	MCTS (Basislinie)	0:3	44	MCTS (Basislinie)
90	MCTS (Basislinie) 91	Amazon Drift-Notizen	3:0	42	MCTS (Basislinie)
	Die Liebe wird verschwinden, oder?	MCTS (Basislinie)	0:3	40	MCTS (Basislinie)
92	MCTS (Basislinie)	Die Liebe wird verschwinden, oder?	3:0	38	MCTS (Basislinie)

Machine Translated by Google

Tabelle A3. Forts.

ID	Erster Spieler	Zweiter Spieler	Ergebnis	Übereinstimmen Punkt Runden	Gewinner
93	Die kleine Su und die Katze	MCTS (Basislinie)	0:3	44	MCTS (Basislinie)
94	MCTS (Basislinie)	Die kleine Su und die Katze	3:0	40	MCTS (Basislinie)
95	Will Marla nicht	MCTS (Basislinie)	0:3	41	MCTS (Basislinie)
96	MCTS (Basislinie)	Will Marla nicht	3:0	37	MCTS (Basislinie)
97	Parameter-Anpassungsteam	MCTS (Basislinie)	0:3	42	MCTS (Basislinie)
98	MCTS (Basislinie)	Parameter-Anpassungsteam	3:0	39	MCTS (Basislinie)
99	AlphaAmazon	MCTS (Basislinie)	0:3	41	MCTS (Basislinie)
100	MCTS (Basislinie)	AlphaAmazon	0:3	47	AlphaAmazon

Verweise

1. **Li, R.; Gao, M. Amazons Suchalgorithmus-Design basierend auf einem CNN-Modell.** Digit. Technol. Appl. **2022,** 40, **164–166.**
2. **Guo, Q.; Li, S.; Bao, H. Forschung zur Bewertungsfunktion von Computerspielen bei Amazon.** Comput. Eng. Appl. **2012,** 48, **50–54.**
3. **Guo, T.; Qiu, H.; Tong, B.; Wang, Y. Optimierung und Vergleich mehrerer Spielalgorithmen in Amazons. In Proceedings of die Chinese Control And Decision Conference (CCDC) 2019, Nanchang, China, 3.–5. Juni 2019; S. 6299–6304.**
4. **Quan, J.; Qiu, H.; Wang, Y.; Li, F.; Qiu, S. Anwendung von UCT-Technologien für Computerspiele von Amazon. In Proceedings of the 2016 Chinesische Kontroll- und Entscheidungskonferenz (CCDC), Yinchuan, China, 28.–30. Mai 2016; S. 6896–6899.**
5. **Ju, J.; Qiu, H.; Wang, F.; Wang, X.; Wang, Y. Forschung zur Thread-Optimierung und zum Öffnen von Bibliotheken basierend auf parallelem PVS Algorithmus in Amazonen. In Proceedings der 33. Chinese Control and Decision Conference (CCDC) 2021, Kunming, China, 28.–30. Mai 2021; S. 2207–2212.**
6. **Li, Z.; Ning, C.; Cao, J.; Li, Z. Amazonen basierend auf dem UCT-PVS-Hybridalgorithmus. In Proceedings der 33. chinesischen Kontrollkonferenz 2021 und Entscheidungskonferenz (CCDC), Kunming, China, 28.–30. Mai 2021; S. 2179–2183.**
7. **Ding, M.; Bo, J.; Qi, Y.; Fu, Y.; Li, S. Design von Amazons Spielsystem basierend auf Reinforcement Learning. In Proceedings of the 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 3.–5. Juni 2019; S. 6337–6342.**
8. **Tong, B.; Qiu, H.; Guo, T.; Wang, Y. Forschung und Anwendung der parallelen Berechnung des PVS-Algorithmus basierend auf Amazon Mensch-Maschine-Spiel. In Proceedings der 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 3–5 Juni 2019; S. 6293–6298.**
9. **Chen, X.; Yang, L. Forschung zur Bewertungsfunktion bei Amazons.** Comput. Knowl. Technol. **2019,** 15, **224–226.**
10. **Wang, C.; Ding, M. Interface-Design und Implementierung personalisierter Amazons.** Intell. Comput. Appl. **2017, 7, 78–80.**
11. **Zhang, L. Forschung zu Amazons Spielsystem basierend auf dem Minimax-Suchalgorithmus. Masterarbeit, Northeast University, Shenyang, China, 2010.**
12. **Metropolis, N.; Ulam, S. Die Monte-Carlo-Methode.** J. Am. Stat. Assoc. **1949,** 44, **335–341.** [CrossRef] [PubMed]
13. **Coulom, R. Effiziente Selektivität und Backup-Operatoren in der Monte-Carlo-Baumsuche. In Proceedings of the Computers and Games, Turin, Italien, 1.–5. April 2007; S. 72–83.**
14. **Gelly, S.; Kocsis, L.; Schönauer, M.; Sebag, M.; Silber, D.; Szepesvári, C.; Teytaud, O. Die große Herausforderung des Computers Go: Monte Carlo-Baumsuche und Erweiterungen.** Commun. ACM **2012,** 55, **106–113.** [CrossRef]
15. **Gelly, S.; Silver, D. Monte-Carlo-Baumsuche und schnelle Aktionswertschätzung im Computer-Go.** Artif. Intell. **2011,** 175, **1856–1875.** [Querverweis]
16. **Vinyals, O.; Babuschkin, I.; Czarnecki, WM; Mathieu, M.; Dudzik, A.; Chung, J.; Choi, DH; Powell, R.; Ewalds, T.; Georgiev, P.; et al. Großmeister-Level in StarCraft II mithilfe von Multi-Agent-Reinforcement-Learning.** Nature **2019,** 575, **350–354.** [CrossRef]
17. **Silver, D.; Huang, A.; Maddison, CJ; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Das Spiel Go mit tiefen neuronalen Netzwerken und Baumsuche meistern.** Nature **2016,** 529, **484–489.** [CrossRef]
18. **Berner, C.; Brockman, G.; Chan, B.; Cheung, V.; D ŷebiak, P.; Dennison, C.; Farhi, D.; Fischer, Q.; Hashme, S.; Hesse, C.; et al. Dota 2 mit Large Scale Deep Reinforcement Learning.** arXiv **2019, arXiv:1912.06680.**
19. **Browne, CB; Powley, E.; Whitehouse, D.; Lucas, SM; Cowling, PI; Rohlfshagen, P.; Tavener, S.; Perez, D.; Samothrakis, S.; Colton, S. Eine Übersicht über Monte-Carlo-Baumsuchmethoden.** IEEE Trans. Comput. Intell. AI Games **2012, 4, 1–43.** [CrossRef]
20. **Kloetzer, J.; Iida, H.; Bouzy, B. Der Monte-Carlo-Ansatz bei Amazonen. In Proceedings of the Computer Games Workshop, Amsterdam, Niederlande, 15.–17. Juni 2007; S. 185–192.**
21. **Shannon, CE Programmierung eines Computers zum Schachspielen.** Lond. Edinb. Dublin Philos. Mag. J. Sci. **1950,** 41, **256–275.** [CrossRef]
22. **Chaslot, G.; Bakkes, S.; Szita, I.; Spronck, P. Monte-Carlo Tree Search: Ein neues Framework für Game AI. In Proceedings of the AAAI-Konferenz zu künstlicher Intelligenz und interaktiver digitaler Unterhaltung, Kalifornien, CA, USA, 22.–24. Oktober 2008; S. 216–217.**
23. **Swiechowski, M.; Godlewski, K.; Sawicki, B.; Ma ´ndziuk, J. Monte Carlo Tree Search: Ein Überblick über aktuelle Modifikationen und Anwendungen.** Artif. Intell. Rev. **2023,** 56, **2497–2562.** [CrossRef]

Machine Translated by Google

24. Kloetzer, J. Monte-Carlo Opening Books for Amazons. In Computers and Games: 7. Internationale Konferenz (CG 2010), Kanazawa, Japan, 24.–26. September 2010, Revised Selected Papers **7; Springer: Berlin/Heidelberg, Deutschland, 2011; S. 124–135.**

25. Song, J.; Müller, M. Ein verbesserter Solver für das Spiel der Amazonen. IEEE Trans. Comput. Intell. AI Games **2014, 7, 16–27. [CrossRef]**

26. Zhang, G.; Chen, X.; Chang, R.; Zhang, Y.; Wang, C.; Bai, L.; Wang, J.; Xu, C. Das Spiel der Amazonen schnell meistern durch Entkopplung des Netzwerklernens. In Proceedings der 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18.–22. Juli 2021; S. 1–8.

27. de Koning, J. INVADER verlängert Amazons Titel. ICGA J. **2011,** 34, **96. [CrossRef]**

28. Kloetzer, J. INVADER gewinnt Amazons Turnier. ICGA J. **2009,** 32, **112–113. [CrossRef]**

29. Lorentz, R. Invader gewinnt die achte Amazonen-Goldmedaille. ICGA J. **2017,** 39, **228–229. [CrossRef]**

30. Guo, Q.; Li, S.; Bao, H. Die Erforschung des Suchalgorithmus in Amazons Spiel. In Proceedings der 2011 Chinese Control and Decision Conference (CCDC), Mianyang, China, 23.–25. Mai 2011; S. 1859–1862.

31. Li, X.; Hou, L.; Wu, L. UCT-Algorithmus in Amazons Mensch-Computer-Spielen. In Proceedings der 26. chinesischen Kontroll- und Decision Conference (CCDC), Changsha, China, 31. Mai–2. Juni 2014; S. 3358–3361.

32. Quan, J.; Qiu, H.; Wang, Y.; Li, Y.; Wang, X. Untersuchung der Leistung von Suchalgorithmen in Amazons. In Proceedings des 27. Chinesische Kontroll- und Entscheidungskonferenz (CCDC), Qingdao, China, 23.–25. Mai 2015; S. 5811–5813.

33. Lieberum, J. Eine Bewertungsfunktion für das Spiel der Amazonen. Theor. Comput. Sci. **2005,** 349, **230–244. [CrossRef]**

34. Chai, Z.; Fang, Z.; Zhu, J. Amazons Evaluierungsoptimierungsstrategie basierend auf dem PSO-Algorithmus. In Proceedings of the 2019 Chinesische Kontroll- und Entscheidungskonferenz (CCDC), Nanchang, China, 3.–5. Juni 2019; S. 6334–6336.

35. Sun, Y.; Yuan, D.; Gao, M.; Zhu, P. GPU-Beschleunigung des Monte-Carlo-Baumsuchalgorithmus für Amazons und seine Bewertungsfunktion. In Proceedings der 2022 International Conference on Artificial Intelligence, Information Processing and Cloud Computing (AIIPCC), Kunming, China, 21.–23. Juni 2022; S. 434–440.

36. Kocsis, L.; Szepesvari, C.; Willemson, J. Verbesserte Monte-Carlo-Suche. Univ. Tartu Est. Techn. Rep. **2006, 1, 1–22.**

37. Finnsson, H.; Björnsson, Y. Wildbaumeigenschaften und MCTS-Leistung. IJCAI **2011,** 11, **23–30.**

38. Phatt, C. Intel® Threading-Bausteine. J. Comput. Sci. Coll. **2008,** 23, **164–166.**

39. Childs, BE; Brodeur, JH; Kocsis, L. Transpositionen und Bewegungsgruppen in der Monte-Carlo-Baumsuche. In Proceedings des 2008 IEEE Symposium on Computational Intelligence and Games, Perth, Australien, 15.–18. Dezember 2008; S. 389–395.

40. Van Eyck, G.; Müller, M. Revisiting move groups in monte-carlo tree search. In Advances in Computer Games: 13. Internationale Konferenz, ACG 2011, Tilburg, Niederlande, 20.–22. November 2011, Revised Selected Papers 13; **Springer: Berlin/Heidelberg, Deutschland, 2012; S. 13–23.**

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Artikel

Farbstandardisierung von Bildern chemischer Lösungen mittels Vorlagenbasiertes Histogramm-Matching in Deep Regression lernen

Patrycja Kwiek und Małgorzata Jakubowska *

Fakultät für Materialwissenschaften und Keramik, AGH-Universität Krakau, al. Mickiewicza 30, 30-059 Krakau, Polen; pakwiek@agh.edu.pl *
Korrespondenz: jakubows@agh.edu.pl

Zusammenfassung: Farbverzerrungen in einem Bild stellen eine Herausforderung für die Klassifizierung und Regression durch maschinelles Lernen dar, wenn die Eingabedaten aus Bildern bestehen. Daher wird ein neuer Algorithmus zur Farbstandardisierung von Fotos vorgeschlagen, der die Grundlage für ein Regressionsmodell mit tiefen neuronalen Netzwerken bildet . Dieser Ansatz verwendet eine selbst entworfene Farbvorlage, die auf der Grundlage einer ersten Reihe von Studien und digitaler Bildgebung entwickelt wurde. Unter Verwendung des ausgeglichenen Histogramms der R-, G- und B-Kanäle der digitalen Vorlage und ihres Fotos wurde eine Farbzuordnungsstrategie berechnet. Durch Anwendung dieses Ansatzes wurden die Histogramme angepasst und die Farben vor Der vorgeschlagene Algorithmus wurde für eine Bildserie entwickelt, bei der die gesamte Oberfläche ungefähr eine einheitliche Farbe aufwies und die Farbunterschiede zwischen den Fotos der einzelnen Objekte gering waren. Dieser optimierte Ansatz wurde im kolorimetrischen Bestimmungsverfahren von Vitamin C validiert. Der Datensatz für das tiefe neuronale Netzwerk in der Regressionsvariante wurde aus Fotos von Proben unter zwei verschiedenen Lichtbedingungen gebildet. Für den Vitamin-C-Konzentrationsbereich von 0 bis 87,72 µg·mLÿ1 lag der RMSE für den Testsatz zwischen 0,75 und 1,95 µg·mLÿ1 . Zur nicht standardisierten Variante **hingegen lag dieser Indikator bei 1,48 bis 2,29 µg·mLÿ1 . Die Übereinstimmung der vorhergesagten Konzentrationsergebnisse mit den tatsächlichen Daten, ausgedrückt als R2 , lag für jede der standardisierten Varianten zwischen 0,9956 und 0,9999. Mit diesem Ansatz können Lichtreflexionen auf glänzenden Oberflächen von Lösungen entfernt werden , was bei Flüssigkeitsproben ein häufiges Problem darstellt. Dieser Farbabgleichalgorithmus hat universellen Charakter und sein Anwendungsbereich ist nicht begrenzt.**

Schlüsselwörter: Farbstandardisierung; Farbvorlagenabgleich; RGB-Histogrammanpassung; Deep Learning; kolorimetrische Bestimmung von Vitamin C

check for updates

Zitat: Kwiek, P.; Jakubowska, M. Farbstandardisierung von Chemikalien Lösungsbilder mithilfe von vorlagenbasiertem Histogramm-Matching in Deep Learning Regression. Algorithms **2024,** 17, **335. https://doi.org/10.3390/a17080335**

Wissenschaftlicher Herausgeber: Marcin Iwanowski

Empfangen: 11. Juli 2024
Überarbeitet: 29. Juli 2024
Akzeptiert: 30. Juli 2024
Veröffentlicht: 1. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/ 4.0/).

1. Einleitung

Die Entwicklung von Messmethoden und Forschungsstrategien in verschiedenen Bereichen der Wissenschaft und Technologie ist eng mit der potenziellen Unterstützung durch Techniken des maschinellen Lernens (ML) und des tiefen Lernens (DL) verbunden. Wenn die Informationen des zu lösenden Problems durch Bilder, Zeichnungen, Skizzen oder Fotos übermittelt werden, ist die Verwendung von DL bei der Datenmodellierung oft der optimale Ansatz. Die Anwendung neuronaler Netzwerke im typischen überwachten Lernen erfordert den Zugriff auf große Datensätze, was die Methode zum Erhalten und Kennzeichnen von Daten zu einer erheblichen Einschränkung macht. Die heutige Technologie bietet viele elektronische Geräte wie Kameras oder Scanner, mit denen Daten zum Erstellen von Modellen schnell und ohne besondere Qualifikationen erhalten werden können . Informationen können oft einfach in Form von Objekten, der Beziehung zwischen Objekten im analysierten **Bereich, Texturen, der Menge kleiner Komponenten und Farben kodiert werden .**

In verschiedenen Branchen werden verschiedene Farbräume für unterschiedliche Anwendungen verwendet [1]. Der RGB-Raum, der auf den Primärfarben des Lichts (Rot, Grün und Blau) basiert, wird häufig für digitale Anzeigen und Bilder verwendet. CMYK (Cyan, Magenta, Gelb und Schwarz) wird im Farbdruck verwendet und basiert auf dem subtraktiven Farbmodell, um eine breite Pal

Machine Translated by Google

auf Papier. HSI/HSV (Farbton, Sättigung, Intensität/Wert) wird aufgrund seiner Übereinstimmung mit der menschlichen Farbwahrnehmung häufig in der Bildanalyse und Computervision verwendet. YUV und YCbCr werden bei der Videokomprimierung und -übertragung verwendet und trennen das Bild in Luminanz- und Chrominanzkomponenten, um die Datenspeicherung und -übertragung zu optimieren. Der CIE Lab-Farbraum , der von der Internationalen Beleuchtungskommission entwickelt wurde, bietet eine wahrnehmungsmäßig einheitliche Farbdarstellung und ist daher ideal für Farbdifferenzierungs- und Transformationsaufgaben . RAL CLASSIC ist ein in Europa weit verbreitetes Farbabstimmungssystem, das aus 213 standardisierten Farben besteht, während RAL DESIGN eine umfangreichere Palette von 1625 Farben für Designer und Architekten bietet. RAL DIGITAL erleichtert die präzise Farbkommunikation in digitalen Anwendungen durch RGB- und CMYK-Werte. Das PANTONE Matching System (PMS) ist weltweit in den Bereichen Grafikdesign, Druck und Mode anerkannt und bietet standardisierte Farbcodes, um Konsistenz über verschiedene Medien und Materialien hinweg zu gewährleisten. Jeder dieser Farbräume hat seine eigenen Vorteile und Anwendungen und erfüllt die spezifischen Anforderungen verschiedener Branchen.

Wenn die Farbe das wichtigste und entscheidende Kriterium bei der Bildanalyse ist, das die gesuchten Informationen kodiert, ist es entscheidend, Fotos so aufzunehmen, dass die entsprechende Qualität dieses Parameters erhalten bleibt. Die wahre Farbe des Bildes ermöglicht die richtige Interpretation verschiedener Merkmale der fotografierten Objekte. Unerwünschte Variabilität kann durch die Verwendung mehrerer Geräte zur Aufnahme von Bildern oder durch die Methode der Aufnahme von Bildern mit einem einzigen Gerät unter unkontrollierten oder sich ändernden Bedingungen entstehen. Zu den wesentlichen Faktoren gehören Änderungen der Lichtverhältnisse, Reflexionen auf glänzenden Oberflächen und Schwankungen wie die Tageszeit. Diese Elemente wirken sich möglicherweise nicht direkt auf die Qualität des Fotos aus, können jedoch seinen Informationswert beeinflussen. Daher sollte die Bedeutung der Farbstandardisierung als entscheidender Schritt bei der Datenvorverarbeitung berücksichtigt werden. In zahlreichen Arbeiten werden verschiedene rechnerische Ansätze zur Erhaltung oder Wiederherstellung der wahren Farbe von Bilde

Diese Studie präsentiert einen originellen Algorithmus zur Standardisierung der Farben von Fotos, die mit einer einzigen Smartphone-Kamera unter verschiedenen Lichtbedingungen aufgenommen wurden. Dabei werden auch die durch zufällige Faktoren im Zusammenhang mit dem Aufnahmeprozess bedingten Variabilitäten und die inhärente fehlende perfekte Wiederholbarkeit der Kamera berücksichtigt. Die Forschung konzentrierte sich auf eine homogene Flüssigkeit, die unter den getesteten Objekten nur minimale Farbabweichungen aufwies . Das gesamte im Modellierungsprozess verwendete Foto wies dieselbe Farbe auf, allerdings mit einem gewissen Grad an zufälliger Variabilität, die nicht vollständig eliminiert werden konnte. In diesem Zusammenhang hatten selbst geringfügige Farbänderungen einen erheblichen Informationswert, was die Bedeutung einer genauen Farbinterpretation für die Verbesserung der Qualität des Deep- Learning-Regressionsmodells un

Es wurde ein Standardisierungsalgorithmus vorgeschlagen, der das Histogramm einer speziell entwickelten Vorlage verwendete, die unter kontrollierten Bedingungen zusammen mit dem getesteten Objekt fotografiert wurde, in Bezug auf das Histogramm der Vorlage, die mit einer Grafikanwendung erstellt wurde, und ihres Fotos. Diese Studie präsentiert auch die Überprüfung des vorgeschlagenen Ansatzes für die kolorimetrische Bestimmung von Vitamin C unter Verwendung eines von DNN (Deep **Neural Networks) definierten Regressionsmodells .**

2. Verwandte Werke

2.1. Farbstandardisierung

Es gibt zahlreiche Kategorien von Farbkalibrierungsproblemen, wobei zwei grundlegende Gruppen entweder mit der Verwendung unterschiedlicher Geräte innerhalb einer Fotoserie oder mit Variationen der Lichtverhältnisse zusammenhängen, die sich auf die fotografierten Motive auswirken. Eine Hardware-Farbkompensation kann durch den Einsatz von Kamerasensoren erreicht werden, die die Einstellungen automatisch an die Beleuchtung der aufgenommenen Szene anpassen. Der alternative Ansatz umfasst die numerische Verarbeitung von Fotos mithilfe verschiedener Strategien, die letztendlich darauf abzielen , **konsistente Farbtöne über mehrere Bilder hinweg zu erzielen .** Zu diesem Zweck werden **zunächst Kalibrierungstabellen verwendet, die die einzelnen Pixel der aufgenommenen Fotos mit den standardisierten Werten der entsprechenden Farbfelder auf der Tabelle** korrelieren und abbilden **[2**

Machine Translated by Google

Kalibrierungstabellen dienen als Referenzwerkzeug in der Fotografie, bei der Filmerstellung, bei Videoaufnahmen oder im Grafikdesign und sorgen für eine genaue und konsistente Farbwiedergabe in Bildern. Normalerweise bestehen sie aus einem Raster aus Farbfeldern mit bekannten und standardisierten Farben. Diese Tabellen werden bei der Kalibrierung von Geräten und bei der Nachbearbeitung angewendet und sorgen so für eine konsistente Farbdarstellung auf verschiedenen Geräten und unter verschiedenen Bedingungen. Im Handel sind verschiedene Kalibrierungstabellen erhältlich Die 24-Farben-Kalibrierungskarte, die oft unter ihrem ursprünglichen Namen Macbeth ColorChecker bekannt ist, hat jedoch in zahlreichen Anwendungen weite Verbreitung gefunden [4]. Darüber hinaus spielt die Entwicklung kundenspezifischer Kalibrierungskarten eine wichtige Rolle bei der Standardisierung von Bildfarben.

Farbstandardisierung spielt in vielen Bereichen von Wissenschaft und Technik eine wichtige Rolle, darunter in der medizinischen Diagnostik, Bioanalyse, Landwirtschaft, Beweisanalyse in der Forensik, analytischen Chemie usw. Bei der landwirtschaftlichen Bildverarbeitung liegt der Schwerpunkt darauf, Veränderungen der Farbe von Bildern in bestimmten Interessensbereichen mit bestimmten Qualitätsmerkmalen wie Pflanzengesundheit, Erntestress und Reifegrad zu verknüpfen. In [5] wurde eine Technik vorgeschlagen, um digitale Bilder zu kalibrieren, ihre Einheitlichkeit sicherzustellen und phänologische Vergleiche in der Landwirtschaft zu verbessern. Die Farbtransformation der Bilder wurde mithilfe einer standardmäßigen 24-Farben-Kalibrierungskarte erreicht. Die Kalibrierung liefert typische und erwartete Ergebnisse, indem die RGB-Werte der Farbfelder im Eingabebild an standardmäßige RGB-Werte angepasst werden. In [6] wurde ein digitales Farbabstimmungsgerät zur Bestimmung der Zahnfarbe entwickelt, da eine genaue Farbbestimmung bei Zahnrestaurationen von entscheidender Bedeutung ist. Bildaufzeichnungen, die mit einer Intraoralkamera gemacht wurden, dienten als Grundlage für die Berechnungen un

Zur Bewertung der Zahnfarbe wurden der Support Vector Machine (SVM)-Algorithmus und der CIEL*a*b*-Farbraum verwendet . Die Farbkorrektur basierte auf der Berechnung der euklidischen

Abstand zwischen den Farbtönen. Die Autoren erstellten eine spezielle Vorlage, die sich von der üblichen durch ein breites Farbspektrum unterscheidet und speziell an die in der Studie untersuchte Bandbreite an Zahnfarbvariationen angepasst ist. Der nächste Vorschlag betrifft die Verbesserung der Genauigkeit der Farbwiedergabe bei Digitalkameras und Smartphones. Die erste Methode verwendet eine Weißabgleichkorrektur, gefolgt von einer Farbraumtransformation. Der zweite Ansatz, der auf offline verarbeitete Bilder anwendbar ist, basiert auf der Durchführung eines vollständigen Farbabgleichs anstelle eines Weißabgleichs und erfordert den Einsatz von maschinellem Lernen zur Vorhersage der Farbabgleichmatrix [7]. Der Übersichtsartikel [8] erörtert die Verwendung verschiedener Algorithmen zur histopathologischen Bilderkennung. Die Farbunterschiede in der Histopathologie resultieren aus der Verwendung verschiedener Geräte oder Gewebefärbetechniken . Typischerweise umfasst die Farbnormalisierung das Übertragen der Durchschnittsfa Der Artikel beschreibt eine umfassende Reihe von Methoden zur Farbstandardisierung, die anhand

histopathologischer Bilder, einschließlich solcher, die bei der Diagnose von Brust- und Dickdarmkrebs gewonnen wurden, überprüft wurden.

Aufgrund der Entwicklung digitaler Bildtechnologie und der daraus resultierenden Entstehung vieler völlig neuer Anwendungen werden in verschiedenen Bereichen Lösungen auf Basis von Smartphone-Bildern vorgeschlagen. Tabelle 1 enthält eine Liste ausgewählter Artikel, in denen Farbbilder als wichtige Informationsträger dienen und Farbkorrektur und -standardisierung die Modellierungsergebnisse und die Qualität der vorgeschla

Tabelle 1. Vergleich verwandter Arbeiten.

Art.-Nr.	Jahr	Problem gelöst	Verwendeter Datensatz	Farbkorrekturstrategie
[9]	2022	Das allgemeine Konzept der relativen Farbkonstanz (RCC), d. h. die Fähigkeit, Farben derselben Objekte zwischen Bildern unabhängig von Beleuchtung und Kamera anzugleichen	Drei Bilddatensätze, die von mehreren Kameras unter verschiedenen Beleuchtungs- und Belichtungsbedingungen (Belfast, Middlebury, Gehler-Shi)	Drei Schritte des Farbabstimmungsrahmens: Kalibrierung der Kamerareaktion mit Farbfeldern, Reaktionslinearisierung, pixelweise Farbintensität und Farbwertanpassung

Machine Translated by Google

Tabelle 1. Fortsetzung

Art.-Nr.	Jahr	Problem gelöst	Verwendeter Datensatz	Farbkorrekturstrategie
[10]	2022	Farbabweichungen innerhalb klinischer Bilder von Kiefer- und Gesichtsprothesen-Silikonproben (In-vitro-Verfahren)	Bilder der pigmentierten Prothesensilikonproben in verschiedenen Umgebungen Lichtverhältnisse	Computergestützte, softwarebasierte Weißabgleichkalibrierung nach der Verarbeitung (PPWBC) mithilfe einer Graukarte und einer Macbeth-Farbkarte
[11]	2023	Nichtinvasive Erhaltung des kulturellen Erbes im Kontext der Materialdegradation	Vier Bildsätze unter unterschiedlichen Lichtbedingungen, aufgenommen mit einer herkömmlichen Handykamera im sRGB-Farbraum	Eine adaptive Übertragungsfunktion, die mit einer selbst erstellten Farbkalibrierungstabelle arbeitet; neuronales Netzwerk-basierte Endabstimmung der Kalibrierungsfunktionen
[12]	2023	Smartphone-basierte Urin-Kolorimetrie	Urinproben von 58 Patienten, fotografiert in einer maßgeschneiderten Fotobox unter fünf simulierten Lichtbedingungen mit fünf Smartphones im RGB-Farbraum	Farbkalibrierung mit SpyderCHEKRTM Farbkarte und die proprietären Algorithmen von Adobe Photoshop zur Skalierung der unkorrigierten RGB-Werte
[13]	2023	Eine Smartphone-basierte Lösung für genaue kolorimetrische Messungen mit einem Augmented-Reality-Leitsystem	Simulierte und reale Bilder, Bilder von Farbstreifen-Kits zur pH-Messung	Ein Farbkorrekturalgorithmus mit dem räumlich variierenden Regressionsmodell erster Ordnung, das sowohl die absolute Farbgröße als auch die Skala nutzt; das System umfasst eine Farbreferenztafel
[14]	2023	Beurteilung des Heilungsprozesses chronischer Wunden	Wundgewebebilder, aufgenommen zu verschiedenen Zeitpunkten, unter verschiedenen Lichtbedingungen, aus unterschiedlichen Entfernungen, mit unterschiedlichen Objektiven und Smartphone-Kameras im RGB-Farbraum	Bilder von Wundgewebe aufgenommen mit einem Macbeth-Diagramm mit 24 quadratischen Farbfeldern und vier ArUco-Markierungen an den Ecken des Diagramms; Berechnung der Transformationsmatrix mit dem Moore-Penrose-Inverse Matrix und die Zielmatrix
[15]	2024	Ein Allzweckalgorithmus	Fotos von Natur- und Stadtlandschaften, Vegetation und von Menschenhand geschaffenen Objekten im RGB-Farbraum	Eine matrixbasierte Transformation Verwendung von Tensorprodukt-B-Splines mit Optimierung des Glättungsparameters ÿ und der Anzahl der Spline-Basisfunktionen
[16]	2024	Reproduktion der Referenzfarbe von Textilbekleidung durch Auswahl geeigneter Dispersionsfarbstoffe	Ergebnisse kolorimetrischer Messungen mit einem Spektralphotometer Typ Spectraflash 600 Plus (Datacolor International, USA)	Ein genetischer Algorithmus zur Optimierung der Auswahl und Konzentration von Farbstoffen für eine exakte Farbabstimmung in Bikomponentenfilamenten (PET/PT
[17]	2024	Optimierung des Aussehens von pflanzlichen Lebensmitteln unter Berücksichtigung der Tatsache, dass tierische Produkte eine Vielzahl unterschiedlicher Pigmente und Strukturelemente enthalten können, die Licht streuen	L*-, a*-, b*-Werte von tierischen Produkten und pflanzlichen Emulsionen, gemessen mit einem Kolorimeter (ColorFlex EZ 45/0-LAV, Hunter Associates Laboratory Inc., Reston, VA, USA)	Das Farbabstimmungsmodell basiert auf Kubelka-Munk-Theorie, die zur Berechnung der spektralen Reflektivität der gemischten Emulsionen aus den spektrale Reflektivität der einzelnen farbhaltigen Emulsionen
[18]	2024	Farbdesign für Mode – Farbauswahl und Abstimmung verschiedener Farbkombinationen	Grafische Objekte entworfen in CorelDraw	Farbabgleich basierend auf einem interaktiven genetischen Algorithmus unter Verwendung von Referenzbildern

Machine Translated by Google

2.2. Vitamin C-Bestimmung

Farbstandardisierungstechniken sind in der analytischen Chemie von entscheidender Bedeutung, insbesondere bei der kolorimetrischen Bestimmung verschiedener Substanzen. Dies wird insbesondere bei Methoden deutlich, bei denen Lösungen getesteter Verbindungen fotografiert werden und aus diesen Fotos mithilfe verschiedener statistischer Tools und multivariater Analysen Konzentrationsinformationen abgeleitet werden. Ein Beispiel für eine solche Methodik ist die kolorimetrische

Ascorbinsäure (AA), also Vitamin C, ist eine natürliche, gut wasserlösliche Verbindung, die eine wichtige Rolle für die menschliche Gesundheit spielt und für eine Vielzahl von Stoffwechselreaktionen und biologischen Prozessen unverzichtbar ist. Sie verfügt über eine starke antioxidative Wirkung und begrenzt Schäden Eine unzureichende Vitamin-C-Konzentration im Körper führt unter anderem zu Skorbut und Anämie und wird mit einigen psychischen Störungen, beispielsweise Depressionen, in Verbindung gebracht. Eine übermäßige Einnahme von Vitamin C kann negative Folgen haben, beispielsweise die Aufnahme von Vitamin B12 behindern und möglicherweise Magen-Darm-Probleme oder Nierensteine verursa Da der Mensch nicht in der Lage ist, Vitamin C selbst zu produzieren, ist die ergänzende Einnahme einer kontrollierten Menge an AA, üblicherweise aus frischem Obst und Gemüse, wichtig **[21,22].**

Die analytische Chemie bietet viele Strategien zur Bestimmung von AA in Lebensmitteln und Vitamin-C-Ergänzungsmitteln. Traditionell werden instrumentelle Analysetechniken zur Messung des Ascorbinsäurespiegels eingesetzt, die einen breiten Konzentrationsbereich mit niedriger Nachweisgrenze bieten . Zu diesen Methoden gehören direkte Titration [23], Ultraperformance-Flüssigkeitschromatographie (UPLC), Hochleistungsflüssigkeitschromatographie (HPLC) [24,25], Fourier-Transform- Infrarotspektrometrie (FT-IR) [26], Chemilumineszenz [27], Raman-Spektroskopie [28], Nahinfrarot (NIR), FT-Raman-Spektroskopie [29] und Kapillarelektrophorese [30]. Darüber hinaus werden verschiedene elektrochemische Methoden zur Bestimmung von AA eingesetzt, die eine extrem hohe Empfindlichkeit, Selektivität, Reproduzierbarkeit und Zuverlässigkeit bieten **[31–33]. Die genannten Methoden der chemischen Instrumentalanalyse erfordern jedoch** spezialisierte **Laborräume, teure Instrumente, viele Hilfsreagenzien, zeitaufwändige Extraktionsschritte und komplizierte Messverfahren, die hochqualifiziertes Personal für Aufgaben wie die Vorbereitung von Reagenzien und Proben, die Durchführung von Untersuchungen oder die Interpretation von Ergebnissen erfo**

Die Entwicklung neuartiger Methoden, die traditionelle Laborstrategien durch kostengünstige Instrumente entweder ersetzen oder unterstützen, erfüllt den Bedarf an hocheffizienten und anpassungsfähigen Ansätzen in der Lebensmittel-, Arzneimittel- und Umweltanalytik. Beliebte und weit verbreitete Smartphones bieten eine vielversprechende Analyseplattform, die sich durch Vollautomatisierung, große lokale Speicherkapazität, Zugriff auf externen Cloud-Speicher und insbesondere hochwertige Kameras auszeichnet und sich daher für Im Vergleich zu herkömmlichen Methoden nutzen Smartphone-basierte Messungen digitale Bildgebung und bieten viele Anwendungsmöglichkeiten in verschiedenen Bereichen und erhebliche Vorteile. Dazu gehören niedrige Kosten, einfache Bedienung, die Möglichkeit, verschiedene spezielle Sensoren zu installieren, präzise Messungen in Echtzeit, eine in vielen Fällen zerstörungsfreie Art der Prüfung, schnelle Analyse und Portabilität mit anderen ähnlichen Geräten. In der analytischen Chemie werden allgemein erhältliche Geräte wie Kameras, Webcams, Scanner und Smartphones verwendet, um Bilder von kolorimetrischen Reaktionen aufzunehmen, bei denen die Intensität der resultierenden Farbe direkt mit der Konzentration des Analyten korreliert . In der Literatur finden sich verschiedene Beschreibungen der Verwendung digitaler Bildgebungsmethoden zur Bestimmung von AA. Die Arbeit [34] stellt eine unkomplizierte, schnelle und kostengünstige Technik vor, bei der Smartphone-Bildanalyse zur Quantifizierung von AA in einer wässrigen Lösung eingesetzt wird. Die Studie basierte auf der Bewertung der störenden Wirkung von AA bei der enzymatischen kolorimetrischen Erkennung von Glucose. Ein anderer Ansatz wurde vorgeschlagen, bei dem digitale Bildkolorimetrie zur Bestimmung des Vitamin-C-Gehalts in natürlichen Fruchtsäften eingesetzt wird . Bei dieser Methode wird die Farbe des Eisen(II)-o-Phenanthrolin-Komplexes gemessen [35]. Im nächsten Artikel entwickelten die Autoren ein neues Zirkulardichroismus-Spektrometer auf Smartphone-Basis für eine hochempfindliche und ultraportable kolorimetrische Analyse, das die Vorteile von Kosteneffizienz und Einfachheit bietet. Dieses Spektrometer verwendete ein HSV-Farbmodell und bestimmte die Absorption durch Summierung der V-Werte benachbarter Positionen, um die Gesamtintensität zu berechnen. Die TMB-MnO2- Nanosheet-Reaktion wurde ebenfalls verwendet und tatsächlich wurde ein hochempfindliches und spezifisches System zur Erkennung von Vitamin

Machine Translated by Google

C [36]. In [37] wurde ein Point-of-Care-Sensor entwickelt, der aus einem fluoreszierenden Papierchip, 3D-gedruckten Teilen und einem Smartphone besteht. Dieses System ermöglichte eine quantitative und hochempfindliche visuelle Echtzeiterkennung von AA. Das Design dieses Geräts beinhaltete die Platzierung eines Fluoreszenzsensors auf einer tragbaren Sensorplattform, die mit einem Smartphone verbunden

Die in dieser Arbeit verwendete analytische Messmethode basierte auf den reduzierenden Eigenschaften von Vitamin C. AA reduziert Fe(III) zu Fe(II) und bildet dann einen Komplex mit o-Phenanthrolin. Der stark rote Komplex aus Fe(II) und o-Phenanthrolin wird häufig zur kolorimetrischen Bestimmung und als Indikator verwendet [38]. Diese Reaktion zeichnet sich durch hohe Geschwindigkeit, Stabilität und Reproduzierbarkeit aus. Die Verwendung dieses kamerabasierten Ansatzes in Verbindung mit der digitalen Bildgebung farbiger Verbindungen erhöht die Genauigkeit und reduziert die Kosten der Messung des Vitamin-C-Gehalts.

2.3. Multivariate Regressionsalgorithmen

Regressionstechniken werden häufig als Methode des überwachten Lernens eingesetzt. Dabei wird die Beziehung zwischen Merkmalen und der Zielvariablen modelliert, um Aufgaben zu bewältigen, bei denen kontinuierliche Werte vorhergesagt werden sollen. Typische Algorithmen, die zu diesem Zweck verwendet werden, sind PCR (Hauptkomponentenregression) [39], PLSR (Partial Least Squares Regression) [40,41] und SVR (Support Vector Regression) [42]. DNN mit verschiedenen Architekturen, die die nichtlineare Struktur multivariater Daten abbilden, können auch auf Regressionsprobleme angewendet werden [43,44]. Convolutional Neural Networks (CNNs) sind eine Art Deep-Learning-Algorithmus, dessen Funktionsweise auf einer mathematischen Operation namens Faltung beruht. Sie bestehen aus mehreren Schichten von Faltungsfiltern, die automatisch hierarchische Darstellungen von Merkmalen direkt aus den Rohpixeldaten lernen, was sie für Aufgaben wie Bildklassifizierung oder -segmentierung und Objekterkennung hocheffektiv macht. Normalerweise wird die Qualität von Regressionsmodellen anhand der mittleren Quadrate der Fehler zwischen tatsächlichen Werten und Vorhersagen (MSE) oder des mittleren a

MSE und MAE sind verschiedene Fehlermaße, die zur Bewertung der Leistung von Vorhersagemodellen verwendet werden. MSE reagiert empfindlicher auf große Fehler, da diese quadriert werden, während MAE ein intuitiveres Verständnis des durchschnittlichen Fehlers bietet und in der Einheit der Zielvariable ausgedrückt wird. Diese Werte, die als Verlustfunktionen während des Trainings neuronaler Netze verwendet werden, minimieren den mittleren Unterschied zwischen den vorhergesagten und den tatsächlichen Werten. Die Trainings- und Vorhersageleistung des Kalibrierungsmodells wird auch anhand der folgenden Werte bewertet: der quadratische Mittelwertfehler der Kalibrierung (RMSEC) und der Vorhersage (RMSEP) sowie der Determinationskoeffizient R2 für Kalibrierung und Vorhersage [45]. Da das Vorhersageergebnis voraussichtlich dem tatsächlichen Wert entspricht, sollten die Punkte mit den Koordinaten (tatsächlich, vorhergesagt) idealerweise auf der Linie y **=** x **liegen . Aufgrund der zufälligen Variabilität und der unvollkommenen Modellanpassung an die Daten sollten jedoch zwei Bedingungen erfüllt sein: Der Wert eins sollte innerhalb des Konfidenzintervalls der Steigung liegen und der Wert null sollte inn**

3. Vorgeschlagene **Methode**

3.1. Farbabgleichalgorithmus

Das Hauptziel dieser Arbeit bestand darin, einen Ansatz zur Bestimmung von Vitamin C zu entwickeln, der auf der Farbänderung der Lösung basiert, die durch ihre Konzentration beeinflusst wird. Ein entscheidender Schritt zur Erstellung eines hochwertigen Regressionsmodells war daher die Durchführung des Farbstandardisierungsverfahrens an den fotografierten Vitamin C-Proben. Der vorgeschlagene Farbstandardisierungsalgorithmus wurde in zwei Teile unterteilt. Im ersten Schritt wurde ein Standard-Histogrammabgleich zwischen einer digitalen Farbvorlage und einem Foto einer gedruckten Vorlage durchgeführt, das unter bestimmten Lichtbedingungen aufgenommen wurde. Im zweiten Schritt wurde ein Histogrammabgleich zwischen einem Foto der Lösung (getestete Probe), das unter derselben Beleuchtung wie das Muster aufgenommen wurde, und der Referenzvorlage durchgeführt. Am Ende wurden 512 kleinere zufällige Quadrate aus dem rekonstruierten Bild ausgeschnitten, um einen ausreichend großen Datensatz für das **anschließende Training des Regressionsneuronalen Netzwerks zu erhalten .** Die Anzahl der zufällig ausgewählten **kleinen Qua**

Machine Translated by Google

führte zu einer geringeren Modellqualität. Außerdem war ein zusätzlicher Schritt erforderlich, bei dem die selbst entworfene Referenzvorlage an den entsprechenden Wertebereich angepasst wurde. Dieser Vorgang führte zur Entwicklung eines angepassten Histogramm-Matching-Algorithmus.

Histogrammabgleich ist eine Operation, bei der die Pixel eines Eingabebildes so manipuliert werden, dass sein Histogramm mit dem Histogramm des Referenzbildes übereinstimmt. In dieser Arbeit wurde der RGB-Farbraum für die Berechnungen gewählt. Er ist weithin verfügbar und beliebt, da die meisten Kameras und Bilderfassungsgeräte Daten direkt im RGB-Format aufzeichnen, wodurch zusätzliche Farbraumtransformationen überflüssig werden. Der RGB-Raum ist außerdem einfach zu implementieren, da er intuitiv ist und direkt mit den physikalischen Eigenschaften des Lichts zusammenhängt, was die Entwicklung von Algorithmen vereinfacht. Darüber hinaus hat sich in unserer spezifischen Anwendung gezeigt, dass RGB zufriedenstellende Ergebnisse liefert. Daher war RGB trotz der Vorteile anderer Farbräume aus praktischen und technischen Gründen am besten geeignet. Im Fall von Farbbildern wurde jeder Kanal (RGB) unabhängig betrachtet. Im Kontext der Bildverarbeitung war dies e



Abbildung 1. Beispiel für Histogramm-Matching zu Illustrationszwecken.

Mathematisch ausgedrückt lässt sich die Histogrammanpassung als Bildtransformation darstellen, sodass die kumulative Verteilungsfunktion (CDF) der Werte in jedem Band der CDF der Bänder in einem anderen Bild entspricht.

Um die Histogramme zweier Bilder anzupassen, ist es notwendig, ihre Histogramme anzugleichen. Anschließend wird jeder Pixel des ersten Bildes geändert und dem zweiten Bild zugeordnet. Dieser Vorgang wurde für jeden der Kanäle (R, G, B) einer fotografierten Vorlage unter bestimmten Lichtbedingungen (natürliches Licht und künstliches Licht) und für die Kanäle einer digitalen Vorlage durchgeführt. Der gesamte Prozess der Bildfarbtransformation ist in Abbildung 2 dargestellt.

Um die Bilder der getesteten chemischen Lösungen zu standardisieren, wurde der richtige Bereich der RGB-Werte aus den Histogrammen des Fotos der getesteten Probe ausgewählt. Wenn der Histogrammwert für eine Probe größer oder gleich 1 % des Maximalwerts des Histogramms der chemischen Lösung war, wurde der Histogrammwert nicht geändert, andernfalls wurde er auf 0 gesetzt. Die Diskrepanz in der Anzahl der Pixel wurde durch ihre gleichmäßige zufällige **Verteilung über den Wertebereich neutralisiert .** Der letzte Schritt im Standardisierungsalgorithmus bestand darin , **die Histogramme eines Bildes einer Lösung mit den angepasst Dies geschah auf die gleiche Weise wie in Abbildung 3 beschrieben, gefolgt von der Rekonstruktion eines Bildes aus den R-, G- und B-Kanälen.**

Machine Translated by Google

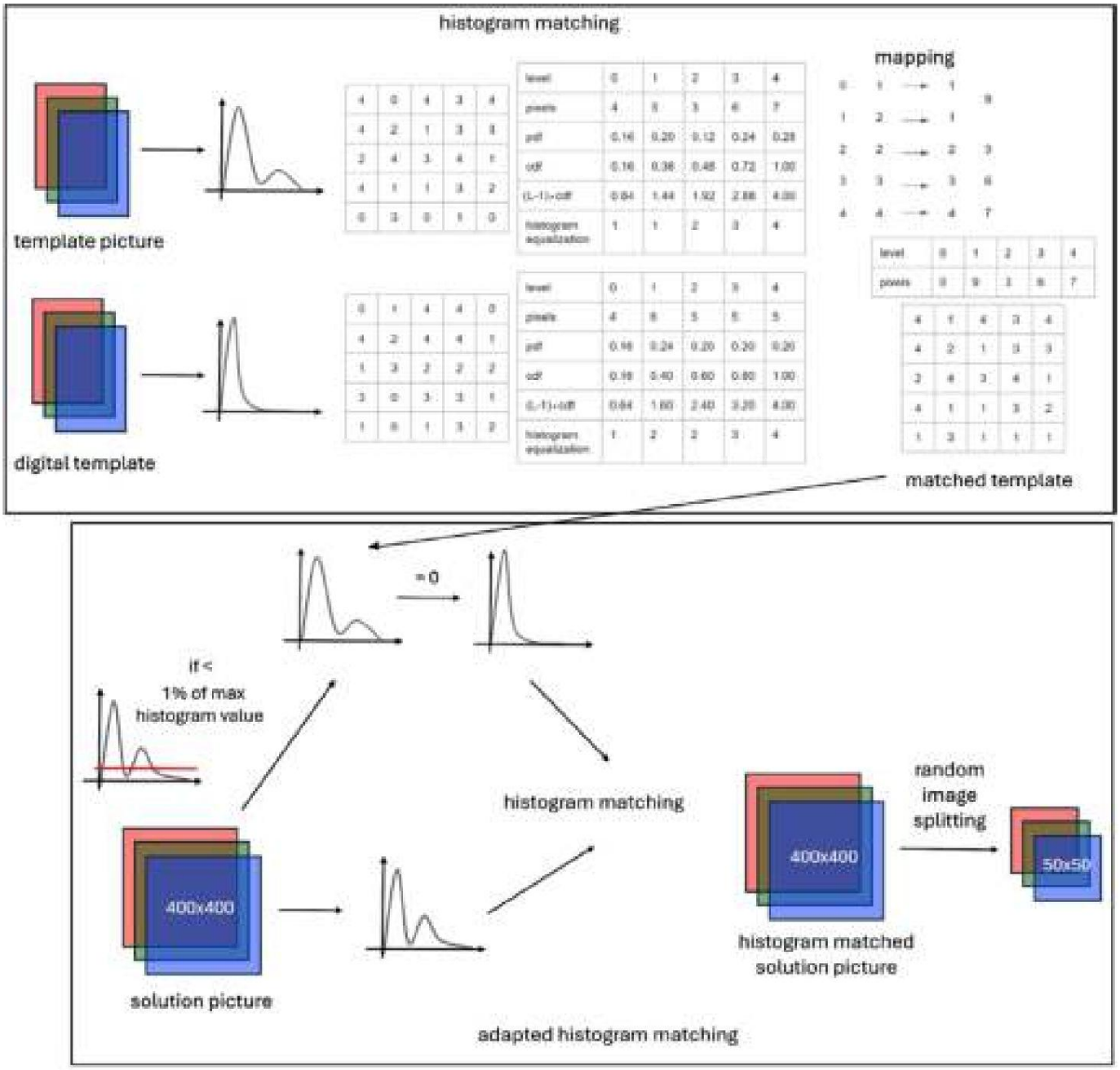
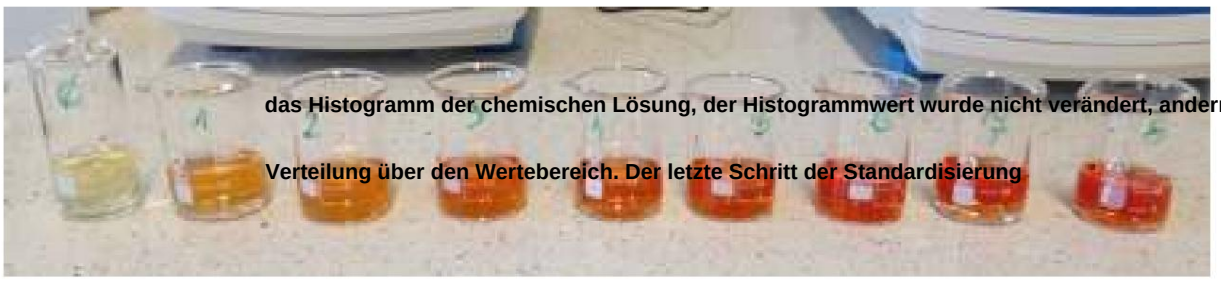


Abbildung 2. Diagramm des Farbabgleichalgorithmus.



das Histogramm der chemischen Lösung, der Histogrammwert wurde nicht verändert, andernfalls wurde er

Verteilung über den Wertebereich. Der letzte Schritt der Standardisierung

Abbildung 3. Vitamin C-Lösungen (0, 4,54, 11,31, 18,05, 22,52, 33,63, 44,64, 66,37 und 87,72 µg·mLÿ1)
in Gegenwart von doppelt destilliertem Wasser, Acetatpuffer, Ammoniumeisen(III)-sulfat und Orthophenanthrol

Machine Translated by Google

3.2. Software

Alle Berechnungen wurden in Python 3.10.0 durchgeführt. Die folgenden Bibliotheken und Pakete wurden verwendet: 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 und keras 2.10.0.

4. Ergebnisse und Diskussion

4.1. Reagenzien, Laborgeräte und Messverfahren

In dieser Arbeit basierte die kolorimetrische Reaktion zur Bestimmung von Ascorbinsäure (AA) auf der Reduktion von Fe(III) zu Fe(II) durch AA, das mit Orthophenanthrolin einen roten Farbkomplex bildet. In diesem Experiment wurden L(+)-Ascorbinsäure rein pa, CAS: 50-81-7 (POCH, Gliwice, Polen), 1,10-Phenanthrolin pa, CAS: 66-71-7 (POCH, Polen) und Ammoniumeisen (III)-sulfat-Dodecahydrat rein pa, CAS: 7783-83-7 (POCH, Polen) verwendet. Der 1 M Acetatpuffer wurde in unserem Labor durch Mischen von 1,0 mol·L (CH3COOH) mit 1,0 mol·L Essigsäure (CH3COONa) (beide Reagenzien von Avantor Performance Materials, Gliwice, Polen, bezogen) und Einstellen auf den gewünschten pH-Wert mit 10N HCl hergestellt. Alle Lösungen, einschließlich AA in einer Konzentration von 2,5 gLÿ1 , **Phenanthrolin in einer Konzentration von 4,0** gLÿ1 , **Ammoniumeisen(III)-sulfat in einer Konzentration von 2,5** gLÿ1 **und 1 M Acetatpuffer (pH 4,6), wurden mit doppelt destilliertem Wasser hergestellt.** Für die Experimente wurden **kleine Laborgeräte verwendet, darunter eine präzise Analysenwaage (RADWAG, Modell AS 60/220.XS, Radom, Polen), ein Magnetrührer (WIGO, Toszek, Polen), automatische Pipetten mit verschiedenen Volumina (Eppendorf, Deutschland) und Glaswaren, d. h. Becher und Petrischalen. Zur Einstellung der pH-Werte der Pufferlösung verwendeten wir das Labor-pH-Meter SevenCompact S210 (Mettler Toledo, Greifensee, Schweiz). Alle Experimente wurden bei Raumtemperatur durchgeführt.**

Zur direkten Farbbeurteilung wurden Lösungen mit unterschiedlichen Vitamin-C-Konzentrationen hergestellt, indem den Bechern zunächst 9 ml destilliertes Wasser, dann 1 ml Acetatpuffer, anschließend die entsprechende Menge Ascorbinsäure und 0,5 ml Ammoniumeisen(III)-sulfat zugegeben wurden. Nach einer Wartezeit von 3 min wurden 0,5 mL Phenanthrolin zugegeben. Die Zugabemengen von AA betrugen 0, 20, 50, 80, 100, 150, 200, 300 und 400 µL. Abbildung 3 zeigt das Foto der für weitere Tests vorbereiteten Proben. Die zur Optimierung untersuchten Parameter waren Reagenzkonzentration, Volumen und die Zeit zwischen den einzelnen Schritten (Einfluss der chemis Nach einer Wartezeit von etwa 20 Minuten wurden 10 ml jeder farbigen Lösung in Petrischalen mit flachem Boden gegossen und mit einer Smartphone-Kamera Fotos aufgenommen.

4.2. Vorbereitung einer Farbvorlage und Bildaufnahme

Das Experiment war in mehrere Schritte unterteilt und der Ablauf ist in Abbildung 4 dargestellt.

Der erste Schritt war die Vorbereitung der Vorlage für die Farbstandardisierung. Um zu bestimmen, welche Farben als Referenz verwendet werden sollten, wurde eine Reihe von Vitamin-C-Lösungen hergestellt, die den in dieser Studie betrachteten Konzentrationsbereich abdeckten. Dies wurde getan, um den Bereich der für diese Berechnungen signifikanten RGB-Werte abzuschätzen. Es wurden Bilder der getesteten Lösungen aufgenommen und mithilfe einer speziellen selbst entwickelten Software 12 Konfigurationen von RGB-Werten (Farben) zur Verwendung in der Vorlage ausgewählt. Die gemittelten RGB-Werte bildeten die Grundlage für die Bestimmung des Farbbereichs in der Vorlage, der dem zu lösenden Problem gewidmet war. Um die Farbpalette etwas zu erweitern, wurden zusätzliche obere und untere Werte geschätzt und die Vorlage so erstellt, dass sie 12 Farben enthielt. Die Vorlage, bei der jede Farbe durch ein Quadrat dargestellt wurde, wurde in mehreren Kopien mit einem Drucker ausgedruckt. Die Größe jedes Qu

Machine Translated by Google

tt

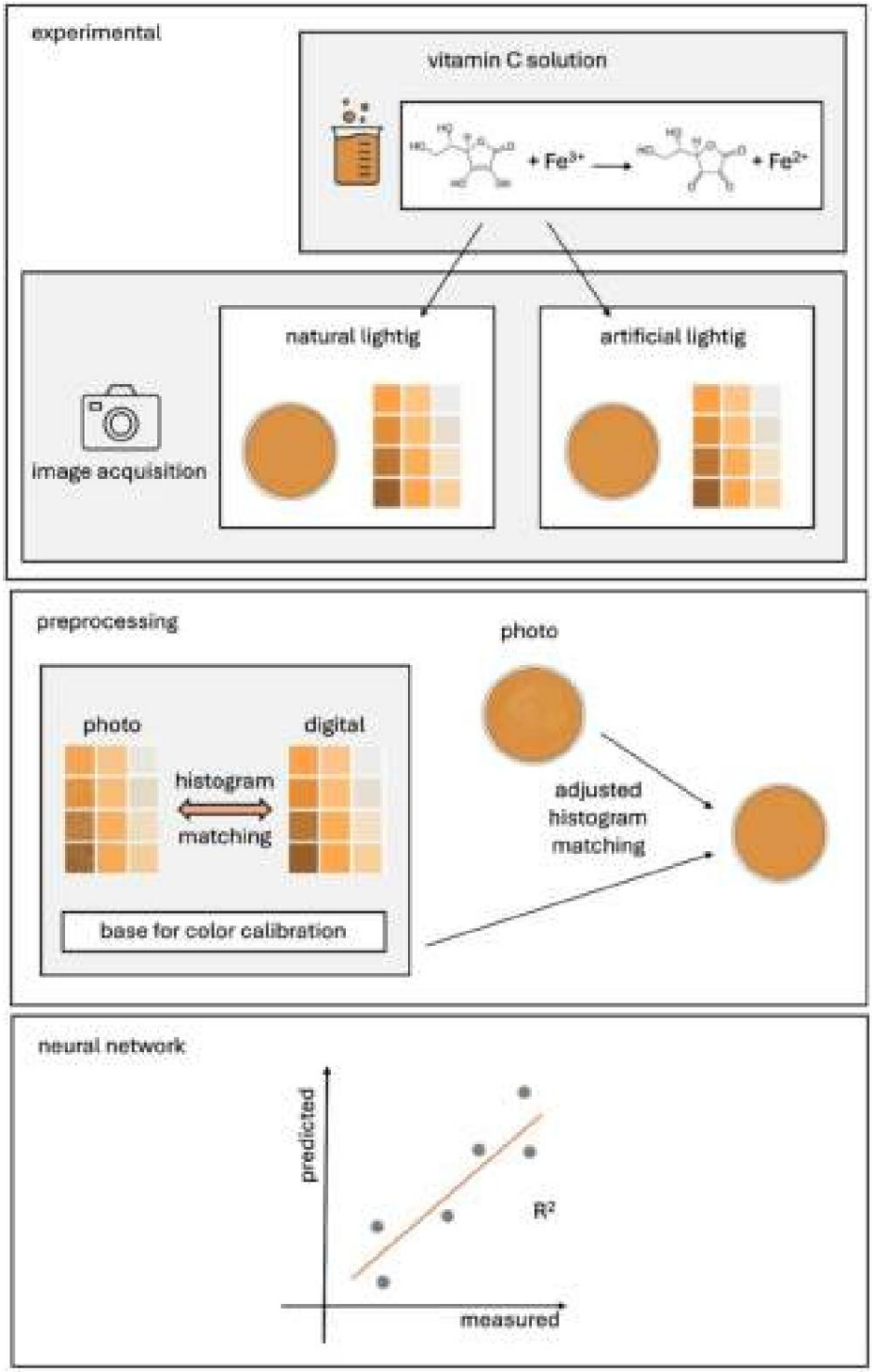


Abbildung 4. Schema der durchgeführten Experimente und Berechnungen: Herstellung von AA-Lösungen mit unterschiedlichen Konzentrationen; Auftragen der Lösungen auf Petrischalen; Bildaufnahme unter verschiedenen Lichtbedingungen (AA-Lösungen mit Farbvorlagen); Bildvorverarbeitung – Farbstandardisierung; Training und Auswertung des Regressionsnetzwerks.

ff

Der Prozess der Aufnahme der Bilder bestand darin, 10 ml der Lösung in einer Petrischale auf einem weißen Hintergrund (wodurch der Kontrast zwischen dem Hintergrund und der Petrischale mit der Lösung hervorgehoben wurde) neben der Vorlage (die als Farbkontrollsatz diente) auf einer schwarzen Oberfläche zu platzieren. Es gab zwei Bedingungen, unter denen die Bilder aufgenommen wurden: natürliches Licht und das Licht der Taschenlampe eines Smartphones. In der

Machine Translated by Google

Die fotografierten Objekte wurden von der Umgebung isoliert und so in den Versuchsaufbau eingebaut, dass sie kein natürliches Licht erhielten. Unter diesen beiden Beleuchtungen konnten Unterschiede in den Farben der Lösungen und der fotografierten Farbtabellen beobachtet werden (Abbildung 5). Der Vorgang wurde für jede Probe mit einer anderen Vitamin-C-Kon Alle Bilder wurden mit einem Xiaomi-Smartphone (Xiaomi Redmi Note 9, Xiaomi, Peking, China) aufgenommen, das über eine 48 MP-Kamera mit einer f/1.79-Blende verfügt. Die Auflösung der Bilder betrug 3984 × 1840 px.

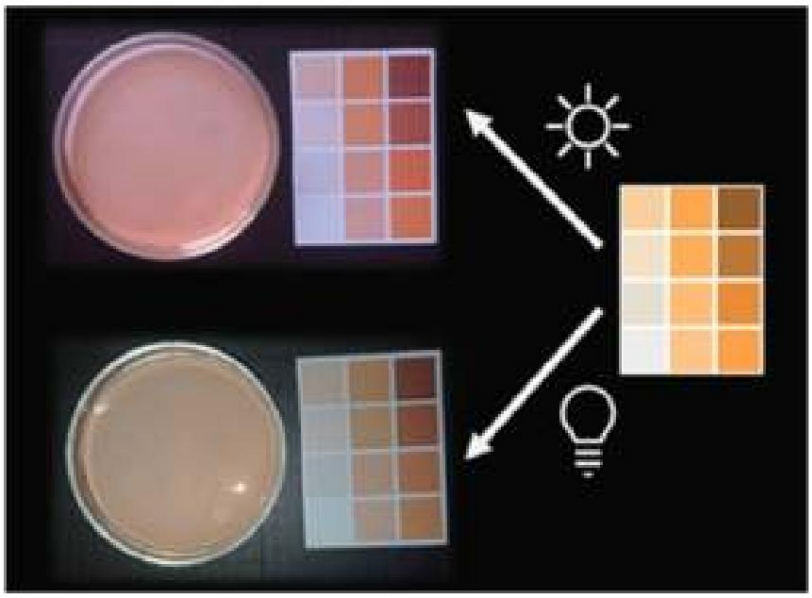


Abbildung 5. Kontrolle der Lichtverhältnisse – zwei Ansätze: natürliches Licht (oben) und Beleuchtung durch Smartphone-Lampe (unten). Rechte Seite eines Bildes – digitale Vorlage, die beim Histogrammabgleich verwendet wird; Pfeile zeigen die Lichtverhältnisse an, unter denen die Bilder aufgenommen wurden.

Um die bei der Standardisierung verwendeten Vorlagen histogrammmäßig abzugleichen, wurden unter beiden Bedingungen zusätzliche Bilder der Vorlage aufgenommen. Aus den Bildern wurden Vorlagen ausgeschnitten und ein Histogrammabgleich zwischen den resultierenden Vorlagen und der Originalvor Der Prozess mit den Ergebnissen ist in Abbildung 6 zu sehen. Jede Vorlage wird als RGB-Bild zusammen mit ihren Histogrammen für jeden der R-, G- und B-Kanäle dargestellt. Diese Abbildung veranschaulicht deutlich die Funktionsweise unseres Algorithmus. Die erste Spalte zeigt das ideale Histogramm der für diese Aufgabe entwickelten digitalen Farbvorlage. Es wird erwartet, dass die Farbhistogramme der Vorlagenfotos den idealen Eigenschaften entsprechen. Unter realen Bedingungen gibt es jedoch erhebliche Unterschiede (zweite und vierte Spalte) zwischen den Histogrammen der Fotos und denen der digitalen Version. Nach Anwendung des in dieser Arbeit vorgeschlagenen Algorithmus kann jedoch die erwartete Form des Farbhistogramms für alle drei Komponenten R, G und B reproduziert werden. Die Diagramme

Die kumulativen Histogramme der übereinstimmenden Bilder sind für jeden Kanal dieselben wie das kumulative Histogramm der digitalen Vorlage (die als Referenz dient). Jeder Ausschlag in einem bestimmten Histogramm entspricht einer Komponente einer bestimmten F Das weniger diskrete Erscheinungsbild der Histogramme der ursprünglichen Vorlagenbilder wird durch Abweichungen beim Drucken und verschiedene Variablen verursacht, die sich darauf auswirken, wie die Kamera das Objekt bei bestimmten Änderungen wahrnimmt (z. B. Schatten und Lichter).

Machine Translated by Google

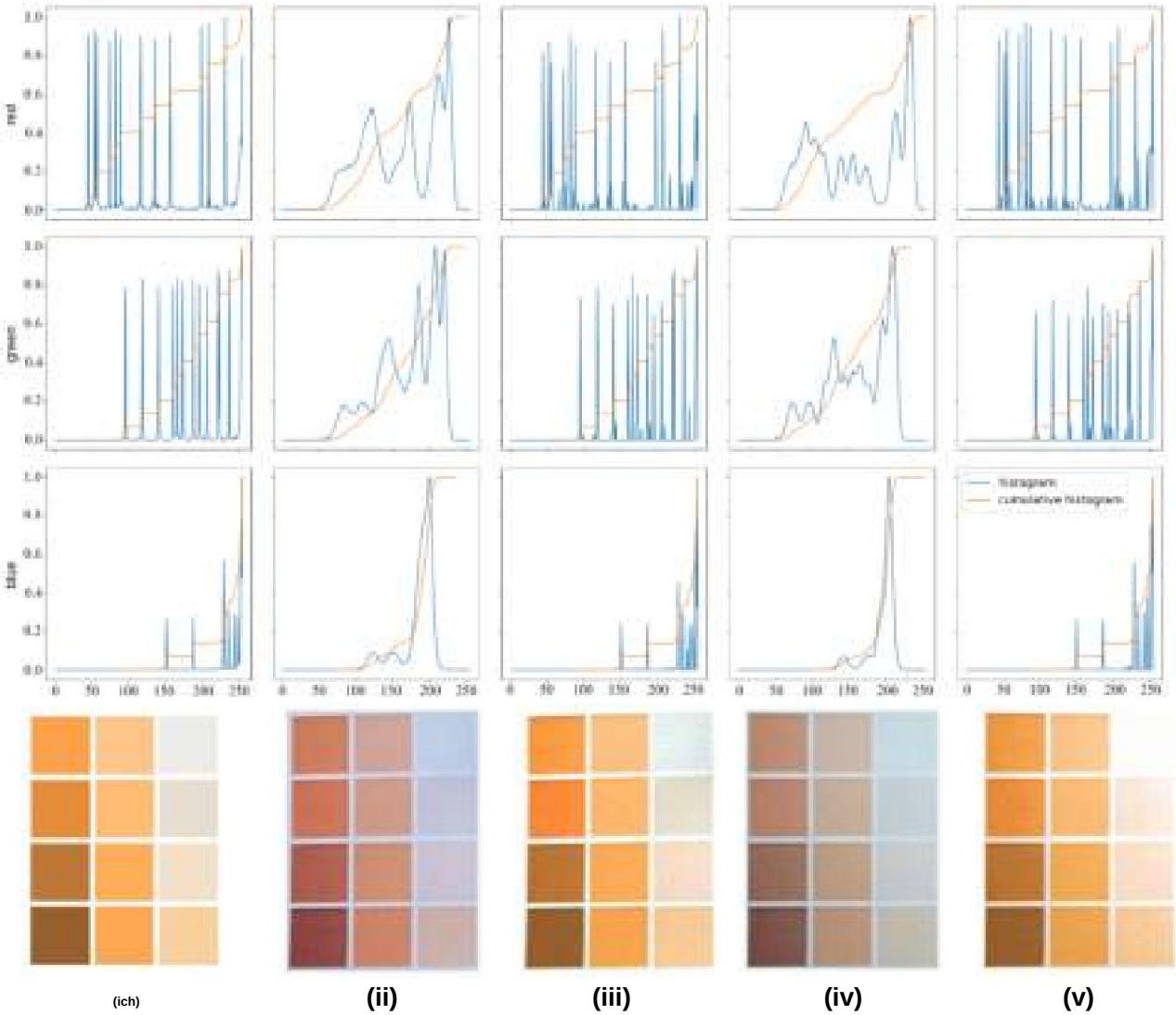


Abbildung 6. Vorlagen und ihre Histogramme für RGB-Kanäle: (i) ursprüngliche digitale Vorlage, (ii) Bild von eine Vorlage unter natürlichem Licht, (iii) Bild einer passenden Vorlage unter natürlichem Licht, (iv) Bild einer Vorlage unter künstlichem Licht und (v) Bild einer passenden Vorlage unter künstlichem Licht.

Der Datensatz für die Berechnungen wurde wie folgt aufbereitet: 198 Bilder wurden aufgenommen für für neun Konzentrationen von Vitamin C unter zwei verschiedenen Lichtbedingungen. Das bedeutet dass für eine gegebene Lösung für jede der Bedingungen 11 Bilder aufgenommen wurden. Jedes der Bilder wurden einem Standardisierungsprozess unterzogen, wie in Abschnitt 3.1 beschrieben. Das exemplarische Der Effekt ist in Abbildung 7 dargestellt . In Abbildung 8 wird der Effekt der Anwendung des Algorithmus auf verschiedene Bilder der Vitamin C-Lösung mit einer Konzentration von 33,63 µg/mL werden präsentiert. Diese wurden Datensätze für zwei Lichtverhältnisse generiert. Die Kombination beider Sets führten zur Erstellung eines gemischten Bedingungssets, bestehend aus:

- Trainingssatz: 60.416 Bilder
- Validierungssatz: 20.480 Bilder
- Testsatz: 20.480 Bilder.

Die vorbereiteten Bilddatensätze dienten als Eingangsdaten für das Training eines Regressionsneurons Netzwerk, das verwendet wurde, um ein Modell zur Bestimmung der AA-Konzentration zu erhalten, basierend auf die aufgenommenen Fotos.

Machine Translated by Google

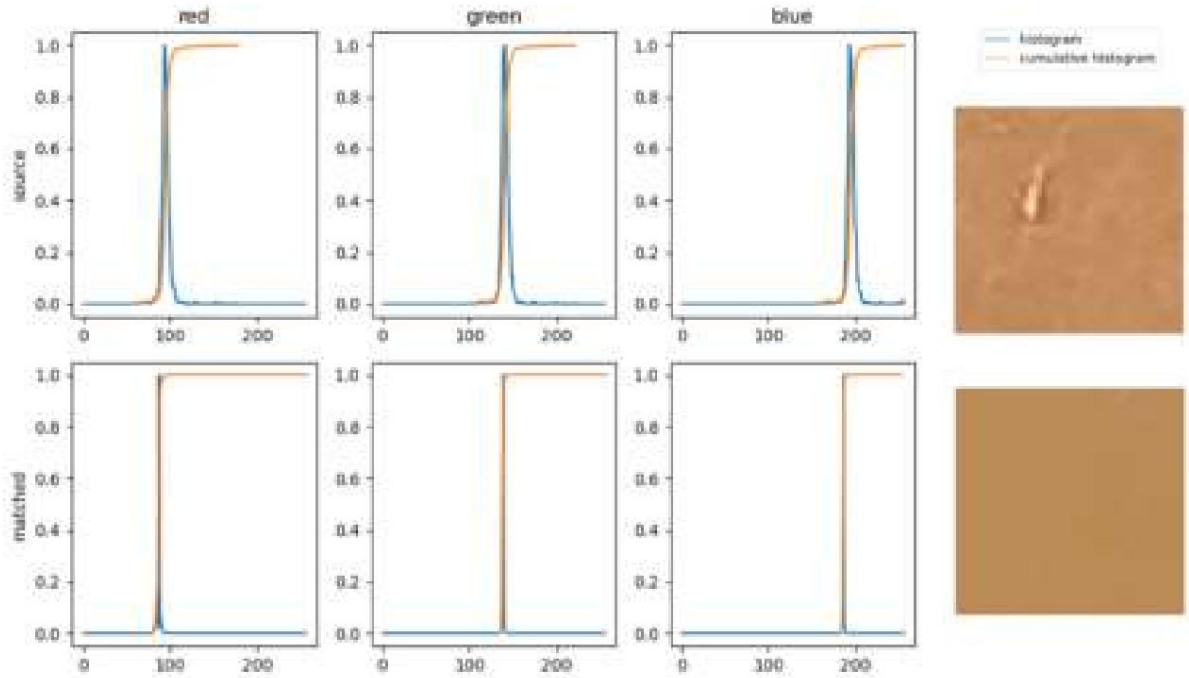


Abbildung 7. Bilder der Vitamin-C-Lösung (33,63 µg/ml) unter künstlichen Lichtbedingungen vor (Quelle) und nach (übereinstimmend) Anwendung des Standardisierungsalgorithmus und entsprechende Histogramme der RGB-Kanäle.

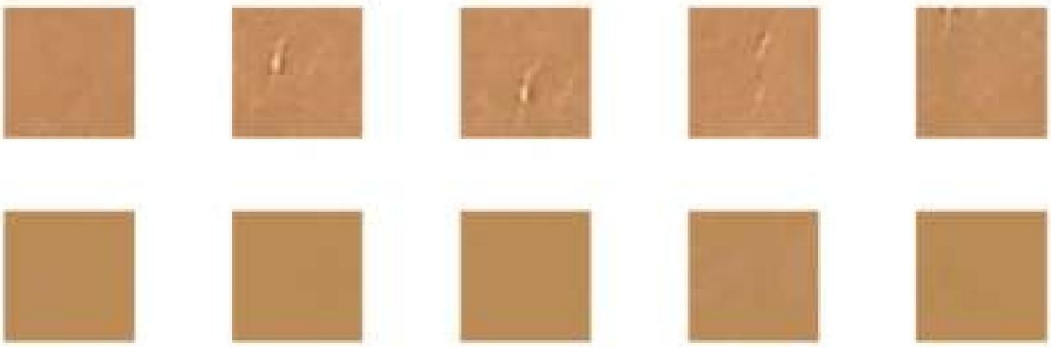


Abbildung 8. Verschiedene beispielhafte Bilder der Vitamin C-Lösung (33,63 µg/mL) unter künstlichen Lichtbedingungen vor (oben) und nach (unten) Anwendung des Standardisierungsalgorithmus

tt tt

4.3. Netzwerkarchitektur

Ein multivariates Regressionsmodell wurde mithilfe tiefer neuronaler Netzwerke mit einer unkonventionellen Architektur definiert, die speziell für diese Aufgabe entwickelt wurde. Das Netzwerkdesign berücksichtigte wesentliche Informationen wie farbbasierte Modellierung, die Ähnlichkeit der Farben, die aufeinanderfolgenden Werten der Zielvariable entsprechen, und die annähernd einheitliche Farbe jedes Fotos.

Die Eingabeschicht des neuronalen Netzwerks akzeptierte Bilder mit den Abmessungen 50 × 50 und drei Farbkanälen (RGB), die zufällig aus jedem Foto ausgewählt wurden. Anschließend wurden Conv2D -Faltungsschichten mit kleinen 1 × 1-Filtern, L2-Regularisierung und der GELU-Aktivierungsfunktion (Gaussian Error Linear Unit) nach jeder Faltungsschicht angewendet. Außerdem wurde eine Normalisierungsschicht verwendet, die entlang der Merkmalsachse für jede Probe operierte, um den Lernprozess zu stabilisieren und zu verbessern. Die Architektur umfasste auch Pooling-Schichten, die die räumliche Größe der Merkmalskarte reduzierten, darunter AveragePooling2D und GlobalMaxPooling2D, die die räumliche Dimension auf einen Wert pro Merkmalskarte reduzierten. Die Kombination aus globalem Pooling und Abflachung ist eine Standardtechnik zur Dimensionsreduzierung vor vollständig verbundenen Schichten. Im nächsten Sc

tt tt

Machine Translated by Google

zu vollständig verbundenen dichten Schichten mit GELU-Aktivierung und L2-Regularisierung, während die Das Hinzufügen einer Dropout-Schicht verhinderte Überanpassung. Die Ausgabe war ein Wert, der zur vorhergesagten Konzentration der Lösung im Bild. Eine detaillierte Beschreibung der Die Netzwerkarchitektur ist in Tabelle 2 aufgeführt und eine Übersicht ist in Abbildung 9 dargestellt. Die Gesamtzahl der Parameter betrug 23.879, und das Training in einem einzigen Zyklus dauerte einige Sekunden (je nach Berechnungsvariante).

Tabelle 2. In dieser Studie verwendete Architektur und Parameter des neuronalen Netzwerks.

Schicht (Typ)	Ausgabeform	Anzahl der Parameter
Eingabeebene	(Keine, 50, 50, 3)	0
Konv2D	(Keine, 50, 50,3 2)	128
Aktivierung	(Keine, 50, 50, 32)	0
Konv2D	(Keine, 50, 50, 64)	2112
Aktivierung	(Keine, 50, 50, 64)	0
DurchschnittlicherPooling2D	(Keine, 46, 46, 64)	0
Ebenennormalisierung	(Keine, 46, 46, 64)	128
Konv2D	(Keine, 46, 46, 64)	4160
Aktivierung	(Keine, 46, 46, 64)	0
DurchschnittlicherPooling2D	(Keine, 44, 44, 64)	0
GlobalMaxPooling2D	(Keine, 64)	0
Abflachen	(Keine, 64)	0
Dicht	(Keine, 150)	9750
Ausfallen	(Keine, 150)	0
Dicht	(Keine, 50)	7550
Dicht	(Keine, 1)	51
	Gesamtparameter: 23.879 (93,28 KB) Trainierbare Parameter: 23.879 (93,28 KB) Nicht trainierbare Parameter: 0 (0,00 Byte)	

In einer Faltungsschicht wie Conv2D arbeitet die angewandte L2-Regularisierung auf der Faltungsfilter (Kernel). Diese Filter sind für die Extraktion von Merkmalen aus die Eingabebilder. Das Hinzufügen einer L2-Regularisierung bedeutet, dass während des Optimierungsprozesses Die Verlustfunktion, die das Modell zu minimieren versucht, enthält einen zusätzlichen Term, der bestraft große Gewichtswerte der Filter. Das Hinzufügen einer L2-Regularisierung zur Faltungsschicht hilft verhindert Überanpassung, stabilisiert den Optimierungsprozess und fördert eine gleichmäßigere und stabile Lösungen, die zu einer besseren Modellgeneralisierung führen.

Die Kernelgröße von 1 × 1 in der Faltungsschicht hat spezifische und nützliche Eigenschaften in neuronale Netze. Die 1 × 1-Filter werden typischerweise verwendet, um räumliche Dimensionen zu reduzieren, ohne räumliche Aggregation, die bei der Reduzierung der Anzahl der Feature-Kanäle hilfreich sein kann oder Einführung von Nichtlinearität ohne Änderung der räumlichen Größe. Diese Konfiguration wird oft genannt eine „punktweise Faltung“. Der 1 × 1-Kernel verändert die Anzahl der Ausgabekanäle (Tiefe) ohne die Breite und Höhe des Eingabebildes zu ändern, was die Kombination von Informationen, die in verschiedenen Eingangskanälen enthalten sind, ohne räumliche Informationen zu integrieren. Jeder Ausgabepunkt ist eine lineare Kombination der Eingangskanalwerte bei einem gegebenen räumlichen Punkt. Wenn wir beispielsweise an einem gegebenen Eingabepunkt (x, y) Werte in den Kanälen (w1, w2, w3), der 1 × 1-Kernel wendet auf diese Werte unterschiedliche Gewichte an und kombiniert sie, neue Ausgabekanäle erstellen. Ein solcher Kernel hat weniger Parameter im Vergleich zu größeren Faltungskernel, was zu weniger Rechenoperationen führt. Dies ist effizient in in Bezug auf Speicher und Rechenzeit. Bei der Modellierung auf Basis von RGB-Komponenten Ein 1 × 1 Kernel bedeutet, die Beziehungen zwischen einzelnen Farbkomponenten zu untersuchen,

Machine Translated by Google

anstatt der Nachbarschaft jedes Pixels. In unserer Aufgabe war es nicht nötig, kleine
und dann größere Details, wie es bei der Formerkennung der Fall ist.

Die GELU-Aktivierungsfunktion [47], die als Alternative zu ReLU (Rectified Linear
Aktivierungsfunktionen in neuronalen Netzwerken verantwortlich und sorgen für ein besseres Gradiente
und effizienteres Lernen in tiefen neuronalen Netzwerken (Abbildung 10). Seine Kontinuität,
Differenzierbarkeit und Fähigkeit, nichtlineare Dateneigenschaften zu bewahren, machen es zu einem wertvollen
Design von Deep-Learning-Modellen. Im Gegensatz zu ReLU, das unter Gradientenflussproblemen leiden kann
Trainings (bekannt als das Problem des verschwindenden Gradienten) bietet GELU einen besseren Gradi
gleichmäßigeren und gleichmäßigeren Durchsatz. GELU bewahrt nichtlineare Merkmale, die entscheidend
Lernen von Datendarstellungen, um Modellen zu helfen, komplexe Abhängigkeiten besser zu erfassen in
Trainingsdaten auf.

Die Kernelgröße von 1 × 1 in der Faltungsschicht weist spezifische und nützliche Eigenschaften für

Faltungsfilter (Kernel). Diese Filter sind für das Extrahieren von Merkmalen aus t Unit) und anderen

Gewichtswerte der Filter. Das Hinzufügen einer L2-Regularisierung zur Faltungsschicht hilft p- während des

verhindert Überanpassung, stabilisiert den Optimierungsprozess und fördert aufgrund seiner Struktur einen

bare Lösungen, was zu einer besseren Modellgeneralisierung führt.

in neuronalen Netzen. Die 1 × 1-Filter werden typischerweise verwendet, um räumliche Dimensionen zu redu



oder die Einführung von Nichtlinearität ohne Änderung der räumlichen Größe. Diese Konfiguration wird oft

Kombination von Informationen aus verschiedenen Eingangskanälen ohne Integration von

Kanäle (w1, w2, w3) wendet der 1 × 1-Kernel unterschiedliche Gewichte auf diese Werte an und

Rechenoperationen. Dies ist e

, als Alternative zu ReLU (Rectified Li-ear Unit) und anderer
Aktivierungsfunktionen in neuronalen Netzwerken eingesetzt, sorgt für besseres Gradientenverhalten
und effizienteres Lernen in tiefen neuronalen Netzwerken (Abbildung 10). Seine Kontinuität, d-

bei der Entwicklung von Deep-Learning-Modellen. Im Gegensatz zu ReLU, das während des Trainings
unter Gradienten-FL-Problemen leiden kann (bekannt als das verschwindende Gradientenproblem), bietet G

cial für das Lernen von Datendarstellungen, hilft Modellen, komplexe Abhängigkeiten besser zu erfassen

Abbildung 9. In dieser Studie verwendete neuronale Netzwerkarchitektur.

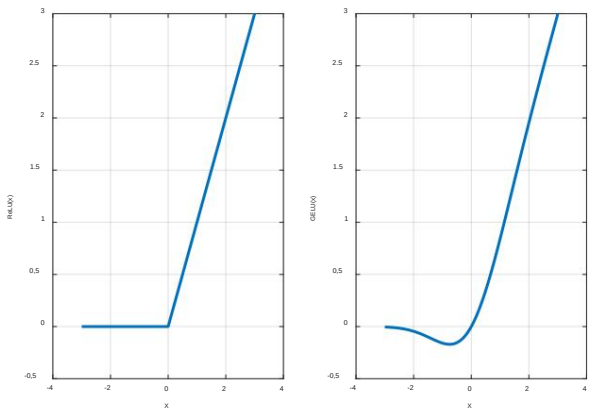


Abbildung 10. ReLU- und GELU-Aktivierungsfunktionen.

Machine Translated by Google

In der Ebene „LayerNormalization“ wird für jede Probe eine Normalisierung entlang der Merkmalsachse durchgeführt. Dies bedeutet, dass die Normalisierung unabhängig von anderen Proben im Stapel is Jeder Merkmalsvektor (z. B. Kanäle in einem Bild) wurde separat normalisiert, wobei nur die Werte innerhalb dieses Vektors und nicht der gesamte Stapel berücksichtigt wurden. Im Gegensatz zur BatchNormalisierung, die je nach Stapelgröße funktioniert, bleibt die LayerNormalisierung auch bei kleinen Stapeln oder einer Stapelgröße von eins stabil. Die Normalisierung kann dazu beitragen, Probleme wie verschwindende oder explodierende Gradienten zu verhindern und die Konvergenz zu beschleunigen. Die Normalisierung der Aktivierungen hilft dabei, Werte in einem vernünftigen Bereich zu halten, was den Lernprozess stabilisiert. Wenn Aktivierungen normalisiert sind, fällt es dem Optimierer (Lernalgorithmus) leichter, Netzwerkgewichte anzupassen, was die Konvergenz zu optimalen Gewichtungswerten beschleunigt.

GlobalMaxPool2D ist eine Schicht, die in neuronalen Netzwerken verwendet wird, um räumliche Datendimensionen nach dem Anwenden von Faltungsschichten oder anderen Schichten, die räumliche Daten verarbeiten, wie z. B. Pooling-Schichten [48,49], zu reduzieren. GlobalMaxPool2D funktioniert, indem der Maximalwert jedes Merkmalskanals über den gesamten räumlichen Bereich (alle Punkte) des Eingabemerkmalstensors ausgewählt wird. GlobalMax-Pool2D ist ein effektives Werkzeug zur Reduzierung der räumlichen Dimensionalität, da nur die signifikantesten Merkmale (höchsten Werte) aus jedem Merkmalskanal beibehalten werden. Dadurch wird die Anzahl der Modellparameter reduziert, was dazu beitragen ka GlobalMaxPool2D hat keine trainierbaren Parameter; sein einziger Zweck besteht darin, den Maximalwert aus jedem Merkmalskanal in der gesamten Merkmalskarte auszuwählen . Für jeden Merkmalskanal wählt die GlobalMaxPool2D-Ebene den höchsten Wert im gesamten räumlichen Bereich (Höhe und Breite) aus. Das Ergebnis von GlobalMaxPool2D ist ein Tensor mit reduzierten Dimensionen, der nur die Maximalwerte für jeden Merkmalskanal enthält. Nach der Anwendung von GlobalMaxPool2D kann der resultierende Tensor an eine vollständig verbundene Ebene oder andere Ebenentypen übergeben werden, die eindimensionale Datenvektoren verarbeiten.

In dieser Studie wurde das Modell mithilfe des Adam-Optimierers mit angepassten Betaparametern und einer definierten Verlustfunktion kompiliert, die die Summe von MSE und MAE war. Wenn die Verlustfunktion eine Kombination aus MSE und MAE ist, wirkt sich dies darauf aus, wie das Modell lernt, Fehler zu minimieren. MSE bestraft große Fehler strenger (aufgrund der Quadrate der Differenzen), während MAE alle Fehler gleich behandelt. Die Kombination dieser Verlustfunktionen bedeutet, dass das Modell darauf abzielt, sowohl große als auch kleine Fehler zu reduzieren. Der Gradient von MSE ist bei größeren Fehlern steiler, wodurch das Modell schneller aus signifikanten Fehlern lernt. Der MAE-Gradient ist konstant, was zu einer konsistenteren Lernrate über alle Fehler hinweg führt. MAE ist resistenter gegenüber Ausreißern als MSE, da es keine quadrierten Differenzen beinhaltet. Eine Verlustfunktion, die MSE und MAE kombiniert, kann in einigen Fällen zu einer schnelleren Konvergenz führen, indem sie die Stärken beider Methoden nutzt. MSE kann die Reduzierung großer Fehler beschleunigen, während MAE Stabilität u

Während des Trainings wurde eine Änderung der Lernrate vorgenommen, nachdem für fünf Epochen ein stabiles val_loss-Niveau erreicht wurde. Der Anfangswert war 0,001. Einzelheiten zu den Lernparametern finden Sie in Tabelle 3.

Tabelle 3. Parameter des Netzwerktrainings.

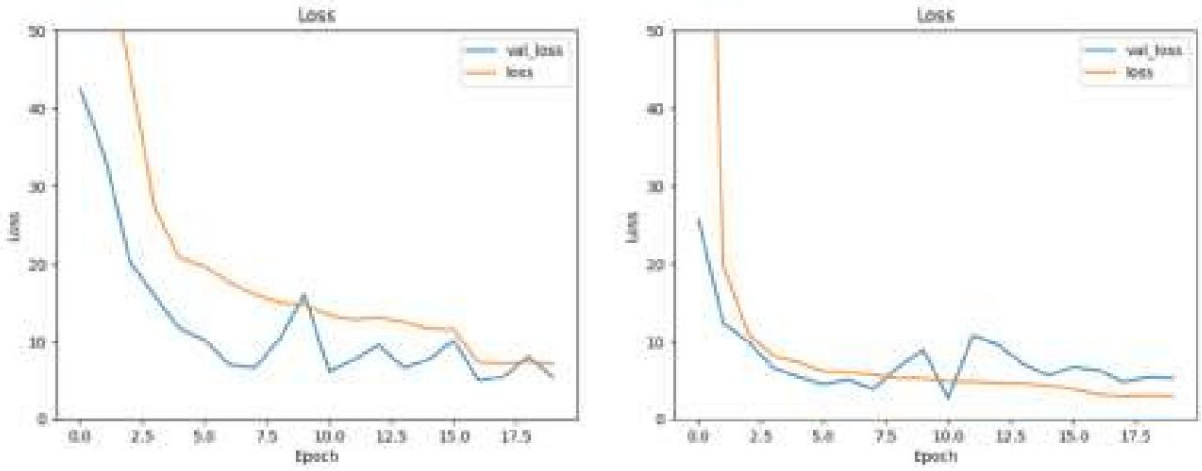
Parameter	Einstellung
Bildgröße	50 × 50 × 3
Verlustfunktion	MSE + MAE
Optimierer	Adam
Anfängliche Lernrate	0,001
Metrisch	RMSE
Batchgröße	30
Epoche	20
Shuffle	Jede Epoche

Der Eingabedatensatz bestand aus Bildern mit den Abmessungen 400 × 400 und wurde in drei Teile aufgeteilt: Trainings-, Validierungs- und Testsätze im Verhältnis 60/20/20. Während des Trainin

Machine Translated by Google

Aus den Bildern wurden dynamisch **50 × 50 Quadrate ausgeschnitten, die als der direkte Input zur ersten Schicht des Netzwerks.**

Es gab sechs verschiedene Varianten des Testens im Deep Learning zum Trainieren einer Regression Algorithmus zur Bestimmung der Vitamin C-Konzentration anhand von Bildern: Originalbilder für künstliche Beleuchtung, angepasste Bilder für künstliche Beleuchtung, Originalbilder für natürliche Beleuchtung abgestimmte Bilder für künstliche Beleuchtung, Originalbilder für gemischte Datensätze und abgestimmte Bilder für gemischte Datensätze. Für jedes davon wurde das Modell trainiert (Abschnitt 3.2) und Vorhersagen wurden gemacht. Trainings- und Validierungsverluste für Berechnungen von künstlichem Licht Daten, sowohl für Originalbilder als auch für Histogramm-angepasste Bilder, sind in Abbildung 11 dargestellt. Die Spanne war in beiden Fällen ähnlich; die Wertschwankungen waren jedoch bei den angepassten Bilder, und unter dieser Bedingung waren Validierungs- und Trainingsverluste einander ähnlicher.



(ich) **(ii)**

Abbildung 11. Trainings- und Validierungsverluste für die Berechnung künstlicher Beleuchtungsbedingungen für beide Original- und (i) und histogrammangepasste (ii) Bilder.

Abbildung 12 zeigt die Ergebnisse der Vorhersage mit dem vorgestellten neuronalen Netzwerk. Es gab drei Sätze von Ergebnispaaren: künstlicher Blitz, natürlicher Blitz, und gemischte Daten. Jeder Satz wurde sowohl für die Originalbilder als auch für die verarbeiteten Es gab eine Verbesserung der höheren Konzentrationswerte für jeden der Bedingungen (Tabelle 4). Die Standardabweichung der Vorhersagen war geringer bei der Anwendung die vorgeschlagene Standardisierungsmethode. Der größte Unterschied war im Fall von künstliches Licht, das durch die Vereinheitlichung der Lichtreflexe entstehen kann, die in die Originalbilder.

Tabelle 4. Auswertung der DL-Regressionsmodelle mit den besten Ergebnissen für sechs verschiedene Testvarianten.

	Künstlich Beleuchtung, Original	Künstlich Beleuchtung, Abgestimmt	Natürlich Beleuchtung, Original	Natürlich Beleuchtung, Abgestimmt	Gemischt Beleuchtung, Original	Gemischt Beleuchtung, Abgestimmt
Neigung	0,9799	0,9961	0,9706	0,9452	0,9784	0,9825
Abfangen	ÿ0,0036	ÿ0,2091	1,1888	0,6051	0,5832	0,4086
RMSE/µg·mLÿ1	1,5769	0,7525	2,2900	1,9542	1,4828	1,3510
R^2 Punktzahl	0,9979	0,9999	0,9953	0,9956	0,9966	0,9968

Machine Translated by Google

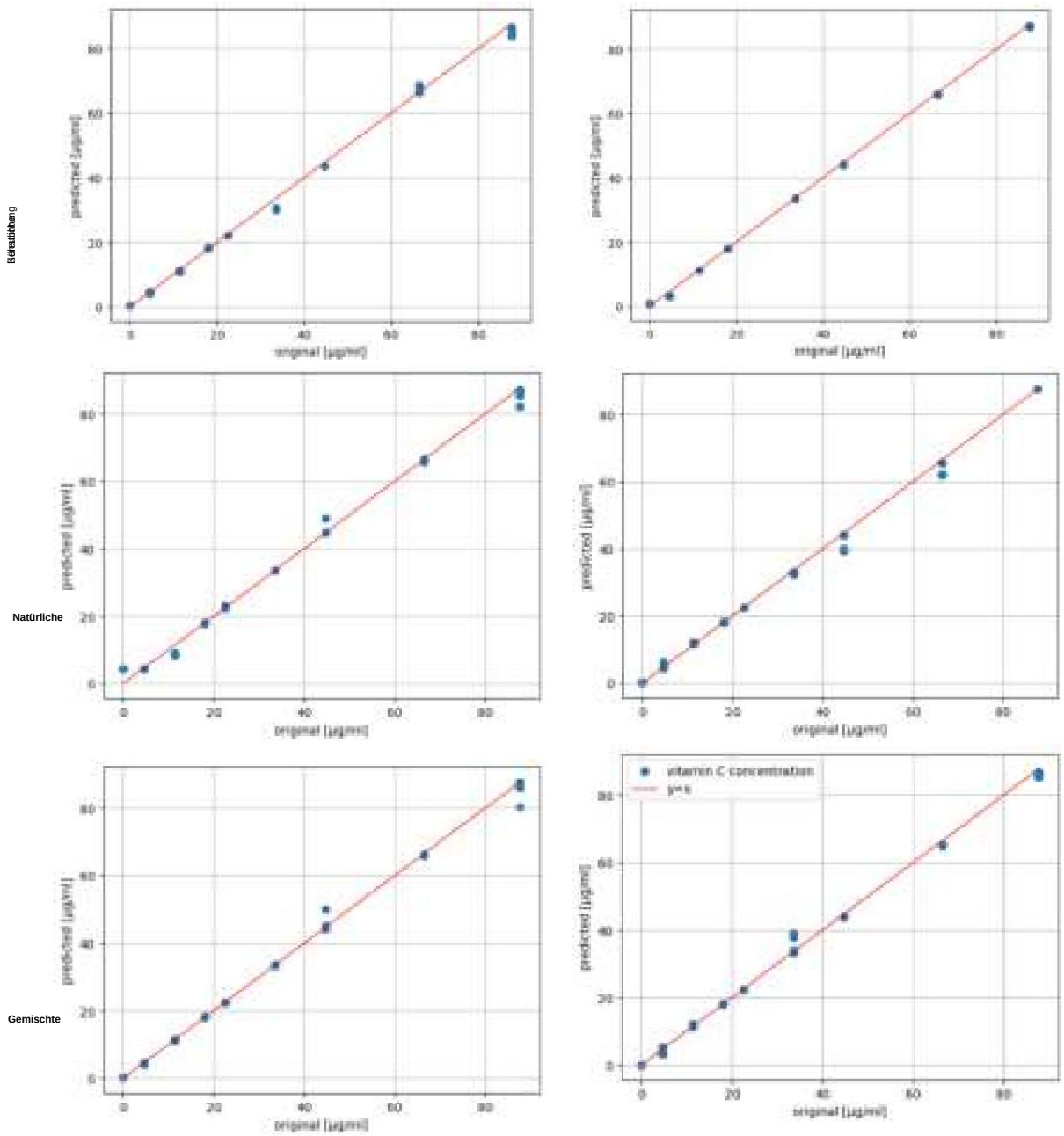


Abbildung 12. Auswertung von DL-Regressionsmodellen für sechs verschiedene Testvarianten vor (erste Spalte) und nach dem Farbabgleich (zweite Spalte).

Die Werte der numerischen Modellbewertung und Anpassungsfaktoren sind in Tabelle 4 aufgeführt. Der R2 -Wert war in allen Fällen sehr hoch (über 0,99); in jedem Fall R^2 war der Wert jedoch höher, wenn eine Standardisierung durchgeführt wurde. Die mittleren quadratischen Fehlerwerte waren bei den **abgeglichenen Bildern niedriger als bei den Originalbildern. Abhängig von den Lichtverhältnissen änderte sich die Lernrate von einem**

4.4. Ablationsstudie **In**

dieser Studie haben wir auch analysiert, wie sich verschiedene Modifikationen der Netzwerkarchitektur auf die Qualität des Regressionsmodells auswirkten. In der ersten Phase haben wir ein Netzwerk ff betrachtet , **das die Besonderheiten des Problems nicht berücksichtigte. Das Modell enthielt sechs Schichten: eine Faltung, eine Pooling-Schicht, eine Abflachungsschicht und drei dichte Schichten. Es wurde unter denselben Bedingungen wie zuvor trainiert. Die detaillierten quantitativen Ergebnisse sind in Tabelle 5 d**

Machine Translated by Google

einer konventionelleren Architektur verschlechterte die Fähigkeit des Modells, die Daten anzupassen, und seine Vorhersagefähigkeiten. Allerdings wurden immer noch Verbesserungen in der Modellqualität beobachtet nach Anwendung des Farbstandardisierungsalgorithmus.

Tabelle 5. Auswertung der in der Ablationsstudie berücksichtigten DL-Regressionsmodelle für sechs verschiedene Testvarianten, abhängig von den Lichtverhältnissen und dem gewählten Datensatz.

	Künstlich Beleuchtung, Original	Künstlich Beleuchtung, Abgestimmt	Natürlich Beleuchtung, Original	Natürlich Beleuchtung, Abgestimmt	Gemischt Beleuchtung, Original	Gemischt Beleuchtung, Abgestimmt
Neigung	0,9058	1,0138	0,9538	1,0071	1,0005	1,0112
Abfangen	0,6891	ÿ0,3598	ÿ0,8347	0,2263	0,8747	0,0203
RMSE/µg·mLÿ1	5,8415	1,7077	3,9504	2,5151	4,3488	2,8211
R^2 Punktzahl	0,9557	0,9962	0,9798	0,9918	0,9755	0,9897

Darüber hinaus wurden vor der endgültigen Version gewählt wurde. Einige davon werden im Vergleich zur endgültigen Version bewertet in Tabelle 6.

Tabelle 6. Auswertung der in der Ablationsstudie berücksichtigten DL-Regressionsmodelle für gemischte Beleuchtung und angepasste Testbedingungen, abhängig von Änderungen im Modell.

Wandel in der Architektur	RMSE/µg·mLÿ1	R^2 Punktzahl	Vergleich
Batchgröße = 120	2,3535	0,9947	Rückgang des R2- Scores und Anstieg des RMSE in Vergleich mit dem endgültigen Modell.
Bildgröße = 16	4,0146	0,9825	Signifikante Abnahme des R2- Scores und Zunahme von RMSE im Vergleich zum finalen Modell. Geringerer Rückgang des Anfangskurses im Vergleich zur endgültigen Version.
Modifikation der Verlustfunktion	3,8203	0,9856	Rückgang des R2- Scores und signifikanter Anstieg von RMSE im Vergleich zum finalen Modell. Minimiert Unterschiede zwischen Trainings- und Validierungsverlust während des Trainings.
Optimierer = Adagrad	3.1625	0,9877	Rückgang des R2- Scores und Anstieg des RMSE in Vergleich mit dem endgültigen Modell.
Optimierer = SGD	-	-	Es gab keine Veränderung im Verlustwert zwischen den Epochen. Das Modell lernte nicht, die Konzentration von Vitamin C-Lösungen.
Anfängliche Lernrate = 0,01 oder 0,0001	-	-	Es wurden keine signifikanten Änderungen der Ergebnisse festgestellt. Die erzielten Ergebnisse variierten nicht signifikant in Vergleich mit den endgültigen Modellergebnissen.

5. Schlussfolgerungen

Diese Arbeit stellt einen neuen Farbstandardisierungsalgorithmus vor, in dem die Hauptschritte bestehen aus der Vorbereitung einer Vorlage auf der Grundlage einer ersten Fotoserie, der Entwicklung einer Farb Abbildungsprinzip mit Hilfe entzerrter digitaler RGB-Histogramme eines idealen Musters und seines Fotos, Farbkonvertierung von Fotografien des untersuchten Objekts unter Verwendung von Informationen über die Abbildung von Vorlagenhistogrammen und die Bildrekonstruktion auf Basis der transformierten Histogramm. Der Algorithmus wurde für Fälle entwickelt, in denen sich die untersuchten Objekte wenig unterscheiden in Farbe und die Farbe jedes fotografierten Objekts ist nahezu einheitlich. Der Betrieb der Das vorgeschlagene Verfahren wurde in einer analytischen Chemieaufgabe demonstriert und verifiziert, die umfasste die kolorimetrische Bestimmung von Vitamin C. Fotografien von AA-Lösungen, unter kontrollierten Bedingungen zeigte sich kein großer Farbunterschied innerhalb der Konzentration Reichweite. Die Untersuchungen wurden unter Tageslicht- und künstlichen Beleuchtungsbedingungen durchgeführt

Machine Translated by Google

um sie zu einem einzigen Datensatz zu kombinieren und ein Modell zu erhalten, das die Bestimmung von Vitamin C unabhängig von den Lichtverhältnissen beim Aufnehmen der Fotos ermöglicht. Sätze standardisierter Bilder bildeten die Grundlage für Deep Machine Learning unter Verwendung der Architektur neuronaler Netzwerke in der Regressionsvariante.

Für den Konzentrationsbereich von Vitamin C von 0 bis 87,72 µg·mLÿ1 beträgt der RMSE für die Der Testsatz lag zwischen 0,75 und 1,95 µg·mLÿ1 im Vergleich zur nicht standardisierten Variante, bei der der RMSE zwischen 1,35 und 2,29 µg·mLÿ1 lag . **Die Kompatibilität der Ergebnisse der Konzentrationsvorhersage mit den tatsächlichen Daten, ausgedrückt durch R2 ,** war **für jede der standardisierten Varianten größer als** 0,99. Die größte Verbesserung wurde bei künstlichem Licht festgestellt, was durch die Beseitigung von Reflexionen auf den Oberflächen **der Lösungen in den Bildern nach dem Standardisierungsprozess erklärt werden kann .** Es sind weitere Studien erforderlich, um **die Nützlichkeit des Algorithmus in verschiedenen Szenarien zu bestimmen (z. B. kolorimetrische Vorhersage der Konzentration verschiedener Verbindungen). Dieser neue Ansatz kann zur Entwicklung eines neuen Bereichs der chemischen Analyse beitragen, der die quantitative Auswertung eines breiten Spektrums chemischer Analyten mithilfe von Mobiltelefonen und Bildanalyse umfasst. Diese vorgestellte Studie ist vielversprechend für Anwendungen in mehreren Bereichen von Wissenschaft und Technologie, wie z. B. Biomedizintechnik, Cl**

Beiträge des Autors: PK: Konzeptualisierung; Methodik; Datenkuratierung; formale Analyse; Untersuchung ; Software; Ressourcen; Visualisierung; Schreiben – Originalentwurf; Schreiben – Überprüfung u MJ: Konzeptualisierung; Formale Analyse; Mittelbeschaffung; Untersuchung; Methodik; Projektverwaltung ; Überwachung; Validierung; Schreiben – Originalentwurf; Schreiben – Überprüfung und Bearbeitung. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und ihr zugestimmt.

Finanzierung: Forschungsprojekt gefördert durch das Programm „Exzellenzinitiative – Forschungsuniversität“ für die AGH-Universität Krakau (ID1479).

Datenverfügbarkeitserklärung: Die in Python implementierten Algorithmen und die Beispieldaten für die Berechnungen sind auf der Github-Plattform verfügbar unter: https://github.com/pwkwiek/vitaminC (abgerufen am: 13. Juli 2024).

Danksagungen: Die Autoren danken Inz. Filip Ciepiela (AGH-Universität Krakau) für wertvolle Anleitung während der Projektimplementierung.

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Verweise

1. **Fairchild, MD** Color Appearance Models, **3. Auflage; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2013; ISBN 9781119967033.**
2. **Minz, PS; Saini, CS Bewertung des RGB-Würfel-Kalibrierungsrahmens und Auswirkung von Kalibrierungstabellen auf die Farbmessung von Mozzarella-Käse.** J. Food Meas. Charact. **2019,** 13, **1537–1546. [CrossRef]**
3. **Ernst, A.; Papst, A.; Ruf, T.; Garbas, JU Check My Chart: Ein robuster Farbdiagramm-Tracker für die kolorimetrische Kamerakalibrierung In Proceedings der 6. Internationalen Konferenz zu Computer Vision/Computergrafik-Kollaborationstechniken und -Anwendungen, MIRAGE'13, Berlin, Deutschland, 6.–7. Juni 2013.**
4. **McCamy, CS; Marcus, H.; Davidson, JG Farbwiedergabetabelle.** J. Appl. Photogr. Eng. **1976, 2, 95–99.**
5. **Sunoj, S.; Igathinathane, C.; Saliendra, N.; Hendrickson, J.; Archer, D. Farbkalibrierung digitaler Bilder für Landwirtschaft und Andere Anwendungen.** ISPRS J. Photogramm. Remote Sens. **2018,** 146, **221–234. [CrossRef]**
6. **Kim, M.; Kim, B.; Park, B.; Lee, M.; Won, Y.; Kim, CY; Lee, S. Ein digitales Farbabstimmungsgerät zur Bestimmung der Zahnfarbe mithilfe des Support Vector Machine-Algorithmus.** Sensors **2018,** 18, **3051. [CrossRef]**
7. **Karaimer, HC; Brown, MS Verbesserung der Farbwiedergabegenauigkeit bei Kameras. In Proceedings der IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake, UT, USA, 18.–23. Juni 2018; S. 6440–6449.**
8. **Roy, S.; Kumar Jain, A.; Lal, S.; Kini, J. Eine Studie über Farbnormalisierungsmethoden für histopathologische Bilder.** Micron **2018,** 114, **42–61. [CrossRef]**
9. **Zhao, Y.; Ferguson, S.; Zhou, H.; Elliott, C.; Rafferty, K. Farbausrichtung für relative Farbkonstanz über nicht-standardisierte Referenzen.** IEEE Trans. Image Process. **2022,** 31, **6591–6604. [CrossRef]**
10. **Rashid, F.; Jamayet, NB; Farook, TH; AL-Rawas, M.; Barman, A.; Johari, Y.; Noorani, TY; Abdullah, JY; Eusufzai, SZ; Alam, MK Farbvariationen während der digitalen Bildgebung von Gesichtsprothesen, die ungefiltertem Umgebungslicht und Bildkalibrierungstechniken in Zahnkliniken ausgesetzt sind: Eine In-vitro-Analyse.** PLoS ONE **2022,** 17, **e0273029. [CrossRef]**
11. Barbero-Álvarez, **MA; Rodrigo, JA; Menéndez, JM Adaptive RGB-Kalibrierung mit minimalem Fehler im Kontext kolorimetrischer Unsicherheit zur Erhaltung des kulturellen Erbes.** Comput. Vis. Image Underst. **2023,** 237, **103835. [CrossRef]**

Machine Translated by Google

12. **Noor Azhar, M.; Bustam, A.; Naseem, FS; Shuin, SS; Md Yusuf, MH; Hishamudin, NU; Poh, K. Verbesserung der Zuverlässigkeit der Smartphone-basierten Urinkolorimetrie mithilfe einer Farbkartenkalibrierungsmethode.** Digit. Health **2023, 9, 20552076231154684.** [Querverweis]
13. **Zhang, G.; Song, S.; Panescu, J.; Shapiro, N.; Dannemiller, KC; Qin, R. Eine neuartige Systemlösung für genaue kolorimetrische Messung durch Smartphone-basierte Augmented Reality.** PLoS ONE **2023,** 18, **e0287099.** [CrossRef]
14. **Chairat, S.; Chaichulee, S.; Dissaneewate, T.; Wangkulangkul, P.; Kongpanichakul, L. KI-gestützte Beurteilung von Wundgewebe mit automatischer Farb- und Maßkalibrierung auf mit einem Smartphone aufgenommenen Bildern.** Healthcare **2023,** 11, **273.** [CrossRef]
15. **Suominen, J.; Egiazarian, K. Kamerafarbkorrektur mit Splines. In Proceedings des IS&T International Symposium on Electronic Imaging, Burlingame, CA, USA, 21.–25. Januar 2024; S. 165-1–165-6.**
16. **Souissi, M.; Chaouch, S.; Moussa, A. Farbabstimmung von Bikomponentenfilamenten (PET/PTT) mit hoher Leistung unter Verwendung eines genetischen Algorithmus.** Sci. Rep. **2024,** 14, **10949.** [CrossRef]
17. **Wannasin, D.; Grossmann, L.; McClements, DJ Optimierung des Aussehens pflanzlicher Lebensmittel durch natürliche Pigmente und Farbabgleichstheorie.** Food Biophys. **2024,** 19, **120–130.** [CrossRef]
18. **Wu, Y. Referenzbildgestütztes Farbabgleichdesign basierend auf einem interaktiven genetischen Algorithmus.** J. Electr. Syst. **2024,** 20, **400–410.** [Querverweis]
19. **Ernährungs- und Landwirtschaftsorganisation; Weltgesundheitsorganisation.** Vitamin- und Mineralstoffbedarf in der menschlichen Ernährung, **2. Auflage; FAO/WHO: Genf, Schweiz, 1998; S. 1–20, ISBN 9241546123.**
20. **Ernährungs- und Landwirtschaftsorganisation; Weltgesundheitsorganisation.** Vitamin- und Mineralstoffbedarf des Menschen; **FAO/WHO: Genf, Schweiz, 2001.**
21. **Lykkesfeldt, J. Über die Wirkung der Vitamin-C-Einnahme auf die menschliche Gesundheit: Wie man die klinischen Beweise (falsch) interpretiert.** Redox Biol. **2020,** 34, **101532.** [CrossRef]
22. **Dosed, M.; Jirkovsk, E.; Kujovsk, L.; Javorsk, L.; Pourov, J.; Mercolini, L.; Remi, F. Vitamin C – Quellen, physiologische Rolle, Kinetik, Mangel, Verwendung, Toxizität und Bestimmung.** Nutrients **2021,** 615, **1–34.**
23. **Suntornsuk, L.; Gritsanapun, W.; Nilkamhank, S.; Paochom, A. Quantifizierung des Vitamin-C-Gehalts in Kräutersaft mittels direkter Titration.** J. Pharm. Biomed. Anal. **2002,** 28, **849–855.** [CrossRef]
24. **Klimczak, I.; Gliszczy ´nska-Swigÿo, A. Vergleich von UPLC- und HPLC-Methoden zur Bestimmung von Vitamin C.** Lebensmittelchemie. **2015,** 175, **100–105.** [CrossRef]
25. **Gazdik, Z.; Zitka, O.; Petrlova, J.; Adam, V.; Zehnalek, J.; Horna, A.; Reznicek, V.; Beklova, M.; Kizek, R. Bestimmung von Vitamin C (Ascorbinsäure) mittels Hochleistungsflüssigkeitschromatographie gekoppelt mit elektrochemischer Detektion.** Sensors **2008, 8, 7097–7112.** [CrossRef]
26. **Bunaciu, AA; Bacalum, E.; Aboul-Enein, HY; Udristioiu, GE; Fleschin, ¸S. FT-IR-spektrophotometrische Analyse von Ascorbinsäure und Biotin und deren pharmazeutischen Formulierungen.** Anal. Lett. **2009,** 42, **1321–1327.** [CrossRef]
27. **Zhu, Q.; Dong, D.; Zheng, X.; Song, H.; Zhao, X.; Chen, H.; Chen, X. Chemilumineszenzbestimmung von Ascorbinsäure unter Verwendung von Graphenoxid@kupferbasierten Metall-organischen Gerüstverbindungen als Katalysator.** RSC Adv. **2016, 6, 25047–25055.** [CrossRef]
28. **Berg, RW Untersuchung von L (+)-Ascorbinsäure mit Raman-Spektroskopie im sichtbaren und UV-Licht.** Appl. Spectrosc. Rev. **2015,** 50, **193–239.** [CrossRef]
29. **Yang, H.; Irudayaraj, J. Schnelle Bestimmung von Vitamin C durch NIR-, MIR- und FT-Raman-Techniken.** J. Pharm. Pharmacol. **2010,** 54, **1247–1255.** [CrossRef]
30. **Zykova, EV; Sandetskaya, NG; Ostrovskii, OV; Verovskii, VE Analysemethoden und Prozesskontrolle Bestimmung von Ascorbinsäure in Arzneimitteln mittels Kapillarzonenelektrophorese und Micellar.** Pharm. Chem. J. **2010,** 44, **463–465.** [CrossRef]
31. **Dodevska, T.; Hadzhiev, D.; Shterev, I. Eine Übersicht über elektrochemische Mikrosensoren zur Ascorbinsäure-Erkennung: Klinische, Pharmazeutische und Lebensmittelsicherheitsanwendungen.** Micromachines **2023,** 14, **41.** [CrossRef]
32. **Huang, L.; Tian, S.; Zhao, W.; Liu, K.; Guo, J. Elektrochemische Vitaminsensoren: Eine kritische Überprüfung.** Talanta **2021,** 222, **121645.** [Querverweis]
33. **Broncová, G.; Prokopec, V.; Shishkanova, TV Potentiometrische elektronische Zunge für pharmazeutische Analytik: Bestimmung von Ascorbinsäure basierend auf elektropolymerisierten Filmen.** Chemosensors **2021, 9, 110.** [CrossRef]
34. **Coutinho, MS; Morais, CLM; Neves, ACO; Menezes, FG; Lima, KMG Kolorimetrische Bestimmung von Ascorbinsäure basierend auf ihrer störenden Wirkung bei der enzymatischen Analyse von Glukose: Ein Ansatz unter Verwendung der Smartphone-Bildanalyse.** J. E Chem . Soc. **2017,** 28, **2500–2505.**
35. **Porto, ISA; Santos Neto, JH; dos Santos, LO; Gomes, AA; Ferreira, SLC Bestimmung von Ascorbinsäure in natürlichen Fruchtsäften mittels digitaler Bildkolorimetrie.** Microchem. J. **2019,** 149, **104031.** [CrossRef]
36. **Kong, L.; Gan, Y.; Liang, T.; Zhong, L.; Pan, Y.; Kirsanov, D.; Legin, A.; Wan, H.; Wang, P. Ein neuartiges Smartphone-basiertes CD-Spektrometer zur hochempfindlichen und kostengünstigen kolorimetrischen Erkennung von Ascorbinsäure.** Anal. Chim. Acta **2020,** 1093, **15** [Querverweis]
37. **Li, C.; Xu, X.; Wang, F.; Zhao, Y.; Shi, Y.; Zhao, X.; Liu, J. Tragbare Smartphone-Plattform mit integriertem Papierstreifen-unterstützten Fluoreszenzsensor zur ultrasensitiven und visuellen Quantifizierung von Ascorbinsäure.** Food Chem. **2023,** 402, **134222.** [CrossRef]
38. **Zhaoa, W.; Caoa, P.; Zhua, Y.; Liua, S.; Gaob, H.-W.; Huang, C. Schnelle Erkennung des Vitamin C-Gehalts in Obst und Gemüse Verwendung einer Digitalkamera und Farbreaktion.** Quim. Nov. **2020,** 43, **1421–1430.** [CrossRef]

Machine Translated by Google

39. Dumancas, GG; Ramasahayam, S.; Bello, G.; Hughes, J.; Kramer, R. Chemometrische Regressionstechniken als neue, leistungsstarke Werkzeuge in genetischen Assoziationsstudien. TrAC Trends Anal. Chem. **2015,** 74, **79–88.** [CrossRef]

40. Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: Ein grundlegendes Werkzeug der Chemometrie. Chemom. Intel. Labor. Syst. **2001,** 58, **109–130.** [Querverweis]

41. Li, B.; Morris, J.; Martin, EB Modellauswahl für partielle Kleinstquadrate-Regression. Chemom. Intell. Lab. Syst. **2002,** 64, **79–89.** [Querverweis]

42. Smola, AJ; Schokopf, B. Ein Tutorial zur Support Vector Regression. Stat. Comput. **2004,** 14, **199–222.** [CrossRef]

43. Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning; **MIT Press: Cambridge, MA, USA, 2016; ISBN 9780262035613.**

44. Lathuiliere, S.; Mesejo, P.; Alameda-Pineda, X.; Horaud, R. Eine umfassende Analyse der tiefen Regression. IEEE Trans. Pattern Anal. Mach. Intell. **2020,** 42, **2065–2081.** [CrossRef]

45. Pascual, L.; Gras, M.; Vidal-Brotóns, D.; Alcañiz, M.; Martínez-Máñez, R.; Ros-Lis, JV Ein voltammetrisches E-Tongue-Tool zur Emulation der sensorischen Analyse und Unterscheidung von Pflanzenmilch. Sens. Actuators B Chem. **2018,** 270, **231–** [Querverweis]

46. Wójcik, S.; Ciepiela, F.; Jakubowska, M. Computer Vision-Analyse von Probenfarben im Vergleich zu Quadruple-Disk Iridium-Platinum Voltammetric e-Tongue zur Erkennung von Verfälschungen von natürlichem Honig. Meas. J. Int. Meas. Confed. **2023,** 209, **112514.**

47. Lee, M. Mathematische Analyse und Leistungsbewertung der GELU-Aktivierungsfunktion im Deep Learning. J. Math. **2023, 1, 4229924.** [CrossRef]

48. Sikandar, S.; Mahum, R.; Alsalman, AM Ein neuartiger hybrider Ansatz für eine inhaltsbasierte Bildsuche mittels Merkmalsfusion. Appl. Sci. **2023,** 13, **4581.** [CrossRef]

49. Hasan, MA; Haque, F.; Sabuj, SR; Sarker, H.; Goni, MOF; Rahman, F.; Rashid, MM Ein End-to-End-leichtgewichtiges Multiskalen- CNN zur Klassifizierung von Lungen- und Dickdarmkrebs mit XAI-Integration. Technologies **2024,** 12, **56.** [CrossRef]

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Artikel

Fallstricke bei der metaheuristischen Lösung stöchiometrischer Optimierungsmodelle für metabolische Netzwerke

Mónica Fabiola Briones-Báez [1,*], Luciano Aguilera-Vázquez [1], Nelson Rangel-Valdez [1], Cristal Zuñiga [2,*], Ana Lidia Martínez-Salazar [1] und Claudia Gomez-Santillan

[1] División de Estudios de Posgrado e Investigación, Instituto Tecnológico de Ciudad Madero (TECNM), Los Mangos 89440, Mexiko; 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.m

[2] Institut für Biologie, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA

* Korrespondenz: d00070469@cdmadero.tecnm.mx (MFB-B.); czuniga2@sdsu.edu (CZ)

Zusammenfassung: Die Flux Balance Analysis (FBA) ist eine auf Einschränkungen basierende Methode, die häufig verwendet wird, um Metaboliten durch restriktive Pfade zu leiten, die oft Bedingungen wie anaplerotische Zyklen wie Calvin, reversible oder irreversible Reaktionen und Knotenpunkte an Stoffwechselpfaden beinhalten, an denen sich diese verzweigen. Die Methode kann die besten Bedingungen für einen Kurs ermitteln, versagt jedoch bei der Behandlung der Pfade mehrerer relevanter Metaboliten. Neuere Studien zum Stoffwechsel halten es für natürlicher, mehrere Metaboliten gleichzeitig zu optimieren, als nur einen; darüber hinaus weisen sie auf die Verwendung von Metaheuristiken als attraktive Alternative hin, die die FBA erweitert, um mehrere Ziele anzugehen. Die Literatur warnt jedoch auch davor, solche Techniken wild einzusetzen. Stattdessen müssen sie sorgfältigen Feinabstimmungs- und Auswahlprozessen unterzogen werden, um die gewünschten Ergebnisse zu erzielen. Diese Arbeit analysiert die Auswirkungen auf die Qualität der mit den Algorithmen NSGAII und MOEA/D und mehreren neuartigen Optimierungsmodellen erstellten Pfade; Es werden zwei Fallstudien untersucht, die Pigmentbiosynthese und der Knotenpunkt im Glutamatstoffwechsel der Mikroalge Chlorella vulgaris, **unter drei Kulturbedingungen (autotroph, heterotroph und mixotroph), während gleichzeitig drei Stoffwechselzwischenprodukte als unabhängige Zielfunktionen optimiert werden. Die Ergebnisse zeigen unterschiedliche Leistungen zwischen NSGAII und MOEA/D und zeigen, dass die Auswahl eines Optimierungsmo**

Schlüsselwörter: Zellstoffwechsel; FBA; Mehrzieloptimierung; NSGAII; MOEA/D

check for updates

Zitat: Briones-Baez, MF; Aguilera-Vazquez, L.; Rangel-Valdez, N.; Zuñiga, C.; Martinez-Salazar, AL; Gomez-Santillan, C. Fallstricke bei der Metaheuristik zur Lösung stöchiometrischer Optimierung Modelle für metabolische Netzwerke. Algorithmen **2024,** 17, **336. https://doi.org/10.3390/a17080336**

Wissenschaftliche Redakteure: Massimiliano Caramia und Frank Werner

Erhalten: 30. März 2024

Überarbeitet: 1. Juli 2024

Akzeptiert: 16. Juli 2024

Veröffentlicht: 1. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/4.0/).

1. Einleitung

Mikroalgen sind photosynthetische zelluläre Mikroorganismen, die seit **jeher bekannt sind .** Sie können entweder in Abwasser oder in sauberem oder salzigem Wasser gezüchtet werden; **einige Stämme, wie** Dunaliella salina, **können in salzigem Wasser gezüchtet werden, und andere Stämme, wie Chlorella** vulgaris, **werden in Süßwasser gezüchtet und können hohe Wachstumstemperatu Mikroalgen benötigen eine Kohlenstoffquelle, um Photosynthese zu betreiben und ihre Biomasse zu produzieren. Kohlenstoff kann als CO2 aus umweltschädlichen Quellen gewonnen werden und wird in Sauerstoff umgewandelt, was auf kreisförmiger Basis zur Verringerung der globalen Erwärmung beiträgt [1].**

Mikroalgen sind große Biomasseproduzenten. In ihrem Inneren befinden sich Metaboliten wie Lipide , die in Zukunft als Biokraftstoffe verwendet werden können, Aminosäuren und Pigmente, die derzeit in der Pharma- und Kosmetikindustrie verwendet werden, sowie Proteine, die als Nahrungsergänzungsmittel eingesetzt werden [2,3]. Darüber hinaus bieten Mikroalgen biotechnologische Anwendungen als Bioremediationsquellen für die Wasserqualität und werden als Alternativen zur Entfernung von Schwermetallen eingesetzt, da einige Stämme wie Chlorella vulgaris **und** Scenedesmus obliquos **Schwermetalle wie Cadmium (Cd) u**

Alle oben genannten Eigenschaften machen das Studium des Stoffwechsels von Mikroalgen für das Stoffwechsel-Engineering attraktiv. Diese Disziplin konzentriert sich auf das Studium der Topograph

Machine Translated by Google

Netzwerk, die Regulierung eines Stoffwechselweges, die Identifizierung von Engpässen, die Bestimmung von Stoffwechselflüssen und die Eliminierung von Nebenreaktionen durch Gendeletion [6].

Insbesondere wurden verschiedene Stoffwechseltechniktechniken eingesetzt, um Stoffwechselwege zu analysieren und Flüsse zu optimieren, um den Stoffwechsel zu manipulieren und die Flüsse in Richtung eines gewünschten Produkts zu modifizieren und so einen kommerziellen Mehrwert zu erzielen. Stoffwechselflüsse können jedoch nicht in vivo gemessen werden; aus diesem Grund sind Modellierungsansätze erforderlich, um sie zu messen oder vorherzusagen [7]. Dazu gehören einzielbasierte, auf Beschränkungen basierende Ansätze wie FBA (Flux Balance Analysis), exakte mathematische mehrzielige und heuristikbasierte Ansä

FBA ist eine der am häufigsten verwendeten Techniken zur Untersuchung des Zellstoffwechsels und basiert auf einem auf Einschränkungen basierenden Einzelziel-FBA-Ansatz. Dieser Ansatz wird häufig bei der Analyse der Flüsse von Stoffwechselnetzwerken verwendet, da er auch dann verwendet werden kann, wenn keine kinetischen Daten verfügbar sind. Er erfordert jedoch Informationen über die stöchiometrischen Daten der im Netzwerk vorhandenen Reaktionen, Wachstumsanforderungen und parameterspezifische Messmethoden des biologischen Systems, insbesondere die Rekonstruktion des Stoffwechselnetzwerks für die Genomskala [8] , die alle bekannten Reaktionen umfasst, die in den untersuchten Organismen vorhanden sind, und die Gene , die jedes Enzym kodieren [9].

Der Zellstoffwechsel wird in der Stoffwechselmodellierung als die Gesamtheit der chemischen Reaktionen beschrieben, die in einem Organismus vorhanden sind. Dies wird mathematisch durch eine stöchiometrische Matrix S der Größe (m × n) dargestellt, wobei n die Reaktionen und m die an jeder Reaktion beteiligten Metaboliten sind . Dabei wird ein negativer Koeffizient zugewiesen, wenn es sich um einen Reaktanten handelt, und ein positiver Koeffizient, wenn es sich um ein Produkt handelt. Ein Koeffizient von Null bedeutet, dass der Metabolit in der Reaktion nicht vorhanden ist. Jede Reaktion hat eine Unter- und eine Obergrenze, die den Lösungsraum oder den Maximal- und Minimalwert des zulässigen Flusses

lineare Optimierung einer Zielfunktion. Diese Funktion stellt die lineare Kombination der Flüsse dar, die im Allgemeinen die Biomasseproduktion repräsentieren [10].

$$\begin{aligned} &\textbf{FBA max } F(v) = v\text{Biomasse} \\ &\quad \textbf{Vorbehaltlich} \\ &\qquad \mathbf{S} \cdot v = \mathbf{0} \\ &\qquad \mathbf{LBj} \ddot{y} vj \ddot{y} \text{UBj} , \ddot{y}j \ddot{y} \{1, \ldots, N\} \end{aligned} \tag{1}$$

Gleichung (1) definiert das zugehörige lineare Optimierungsproblem FBA [11], wobei v der Flussvektor über die Reaktionen ist . Die stöchiometrische Matrix Sm×n stellt das metabolische Netzwerk dar, in dem pro Zeile ein Metabolit und pro Spalte eine Reaktion vorhanden ist. Der Wert der Zelle Sij ist der stöchiometrische Koeffizient des an Reaktion j beteiligten Metaboliten i [9], und LBj und UBj sind die Unter- und Obergrenzen für die im metabolischen System zulässigen Flüsse. Die Annahme eines stationären Zustands wird durch Sv = 0 [12] festgelegt.

Aufgrund seiner Vielseitigkeit wird FBA häufig bei verschiedenen Organismen eingesetzt, darunter Mikroalgen, Bakterien, Mikroorganismenkonsortien usw. Ein Beispiel hierfür ist die Vorhersage des Zellwachstums der Cyanobakterien Synechocystis in [13]. Außerdem wird es beim Abbau von Glucose durch anaerobe Vergärung eingesetzt, um die Verteilung von Metaboliten vorherzusagen und die Umwandlung von Kohlenstoff aufzudecken, um die Umwandlung von Ethanol, Propionsäure und Buttersäure in Essigsäure zu bewerten [14]. FBA wurde in der Medizin als Studienmaterial verwendet, bei denen Flussaktivitäten berechnet wurden, um Unterschiede in Stoffwechselwegen zu untersuchen und Brustkrebssubtypen zu vergleichen [15].

Die FBA-Methode wurde verwendet, um Stoffwechselflüsse in verschiedenen Mikroalgenstämmen unter drei heterotrophen und mixotrophen autotrophen Wachstumsbedingunge Die erste Mikroalge, die FBA verwendete, war die Mikroalge Chlamydomonas reinhardtii [16], mit der die erste Stoffwechselkarte erstellt wurde. Später wurde die Mikroalge Chlorella vulgaris zur Untersuchung der Lipidproduktion verwendet [17] . Beide hatten die Biomasseproduktion als einzige Zielfunktion.

Obwohl dieser Ansatz weit verbreitet ist, wird die Verteilung der Metaboliten durch die Stoffwechselwege durch die in der Zielfunktionsgleichung und in den experimentellen Parametern vorhandenen Metaboliten bestimmt; die Blockierung einiger Stoffwechselwege führt zu

Machine Translated by Google

eine Verteilung von Flüssen mit Nullwerten. Darüber hinaus wird FBA häufig verwendet, um die Produktion von interessanten Verbindungen zu maximieren, aber der Zellstoffwechsel in seinem natürlichen Zustand lenkt Stoffwechselwege nicht in Richtung der Produktion eines bestimmten Metaboliten . Auf der Suche nach einem besseren Verständnis des Mikroalgenstoffwechsels sind neue Optimierungstechniken entstanden, wie beispielsweise Metaheuris Diese Techniken streben eine gleichmäßigere Verteilung an und kommen der Realität des Geschehens auf Stoffwechselebene näher, indem sie verschiedene Funktionen mit widersprüchlichen Zielen gleichzeitig optimieren.

Die Mehrzieloptimierung basiert im Allgemeinen auf der Suche nach Lösungen für verschiedene widersprüchliche Ziele, die gleichzeitig optimiert werden müssen. Die Mehrzieloptimierung ist von großer Bedeutung und wurde auf technologischer und wissenschaftlic Einige Beispiele aus der chemischen Industrie sind die Optimierung von Betriebsabläufen , Bioraffinerien, Reaktionstechnik, Prävention und Kontrolle usw. [19]. Auch im Bereich Biologie und Medizin [20] sowie in der Stoffwechseltechnik [18] wird das Verfahren eingesetzt.

Die Mehrzieloptimierung steht im Gegensatz zum frei zugänglichen Cobrapy FBA, bei dem nur eine Zielfunktion maximiert oder minimiert wird und nur eine Lösung erzielt wird. Bei der Mehrzieloptimierung erhält man eine Menge von Lösungen. Die erhaltenen Lösungen heißen nichtdominiert, weil keine andere Lösung im Suchraum besser ist als die anderen, wenn alle Ziele gleichzeitig berücksichtigt werden. Diese Menge von Lösungen nennt man Pareto- optimale Lösungen [21].

Zu den Methoden, die für die multiobjektive Optimierung von Mikroorganismen verwendet wurden, gehören evolutionäre Algorithmen wie NSGAII [18], MOMO, basierend auf dem Bio-Objektiv-Modell, und exakte mathematische Methoden, die einen großen Rechenaufwand erfordern [22].

Metaheuristische Algorithmen haben ihren Ursprung in der natürlichen Evolution biologischer Gruppen; sie sind Teil der künstlichen Intelligenz und sind eine Weiterentwicklung des Natural Computing und heuristischer Methoden (partielle Suchalgorithmen). Im Vergleich zu mathematischen Methoden, die einen hohen Rechenaufwand verursachen, liefern diese Methoden **bei akzeptablem Rechenaufwand und -platz ausreichend gute Lösungen für** einen Optimierungsalgorithm

Multi-Objective Evolutionary Algorithms (MOEAs) werden in der wissenschaftlichen Gemeinschaft allgemein als Ansatz zur Lösung multi-objektiver Optimierungsprobleme anerkannt. Insbesondere hat sich das von [24] vorgeschlagene NSGAII (multi-objective EA based on non-dominated classification) **bei der Behandlung von zwei oder drei Zielen als recht effektiv erwiesen [25,26] Der auf Zerlegung basierende Evolutionsalgorithmus MOEA/D wird getestet, da er die Eigenschaft besitzt, auch bei mehr als drei Zielfunktionen korrekt zu funktionieren.**

Zuvor wurden metaheuristische Methoden zur Untersuchung des Stoffwechsels in [18] verwendet. Der NS- GAII-Algorithmus wurde entwickelt, um drei Zielfunktionen, Proteine, Kohlenhydrate und CO2, in einem Stoffwechselnetzwerk der Mikroalge Chlamydomonas reinhardtii **unter Verwendung des NSGAII-Algorithmus, einem auf Flussbilanzanalyse (FBA) basierenden Kodierungsschema, zu optimieren. Die algorithmische Lösung kann jedoch für jedes Optimierungsproblem unterschiedlich sein, ein Unterschied, der sich vergrößern kann, je nachdem, wie gut d Stoffwechselnetzwerke sind von solchen Problemen möglicherweise nicht ausgenommen, und diese Arbeiten analysieren verschiedene Algorithmen (NSGAII und MOEA/D) und verschiedene Optimierungsmodelle für Stoffwechselnetzwerke (vier multiobjektive Optimierungsansätze) im Hinblick auf die Qualität der Lösungen, die mit ihnen erreicht werden können, und die Zweckmäßigkeit der be Die Studie wird anhand von zwei Fallstudien durchgeführt, die komplizierte Bedingungen mit Zyklen, Verzweigungen sowie reversiblen und irreversiblen Reaktionen beinhalten. Die wichtigsten Beiträge dieser Arbeit umfassen die Entwicklung von drei neuen Mehrzieloptimierungsmodellen für Stoffwechselnetzwerke, eine neue algorithmische Lösung auf Basis von Zerlegung und eine Analyse, die die richtige Identifizierung von Metaheuristiken und Modellen zur Lösung des Optimierungsprozesses hinter Stoffwechseln unterstützt.**

Der Rest des Dokuments ist wie folgt strukturiert. Abschnitt 2 beschreibt die Metaheuristiken, die für die Analyse in dieser Forschung verwendet werden; insbesondere werden die allgemeine Definition und die Bestandteile vorgestellt. Abschnitt 3 beschreibt die neuen Optimierungsmodelle für die

Machine Translated by Google

in dieser Arbeit vorgeschlagene Stoffwechselnetzwerke. Abschnitt 4 erörtert die ursprünglichen Merkmale, die im Design von NSGAII und MOEA/D enthalten sind, die zur Lösung der vorgeschlagenen Optimierungsmodelle verwendet wurden. In den Abschnitten 5 und 6 wird das Design der Experimente beschrieben, die zur Prüfung der vorgeschlagenen Optimierungsmodelle und ihrer metaheuristischen Lösungen durchgeführt wurden. Sie enthalten die Definitionen der untersuchten Fälle zusammen mit den Experimenten und den Ergebnissen und schließen mit einer kurzen Diskussion der beobachteten Daten.

2. Materialien und Methoden

In dieser Arbeit werden NSGAII- und MOEA/D-Algorithmen als Metaheuristiken zur Lösung des Optimierungsproblems in metabolischen Netzwerken betrachtet. Die allgemeine Vorstellung des Entwurfs solcher Algorithmen wird im Rest dieses Abschnitts erläutert.

2.1. NSGAII-Algorithmus

NSGAII ist ein Multi-Objective Evolutionary Algorithm (MOEA), der nichtdominierte Sortierung und Crowding-Distanz nutzt, um Selektionsdruck in Richtung der Pareto-Front auszuüben [24]. Die Metaheuristik entwickelt eine anfängliche Population von P0- Eltern unter Verwendung allgemeiner berechenbarer genetischer Operatoren wie Mutation, Crossover und Turnierauswahl, die in jeder Generation **einen neuen Nachkommen,** Qt , **erzeugen. Um den Elitismus aufrechtzuerhalten, wird die aktuelle Population,** Pt , **mit dem Nachkommen** Qt kombiniert , **und eine neue Population** Pt+1 **wird basierend auf Nichtdominanzrängen und Crowding-Distanz ausgewählt, um zu diversifizieren und bei Bedarf Gleichstände aufzulösen.**

Die drei Hauptkomponenten von NSGAII sind sein schneller nicht dominierter Sortieransatz, das schnelle, überfüllte Distanzschätzungsverfahren und der einfache überfüllte Vergleichsoperator . Die aus NSGAII abgeleitete allgemeine Methode kann wie folgt dargestellt werden:

1. **Initialisierung einer Population** Po **der Größe** N **unter Verwendung einer gleichmäßigen Verteilung.**
2. **Erstellen Sie eine Nachkommenpopulation** Qt **mithilfe einer binären Turnierauswahl basierend auf einem Crowding-Vergleichsoperator, Crossover und einer an der Elternpopulation (Pt) durchgeführten Mutation ,** wobei **der Index** „t“ **die Anzahl der Generationen bezeichnet. Die Nachkommenpopulation und ihre Elternpopulation werden kombiniert, um die Gesamtpopulation** Rt zu erzeugen , Rt **=** Pt **+** Qt . **Die Population** Rt **hat eine Größe von 2N.**
3. **Führen Sie einen schnellen nichtdominierten Sortieransatz für die gesamte Population** Rt durch, um **verschiedene Fronten der Zielfunktionen zu identifizieren .** F = schnell **ÿ** nichtdominiert **ÿ** sort(Rt), **wobei** F = **(F0, F1, F2, ...)** in **F0 die nichtdominierte Menge an Lösungen von** Rt **enthält , die die Pareto-Grenze am besten approximiert.**
4. Konstruieren Sie **aus den erhaltenen Fronten** (Fi) **eine neue Elternpopulation** (Pt+1) der Größe N. **Diese Population der Größe N wird nun für Selektion, Crossover und Mutation verwendet, um eine neue Population** (Qt+i) **der Größe N** zu erstellen.
5. **Der Vorgang muss wiederholt werden, bis die maximale Anzahl an Iterationen erreicht ist.**

2.2. MOEA/D

MOEA/D ist eine Strategie, die auf der Zerlegung des multiobjektiven Optimierungsproblems MOP (wie in [27] definiert) in eine bestimmte Anzahl skalarer Optimierungsunterprobleme basiert, die gleichzeitig optimiert werden. Jedes Unterproblem wird unter Verwendung von Informationen optimiert, die ausschließlich aus seinen benachbarten Unterproblemen stammen, wodurch in jeder Generation eine geringere Rechenkomplexität erreicht wird . Es gibt mehrere Ansätze, um ein multiobjektives Problem in eine skalare Anzahl von Optimierungsproblemen umzuwandeln. Einer der beliebtesten Ansätze ist das von **[28] vorgeschlagene MOEA/D ,** bei dem die skalaren Optimierungsprobleme **wie folgt formuliert werden können :**

$$G^{te}\ (y \mid wi\ ,\ z) = \mathbf{max\{wi \mid}\ fk\ (x)\ \ddot{y}\ zk\ \mathbf{|\}}, \quad \textbf{Betrifft:}\ x\ \ddot{y}\ \ddot{y} \tag{2}$$

wobei w = **(w1, w2, . . . ,** wN) **ein Vektor von Gewichten ist und** wi **ÿ 0 für alle** i = **1, . . . , k, und die Menge** z = **{z1, z2, . . . ,** zk} **der Referenzpunkt ist, wobei** zi = max{ fi(x) **| × ÿ ÿ} für** i = **1, . . . ,** **M.**

Machine Translated by Google

Für jeden Pareto-optimalen Punkt x ý, **Es gibt einen Vektor von Gewichten w, wobei** x ý **ist die optimale Lösung von 2.**

MOEA/D funktioniert über eine bestimmte Anzahl von Generationen und übt während jeder Generation auch selektiven Druck in Richtung der Pareto-Front aus, indem es genetische Operatoren wie Mutation und Crossover verwendet. Das Schlüsselelement dieser Strategie besteht darin, dass die Nachkommen die Eltern auf der Grundlage der Skalarfunktion und ihrer Nähe bei jeder Iteration ersetzen, bis eine er vom Algorithmus geliefert.

3. Vorgeschlagene Ansätze

Diese Arbeit schlägt drei neue Optimierungsmodelle für Stoffwechselnetzwerke vor, die FBA auf ein Mehrzieloptimierungsproblem erweitern. Die Modelle, MOFBA2, MOFBA3 und MOFBA4 genannt, stellen Verbesserungen gegenüber dem in [18] **vorgeschlagenen und in Gleichung (3) dargestellten MOFBA1 dar. MOFBA1 optimiert gleichzeitig eine Reihe von Bioprodukten** {vb1 , . . . vbm} **statt nur eines, hält die Reaktionsflüsse innerhalb der Grenzen und erfüllt die Steady-State-Bedingung, d. h. sie stellen sicher, dass** Sv **= 0 (wobei** S **die stöchiometrische Matrix und** v **der Flussvektor ist).**

$$\begin{aligned} &\textbf{MOFBA max } F(v) = \{vb1 , \ldots , vbm \} \\ &\quad \textbf{Vorbehaltlich} \\ &\qquad \textbf{S} \cdot v = \textbf{0} \\ &\qquad \textbf{LBj ÿ } vj \textbf{ ÿ UBj} , \ \ddot{y}j \textbf{ ÿ \{1, . . . , N\}} \end{aligned} \tag{3}$$

MOFBA1 ist das Optimierungsproblem, das sich aus der direkten Implementierung des in Gleichung (3) definierten Problems ergibt. Es berücksichtigt so viele Zielfunktionen wie **interessierende Metabolitensätze .** Ebenso berücksichtigt es so viele Entscheidungsvariablen, wie Reaktionsflüsse **zur Definition des Stoffwechselsystems erforderlich sind. Beachten Sie, dass bei einem Optimierungsansatz der Suchraum von den Entscheidungsvariablen abhängt und es basierend auf dieser Definition für jeden möglichen Fluss eine gibt ,** d. h. eine Metaheuristik muss innerhalb der vorgegebenen Grenzen von n **unterschiedlichen Entscheidungsvariablen nach geeigneten Flusswerten suchen .**

Die im vorigen Abschnitt beschriebenen Bedingungen kennzeichnen eine häufige Falle bei der Entwicklung von Lösungsstrategien für Optimierungsprobleme. Die Schwierigkeit liegt darin, dass Metaheuristiken bei **einer großen Anzahl von Entscheidungsvariablen möglicherweise längere Laufzeiten benötigen, um praktikable Lösungen zu finden .** Diese Arbeit berücksichtigt diese Situation und **schlägt drei neue Optimierungsmodelle für metabolische Netzwerke vor, die den Suchraum (d. h. die Anzahl der Entscheidungsvariablen, die eine Metaheuristik bei der Suche verwendet) reduzieren. Diese Modelle werden in ein geeignetes experimentelles Design integriert, um zu zeigen, dass es aus Qualitätsgründen wichtig ist, welches Modell man zur Lösung bestimmter Probleme wählt.**

Während Einzelheiten zu den Experimenten in späteren Abschnitten bereitgestellt werden, enthält der Rest dieses Abschnitts eine ausführliche Beschreibung der drei in dieser Arbeit vorgeschlagenen neuen MOFBAs mit einer Zusammenfassung ihrer Relevanz und Auswirkungen am Ende.

3.1. MOFBA2

In MOFBA2 sind wie in MOFBA1 die zu optimierenden Zielfunktionen die Sätze der interessierenden Metabolite; daher ist die Anzahl der Ziele, m, dieselbe. Andererseits berücksichtigt **MOFBA2 einen reduzierten Satz von Entscheidungsvariablen, der nur aus den Reaktionsflüssen** {vb1 , ... vbm} besteht , **die mit den gleichen interessierenden Metaboliten in der Zielfunktion verbunden sind , und einer zusätzlichen Variable, v, die angibt, welcher der interessierenden Metabolite die Suche leitet. Mit anderen Worten, MOFBA2 hat** m + **1 Entscheidungsvariablen statt n. Gleichung (4) definiert MOFBA2 formal.**

$$\begin{aligned} &\textbf{MOFBA max } F(v) = \{vb1 , \ldots , vbm \} \\ &\quad \textbf{Vorbehaltlich} \\ &\qquad \text{FBA(vbk, } \textbf{v)ist möglich} \\ &\qquad \textbf{LBj ÿ } vbi \ \ddot{y} \textbf{ UBj , 1 ÿ } \ddot{y}i \textbf{ ÿ \{1, . . . , m\}} \\ &\qquad k \textbf{ ÿ } m \end{aligned} \tag{4}$$

Machine Translated by Google

MOFBA2 beschreibt ein zweistufiges Optimierungsmodell, bei dem das innere Modell den führenden Metaboliten von Interesse, v, mithilfe von FBA optimiert und die Grenzen der interessierenden Metaboliten auf die im äußeren Modell definierten begrenzt. Die Grenzen der verbleibenden Reaktionen werden als bekannt angenommen und sind gemäß dem analysierten Stoffwechs

3.2. MOFBA3

MOFBA3 schlägt ein Ersatzmodell zur Optimierung der Metabolite von Interesse vor. Das Ersatzmodell sucht nach der Verbesserung zweier bekannter Indikatoren: dem Hypervolumen (HV) und der Generationsdistanz (GD). Diese Indikatoren spiegeln wider, wie gut eine Lösung zur Pareto-Front konvergiert. Während das HV maximiert werden muss, muss die GD minimiert werden

MOFBA3 hat m + **1 Entscheidungsvariablen, dieselben wie MOFBA2, d. h. die Reaktionsflüsse { vb1 , ...** vbm}, **die mit denselben Metaboliten von Interesse in der Zielfunktion verbunden sind , und den führenden Metabolitenfluss v. Andererseits ist die Anzahl der Ziele immer 2, unabhängig davon, wie viele Metaboliten von Interesse berücksichtigt werden. Das besondere Merkmal dieses Modells ist seine Surrogatisierung; anstatt direkt nach den richtigen Flusswerten für die Metaboliten von Interesse zu suchen, verwendet es Leistungsindikatoren im Kontext mehrerer Ziele (d. h. die HV- und GD-Indikatoren). Gleichung (5) definiert MOFBA3 formal.**

$$\begin{aligned} &\textbf{MOFBA min } F(v) = ÿ1HV(\{vb1, \ldots, vbm\}, R), DG(\{vb1, \ldots, vbm\}, Zr) \\ &\quad \textbf{Vorbehaltlich} \\ &\qquad \text{FBA(vbk, } \textbf{v)ist möglich LBj} \\ &\qquad ÿ \text{ vb i } ÿ \text{ UBj , ÿi } ÿ \textbf{ \{1, . . . , m\} 1 } ÿ \text{ k } ÿ \text{ m} \end{aligned} \tag{5}$$

Dieses Optimierungsproblem setzt voraus, dass ein Referenzpunkt R und ein Referenzsatz Zr vorhanden sind . **Da FBA unter dem COBRApy-Paket verwaltet wurde und die Einschränkungen des Pakets berücksichtigt wurden, ist der für diese Arbeit berücksichtigte Referenzpunkt** R = **{1000, 1000, ... , 1000}. Angesichts der Verfügbarkeit einer FBA-Implementierung aufgrund desselben Pakets wird der Referenzsatz** Zr **durch den Satz optimaler Lösungen gebildet, die durch die Lösungen gebildet werden, die bei der optimalen Lösung von FBA erzielt werden, wobei jeder interessierende Metabolit als optimierte Biomasse gilt.**

3.3. MOFBA4

MOFBA4 ist das letzte vorgeschlagene Optimierungsproblem, das die Ideen von MOFBA2 und MOFBA3 kombiniert. Das heißt, es schlägt vor, nicht nur die Metabolite von Interesse, sondern auch die Indikatoren HV und GD zu optimieren. Die Anzahl der Entscheidungsvariablen für dieses Modell bleibt m + **1, und die Anzahl der Ziele ist** m + **3. Gleichung (6) definiert dieses Optimierungsproblem formal.**

$$\begin{aligned} &\textbf{MOFBA min } F(v) = ÿ1HV(\{vb1, \ldots, vbm\}, R), DG(\{vb1, \ldots, vbm\}, Zr)(vb1, \ldots, vbm) \\ &\quad \textbf{Vorbehaltlich} \\ &\qquad \text{FBA(vbk, } \textbf{v)ist möglich} \\ &\qquad \textbf{LBj } ÿ \text{ vbi } ÿ \textbf{ UBj , } \quad ÿi \ ÿ \textbf{ \{1, . . . , m\}} \\ &\qquad \textbf{1 } ÿ \text{ k } ÿ \text{ m} \end{aligned} \tag{6}$$

3.4. Analyse

Tabelle 1 fasst die bemerkenswertesten Merkmale der in dieser Arbeit vorgeschlagenen Optimierungsmodelle zusammen und vergleicht sie mit dem in [18] vorgeschlagenen MOFBA1. Der Suchraum ist in den neuen Modellen stark reduziert, und einige von ihnen verwenden Konvergenzinformationen in ihrer Definition. Die einzigartigen Merkmale zeigen den Reichtum der Modelle, die zur Lösung eines bestimmten Optimierungsproblems entwickelt werden können.

Machine Translated by Google

Tabelle 1. Relevante Merkmale von MOFBA-Modellen für metabolische Netzwerke.

Modell	Nr. Entscheidungsvariablen	Nr. Ziele	Surrogat
MOFBA1	N	M	**NEIN**
MOFBA2	m + **1**	M	**NEIN**
MOFBA3	m + **1**	**2**	**Ja**
MOFBA4	m + **1**	m + **2**	**Ja**

Die vorgeschlagenen MOFBAs können nicht mit herkömmlichen linearen Lösern wie FBA gelöst werden. Die Alternativen sind die Verwendung von Aufzählungsschemata oder Näherungsansätzen, die es ermöglichen Lösungen zu erhalten, die der Pareto-optimalen Grenze angehören. In diesem Sinne ist diese Forschung analysierten den Einsatz von Metaheuristiken, die FBA in ihren Suchprozess integrieren, als geeigneten Lösungsansatz, da sie ihre Annäherung an die Pareto-Front verbessern in jede Iteration.

4. Metaheuristische Designs

In diesem Abschnitt werden die besonderen Details vorgestellt, die in dieser Forschung für die Implementierung der NSGAII- und MOEA/D-Metaheuristiken zur Lösung der vier Optimierungsaufgaben erforderlich sind. Probleme MOFBA1, MOFBA2, MOFBA3 und MOFBA4. Diese Metaheuristiken basieren auf die NSGAII- und MOEA/D-Frameworks.

Die betrachteten Metaheuristiken erfordern die Definition folgender Eigenschaften: (1) Kodierungsschemata; (2) Fitnessbewertungsfunktion; (3) genetische Operatoren; und (4) Einschränkung Managementstrategie. Die Populationsinitialisierungsmethode für beide Strategien (NSGAII und MOEA/D) ist zufällig. Das vorgeschlagene Design für den Rest dieser Komponenten zur Handhabung Die neuen MOFBAs werden im Rest dieses Abschnitts ausführlich beschrieben. Die neuen Anpassungen für Die NSGAII- und MOEA/D-Frameworks umfassen die clevere berechenbare Darstellung von Lösungen im Zusammenhang mit den Kodierungsschemata.

4.1. Kodierungsschemata

Diese Arbeit schlägt die Verwendung von unterschiedlichen Lösungscodierungssätzen für jedes MOFBA vor (wie definiert in den Gleichungen (3)–(6). Die Kodierungsschemata beinhalten die Definition einer Datenstruktur das eine Lösung des metabolischen Netzwerks darstellt. Das für Experimente entwickelte Skript befindet sich in diesem Github-Repository (https://github.com/multiobjectiveoptimization2 /MOFBAs, abgerufen am 24. Juni 2024).

Für MOFBA1 ist die Datenstruktur ein realwertiger Vektor, V. Das Kodierungsschema berücksichtigt ein metabolisches Netzwerk, MN, **das aus einer Menge von Reaktionen, V, und zwei Teilmengen** besteht. **V M, V**^{B} **ÿ V, wobei V** ^{M} **ÿ** Vb = **ÿ, die die Reaktionen der Metaboliten von Interesse für einen Entscheidungsträger. Weiterhin sei** v = **(v1, . . . ,** vn) **der Flussvektor für V** und nehmen wir an, dass es **für jedes** vi , **1 ÿ** i **ÿ n, anfängliche Unter- und Obergrenzen** LBi , UBi gibt. **Dann schlägt das W- Kodierungsschema vor, die Grenzen jedes** vi neu zu definieren , **das mit einer Reaktion in V** ^{M} **ÿ** Vb **mit zwei Werten** (li , ÿi). **Die neuen Grenzen berechnen sich wie folgt:**

LBneu = li **und** UBneu = (UBi **ÿ** li)ÿi + li . **Alle übrigen Flussmittel behalten ihre Grenzen unverändert. Mit anderen Worten, die Lösung kodiert Grenzänderungen für FBA, um** MN zu lösen **unter Verwendung eines vorab festgelegten Bioprodukts, das in dieser Arbeit als** v angenommen wird. ^{B} **Die daraus resultierende Der Kodierungsvektor** W **hat die Größe** O(n), **asymptotisch in der Anzahl der Reaktionen.**

Für MOFBA2, MOFBA3 und MOFBA4 wird **ein Vektor der Größe** m + **1 als Kodierungsschema, wobei** m **die Anzahl der Zielfunktionen ist. Die ersten** m **Elemente des Vektors sind reelle Variablen, deren Wert die Obergrenze des Flusses für jede Metabolit von Interesse in den damit verbundenen Reaktionen. Das zusätzliche Element ist ein Optimierungsselektor, der Werte zwischen 1 und m annimmt und angibt, welcher Metabolit wird wiederum optimiert. Wenn es sich um eine reversible Reaktion handelt, wird der Wert in der angegebenen Die Variable ist für die Untergrenze gleich (aber negativ).**

Machine Translated by Google

4.2. Fitness-Bewertungsfunktion

Bei allen MOFBA-Optimierungsproblemen werden die Fitnessbewertungsfunktionen (oder FEA-Funktionen) als abgeleitete Teilmenge der Menge betrachtet, die aus den Werten der betreffenden Metabolitenflüsse besteht, wie z. B. den Metriken Hypervolumen und Generationsdistanz. Da die erforderlichen Informationen zu Bioprodukten mit bestimmten Reaktionen verknüpft sind, wird die Eignung einer durch Metaheuristiken für MOFBAs erhaltenen Lösung unter Berücksichtigung ihrer Flusswerte bewertet . In MOFBA1 und sind die zu optimierenden Kriterien oder Zielfunktionen b **und de die Reaktionsflüsse,MOFBA2 die den in** $v_{1}^{B}, \ldots, v_{m}^{b}$). **In MOFBA3 werden das Hypervolumen und die Generationsdistanz wie V** ausgewählten **und als** (v) bezeichneten Bioprodukten von Interesse **folgt definiert:** entsprechen , **aus einer Lösung, dem Referenzpunkt R und dem Referenzsatz** Zr , **optimiert. Der** Punkt R **ist der schlechtestmögliche Extremwert der Flüsse, der für jeden interessierenden Metaboliten unter Berücksichtigung seiner Definition für FBA in weit verbreiteten Plattformen wie CobraPy 1000 beträgt. Der Satz** Zr **besteht aus drei Punkten, die die optimalen Flüsse umfassen, die durch Lösen des betreffenden Falls mit der FBA-Methode durch individuelle Optimierung jedes interessierenden Metaboliten erzielt werden; wenn also n Zielfunktionen vorhanden sind, hat** Zr eine Kardinalität von n. Es ist erwähnenswert, dass, wenn in MOFBA2 **bis** MOFBA4 **ein führendes Bioprodukt erforderlich ist , dieses abgeleitet vom Wert einer der berücksichtig**

4.3. Genetische Operatoren

Diese Operatoren erzeugen neue Lösungen, indem sie die Werte der Entscheidungsvariablen in den bestehenden Lösungen dynamisch und zufällig variieren. Diese Auswahl erfolgte aufgrund ihres Erfolgs bei der Lösung von Problemen mit Entscheidungsvariablen mit realen Werten [29]. Die für NSGAII ausgewählten Operatoren sind Mutation, Crossover bzw. eine einfache, aber zuverlässige Zufallsauswahl . Die spezifischen Werte dieser Parameter wurden der Literatur entnommen und sind in Tabelle 2 aufgeführt.

Tabelle 2. Implementierungsspezifische Parameter von NSGAII.

Parameter	Wert
Polynommutation	**Wahrscheinlichkeit = 1,0/d, wobei d die Anzahl der Entscheidungsvariablen ist. Verteilungsindex = 20**
SBXCrossover	**Wahrscheinlichkeit = 100 % Verteilungsindex = 20**
Abbruchkriterium	**bis zum Erreichen von 100.000 Bewertungen**
Bevölkerungsgröße	**100**

Die für MOEA/D für Mutation und Crossover gewählten Operatoren waren Polynommutation [30] und Crossover durch differentielle Evolution. Die Auswahlstrategie ist einfach , aber zuverlässig, und die verwendete Aggregationsfunktion war die Tshebycheff-Distanz. Eine ausführlichere Referenz der Operatoren finden Sie in [28]. Die spezifischen Werte dieser Parameter wurden der Literatur entnommen und sind in Tabelle 3 aufgeführt.

Tabelle 3. Implementierungsspezifische Parameter für MOEA/D.

Parameter	Werte
Polynommutation	**Wahrscheinlichkeit = 1,0/d, wobei d die Anzahl der Entscheidungsvariablen ist. Verteilungsindex = 20**
Differenzielle Evolution	**CR = 1** **F = 0,5** **K = 0,5**
Abbruchkriterium	**Nach Abschluss von 100.000 Bewertungen**
Bevölkerungsgröße	**100**

Machine Translated by Google

4.4. Strategie zur

Einschränkungsverwaltung Diese Arbeit verwendet die in [31] **vorgeschlagene Methode zur Einschränkungsverwaltung , um Selektionsdruck in Richtung realisierbarer Lösungen zu erzeugen. Mit der Entwicklung von Generationen in beiden Metaheuristiken wird der Wettbewerb zwischen Lösungen trotz des Nichtdominanzzustands immer die realisierbare Lösung bevorzugen . Auf lange Sicht tendiert eine solche Strategie dazu, nicht realisierbare Lösungen im endgültigen Algorithmusbericht zu eliminieren. Mehrzieloptimierung wird verwendet, wenn mehrere Ziele gleichzeitig optimiert werden müssen. Es gibt mehrere mehrzielige Evolutionsalgorithmen (MOEAs), z. B. NSGAII und MOEA/D. Obwohl sie zur Optimierung mehrzieliger Probleme verwendet werden, unterscheiden sie sich erheblich. NSGAII ist ei**

5. Versuchsdesign Dieser

Unterabschnitt stellt die Reihe von Experimenten vor, die durchgeführt wurden, um die Anwendung der vorgeschlagenen Optimierungsmodelle als Werkzeuge zur Verbesserung des Verständnisses des Mikroalgenstoffwechsels zu validieren. Im Bereich der Forschung zu effektiven Lösungen für multikriterielle Probleme wurden Experimente an zwei Netzwerken der Mikroalge Chlorella vulgaris durchgeführt , **um die Leistung verschiedener Algorithmen und der jeweiligen MOFBAs zu bewerten. Die Unterabschnitte stellen den Fall der Studien, das Versuchsdesign und die bei der Implementierung verwendeten Softwaredetails vor, um die folge**

Hypothese 0 (H0). Für die Auswahl des Modells und/oder des Lösungsalgorithmus ist es nicht relevant, Flüsse in einem metabolischen Netzwerk zu optimieren.

5.1. Studienfälle **Im**

Vergleich zu einer früheren Untersuchung [18] wurden in die beiden Fallstudien zwei Netzwerke einbezogen: Glutamatstoffwechsel und Pigmentflussverteilung der Mikroalge Chlorella vulgaris ; **reversible und irreversible Reaktionen wurden hinzugefügt, die Darstellung einer reversiblen Reaktion mit Intervallen der unteren Flussgrenze von -1000 und der oberen Flussgrenze von 1000 und die irreversiblen Reaktionen mit Intervallen der unteren Flussgrenze von 0 bis zur oberen Flussgrenze von 1000. Darüber hinaus wurden Knotenpunkte einbezogen, an denen sich die Metaboliten in verschiedene Richtungen verzweigen, sowie Zyklen, die im Stoffwechsel der Zellen häufig vorkommen.**

5.1.1. Fallstudie 1: Stoffwechselnetzwerk Chlorella vulgaris

Das metabolische Netzwerk der Mikroalge Chlorella vulgaris **[17] wurde unter Verwendung von NS- GAII- und MOEA/D-Algorithmen für drei verschiedene Kulturbedingungen untersucht: Photoautotrophie (Licht + Komponenten), Heterotrophie (Komponente) und Mixotrophie (CO2 + Licht + Komponente). Zu den Verbindungen, die als Nährstoffe für die Kultivierung verwendet wurden, gehörten die Zugabe von Stickstoffquellen wie NO3 und NH4 sowie Sulfaten wie SO4, Fe2 und Magnesium . Die verschiedenen Erntequellen beeinflussen die Produktion von Metaboliten. Die folgende Abbildung zeigt die Verteil**

In diesem Fall wird ein Teil des Pigmentverteilungsnetzwerks untersucht, da die Verteilung in Mikroalgen wie C. vulgaris **für die Untersuchung der Pigmentsynthese von großer Bedeutung ist . Die am Stoffwechsel beteiligten Reaktionen sind in Anhang A.1, Tabelle A1, aufgeführt. Dieses Netzwerk umfasst die Komplexität reversibler und irreversibler (FRDPth, GRDPth) Reaktionen, Knoten und Zyklen und ist in Abbildung 1 dargestellt.**

Machine Translated by Google

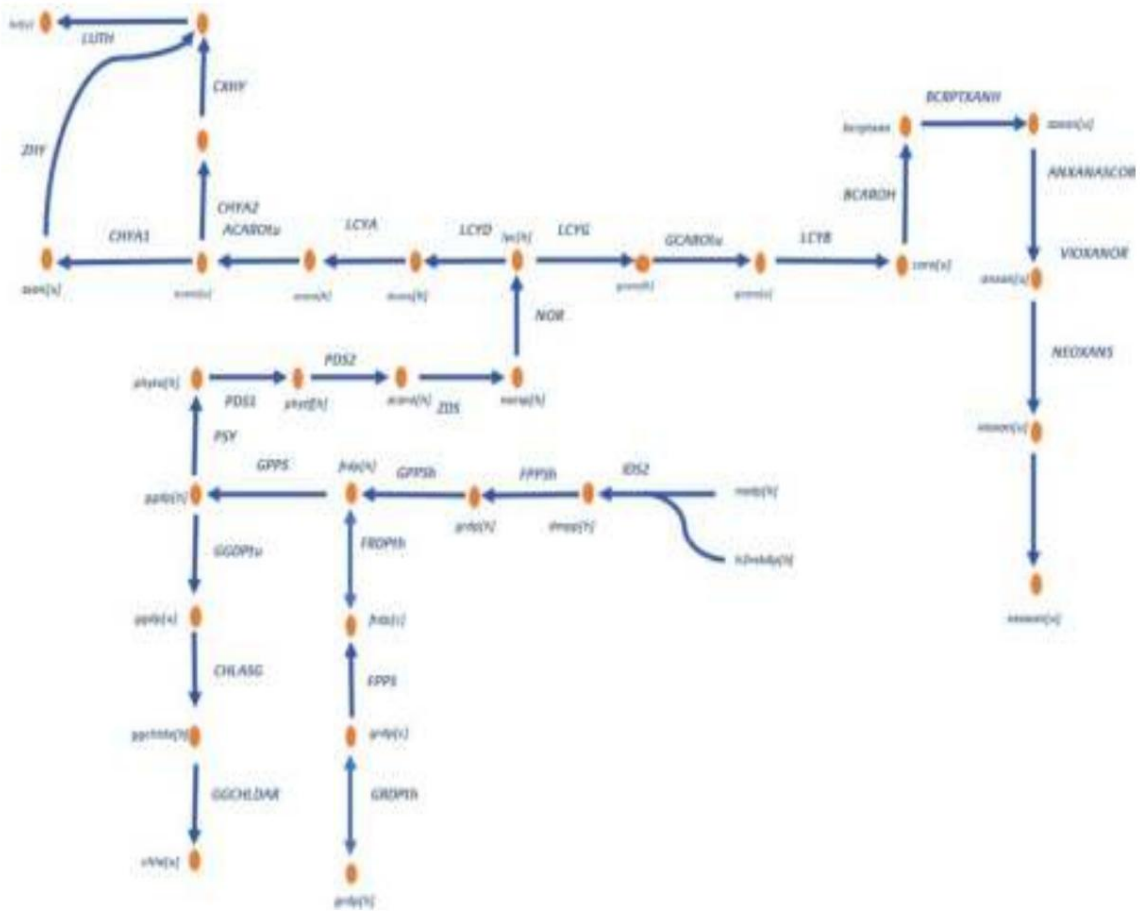


Abbildung 1. Flussverteilung der Pigmentbyosinthesewege [17].

5.1.2. Fallstudie 2: Multiobjektive Optimierung des metabolischen Netzwerks des Glutamatstoffwechsels der Mikroalge Chlorella vulgaris

Der Stoffwechsel besteht aus verschiedenen Stoffwechselwegen, die miteinander verflochten sind und ein komplexeres Ganzes bilden. Ein solcher Weg ist die Verteilung der Flüsse des Glutamatstoffwechsels, der verschiedene Funktionen erfüllt, darunter die Aminosäuresynt

Dieses Stoffwechselnetzwerk weist aufgrund der Anzahl der Metaboliten , die sich am zentralen Knoten verzweigen, und des Vorhandenseins reversibler Reaktionen wie ASPATh und ASPNA1Th eine hohe Komplexität auf. Dieses Netzwerk wurde unter drei verschiedenen Wachstumsbedingungen (Autotrophie, Heterotrophie und Mixotrophie) mithilfe der Algorithmen NSGAII und MOEA/D mit ihren vier MOFBAs ausgewertet. Dieser Stoffwechsel ist von großer Bedeutung, da er die Wege zur Herstellung verschiedener interessanter Produkte, Aminosäuren wie Tyrosin, Valin, Leucin usw., umfasst, die später zur Herstellung von Proteinen verwendet werden können. Die am Stoffwechsel beteiligten Reaktionen sind in Anhang A.1, Tabelle A2, dargestellt. Abbildung 2 stellt die Verteilung der **mit dem Glutamatstoffwechsel verbundenen Flüsse im C**

Machine Translated by Google

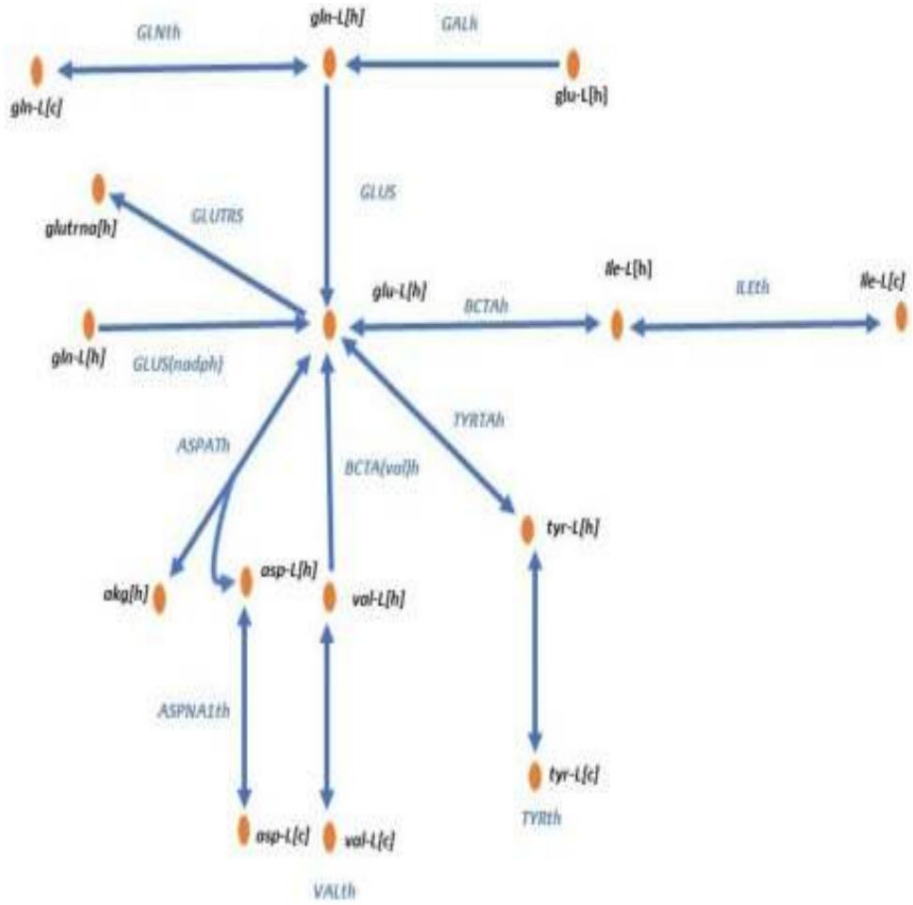


Abbildung 2. Flussverteilung im Zusammenhang mit dem Glutamatstoffwechsel im Chloroplasten und Zytoplasma [17].

5.2. Definition der Experimente :

Es ist möglich , **eine Methodologie auf Grundlage von Metaheuristiken zu definieren, die die von der FBA-Methode bereitgestellten Informationen zum Stoffwechselnetzwerkfluss verbessert, indem die Flussbilanzanalyse als multikriterielles Optimierungsproblem neu definiert wird. Es wurden vier Experimente vorgeschlagen, die in der folgenden Tabelle 4 zusammengefasst sind.**

Tabelle 4. Versuchsaufbau zur Demonstration metaheuristischer Unterstützung für das Verständnis des Stoffwechsels in Mikroalgen.

Ziel des Experiments		**Beteiligte Variablen**
Experiment 1	**Die Verwendung unterschiedlicher Optimierungsmodelle führt zu unterschiedlichen Ergebnissen.**	C. vulgaris, **NSGAII, MOFA1 , MOFBA2 , MOFBA3 , MOFBA4 .**
Experiment 2	**Zeigen Sie, dass der Einsatz von Metaheuristiken das Verständnis des Mikroalgenstoffwechsels unterstützt.**	C. vulgaris, **FBA, NSGAII, MOFBA4**
Experiment 3	**Es gibt Optimierungsalgorithmen, die für die Lösung spezifischer Probleme besser geeignet sind**	C. vulgaris, **NSGAII, MOEA/D, MOFBA4 .**
Experiment 4: Bestätigen Sie, dass eine zufällige Auswahl nicht ausreicht.		C. vulgaris, **NSGAII, FBA, zufällig.**

Machine Translated by Google

Experiment 1 zeigte, dass verschiedene Modelle zur Mehrzieloptimierung unterschiedliche Ergebnisse liefern. Der Algorithmus wurde auf NSGAII eingestellt, die Mikroalgen wurden modifiziert, um die Vielseitigkeit des Ansatzes zu demonstrieren, und schließlich wurden die vier vorgeschlagenen Optimierungsmodelle analysiert. Für die untersuchten Fälle wurde beobachtet, dass das durchweg beste Modell MOFBA4 war, das in den nachfolgenden Experimenten verwendet wurde.

In Experiment 2 wurden der Algorithmus, die Mikroalgen und das eindimensionale FBA-Modell mit dem durch MOFBA4 definierten mehrdimensionalen Modell verglichen. Es wurde beobachtet, dass der Stoffwechsel einer Mikroalge mit verschiedenen Flüssen beschrieben werden kann, nicht nur mit einem, und dass diese kontrolliert werden können, um Informationen über verschiedene Metaboliten von Interesse zu erhalten. Dies bestätigt, dass die vorgeschlagenen Methoden unter Berücksichtigung verschiedener Konfigurationen leicht an verschiedene Arten von Stoffwechselnetzwerken angepasst werden können .

In Experiment 3 wurde die Leistung der Algorithmen NSGAII und MOEA/D bei demselben Optimierungsproblem bewertet. Es wurde festgestellt, dass NSGAII für die drei betrachteten Ziele (d. h. drei Metabolite von Interesse) am besten war. Dieses Ergebnis steht im Einklang mit der Literatur, da NSGAII bei zwei oder drei Zielen eine bessere Leistung zeigt als MOEA/D. Dies lässt die Frage offen, ob MOEA/D bei einer größeren Anzahl von Zielen besser abschneidet , was eine offene Untersuchungslinie für seine Anwendung in der Untersuchung von Stoffwechselnetzwerken von Mikroalgen darstellt.

Experiment 4 zeigt, dass einfache Zufallsstichproben nicht ausreichen, um ein besseres Verteilung von Lösungen, die durch den Einsatz von Metaheuristiken möglich ist.

Die Metaheuristiken NSGAII und MOEA/D wurden mit Hilfe des jMetalPy-Frameworks [32] implementiert. Die Optimierungsmodelle wurden in Python entwickelt und als Teil der FBA-Implementierung verwendet, die vom Paket COBRApy bereitgestellt wird [33]. Grafiken wurden mit der Schnittstelle pyplot von matplotlib nachgebildet [34]. Der für die Durchführung der Experimente verwendete Computer verfügt über einen 64-Bit-Prozessor mit 2,6 GHz und 32 RAM-Speicher.

6. Ergebnisse

Dieser Abschnitt fasst die Daten zusammen, die durch die Durchführung der in Abschnitt 3.2 beschriebenen Experimente gewonnen wurden. Am Ende werden die **erreichten Forschungsziele diskutiert .** Zur Visualisierung der Ergebnisse in den Abbildungen **3–6** wurde **die Bibliothek matplotlib importiert.**

6.1. Experiment 1

En [18] hat gezeigt, dass NSGAII **für drei Zielfunktionen qualitativ bessere Lösungen bietet als die klassische FBA .** Abbildung **3 zeigt die Ergebnisse des Experiments, in dem die vier oben beschriebenen Varianten MOFBA1, MOFBA2, MOFBA3 und MOFBA4 für den NSGAII-Algorithmus getestet wurden, mit der Optimierung von drei Zielfunktionen der Pigmentflussverteilung der Mikroalge** Chlorella vulgaris. **Es ist zu beobachten, dass die verschiedenen MOFBAs unterschiedliche Verhaltensweisen zueinander aufweisen; MOFBA1, MOFBA3 und MOFBA4 verbessern die FBA. Das beste Lösungsverhalten zeigte jedoch MOFBA4 in Abbildung 3d, da es mehr nicht dominierte Lösungen bietet und eine gute Populationsdiversität in einer Umgebung gleicher Populationsgröße aufrechterhält.**

Abbildung 3 zeigt die Pigmentflussverteilung in der Mikroalge C. vulgaris. **Es ist ersichtlich, dass jede Variante von MOFBA ein unterschiedliches Verhalten aufweist und alle die FBA in Abbildung 3a–c verbessern, das beste Verhalten in den Lösungen ist jedoch bei der Variante von MOFBA4 in Abbildung 3d zu beobachten.**

Machine Translated by Google

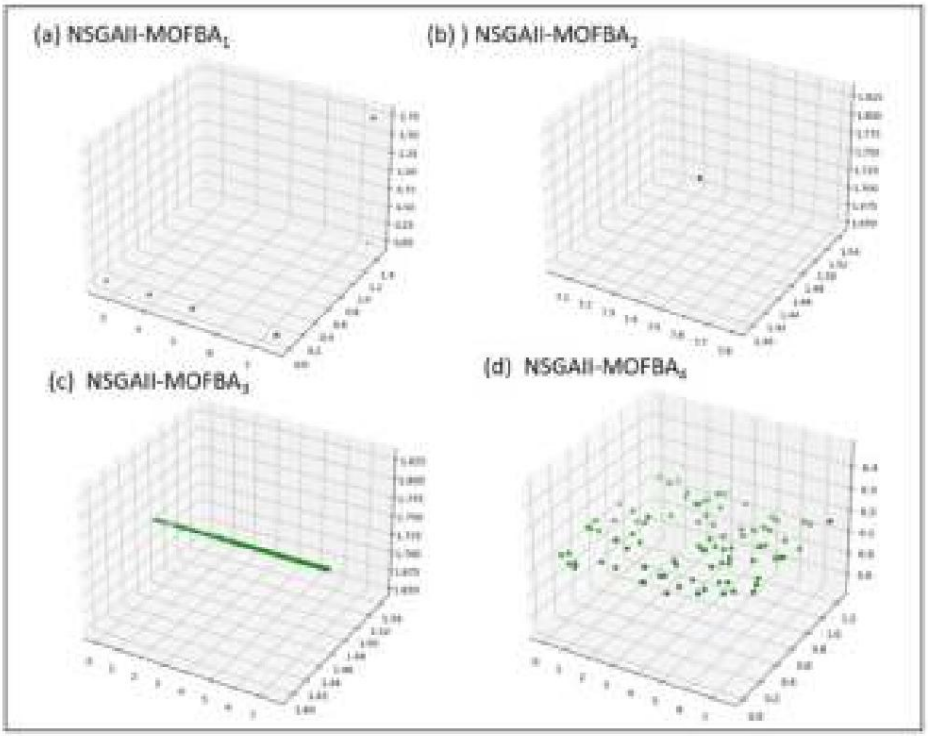


Abbildung 3. Vergleich zwischen dem NSGAII-Algorithmus und den Varianten (a) MOFBA1 , (b) MOFBA2 , (c) MOFBA3 und (d) MOFBA4 in der Verteilung der Pigmentflüsse.

6.2. Experiment

2 **Beim Vergleich der Leistung des NSGAII-Algorithmus mit der MOFBA4- Variante (Abbildung 4a) und der klassischen Einzelzieloptimierung FBA (Abbildung 4b) kann beobachtet werden, dass der NSGAII-MOFBA4- Algorithmus überlegen ist, da er zeigt , dass die bereitgestellten Informationen durch die Bereitstellung von mehr Lösungen und, was wichtig ist, einer deutlich verbesserten Verteilung verbessert werden. Diese verbesserte Verteilung ist besonders in den Abbildungen 3d und 4a deutlich, wo zwei verschiedene Stoffwechselnetzwerke untersucht werden, wie der Glutamatstoffwechsel und die Ve**

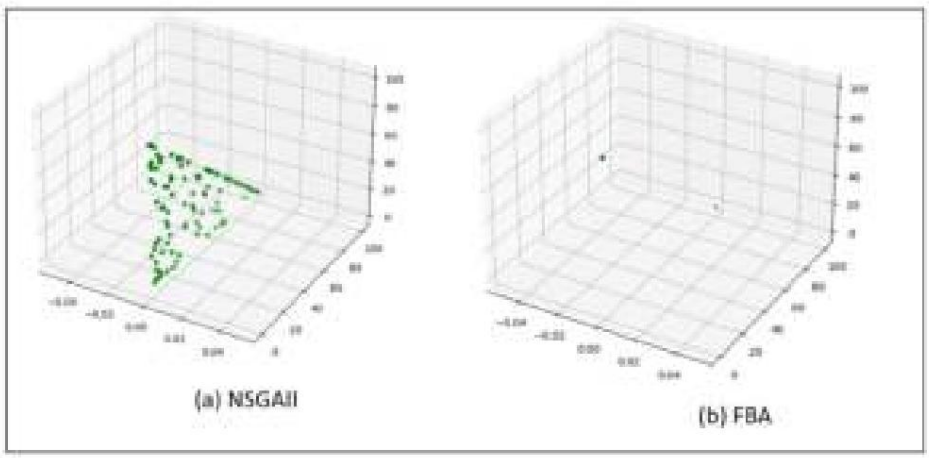


Abbildung 4. Vergleich zwischen (a) NSGAII-MOFBA4 und (b) FBA in der Verteilung von Flüssen, die mit dem Glutamatstoffwechsel der Mikroalge Chlorella vulgaris verbunden sind.

6.3. Experiment 3

Abbildung 5 zeigt die Auswertung zwischen den Algorithmen NSGAII (Abbildung 5a,c) und MOEA/D (Abbildung 5b,d) bei der Verteilung der mit dem Glutamatstoffwechsel verbundenen Fl

Machine Translated by Google

und das Verteilungsnetzwerk und die Pigmentflüsse mit NSGAII, Abbildung 5c, und MOEA/D, Abbildung 5d, aus der Mikroalge C. vulgaris **mit der Variante des MOFBA4- Algorithmus. Es zeigte sich, dass der NSGAII-Algorithmus im Vergleich zu MOEA/D mehr Lösungen und eine bessere Populationsdiversität aufweist . Metaheuristiken sind wichtig; in diesem Fall ist NSGAII das Beste, was mit der Literatur übereinstimmt [35], da dieser Algorithmus mit 2 und 3 Z**

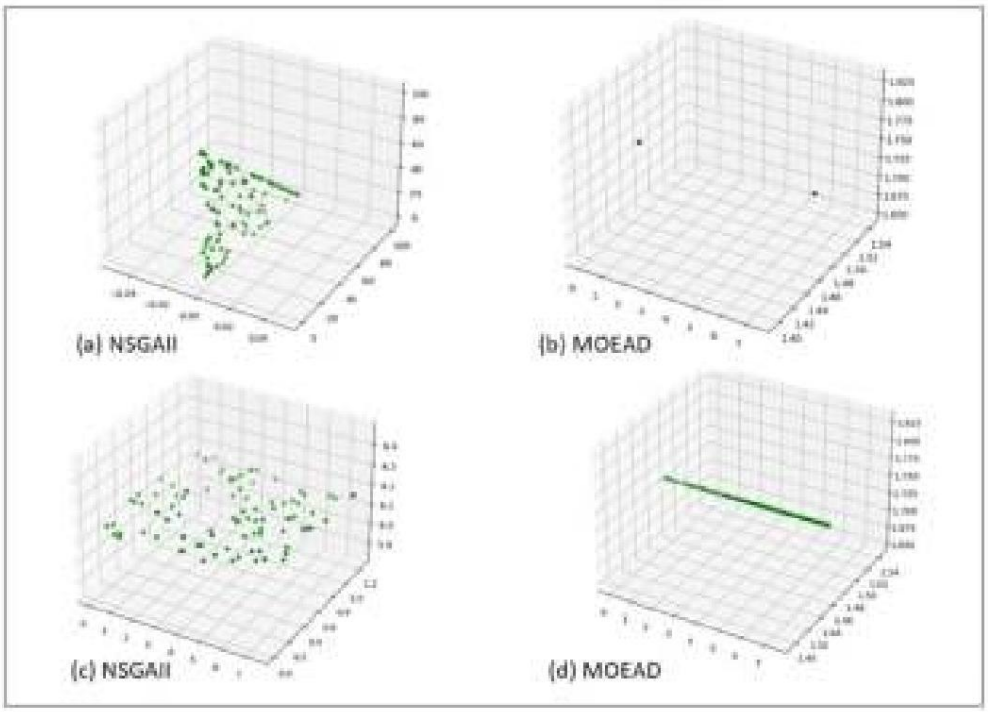


Abbildung 5. Vergleich zwischen den (a,c) NSGAII- und (b,d) MOEA/D-Algorithmen im Pigmentverteilungsnetzwerk und in der Verteilung der mit dem Glutamatstoffwechsel der Mikroalge Chlorella vulgaris verbundenen Flüsse.

6.4. Experiment

4 **Zusätzlich zum Testen der verschiedenen Fallstudien mit den Ansätzen FBA, NSGAII und MOEA/D wurde ein Experiment mit einem schnellen Zufallsansatz, den wir Zufall nennen, zur Verteilung des Pigmentflusses in der Mikroalge** C. vulgaris durchgeführt. **Abbildung 6 zeigt den Vergleich zwischen FBA, Abbildung 6a, Zufall, Abbildung 6c, und NSGAII, A Obwohl die Zufallsmethode schnell ist, konnte sie keine besseren Ergebnisse liefern (Abbildung 6b) als** die in Abbildung **6c, NSGAII, dargestellten Ergebnisse .**

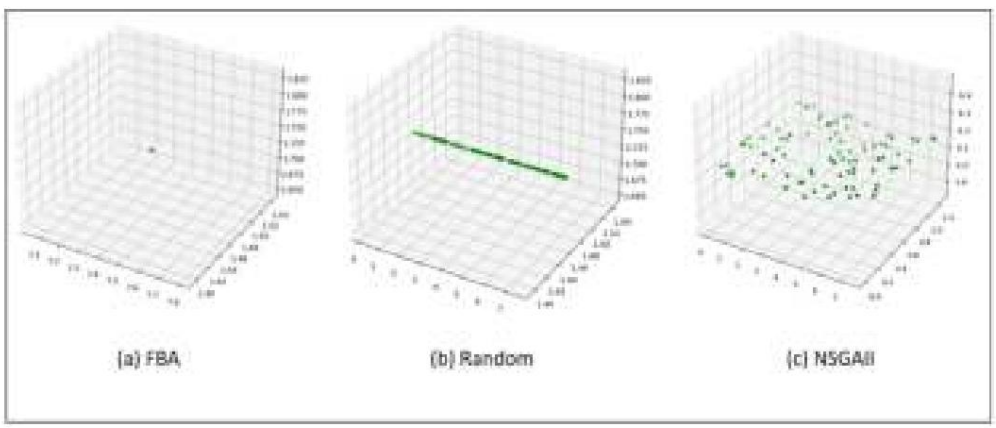


Abbildung 6. Vergleich zwischen den Varianten (a) FBA, (b) zufällig und (c) NSGAII in der Mikroalge C. vulgaris.

Machine Translated by Google

6.5. Statistische Analyse

Die Tabellen 5 und 6 zeigen, dass die vorgeschlagenen Methoden praktikable Lösungen liefern. Es wird gezeigt, dass Sie erfüllen die in **[17]** und FBA **genannten Bedingungen und zeigen damit den Zusammenhang von die Lösungen in silico und ihre Fähigkeit, Ergebnisse unter verschiedenen Kulturbedingungen zu emulieren. In Tabelle 6** kann man einen Vergleich zwischen den Flüssen in mmol **hÿ1 sehen, die mit FBA und NSGAII- MOFBA4** erhalten wurden , **und der euklidischen Distanz zwischen ihnen. Tabelle 5 zeigt, NSGAII bietet eine große Vielseitigkeit bei der Begrenzung der Parameter in den Wachstumsbedingungen durch die unteren und oberen Grenzwerte, zusätzlich zur Möglichkeit, Zyklen zu simulieren und Bifurkationen zwischen metabolischen Netzwerken. Ebenso gibt es einige mögliche Lösungen, die** Es werden **NSGAII mit der MOFBA4- Variante und aus dem klassischen FBA gewonnene Lösungen vorgestellt.**

Tabelle 5. Verteilung der Flüsse in mmol hÿ1 , erhalten durch NSGA II, verbunden mit der Synthese von Pigmente im Stoffwechsel der Mikroalge C. vulgaris.

Reaktion	LB	UB	Staffel 1	Staffel 2	S3	Staffel 4
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 × 10ÿ5	4,33 × 10ÿ7
FPPSh	0	0,000487	7,42 × 10ÿ5	4,72 × 10ÿ5	2,58 × 10ÿ5	4,33 × 10ÿ7
GPPSh	0	0,000487	7,42 × 10ÿ5	4,72 × 10ÿ5	2,58 × 10ÿ5	4,33 × 10ÿ7
FRDPth	0	0	0	0	0	0
FPPS	0	0	0	0	0	0
GRDPth	0	0	0	0	0	0
GRDPH	0	0	0	0	0	0
GGPPS	0	0,000487	7,42 × 10ÿ5	4,72 × 10ÿ5	2,58 × 10ÿ5	4,33 × 10ÿ7
Version 1	0	0,000402078	0	0	0	0
GGDPtu	ÿ0,0000849	0,0000849	7,42 × 10ÿ5	4,72 × 10ÿ5	2,58 × 10ÿ5	4,33 × 10ÿ7
CHLASG	0	0,0000742	7,42 × 10ÿ5	4,72 × 10ÿ5	2,58 × 10ÿ5	4,33 × 10ÿ7
GGCHLDAR 0		0,0000742	7,42 × 10ÿ5	4,72 × 10ÿ5	2,58 × 10ÿ5	4,33 × 10ÿ7
CHLA	0	0,0000742	7,42 × 10ÿ5	4,72 × 10ÿ5	2,58 × 10ÿ5	4,33 × 10ÿ7
PSY	0	2,24 × 10ÿ8	0	0	0	0
PDS1	ÿ2,24 × 10ÿ8	2,24 × 10ÿ8	0	0	0	0
PDS2	ÿ2,24 × 10ÿ8	2,24 × 10ÿ8	0	0	0	0
ZDS	ÿ2,24 × 10ÿ8	2,24 × 10ÿ8	1,59 × 10ÿ8	1,54 × 10ÿ8	1,06 × 10ÿ8	3,58 × 10ÿ9
NOCH	ÿ2,24 × 10ÿ8	2,24 × 10ÿ8	1,59 × 10ÿ8	1,54 × 10ÿ8	1,06 × 10ÿ8	3,58 × 10ÿ9
Version 2	0,00	8,00 × 10ÿ11	0	0	0	0
LCYG	ÿ1,59 × 10ÿ8	1,59 × 10ÿ8	1,59 × 10ÿ8	1,54 × 10ÿ8	1,06 × 10ÿ8	3,58 × 10ÿ9
GCAROtu	ÿ1,59 × 10ÿ8	1,59 × 10ÿ8	1,59 × 10ÿ8	1,54 × 10ÿ8	1,06 × 10ÿ8	3,58 × 10ÿ9
LCYB	ÿ1,59 × 10ÿ8	1,59 × 10ÿ8	1,59 × 10ÿ8	1,54 × 10ÿ8	1,06 × 10ÿ8	3,58 × 10ÿ9
Version 3	0,00	1,24 × 10ÿ8	1,24 × 10ÿ8	1,20 × 10ÿ8	7,10 × 10ÿ9	1,00 × 10ÿ10
BCAROH	ÿ3,48 × 10ÿ9	3,48 × 10ÿ9	3,48 × 10ÿ9	3,48 × 10ÿ9	3,48 × 10ÿ9	3,48 × 10ÿ9
BCRPTXANH ÿ3,48 × 10ÿ9		3,48 × 10ÿ9	3,48 × 10ÿ9	3,48 × 10ÿ9	3,48 × 10ÿ9	3,48 × 10ÿ9
Version 4	0,00	6,16 × 10ÿ9	6,16 × 10ÿ9	6,16 × 10ÿ9	6,16 × 10ÿ9	6,16 × 10ÿ9
ANXANAS	ÿ2,68 × 10ÿ9	2,68 × 10ÿ9	ÿ2,68 × 10ÿ9	ÿ2,68 × 10ÿ9	ÿ2,68 × 10ÿ9	ÿ2,68 × 10ÿ9
Version 5	0,00	3,88 × 10ÿ9	3,88 × 10ÿ9	3,88 × 10ÿ9	3,88 × 10ÿ9	3,88 × 10ÿ9
VIOXANOR ÿ2,41 × 10ÿ9		2,41 × 10ÿ9	ÿ2,41 × 10ÿ9	ÿ2,41 × 10ÿ9	ÿ2,41 × 10ÿ9	ÿ2,41 × 10ÿ9
Version 6	0,00	2,70 × 10ÿ10	2,70 × 10ÿ10	2,70 × 10ÿ10	2,70 × 10ÿ10	2,70 × 10ÿ10
NEOXANS	ÿ1,47 × 10ÿ9	1,47 × 10ÿ9	1,47 × 10ÿ9	1,47 × 10ÿ9	1,47 × 10ÿ9	1,47 × 10ÿ9
NEOXANU 0		1,47 × 10ÿ9	1,47 × 10ÿ9	1,47 × 10ÿ9	1,47 × 10ÿ9	1,47 × 10ÿ9
LCDYD	ÿ6,42 × 10ÿ9	6,42 × 10ÿ9	0	0	0	0
LCYA	ÿ6,42 × 10ÿ9	6,42 × 10ÿ9	0	0	0	0

Machine Translated by Google

Tabelle 5. Fortsetzung

Reaktion	LB	UB	Staffel 1	Staffel 2	S3	Staffel 4
Version 7	0,00	1,33 × 10ÿ9	0	0	0	0
ACAROtu	ÿ5,09 × 10ÿ9	5,09 × 10ÿ9	0	0	0	0
CHYA1	0	0	0	0	0	0
ZHY	0	0	0	0	0	0
CHYA2	0	5,09 × 10ÿ9	0	0	0	0
CXHY	0	5,09 × 10ÿ9	0	0	0	0
Version 8	0	3,35 × 10ÿ9	ÿ1,74 × 10ÿ9	ÿ1,74 × 10ÿ9	ÿ1,74 × 10ÿ9	ÿ1,74 × 10ÿ9
LUTH	ÿ1,74 × 10ÿ9	1,74 × 10ÿ9	1,74 × 10ÿ9	1,74 × 10ÿ9	1,74 × 10ÿ9	1,74 × 10ÿ9
LOROXANU 0		1,74 × 10ÿ9	1,74 × 10ÿ9	1,74 × 10ÿ9	1,74 × 10ÿ9	1,74 × 10ÿ9
FRDPth	0	1	0	0	0	0

Tabelle 6. Verteilung der Flüsse in mmol hÿ1, erhalten durch FBA und den NSGAII-MOFBA4- Algorithmus.

	Euklidische Distanz zwischen FBA und NSGAII-MOFBA4			
Reaktion.	————	MOFBA4	MOFBA4	MOFBA4
ACAROtu'	5,09 × 10ÿ9	0	0	0
ANXANASCOR'	ÿ2,68 × 10ÿ9	ÿ2,68 × 10ÿ9	ÿ2,68 × 10ÿ9	ÿ2,68 × 10ÿ9
BCAROH'	3,48 × 10ÿ9	3,48 × 10ÿ9	3,48 × 10ÿ9	3,48 × 10ÿ9
BCRPTXANH'	3,48 × 10ÿ9	3,48 × 10ÿ9	3,48 × 10ÿ9	3,48 × 10ÿ9
CHLASG'	0	1,36106 × 10ÿ9	2,35109 × 10ÿ5	0,0000742
CHYA1'	0	0	0	0
CHYA2'	5,09 × 10ÿ9	0	0	0
CXHY'	5,09 × 10ÿ9	0	0	0
FPPS'	0	0	0	0
FRDPth'	0	0	0	0
GCAROtu'	1,59 × 10ÿ8	7,94367 × 10ÿ9	1,35944 × 10ÿ8	1,59 × 10ÿ8
GGCHLDAR'	0	1,36106 × 10ÿ9	2,35109 × 10ÿ5	0,0000742
GGDPtu'	0,000148102	1,36106 × 10ÿ9	2,35109 × 10ÿ5	0,0000742
GGPS‘	0,000486596	1,36106 × 10ÿ9	2,35109 × 10ÿ5	0,0000742
GRDPth'	0	0	0	0
IDS1'	0,001459788	1,36106 × 10ÿ9	2,35109 × 10ÿ5	0,0000742
LCYA'	6,42 × 10ÿ9	0	0	0
LCYB‘	1,59 × 10ÿ8	7,94367 × 10ÿ9	1,35944 × 10ÿ8	1,59 × 10ÿ8
'LCYD'	6,42 × 10ÿ9	0	0	0
LCYG'	1,59 × 10ÿ8	7,94367 × 10ÿ9	1,35944 × 10ÿ8	1,59 × 10ÿ8
LUTH'	1,74 × 10ÿ9	1,74 × 10ÿ9	1,74 × 10ÿ9	1,74 × 10ÿ9
NEOXANS'	1,47 × 10ÿ9	1,47 × 10ÿ9	1,47 × 10ÿ9	1,47 × 10ÿ9
NOCH'	2,24 × 10ÿ8	7,94367 × 10ÿ9	1,35944 × 10ÿ8	1,59 × 10ÿ8
PDS1'	2,24 × 10ÿ8	0	0	0
PDS2'	2,24 × 10ÿ8	0	0	0
PSY'	2,24 × 10ÿ8	0	0	0
VIOXANOR'	ÿ2,41 × 10ÿ9	ÿ2,41 × 10ÿ9	ÿ2,41 × 10ÿ9	ÿ2,41 × 10ÿ9
ZDS'	2,24 × 10ÿ8	7,94367 × 10ÿ9	1,35944 × 10ÿ8	1,59 × 10ÿ8
ZHY'	0	0	0	0
Euklidische Distanz		0,001545861	0,001514586	0,001451344

Machine Translated by Google

Dieser Abschnitt bestätigt statistisch, dass es einen Unterschied gibt, wenn man unterschiedliche Optimierungsprobleme oder Algorithmen, um zu zeigen, dass die Wahl relevant ist. Dazu fasst es die Ergebnisse durch Vergleich nach Hypervolumen zusammen (dem Näherungsindikator zu die Pareto-Optimal-Front), um zu sehen, ob ein signifikanter Unterschied besteht zwischen Optimierungsmodelle MOFBA3 und MOFBA4 und die Algorithmen NSGAII und MOEA/D.

Die erste Analyse betrachtet die Modelle MOFBA3 und MOFBA4, setzt die Lösung Algorithmus nach NSGAII und wertet alle Netzwerke aus. Für die Analyse wurde das Hypervolumen aus jedem der 30 Durchläufe des Algorithmus pro Problem erhalten. Unter Verwendung der nichtparametrischen Wilcoxon-Vorzeichen-Rang-Test, mit einem Konfidenzniveau von 95% wurde der H0 validiert, der gibt an, dass es unmöglich ist, eine auf Metaheuristiken basierende Methodik zu definieren, die die Informationen über metabolische Netzwerkflüsse, die die FBA-Methode liefert, durch Neudefinition die Flussbilanzanalyse als multikriterielles Optimierungsproblem. Tabelle 7 fasst zusammen die Ergebnisse, die den Hypervolume-Wert auf logarithmischer Skala pro Lauf für jedes Netzwerk zeigen und der Akzeptanzstatus des H0 in der letzten Zeile. Man sieht, dass die Hypothese in fast allen analysierten metabolischen Netzwerken abgelehnt. Außer bei /textitChlorella kann es Es sei angemerkt, dass MOFBA4 das beste Optimierungsmodell ist.

Tabelle 7. Die Daten wurden statistisch analysiert, um die Unterschiede zwischen MOFBA3 und MOFBA4 zu bestätigen . **Die Nullhypothese H0 wurde akzeptiert, wenn der erhaltene p-Wert kleiner als 0,05 war.**

Glutamatstoffwechsel in C. vulgaris MOFBA4		Pigmentnetzwerk in C. vulgaris	
MOFBA3		MOFBA4 MOFBA3	
34.539	20.723	20,65	20.647
34.539	20.723	20,65	20.648
34.539	20.723	20,65	20.649
34.539	20.723	20,65	20.649
34.539	20.723	20,65	20.651
34.539	20.723	20,65	20.651
34.539	20.723	20,65	20.651
34.539	20.723	20,65	20.651
34.539	20.723	20,65	20.652
34.539	20.723	20,65	20.648
34.539	20.723	20,65	20.650
34.539	20.723	20,65	20.651
34.539	20.723	20,65	20.648
34.539	20.723	20,65	20.653
34.539	20.723	20,65	20.654
34.539	20.723	20,65	20.652
34.539	20.723	20,65	20.651
34.539	20.723	20,65	20.649
34.539	20.723	20,65	20.651
34.539	20.723	20,65	20.651
34.539	20.723	20,65	20.650
34.539	20.723	20,65	20.650
34.539	20.723	20,65	20.651
34.539	20.723	20,65	20.650
34.539	20.723	20,65	20.650
34.539	20.723	20,65	20.647
34.539	20.723	20,65	20.651
34.539	20.723	20,65	20.651
34.539	20.723	20,65	20.652
34.539	20.723	20,66	20,655
H0	AKZEPTIERT	H0	ABGELEHNT

Machine Translated by Google

Die zweite Analyse berücksichtigt die Algorithmen NSGAII und MOEA/D, setzt das Optimierungsmodell auf MOFBA4 und wertet alle Netzwerke aus. Für die Analyse wurde das Hyperv erhalten aus jeder der 30 Ausführungen jedes Algorithmus für das gelöste Problem. der nichtparametrische Wilcoxon-Vorzeichen-Rang-Test mit einem Konfidenzniveau von 95%, der H0 war validiert, was bedeutet, dass die Differenz der Mittelwerte zwischen den Stichproben gleich ist. Tabelle 8 fasst die Ergebnisse zusammen und zeigt den Hypervolume-Wert jedes Netzwerks pro Lauf bei die logarithmische Skala und der Akzeptanzstatus von H0 in der letzten Zeile. Es ist ersichtlich, dass Die Hypothese wird in allen analysierten Stoffwechselnetzwerken abgelehnt. Diese Ergebnisse aus beiden Analysen bestätigen die Erwartung, dass es relevant ist, zu überlegen, welches Optimierungsmodell zu verwenden, und welcher Algorithmus, weil ihre Leistungen beim Erhalten von Lösungssätzen können anders sein.

Tabelle 8. Die Daten wurden statistisch analysiert, um Unterschiede zwischen der Verwendung von NSGAII und MOEA/D. Die Nullhypothese H0 wurde akzeptiert, wenn der erhaltene p-Wert unter 0,05 lag.

Glutamatstoffwechsel in C. vulgaris		Pigmentnetzwerk in 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	ABGELEHNT	H0	ABGELEHNT

Durch die grafischen Ergebnisse beobachtet, vor allem aufgrund des Volumens und der Streuung von Die in allen untersuchten Stoffwechselnetzwerken der Algen erhaltenen Lösungen zeigen, dass die vorgeschlagene Methode, die auf einer Mehrzieloptimierung basiert und durch Metaheuristiken gelöst wird, die Analyse besser unterstützt. Andererseits ist die statistische Analyse Die in diesem Abschnitt vorgestellten Ergebnisse zeigen, dass es wichtig ist, die Optimierung zu berücksichtigen Modell und Algorithmus, da diese zu verschiedenen Arten von Verbesserungen beitragen können.

Machine Translated by Google

Die hier vorgestellten statistischen Analysen zeigen, dass es **zwischen Optimierungsmodellen und metaheuristischen Algorithmen erhebliche Unterschiede geben k**

6.6. Diskussion

Bisher wurden **mit Ausnahme von [18] fast keine Experimente zum Stoffwechsel von Mikroalgen durchgeführt . Obwohl es genaue Methoden gibt, erfordern evolutionäre Ansätze im Bereich mehrerer Ziele weniger Rechenressourcen; sie bieten beispielsweise den Vorteil, dass weniger Zeit und Speicher benötigt werden. Ansätze wie NSGAII und MOEA/D ermöglichen eine größere Entscheidungsfreiheit im Entscheidungsprozess aufgrund der Vielfalt und Anzahl der Lösungen und der Möglichkeit, die wichtigsten Flüsse in einem Netzwerk und ihren Einfluss und ihre Auswirkungen leichter zu erkennen, als wenn man keine Methode hätte.**

Aus den obigen Ergebnissen ergeben sich einige zusätzliche Erkenntnisse. Experiment 2 zeigt die Vielseitigkeit von NSGAII bei der Anpassung an unterschiedliche Umstände und seine Fähigkeit, die Analyse des Stoffwechselnetzwerks zu verbessern, da es für jede dieser Umstände eine größere Anzahl von Lösungen liefert. Wie in den Experimenten 1 bis 4 gezeigt, wird die Analysekapazität eines Stoffwechselnetzwerks durch die Einführung des NSGAII-Algorithmus verbessert.

Die in der Literatur vorhandenen Mehrzieloptimierungsprobleme berücksichtigen derzeit unterschiedliche Lösungsmetriken. Experiment 1 vergleicht die Verwendung des NSGAII-Algorithmus mit den vier Varianten von Optimierungsproblemen, wobei das MOFBA4-Optimierungsproblem das vielversprechendste ist, indem verschiedene Optimierungsstrategien eingeführt werden, wie z. B. die Optimierung nicht nur der interessierenden Metaboliten, sondern auch de Im Vergleich zu MOFBA3, das das Hypervolumen minimiert, sind bei MOFBA2 die Entscheidungsvariablen nur die Flüsse der interessierenden Reaktionen.

Obwohl NSGAII in den Fallstudien eine bessere grafisch erkennbare Leistung als MOEA/D hatte, wie in Experiment 3, da die Fallstudien drei Zielfunktionen hatten und NSGAII laut Literatur besser ist als MOEA/D, wenn drei Zielfunktionen vorhanden sind, eröffnet sich die Möglichkeit, MOEA/D in Netzwerken zu verwenden, in denen mehr als drei Zielfunktionen optimiert werden müssen. Durch Experiment 4 kann auch beobachtet werden, dass die einfache Verwendung einer Zufallsstichprobenlösung nicht ausreicht, um einen guten Satz von Lösungen zu erhalten, wie dies bei der Verwendung von Metaheuristiken der Fall ist. Es müssen jedoch besondere Überlegungen angestellt werden, um die Einhaltung von Einschränkungen oder Kontrollinformationen zu ermöglichen, die von einer interessierten Person gewünscht werden.

Die Algorithmen wurden an verschiedenen Mikroalgenstämmen getestet, wie in [18] mit C. reinhardti zu sehen ist, **und in dieser Forschung mit** C. vulgaris **wird ihre Durchführbarkeit in komplexen Stoffwechselnetzwerken, die Zyklen, Verzweigungen und reversible Reaktionen enthalten, in verschiedenen Stoffwechselnetzwerken überprüft und bestätigt, dass sie nicht nur in einer einzigen Mikroalge verwendet werden können. Dies lässt die Möglichkeit offen, sie in anderen Artenarten zu verwenden, bei denen mehr als eine Zielfunktion optimiert werden muss.**

7. Schlussfolgerungen

In der vorliegenden Forschungsarbeit wurde eine Studie der Stoffwechselflüsse in grünen Mikroalgen durchgeführt. Ziel war es, die Eignung von In-silico-Methoden als unterstützende Strategien zur Verbesserung der Analyse des Stoffwechsels in Mikroalgen zu überprüfen. Durch die durchgeführten Experimente wurden Beweise erhalten, die die folgenden Schlussfolgerungen stützen:

Die Untersuchung metabolischer Flüsse in Mikroalgen wird durch die Erhöhung der Anzahl von Lösungen verbessert, die die Bedingungen einer Mikroalge erfüllen, damit diese leben kann. Dies ist beobachtbar, weil es im Gegensatz zu herkömmlichen Methoden wie FBA, die nur eine Lösung bieten, die durch Sensitivitätsanalysen in begrenztem Umfang erweitert wird, durch die Integration in eine Methodik, die auf Metaheuristiken und multikriteriellen Optimierungsproblemen basiert, erheblich begünstigt wird ; **es erhöht sowohl die Anzahl der Flüsse, die die gesuchten Bedingungen im metabolischen Netzwerk erfüllen, als auch gleichzeitig die Op von Interesse.**

Es gibt mehrere Alternativen zur Analyse eines Stoffwechselnetzwerks durch die Optimierung mehrerer interessanter Metaboliten, wobei die FBA-Methode den Kern des Optimierungspro

Machine Translated by Google

In der vorliegenden Arbeit wurden vier Optimierungsmodelle vorgeschlagen, die dieses Ergebnis demonstrieren. Jedes Modell bietet andere Analysewinkel als FBA.

Es ist möglich, die Optimierungsprobleme, die die Stoffwechselstudie unterstützen, zu lösen, indem man verschiedene evolutionäre Metaheuristiken in Betracht zieht und signifikante Ergebnisse für die Analyse erhält . Dies wird anhand der Verwendung von NSGAII und MOEA/D zur Lösung der vorgeschlagenen Optimierungsprobleme demonstriert . In dieser Studie zeigte NSGAII im Allgemeinen die beste Leistung, was mit der Literatur übereinstimmt, indem es sich ausschließlich mit der gleichzeitigen Optimierung von drei Zielen befasste. Dies zeigt, dass für zukünftige Arbeiten die Analyse der besten Metaheuristik vor der Studie durchgeführt werden muss.

Die Parameter für die Lösungssuche können während der Analyse eines Stoffwechselnetzwerks durch Anpassen der Reaktionsgrenzen gesteuert werden. Dies trägt dazu bei, die Untersuchung von Mikroalgen weiter zu verbessern, da die Definition kontrollierter Umgebungen möglich ist. Dies wird im Validierungsprozess beobachtet, bei dem die Parameter zur Generierung von Lösungen auf die Werte beschränkt wurden, die bei der Arbeit an In-vivo-Proben gefunden wurden.

Die Auswahl eines Algorithmus oder Modells zur Optimierung eines bestimmten Stoffwechselnetzwerks kann schwierig sein und erfordert Feinabstimmung, um die Konfiguration zu ermitteln, die am besten zu den Forschungsinteressen des durchgeführten Stoffwechsel-Engineerings passt. Dies wird deutlich, wenn man die Leistungsunterschiede zwischen Algorithmen und Optimierungsmodellen oder den verschiedenen Kombinationen betrachtet , **die in dieser Forschungsarbeit getestet wurden.**

Ein Entscheidungsträger, beispielsweise ein Forscher im Bereich Stoffwechseltechnik, verbessert seine Entscheidungsfähigkeit durch die Visualisierung einer Reihe von Stoffwechselflüssen, die die für das untersuchte Stoffwechselnetzwerk festgelegten Bedingungen erfüllen.

Beiträge der Autoren: Konzeptualisierung, MFB-B. und LA-V.; Methodik, MFB-B. und NR-V.; Software, MFB-B. und NR-V.; Validierung, LA-V., ALM-S. und CZ; formale Analyse, MFB-B., LA-V. und NR-V.; Untersuchung, MFB-B.; Ressourcen, MFB-B. und LA-V.; Datenkuratierung, LA-V., ALM-S. und CZ; Schreiben – Vorbereitung des Originalentwurfs, MFB-B., NR-V. und CG-S.; Schreiben – Überprüfung und Bearbeitung, LA-V., ALM-S. und CZ; Visualisierung, MFB-B.; Überwachung, LA-V. und ALM-S.; Projektverwaltung, MFB-B. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und ihr zugestimmt .

Finanzierung: CZ wurde von der National Science Foundation unterstützt (Zuschuss-Nr. 2313313). Alle in diesem Material geäußerten Meinungen, Erkenntnisse und Schlussfolgerungen oder Empfehlungen sind die des Autors bzw. der Autoren und spiegeln nicht unbedingt die Ansichten der National Science Foundation wider.

Datenverfügbarkeitserklärung: Das Skript der modifizierten metaheuristischen Algorithmen NSGAII und MOEA/D mit ihren jeweiligen MOFBAs wurde im GitHub-Repository gespeichert: https://github. com/multiobjectiveoptimization2/MOFBAs (abgerufen am 24. Juni 2024).

Danksagungen: Die Autoren danken den CONAHCYT-Projekten Nr. für die Unterstützung. 3058, A1-S-11012 und das Laboratorio Nacional de Tecnologías de la Información (LANTI). Sie danken auch dem ITCM und dem TECNM-Projekt 14612.22-P für die Unterstützung. Mónica Fabiola Briones Báez dankt für das Stipendium Nr. 784358 von CONAHCYT zur Fortsetzung s

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Anhang A

Anhang A.1 **Wie**

bereits erwähnt, ist ein Stoffwechselnetzwerk die Gesamtheit der im Mikroorganismus vorhandenen Reaktionen. Die folgende Reaktionsliste stellt das Stoffwechselnetzwerk für die Synthese von Pigmenten in der Mikroalge Chlorella vulgaris dar.

Machine Translated by Google

Tabelle A1. Reaktionen, die sich aus der Synthese von Pigmenten im Stoffwechsel der Mikroalgen ergeben
C. vulgaris **[17].**

Name	Formel
FRDPth	frdp[h] ÿ frdp[c]
GRDPth	grdp[h] ÿ grdp[c]
ACAROtu	acaro[h] ÿ acaro[u]
GCAROtu	gcaro[h] ÿ gcaro[u]
GGDPtu	ggdp[h] ÿ ggdp[u]
FPPS	grdp[c] + ipdp[c]ÿ frdp[c] + ppi[c]
FPPSh	grdp[h] + ipdp[h]ÿ frdp[h] + h[h] + ppi[h]
GGPS	frdp[h] + ipdp[h]ÿ ggdp[h] + h[h] + ppi[h]
GPPSh	dmpp[h] + ipdp[h]ÿ grdp[h] + h[h] + ppi[h]
IDS2	h[h] + h2mb4p[h] + nadph[h]ÿ dmpp[h] + h2o[h] + nadp[h]
ANXANASCOR	anxan[u] + ascb-L[u] ÿ dhdascb[u] + h2o[u] + zaxan[u]
BCAROH	caro[u] + h[u] + nadph[u] + o2[u] ÿ bcrptxan[u] + h2o[u] + nadp[u]
BCRPTXANH	bcrptxan[u] + h[u] + nadph[u] + o2[u] ÿ h2o[u] + nadp[u] + zaxan[u]
CHYA1	acaro[u] + h[u] + nadph[u] + o2[u]ÿ h2o[u] + nadp[u] + zxan[u]
CHYA2	acaro[u] + h[u] + nadph[u] + o2[u]ÿ crpxan[u] + h2o[u] + nadp[u]
CXHY	crpxan[u] + h[u] + nadph[u] + o2[u]ÿ h2o[u] + lut[u] + nadp[u]
LCYB	gcaro[u] ÿ caro[u]
LCYA	lyc[h] ÿ dcaro[h]
LCYG	lyc[h] ÿ gcaro[h]
NEOXANS	Vioxan[u] ÿ Neoxan[u]
NOCH	norsp[h] + o2[h] + pqh2[h] ÿ 2 h2o[h] + lyc[h] + pq[h]
PDS1	phyto[h] + pq[h] ÿ phytfl[h] + pqh2[h]
PDS2	phytfl[h] + pq[h] ÿ pqh2[h] + zcaro[h]
PSY	2 ggdp[h]ÿ 2 h[h] + phyto[h] + 2 ppi[h]
VIOXANOR	ascb-L[u] + vioxan[u] ÿ anxan[u] + dhdascb[u] + h2o[u]
ZDS	o2[h] + pqh2[h] + zcaro[h] ÿ 2 h2o[h] + norsp[h] + pq[h]
ZHY	h[u] + nadph[u] + o2[u] + zxan[u]ÿ h2o[u] + lut[u] + nadp[u]
CHLASG	chlda[u] + ggdp[u]ÿ ggchlda[u] + h[u] + ppi[u]
GGCHLDAR	ggchlda[u] + 3 h[u] + 3 nadph[u]ÿ chla[u] + 3 nadp[u]
GGDR	ggdp[h] + 3 h[h] + 3 nadph[h] ÿ 3 nadp[h] + pdp[h]
CHLBSG	chldb[u] + ggdp[u]ÿ ggchldb[u] + h[u] + ppi[u]

Tabelle A2. Aus der Flussverteilung abgeleitete Reaktionen im Zusammenhang mit dem Glutamatstoffwechsel [17].

Name	Formel
GLN.	gln-L[c] + h[c] ÿ gln-L[h] + h[h]
GALh	atp[h] + glu-L[h] + nh4[h] ÿ adp[h] + gln-L[h] + h[h] + pi[h]
GLUS	glu-L[h] + h2o[h] + nad[h] ÿ akg[h] + h[h] + nadh[h] + nh4[h]
GLUTRS	atp[h] + glu-L[h] + trnaglu[h] ÿ amp[h] + glutrna[h] + h[h] + ppi[h]
GLUS (nadph)	akg[h] + gln-L[h] + h[h] + nadph[h] ÿ 2 glu-L[h] + nadp[h]
ASPATh	akg[h] + asp-L[h] ÿ glu-L[h] + oaa[h]
ASPNA1.	asp-L[c] + na1[c] ÿ asp-L[h] + na1[h]
VALth	h[c] + Wert-L[c] ÿ h[h] + Wert-L[h]
BCTA(Wert)h	akg[h] + val-L[h] ÿ 3mob[h] + glu-L[h]
TYRTAh	34hpp[h] + glu-L[h] ÿ akg[h] + tyr-L[h]
TYRth	h[c] + tyr-L[c] ÿ h[h] + tyr-L[h]
BCTAh	3mop[h] + glu-L[h] ÿ akg[h] + ile-L[h]
ILEth	h[c] + ile-L[c] ÿ h[h] + ile-L[h]

Verweise

1. **Esteves, AF; Soares, OS; Vilar, VJ; Pires, JC; Gonçalves, AL Der Einfluss der Lichtwellenlänge auf die CO2- Abscheidung und Biomasse Produktion und Nährstoffaufnahme durch grüne Mikroalgen: Ein Fortschritt bei der Prozessintegration und -optimierung.** Energien **2020,** 13, **333.** [CrossRef]
2. **Wang, Y.; Tibbetts, SM; McGinn, PJ Mikroalgen als Quellen hochwertigen Proteins für menschliche Nahrungsmittel und Proteinergänzungsmittel.** Lebensmittel **2021,** 10, **3002.** [CrossRef] [PubMed]
3. **Brown, MR; Jeffrey, S. Biochemische Zusammensetzung von Mikroalgen aus den Grünalgenklassen Chlorophyceae und Prasinophyceae. 1. Aminosäuren, Zucker und Pigmente.** J. Exp. Mar. Biol. Ecol. **1992,** 161, **91–113.** [CrossRef]
4. **Kholssi, R.; Ramos, PV; Marks, EA; Montero, O.; Rad, C. 2Biotechnologische Nutzung von Mikroalgen: Ein Überblick über den aktuellen Stand der Technik und Herausforderungen für die Kreislaufwirtschaft.** Biocatal. Agric. Biotechnol. **2021,** 36, **102114.** [CrossRef]
5. **Romero, DCV; Cardozo, AP; Montes, VD Verwendung von Mikroalgen als Alternative zur Entfernung von Metallen.** RIAA **2022,** 13, **10.**
6. **Woolston, BM; Edgar, S.; Stephanopoulos, G. Metabolic engineering: Vergangenheit und Zukunft.** Annu. Rev. Chem. Biomol. Eng. **2013, 4, 259–288.** [CrossRef] [PubMed]
7. **Kaste, JA; Shachar-Hill, Y. Modellvalidierung und -auswahl in der Stoffwechselflussanalyse und Flussbilanzanalyse.** Biotechnol. Prog. **2024,** 40, **e3413.** [CrossRef] [PubMed]
8. **Anand, S.; Mukherjee, K.; Padmanabhan, P. Ein Einblick in die Flussbilanzanalyse für biochemische Netzwerke.** Biotechnol. Genet. Eng. 2020 , 36, **32–55.** [CrossRef] [PubMed]
9. **Orth, JD; Thiele, I.; Palsson, B.Ø. Was ist Flussbilanzanalyse?** Nat. Biotechnol. **2010,** 28, **245–248.** [CrossRef]
10. **Liu, Y.; Westerhoff, HV Wettbewerbsorientierte, multiobjektive und kompartimentierte Flux Balance Analyse zur Behandlung gewebespezifischer angeborene Stoffwechselstörungen.** J. Inherit. Metab. Dis. **2023,** 46, **573–585.** [CrossRef]

Machine Translated by Google

11. Raman, K.; Chandra, N. Flussbilanzanalyse biologischer Systeme: Anwendungen und Herausforderungen. Kurz. Bioinform. **2009,** 10, **435–449.** [Querverweis]

12. Lewis, N.; Nagarajan, H.; Palsson, B. Einschränkung der metabolischen Genotyp-Phänotyp-Beziehung durch eine Phylogenese von in silico Methoden. Nat. Rev. Microbiol. **2012,** 10, **291–305.** [CrossRef]

13. Toyoshima, M.; Toya, Y.; Shimizu, H. Flussbilanzanalyse von Cyanobakterien zeigt selektive Nutzung photosynthetischer Elektronentransportkomponenten unter unterschiedlichen spektralen Lichtbedingungen. Photosynth. Res. **2020,** 143, **31–43.** [CrossRef] [PubMed]

14. Huang, J.; Hou, J.; Li, L.; Wang, Y. Flussbilanzanalyse des Glukoseabbaus durch anaerobe Vergärung bei Unterdruck. Int. J. Hydrogen Energy **2020,** 45, **26822–26830.** [CrossRef]

15. Trilla-Fuertes, L.; Gámez-Pozo, A.; Díaz-Almirón, M.; Prado-Vázquez, G.; Zapater-Moros, A.; López-Vacas, R.; Nanni, P.; Zamora, P.; Espinosa, E.; Fresno Vara, JA Computergestützte Stoffwechselmodellierung sagt das Risiko eines rezidivfreien Fernüberlebens bei Brustkrebspatientinnen voraus . Zukünftige Onkol. **2019,** 15, **3483–3490.** [CrossRef]

16. Boyle, NR; Morgan, JA Flussbilanzanalyse des Primärstoffwechsels in Chlamydomonas reinhardtii. BMC Syst. Biol. **2009, 3, 4.** [Querverweis] [PubMed]

17. Zuñiga, C.; Li, CT; Huelsman, T.; Levering, J.; Zielinski, DC; McConnell, BO; Long, CP; Knoshaug, EP; Guarnieri, MT; Antoniewicz, MR; et al. Genomweites Stoffwechselmodell für die Grünalge Chlorella vulgaris **UTEX 395 sagt Phänotypen unter autotrophen, heterotrophen und mixotrophen Wachstumsbedingungen genau voraus.** Plant Physiol. **2016,** 172, **589–602.** [CrossRef] [PubMed]

18. Briones-Baez, MF; Aguilera-Vazquez, L.; Rangel-Valdez, N.; Martinez-Salazar, AL; Zuñiga, C. Multi-Objective Optimization of Microalgae Metabolism: Ein evolutiver Algorithmus basierend auf FBA. Metabolites **2022,** 12, **603.** [CrossRef] [PubMed]

19. Rangaiah, GP; Petriciolet, A. Multi-Objective Optimization in Chemical Engineering: Entwicklungen und Anwendungen; **Rangaiah, GP, Bonilla-Petriciolet, A., Hrsg.; Wiley: New York, NY, USA, 2013.**

20. Liu, X.; Tian, J.; Duan, P.; Yu, Q.; Wang, G.; Wang, Y. GrMoNAS: Ein granularitätsbasiertes multi-objektives NAS-Framework für effiziente medizinische Diagnose. Comput. Biol. Med. **2024,** 171, **108118.** [CrossRef]

21. PDE: Ein Pareto-Frontier-Differential-Evolutionsansatz für Mehrzieloptimierungsprobleme. In Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No. 01TH8546), Seoul, Republik Korea, 27.–30. Mai 2001; Band 2, S. 971–978.

22. Andrade, R.; Doostmohammadi, M.; Santos, J.; Sagot, MF; Mira, NP; Vinga, S. MOMO—multi-objektive metabolische gemischte Integer-Optimierung: Anwendung in der Hefestammtechnik. BMC Inform. **2020,** 21, **69.** [CrossRef]

23. Wang, GG; Zhao, X.; Li, K. Metaheuristische Algorithmen: Theorie und Praxis; **CRC Press: Boca Raton, FL, USA, 2024.**

24. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. Ein schneller und elitärer multiobjektiver genetischer Algorithmus: NSGA-II. IEEE Trans. Evol. 2002 , **6, 182–197.** [CrossRef]

25. Cruz, L.; Fernandez, E.; Gomez, C.; Rivera, G.; Perez, F. Mehrziel-Portfoliooptimierung von interdependenten Projekten mit „a priori“-Einbeziehung der Präferenzen der Entscheidungsträger. Appl. Math. Inf. Sci. **2014, 8, 1517–1531.** [CrossRef]

26. Rivera, G.; Florencia, R.; Guerrero, M.; Porras, R.; Sanchez-Solis, J. Online-Multikriteriell-Portfolioanalyse durch Kompromissprogrammierungsmodelle, die auf den Grundprinzipien des Fuzzy-Outrankings basieren. Inf. Sci. **2021,** 580, **734–755.** [CrossRef]

27. Chang, KH Multiobjektive Optimierung und fortgeschrittene Themen. In Design Theory and Methods Using CAD/CAE; **Elsevier: Amsterdam, Niederlande, 2015; S. 325–406.**

28. Zhang, Q.; Li, H. MOEA/D: Ein multiobjektiver evolutionärer Algorithmus basierend auf Zerlegung. IEEE Trans. Evol. Comput. **2007,** 11, **712–731.** [CrossRef]

29. Cruz-Reyes, L.; Fernandez, E.; Rangel-Valdez, N. Eine metaheuristische, optimierungsbasierte indirekte Ermittlung von Präferenzparametern zur Lösung von Mehrzielproblemen. Int. J. Comput. Intell. Syst. **2017,** 10, **56–77.** [CrossRef]

30. Deb, K.; Goyal, M. Eine kombinierte genetische adaptive Suche (GeneAS) für technisches Design. Comput. Sci. Inform. **1996,** 26, **30–45.**

31. Deb, K. Eine effiziente Methode zur Handhabung von Constraints für genetische Algorithmen. Comput. Methods Appl. Mech. Eng. **2000,** 186, **311–338.** [Querverweis]

32. Benítez-Hidalgo, A.; Nebro, AJ; García-Nieto, J.; Oregi, I.; Del Ser, J. jMetalPy: Ein Python-Framework für mehrere Ziele Optimierung mit Metaheuristiken. Swarm Evol. Comput. **2017,** 12, **e0171744.** [CrossRef]

33. Ebrahim, A.; Lerman, JA; Palsson, BO; Hyduke, DR COBRApy: Auf Einschränkungen basierende Rekonstruktion und Analyse für Python. BMC Syst. Biol. **2013, 7, 74.** [CrossRef]

34. Hunter, JD Matplotlib: Eine 2D-Grafikumgebung. Comput. Sci. Eng. **2007, 9, 90–95.** [CrossRef]

35. Liu, D.; Huang, Q.; Yang, Y.; Liu, D.; Wei, X. Bi-objektiver Algorithmus basierend auf dem NSGA-II-Framework zur Optimierung des Reservoirbetriebs. J. Hydrol. **2020,** 585, **124830.** [CrossRef]

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Artikel

Hyperspektrales Python: HypPy

Wim Bakker *, Frank van Ruitenbeek, Harald van der Werff *, Christoph Hecker, Arjan Dijkstra und Freek van der Meer

Fakultät ITC, Universität Twente, 7522 NH Enschede, Niederlande; fjavanruitenbeek@utwente.nl (FvR); cahecker@utwente.nl (CH); ahdijkstra@utwente.nl (AD); fdvandermeer@utwente.nl (FvdM)
* Korrespondenz: whbakker@utwente.nl (WB); hmavanderwerff@utwente.nl (HvdW)

Zusammenfassung: In diesem Dokument werden das Design, die Implementierung und die Verwendung eines Python-Pakets namens Hyperspectral Python (HypPy) beschrieben. Proprietäre Software zur Verarbeitung hyperspektraler Bilder ist teuer und mit diesen Paketen entwickelte Tools können nicht frei verteilt werden. Die Idee von HypPy besteht darin, hyperspektrale Bilder mit kostenloser Open-Source-Software verarbeiten zu können. HypPy wurde mit Python entwickelt und basiert auf den Array-Verarbeitungsfunktionen von Paketen wie NumPy und SciPy. HypPy wurde mit Blick auf die praktische Bildspektrometrie entwickelt und hat eine Reihe neuer Ideen implementiert. Um nur einige dieser Ideen zu nennen: HypPy verfügt über BandMath- und SpectralMath-Tools zur Verarbeitung von Bildern und Spektren mit Python-Anweisungen, kann Spektralbibliotheken verarbeiten, als wären sie Bilder, und kann Bänder nach Wellenlänge statt nach Bandnummer ansprechen. Wir erwarten, dass HypPy für Forschung, Bildung und Projekte mit hyperspektralen Daten von Nutzen ist, da es flexibel **und vielseitig ist .**

Schlüsselwörter: Hyperspektralbildgebung; Bildverarbeitung; Python; Mineralien

check for updates

Zitat: Bakker, W.; van Ruitenbeek, F.; van der Werff, H.; Hecker, C.; Dijkstra, A.; van der Meer, F. Hyperspektrales Python: HypPy. Algorithmen 2024, 17, 337. https:// doi.org/10.3390/a17080337

Wissenschaftliche Redakteure: Laura Antonelli und Lucia Maddalena

Empfangen: 27. Juni 2024
Überarbeitet: 22. Juli 2024
Akzeptiert: 30. Juli 2024
Veröffentlicht: 1. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/ 4.0/).

1. Einleitung

Bildverarbeitungspakete gibt es seit der Erfindung der ersten Computer. Zu Beginn dieses Jahrhunderts waren jedoch nicht viele Programme in der Lage, hyperspektrale Bilddaten zu verarbeiten. Hyperspektrale Bilder sind in der Regel riesig und Standardoperationen wie die Bildklassifizierung sind für die Verarbeitung der hohen Dimensionalität hyperspektraler Daten un Traditionell werden Hyperspektralbilder mit proprietären Paketen wie ENVI (ENVI und IDL sind Warenzeichen von NV5 Geospatial Solutions, Inc.) (Environment for Visualizing Images) [1] und anderen verarbeitet. Mit dem Aufkommen kostenloser Open-Source-Software [2] entstand jedoch der Bedarf, Hyperspektralbilder unabhängig von proprietären Paketen verwenden und verarbeiten zu können.

Mit Paketen wie numpy [3], scipy [4] und matplotlib [5] ist die Sprache Python [6] zu einer Datenverarbeitungs- und Prototyping-Umgebung geworden, die mit Systemen (proprietären und kostenlosen) wie MATLAB (MatLab ist ein eingetragenes Warenzeichen von The MathWorks, Inc.), IDL (ENVI und IDL sind Warenzeichen von NV5 Geospatial Solutions, Inc.), Octave (Octave ist ein Warenzeichen von Octave Technologies, Inc.) und R konkurriert.

Python und die bereits erwähnten Pakete Numpy und Scipy sind kostenlos und Open Source. Benutzer können auf den Quellcode zugreifen, Änderungen vornehmen und ihre eigenen Versionen verteilen. Im Prinzip kann jeder in Python geschriebene Code frei verteilt werden.

Wir verwenden Python für die Implementierung des Pakets Hyperspectral Python, HypPy (ausgesprochen [hip-ee], wie die Subkultur in den 1960er Jahren), einem Softwarepaket zum Lesen und Schreiben, Visualisieren, Verarbeiten und Analysieren von Hypersp

HypPy begann mit einer Reihe von Bildklassen und Factory-Funktionen, die das Lesen und Schreiben von Bildern im ENVI-Format ermöglichten. Die Entwicklung weiterer Tools in HypPy gewann an Dynamik, als unsere Forschungsgruppe 2009 und 2010 am Mars Planetary Mapping Pilot Project [7] beteiligt wurde . **Anschließend wurden verschiedene allgemeine Tools und spezielle Tools für den Umgang mit hyperspektralen Marsdaten geschrieben.**

Machine Translated by Google

Plattform zur Umsetzung experimenteller Ideen. Nach und nach wurde das Paket zu dem erweitert, was es heute ist. Dieses Dokument beschreibt die Designüberlegungen und die Implementierung von HypPy anhand mehrerer Beispiele.

2. Design und Implementierung

In diesem Abschnitt möchten wir zeigen, dass HypPy über eine Reihe von Funktionen verfügt, die es von anderen Paketen unterscheiden und es zu einem vielseitigen und flexiblen Werkzeug für die Verarbeitung hyperspektraler Date

HypPy hat die folgenden Funktionen: • HypPy unterstützt drei Schnittstellenebenen: eine grafische Benutzeroberfläche (GUI), eine Befehlszeilenschnittstelle (CLI) und eine Programmierschnittstelle (API). Die Benutzeroberflächen von HypPy können über ein Menü der obersten Ebene aufgerufen werden. • Das Menü der obersten Ebene ist konfigurierbar und kann für verschiedene Zwecke angepasst werden. • HypPy verfügt über Viewer für Hyperspektralbilder und Spektralbibliotheken. Diese Viewer haben einfache Schnittstellen, bieten aber dennoch ausreichend Funktionalität, um die Qualität und den Informationsgehalt der Daten zu untersuchen.

• HypPy bietet Lese- und Schreibfunktionen für ENVI-Datendateien. • Hyperspektrale Bilder sind in der Regel riesig, dennoch ist HypPy speicherfreundlich.

• HypPy unterstützt die drei nativen Bildformate von ENVI: BSQ, BIL und BIP. Intern übersetzt HypPy diese in ein BIP-Format, sodass der Programmierer nicht mit drei verschiedenen Formaten arbeiten muss.

• HypPy bietet Funktionen zum Arbeiten mit Wellenlängen anstelle von Bandnummern. Bandnummern haben keine physikalische Bedeutung, Wellenlängen hingegen schon. • Rohbilddaten haben möglicherweise keine Bänder in der Reihenfolge zunehmender Wellenlängen. Dah HypPy kann Wellenlängen im Bild im laufenden Betrieb sortieren.

• HypPy kann die Bad Band List (bbl) verwenden und als schlecht markierte Bänder überspringen. • HypPy kann ENVI-Spektralbibliotheken verarbeiten, als wären sie ENVI-Bilder.

• Bei Bedarf werden Spektren im laufenden Betrieb neu abgetastet. Dadurch entfällt die Notwendigkeit, vor der Kombination von Spektren aus verschiedenen Quellen eine erneute Spektralabtastung durchzuführen.

• HypPy ist eine kostenlose Open-Source-Software und daher ein ideales Werkzeug für den Bildungsbereich. Studierende können das Paket kostenfrei abholen und auch zu Hause nutzen.

In den folgenden Abschnitten werden die oben genannten Punkte genauer erläutert.

2.1. Die Nutzungsebenen: GUI, CLI und API

Bildverarbeitungspakete unterstützen verschiedene Arten von Schnittstellen zur Software. Das Paket Khoros [8] unterstützt beispielsweise eine GUI, CLI und API. Darüber hinaus bietet Khoros eine Schnittstelle für visuelle Programmierung namens Cantata [9]. Der Aufbau einer visuellen Programmierumgebung ist komplex und liegt nicht im Rahmen von HypPy. Ähnlich wie Khoros bietet HypPy jedoch eine GUI, CLI und API.

Der Vorteil dieser drei Ebenen ist ihre Flexibilität. GUIs machen die Verwendung der Tools auch für Anfänger einfach. Anhand der Benutzeroberfläche ist klar ersichtlich, welche Ein- und Ausgaben ein bestimmtes Tool benötigt. Darüber hinaus können mit einer GUI mehrere Aufgaben gleichzeitig aufgerufen werden. Auf diese Weise können zwei oder mehr Programme parallel ausgeführt werden, wodurch die Multi-Processing-Funktionen des Computers genutzt werden können, ohne dass parallele Codieru

Mit der CLI können Tools auf der Befehlszeile ausgeführt werden. Die wahre Stärke der CLI besteht jedoch darin, dass damit Skripte erstellt werden können, die eine Abfolge von Tools ausführen, um eine vollständige Verarbeitungskette auszuführen. Darüber hinaus kann Skripting auch für zeitaufwändige Aufgaben verwendet werden, die lange ausgeführt werden oder die Verarbeitung großer Bildsät

Die meisten Funktionen von HypPy können mithilfe der API auch in andere Python-Programme importiert werden.

Die GUI eines HypPy-Tools besteht aus einem separaten Modul, das die HypPy- Funktion importiert, die GUI erstellt und die Parameter an die HypPy-Funktion übergibt. Ein Modul enthält die GUI, während ein anderes Modul die API und die CLI enthält. Beispielsweise enthält das Modul edgy.py die Funktionen für die hyperspektrale Kantenfilterung sowie den Befehl

Machine Translated by Google

Zeilenschnittstelle. Die GUI wird über das Modul tkEdgy.py aufgerufen. In diesem Fall sind die drei Ebenen wie folgt:

• GUI-Programm: tkEdgy.py; • CLI-Programm: edgy.pyl • API-Funktion: edgy().

Über das HypPy-Hauptmenü kann das GUI-Programm tkEdgy.py gestartet werden. Das Edgy Der Filter ist als „Räumlich-Spektraler Kantenfilter“ im Menü „Filter“ zu finden.

2.1.1. Benutzeroberfläche (GUI)

Die meisten HypPy-Tools sind in zwei Programme aufgeteilt, eines für die GUI und eines für die API und die CLI. Die meisten GUIs sind einfache Schnittstellen zum Festlegen von Eingabe- und Ausgabeparametern, die an die API übergeben werden. Ein Beispiel für eine GUI finden Sie in Abschnitt 2.4.1, in dem das Band Math Tool erläutert wird.

Die Benutzeroberflächen für die GUI-Verarbeitungstools werden mit Tkinter erstellt, dem GUI-Paket (Graphical User Interface), das Teil der offiziellen Python-Distribution ist. Tkinter ist eine dünne objektorientierte Schicht über Tcl/Tk [10]. Es gibt viele Pakete, mit denen GUIs in Python entwickelt werden können . Beispielsweise ist wxPython ein umfangreicheres Paket (z. B. ein Widget-Set) als Tkinter. Das Tkinter-Modul (Tk-Schnittstelle) ist jedoch die Standard-Python-Schnittstelle zum Tk-GUI-Toolkit [11] und wurde aufgrund seiner Einfachheit und Verfügbarkeit für HypPy ausgewählt. Alle Namen der HypPy-Tools beginnen mit „tk“, um anzuzeigen, dass sie über eine GUI verfügen.

2.1.2. CLI und Skripting Die

Befehlszeilenschnittstelle ist zum Ausführen von Prozessen an der Eingabeaufforderung nützlich. Der wichtigste Grund für eine CLI ist jedoch die Möglichkeit, eine Verarbeitungskette über ein Skript auszuführen . Skripte sind praktisch, um eine Reihe von Befehlen iterativ zu wiederholen. Auf diese Weise kann eine potenziell große Menge an Bildern auf genau dieselbe Weise verarbeitet werden. Das Ausführen von Skripten reduziert Fehler und bietet die Möglichkeit, Befehle, die viel Verarbeitungszeit benötigen, über Nacht auszuführen .

HypPy hält sich an die Konventionen, die für die Bereitstellung von Befehlszeilenoptionen sowie für die Eingabe und Ausgabe auf der Befehlszeile verwendet werden. Zu diesem Zweck verwendet es das Python-Standardmodul namens argparse, das eine praktische Möglichkeit bietet, Befehlszeilenschnittstellen zu erstellen, die sowohl in Unix- als auch in Windows-Shells verwendet werden können.

2.1.3. Die API

Eine Anwendungsprogrammierschnittstelle (API) ist eine Softwareschnittstelle, über die eine Software einer anderen Software Funktionalität bietet. In Python kann die Funktionalität eines Programms mithilfe der Importanweisung in ein anderes Programm importiert werden. Ein wichtiger Teil der API ist die API-Spezifikation, die die Funktionen und Datenstrukturen beschreibt, die zur Nutzung der verfügbaren Funktionalität erforderlich sind.

Derzeit ist in HypPy kein spezielles API-Handbuch vorhanden, aber wichtige Klassen wie Image und Spectrum werden im Programmierhandbuch bzw. im Spectral Math Manual behandelt. Darüber hinaus werden viele HypPy-Funktionen im HypPy Scripting Manual (CLI) und im HypPy User Manual (GUI) beschrieben, da CLI und GUI oft als dünne Hülle um diese Funktionen fungieren.

Da Datenobjekte ein wichtiger Bestandteil der API sind, sind HypPys Header und Image Objekte werden weiter unten besprochen, siehe Abschnitt 2.2.

2.1.4. Das Top-Level-Menü Das Top-

Level-Menü von HypPy kann mithilfe der Menükonfigurationsdatei hyppymenu.cfg konfiguriert werden. Funktionen können Python-Skripte (normalerweise die HypPy-GUI-Programme), PDF- oder HTML-Dateien oder vollständige URLs sein. Auf diese Weise kann das Top-Menü zum Starten von Skripten sowie zum Anzeigen von Dokumenten wie Handbüchern oder Webseiten verwendet werden.

Machine Translated by Google

Die Tools werden als separate Prozesse ausgeführt. Daher können mehrere HypPy-Tools gleichzeitig gestartet und parallel ausgeführt werden. Abbildung 1 zeigt das obere Menü von HypPy.

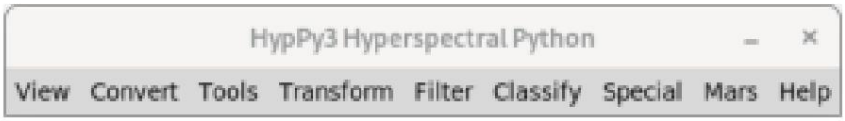


Abbildung 1. Das oberste Menü von HypPy.

Python-Skripte werden von Python als Unterprozesse ausgeführt. Die PDF-, HTML- oder URL-Datei wird an den Standard-Reader des Systems delegiert, der den PDF-Reader oder Webbrowser auf dem System startet .

2.2. Bildformate

HypPy hat die ENVI-Formate als seine nativen Formate übernommen. Die ENVI-Bild- und ENVI-Spektralbibliotheksformate sind Binärdateien mit einer begleitenden Textheaderdatei. ENVI unterstützt in seinem nativen Format drei verschiedene Arten von Interleaving: Band-Sequential (BSQ), Band-Interleaved by Pixel (BIP) und Band-Interleaved by Line (BIL).

HypPy unterstützt alle drei ENVI-Formate: BSQ, BIP und BIL. Darüber hinaus kann HypPy eine ENVI-Spektralbibliothek auf ein Bild abbilden. Die Spektralbibliothek wird dann als y x x z - Bild dargestellt, wobei y die Anzahl der in der Spektralbibliothek gespeicherten Spektren ist, x immer 1 ist (die Spektralbibliothek ist als Bild mit nur einer Pixelspalte sichtbar) und z die Anzahl der Bänder in der Spektralbibliothek ist. Der Vorteil besteht darin, dass Spektralbibliotheken genauso wie Spektralbilder angezeigt werden können. Darüber hinaus kann eine für Spektralbilder vorgesehene Verarbeitungskette zunächst an einer Spektralbibliothek getestet werden, um zu sehen, welche Auswirkungen die Verarbeitungskette auf die verschiedenen in der Spektralbibliothek vorhandenen Mineralien hat.

Aus Gründen der Konsistenz bei der Kodierung werden auch Einzelbandbilder auf drei Dimensionen abgebildet , wobei die Banddimension immer 1 ist. Es gibt eine Ausnahme von dieser Regel: Klassifizierungsbilder werden als zweidimensionale Einzelbandbilder abgebildet. Klassifizierungsbilder haben eine andere Verwendung und Semantik als Hyperspektraldaten, und aus diesem Grund werden Klassifizierungsbilder in HypPy anders behandelt.

Intern bildet HypPy alle drei Formate, BSQ, BIP und BIL, auf ein Array mit einem BIP ab. Struktur und entlastet den Programmierer somit von der Arbeit mit drei verschiedenen Formaten.

2.2.1. Header- und Bildobjekte

Das Header-Modul von HypPy wird zum Lesen und Schreiben von ENVI-Header-Dateien verwendet. ENVI-Bilder und Spektralbibliotheken bestehen normalerweise aus zwei Dateien, den Daten und dem Header. Die Datendateien enthalten rohe Binärdaten. Die Header-Dateien enthalten für Menschen lesbaren Text.

In HypPy wird der ENVI-Header als Python-Objekt der Klasse Header dargestellt. Objekte dieser Klasse enthalten Attribute, die die Attribute widerspiegeln, wie sie in den ENVI-Headerdateien zu finden sind. Im Python-Code gibt es einige Unterschiede. Erstens können die Namen der Attribute im ENVI-Header Leerzeichen enthalten, was in einer Programmiersprache wie Python unpraktisch ist, da Variablennamen keine Leerzeichen enthalten dürfen. Es wurde entschieden, die Leerzeichen in ENVI-Attributnamen durch einen Unterstrich zu ersetzen. Beispielsweise wird das Attribut „data-type“ im ENVI-Dateiheader i Zweitens kann sich das Erscheinungsbild einiger Werte im ENVI-Header in Python ändern. Beispielsweise werden IDL-Datensätze, wie sie in ENVI-Headern erscheinen, wie etwa Werte des Wellenlängenattributs, in den Standard-Python-Listendatentyp geändert. Die Übersetzung zwischen IDL-Datensätzen und Python-Listen wird intern von HypPy gehandhabt.

HypPy unterstützt alle IDL/ENVI-Binärrohdatentypen von vorzeichenlosen 8-Bit-Ganzzahlen bis hin zu vorzeichenbehafteten 64-Bit-Ganzzahlen und von einfachen 32-Bit-Gleitkommazahlen bis hin zu komplexen 64-Bit-Zahlen mit doppelter Genauigkeit und einem Real-Imaginär-Paar mit doppelter Genauigkeit. Darüber hinaus unterstützt HypPy durch die Verwendung des Feldnamens „Byte-Order“ sowohl Big-Endian- als auch Little-Endian-Byte-Order-Dateien.

Machine Translated by Google

HypPy hat zwei Factory-Funktionen in seinem Bildmodul, die Open()-Funktion zum Öffnen vorhandene Bilder und die New()-Funktion zum Einrichten neuer (leerer) Bilder.

Die Factory-Funktion Open() gibt eines der Bildobjekte der Klasse ImageBIP, Im-ageBIL, ImageBSQ, Classification oder ImageSL (eine als Bild zugeordnete Spektralbibliothek) zurück. Bilder, auf die mit der Funktion Open() zugegriffen wird, können nur gelesen, nicht geschrieben werden. Um bestehende Bilder zu schützen, werden diese immer schreibgeschützt geöffnet.

Die Factory-Funktion New() gibt eines der Bildobjekte der Klasse ImageBIP, Image- BIL, ImageBSQ oder Classification zurück. Neben diesen Klassen wird auch die Klasse ImageSL unterstützt, die eine Bilddarstellung einer ENVI-Spektralbibliothek ist.

2.2.2. Speicherzuordnung

HypPy verwendet die Speicherzuordnung von Numpy, „numpy.memmap“, um binäre Datendateien auf Arrays abzubilden. Speicherzugeordnete Dateien werden verwendet, um auf große Dateien auf einer Festplatte zuzugreifen, ohne die gesamte Datei zuerst in den Speicher lesen zu müssen. NumPy-Memmap-Objekte sind arrayähnliche Objekte; sie können auf die gleiche Weise wie Numpy-Arrays bearbeitet werden. Da in HypPy die BIP-Struktur für das interne Array gewählt wurde, müssen die Achsen der beiden anderen Formate, BSQ und BIL, vertauscht werden, damit sie in der zurückgegebenen Speicherzuordnung wie BIP aussehen. Dies wird durch Aufrufen der Transponierungsfunktion der Speicherzuordnung erreicht. BSQ wird mithilfe einer Transponierung (1, 2, 0) abgebildet. BIL wird mithilfe einer Transponierung (0, 2, 1) abgebildet. Das

Man könnte meinen, dass die Verwendung der Transponierungsfunktion eine Kopie der Daten mit auf andere Weise vertauschten Datenachsen erstellen würde. Dies ist jedoch nicht der Fall, da Numpy für bestimmte Vorgänge eine sogenannte „Ansicht“ der Daten unterstützt. Eine Ansicht erstellt einen Verweis auf ein vorhandenes Array, wobei die erforderliche Neuorganisation der Daten zu den Metadaten hinzugefügt wird . Dies hat zwei Vorteile: Erstens erstellt die Ansicht keine Kopie der Daten im Speicher, und zweitens werden alle Änderungen an der Ansicht sofort in den Originaldaten widergespiegelt, wenn die Ansicht im Lese-/Schreibmodus geöf

Darüber hinaus können Bytereihenfolge und Byte-Reihenfolge durch Erstellen einer Ansicht geändert werden. Die drei Formate (BSQ, BIP und BIL), Bytereihenfolge (Little Endian und Big Endian) und Datentypen (von vorzeichenlosen 8-Bit-Ganzzahlen bis zu 64-Bit-Gleitkommazahlen usw.) können durch Erstellen einer Ansicht der Daten in der Datei verwaltet werden. Für die Eingabe- und Ausgabebilddaten muss keine Kopie im Speicher des Computers vorhanden sein. Auf diese Weise kann HypPy große Dateien verarbeiten, die normalerweise nicht in den Speicher passen würden .

Zum Erstellen von Ansichten werden die folgenden Funktionen angewendet: • Erstellen einer Speicherzuordnung: numpy.memmap(fname, mode = 'r', shape = shape, dtype = data_type); • Ändern der Byte-Reihenfolge: Methode newbyteorder(); • Vertauschen der Achsen von BSQ nach BIP: Methode transpose(1, 2, 0); • Vertauschen der Achsen von BIL nach BIP: Methode transpose(0, 2, 1).

Leider erzeugt das Ändern des Datentyps einer Eingabedatei eine Kopie im Speicher, nicht eine Ansicht:

• Ändern Sie den Datentyp der Eingabe: Methode astype(dtype).

Es gibt jedoch einige Einschränkungen bei der Verwendung von Numpy-Speicherzuordnungen. Nicht alle Operationen und Funktionen, die auf Numpy-Arrays verwendet werden können, können auf Speicherzuordnungen verwendet werden, und nur eine begrenzte Anzahl von Operationen oder Funktionen auf einer Speicherzuordnung kann eine Ansicht zurückgeben.

Darüber hinaus hat die Verwendung einer Speicherzuordnung selbst auch Nachteile. Das Einrichten einer Speicherzuordnung verbraucht etwas zusätzlichen Speicher, und die gesamte Neuorganisation einer Ansicht muss bei jedem Zugriff auf die Daten in der Datei sofort ausgeführt werden. Auf diese Weise ist die Speicherzuordnung weniger effizient als das Erstellen einer Array-Kopie im Speicher.

Die Vorteile einer Speicherzuordnung liegen auf der Hand: Sie ermöglicht die Handhabung großer Datensätze , die nicht in den physischen Speicher des Computers passen. Darüber hinaus werden die Daten intern immer auf dieselbe BIP-Struktur abgebildet; in einem Programm ist es nicht erforderlich,

Machine Translated by Google

um das tatsächliche Interleave-Format der Datendatei zu verfolgen. All dies wird transparent von der Speicherzuordnung und den zugehörigen Ansichten gehandhabt.

2.2.3. Spektralbibliotheksformate

HypPy unterstützt zwei grundlegende Arten von Spektralbibliotheken: ENVI-Spektralbibliotheken und ASCII- oder textbasierte Spektralbibliotheken. Bei ENVI-Spektralbibliotheken besteht ein großer Nachteil darin, dass die Liste der Wellenlängen für alle Spektren in der Spektralbibliothek dieselbe sein muss. Daher müssen die Anzahl der Bänder und die zentralen Wellenlängen aller enthaltenen Spektren für alle enthaltenen Spektren gleich sein . Ein Vorteil dieses Formats besteht darin, dass es ein Binärformat ist und wie ein Bild angesprochen werden kann, was die Möglichkeit bietet, Spektralbibliotheken auf dieselbe Weise wie Spektralbilder zu verarbeiten. Dies kann verwendet werden, um Verarbeitungsketten an Spektralbibliotheken zu testen, bevor sie auf Bildern ausgeführt werden. Ein Verfahren zum Erkennen bestimmter Mineralien kann beispielsweise an der USGS-Spektralbibliothek [12] getestet werden , bevor es auf ein Bild angewendet wird. Probleme mit falschen Positivwerten bestimmter Mineralien können in den frühen Phasen der Entwicklung einer Verarbeitungskette er

Andererseits bieten die ASCII-Spektralbibliotheken mehr Flexibilität. Textdateien können in andere Programme importiert werden und die Spektralbereiche der einzelnen Spektren müssen nicht alle gleich sein. Erschwerend kommt hinzu, dass Spektren vor der Verwendung neu abgetastet werden müssen HypPy resampelt Spektren im laufenden Betrieb neu und verwendet dabei die Funktion interp1d() aus dem Modul scipy.interpolate, um beide Wellenlängensätze zu vereinen. Ein Zwischenschritt zum erneuten Sampling von Spektren ist nicht erforderlich.

2.2.4. Schlechte Bandlisten

Das ENVI-Bildformat unterstützt das Konzept einer sogenannten Bad Band List (BBL), die in einer ENVI-Headerdatei durch das Schlüsselwort BBL gekennzeichnet ist. Die BBL muss so viele Elemente enthalten, wie das Bild Bänder enthält. Die Bänder werden nacheinander als schlecht oder gut markie Eine Null (0) für ein bestimmtes Band bedeutet, dass die Daten in diesem Band aufgrund eines Sensorfehlers oder übermäßigen Rauschens nicht vertrauenswürdig sind. Eine Eins (1) für ein Band bedeutet g HypPy kann die Liste der fehlerhaften Bänder berücksichtigen, sodass die fehlerhaften Bänder nicht in den Array-Daten auftauchen. In den meisten HypPy-Tools kann dies angezeigt werden, indem die Option use_bbl auf true ges Wenn diese Option auf „false“ gesetzt ist, sind alle Daten im Array sichtbar, einschließlich der fehlerhaften Bänder. Ein solches virtuelles Image, bei dem use_bbl auf „true“ gesetzt ist, kann weniger Bänder aufweisen als die ursprüngliche Datendatei.

Intern wird dies von einem Index-zu-Index-Array gehandhabt, das als Indexierungsarray verwendet wird, um auf mehrere Elemente gleichzeitig zuzugreifen. Der Nachteil besteht darin, dass die Handhabung solcher virtuellen Bilder langsamer ist als die Verwendung des vollständigen Bildes. Ein weiterer Nachteil besteht darin, dass die Bandnummern im virtuellen Bild nicht mit den Bandnummern im Originalbild übereinstimmen. Eine der Designideen von HypPy besteht jedoch darin, die Wellenlänge anstelle der Bandnummer zu verwenden, um ein Spektralband eines Bildes anzusprechen (siehe nächsten Abschnitt).

2.2.5. Sortieren von Wellenlängen

Sensoren bestehen oft aus mehreren Detektorbänken, die jeweils einen bestimmten Wellenlängenbereich abdecken . Die Daten dieser Detektoren können sich in ihren Wellenlängenbereichen überlappen oder sogar in unerwarteter Reihenfolge auftreten. So enthalten die Rohdaten des OMEGA- Instruments der europäischen Raumsonde Mars Express beispielsweise die Daten der Detektorbänke SWIR1 (0,93 µm bis 2,69 µm) und SWIR2 (2,53 µm bis 5,09 µm) vor den Daten des VNIR-Detektors (0,36 µm bis 1,07 µm) [13]. Ein anderes bekanntes Hyperspektralinstrument, AVIRIS, besteht aus vier Spektrometern mit sich überlappenden Bereichen [14]. Beachten Sie, dass redundante Bänder in der spektralen Überlappung der AVIRIS-Spektrometer auch entfernt oder auf die Liste der fehlerhaften Bänder gesetzt werden können , siehe beispielsweise [15].

Offensichtlich können die vertauschte Reihenfolge und die überlappenden Bereiche beim Plotten oder Verarbeiten solcher Daten zu Problemen führen. Daher war eines der Designkriterien von HypPy, die Wellenlängen eines hyperspektralen Datenwürfels im Handumdrehen sortieren zu können. In einem Kompromiss zwischen Speichernutzung und CPU-Nutzung konvertiert HypPy das Eingabebi

Machine Translated by Google

Es muss das Bild im laufenden Betrieb konvertieren. Viele HypPy-Tools bieten beim Öffnen eines Bildes die Option „sort_wavelengths“.

HypPy sortiert die Wellenlängen, indem es eine Index-zu-Index-Tabelle als Attribut der Bildobjekte führt. Diese Tabelle übersetzt die Bandnummern eines Eingabebildes in Bänder, die nach Wellenlänge sortiert sind. Der Vorteil besteht darin, dass die Daten intern so aussehen, als wären ihre Bänder nach Wellenlänge sortiert. Der offensichtliche Nachteil besteht darin, dass die Bandnummern des Eingabebildes möglicherweise nicht mit den Bandnummern des ursprünglichen Bildobjekts übereinstimmen.

Wenn Bänder jedoch über Wellenlängen und nicht über Bandnummern adressiert werden, ist dieser letzte Punkt irrelevant. Wellenlängen haben eine physikalische Bedeutung und sollten die bevorzugte Methode zur Adressierung von Bildbändern sein. Bandnummern können sich ändern und von einem Remote-Sensor zum anderen sogar eine andere Bedeutung haben. Leider haben nicht alle Bilder Wellenlängen, entweder weil die Wellenlängen fehlen oder die Bilder keine Spektraldaten enthalten, weshalb Bandnummern in HypPy weiterhin verw

2.3. Betrachter

2.3.1. Bildbetrachter Der

Bildbetrachter von HypPy bietet eine schnörkellose Anzeige zum Betrachten von Hyperspektralbildern und Abrufen von Spektren. Der Bildbetrachter wurde für eine schnelle interaktive Beurteilung der Bildqualität und den Abruf von Endmember-Spektren entwickelt. Seine Einfachheit macht ihn auch für unerfahrene Benutzer von Hyperspektraldaten leicht zu verwenden. Mit Schiebereglern kann die Anzeige von einem Band zum anderen geändert werden. Im Farbmodus stehen drei Schieberegler zur Verfügung, einer für jeden der RGB-Kanäle der Anzeige. Eine Statusleiste am unteren Rand des Fensters zeigt Informationen zu Wellenlängen der Bänder, Cursorposition und mehr. Mit den Schiebereglern und der linken Maustaste können die Bänder und Spektren des Eingabebildes hinsichtlich Qualität und Informationsgehalt beurteilt werden. Abbildung 2 zeigt das Hyperspektralbild einer Gesteinsprobe. Diese spezielle Gesteinsprobe 15-91-FE ist Teil des Fen-Komplexes in Norwegen, der aus einer Karbonatitintrusion besteht [16].

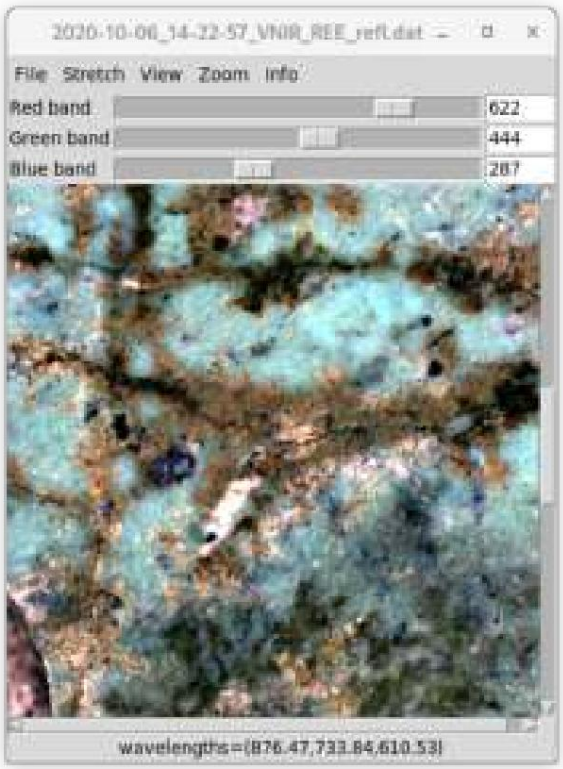


Abbildung 2. Bildbetrachter zeigt eine Falschfarbenkomposition der Probe 15-91-FE. Roter Kanal 876 nm, grüner Kanal 734 nm und blauer Kanal 611 nm. Die Schieberegler für Rot, Grün und Blau können verwendet werden, um die Bänder für die RGB-Farbkanäle zu ändern.

Machine Translated by Google

Unter dem Dateimenü können zwei Optionen eingestellt werden, um zu steuern, ob die Bänder vor dem Lesen sortiert werden müssen oder ob die BBL eines Bildes verwendet werden soll.

Das Anzeigeprogramm kann zwei Bilder gleichzeitig öffnen. Das erste Bild ist im Viewer immer sichtbar. Das zweite Bild, das sogenannte „Schattenbild“, ist nie sichtbar.
Wenn jedoch das Schattenbild vorhanden ist, werden Spektren aus dem Schattenbild und nicht aus dem sichtbaren Bild gewonnen. Dies ist wie ein GIS, bei dem die erste Ebene oben liegt und die zweite Ebene unter dem ersten Bild verborgen ist. Diese Option ist nützlich, um ein Klassifizierungsbild oder ein Maskenbild in die erste Ebene zu laden, während Spektren aus dem Schattenbild entnommen werden. Dadurch lässt sich beurteilen, ob ein berechnetes Klassifizierungs- oder Maskenbild sinnvoll ist . Darüber hinaus ist diese Option nützlich, um eine Wellenlängenkarte zusammen mit einem Hyperspektralbild zu untersuchen, aus dem die Wellenlängenkarte abgeleitet wurde.

Über das Streckmenü lassen sich verschiedene Standardmethoden zur Streckung des Bildes nutzen [17]. Es sind Methoden wie 1–99 % und Histogrammausgleich implementiert. Darüber hinaus kann das Bild invertiert, die Sättigung im Handumdrehen verbessert oder mit einer Reihe verfügbarer Farbschemata pseudogefärbt werden.

Ein Linksklick irgendwo auf das Bild öffnet ein Plot-Fenster mit einem Diagramm, das das Spektrum des Pixels an dieser Stelle darstellt, siehe Abbildung 3. Weiteres Klicken auf das Bild öffnet weitere Diagramme im Plot. Durch Entfernen des Plot-Fensters beginnt der Vorgang von vorne mit einem leeren Plot-Fenster. Auf diese Weise können die Spektren verschiedener Punkte im Bild gemeinsam visualisiert werden.

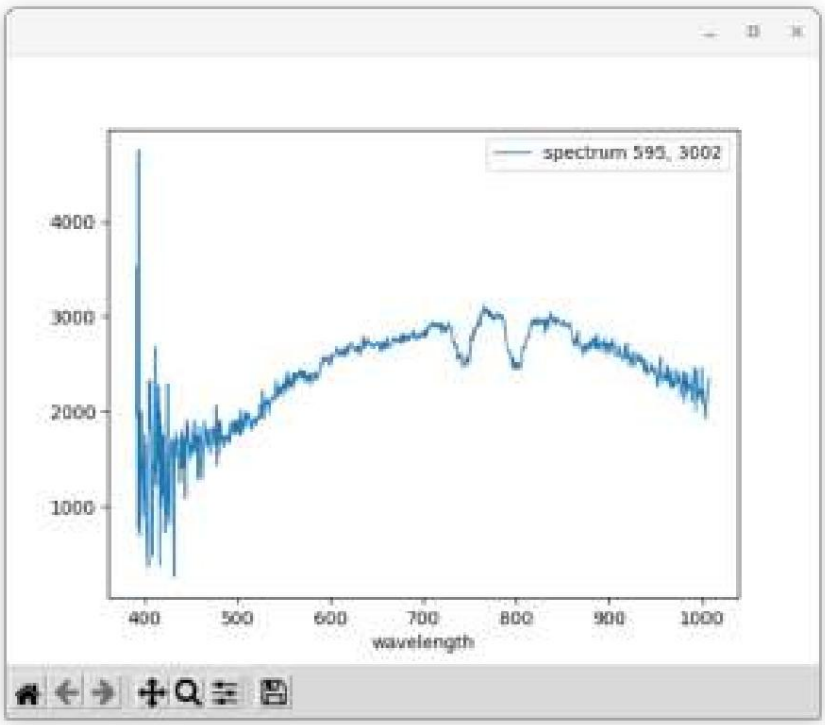


Abbildung 3. Plotfenster mit dem Spektrum der Pixelposition 595, 3002. Das Spektrum zeigt zwei Absorptionsmerkmale bei 740 und 800 nm. Diese Merkmale sind auf Seltenerdelemente (insbesondere Neodym) zurückzuführen, die in Monazitkristallen enthalten sind. Das Spektrum unter 450 nm ist aufgrund der geringen Empfindlichkeit der hyperspektralen VNIR-Kamera im Blaubereich verrauscht. Die Symbolleiste am unteren Rand des Fensters ist eine standardmäßige Matplotlib-Abbildungssymbolleiste und verfügt über Schaltflächen zum Navigieren in der Abbildung: Home, Zurück, Vorwärts, Schwenken/Zoomen, Auf Rechteck zoomen und Subplot-Konfiguration. Mit der Schaltfläche „Speichern“ auf der rechten Seite können Sie die Abbildung in verschiedenen Formaten w

Mithilfe des „Wertefensters“ können diese Spektren auch als Textdatei, Endmember oder Spektralbibliothek gespeichert oder zur Visualisierung und Weiterverarbeitung in andere Software importie
Ein Verzeichnis mit Endmember-Spektren kann als Spektralbibliothek in den Klassifikatoren, dem Spektralwinkel-Mapper und der linearen Entmischung von HypPy verwendet werden.

Machine Translated by Google

Das Bildanzeigeprogramm HypPy verwendet die Module Image und ImageTk aus der Python Imaging Library (PIL, jetzt Pillow [18] genannt) zur Anzeige von Bildern und matplotlib zur Erstellung von 2D-Graphen von Spektren.

2.3.2. Spektralbibliothek-Viewer

Der Spektralbibliotheks-Viewer ist ein Viewer für Spektralbibliotheken vom Typ ENVI und ASCII. Nach Auswahl der Eingabespektrenbibliothek werden die Namen der in der Spektralbibliothek enthaltenen Spektren in einer Liste angezeigt. Durch Doppelklicken auf einen Namen wird ein Plot-Fenster geöffnet, in dem das Spektrum des ausgewählten Minerals oder Materials angezeigt wird. Durch Auswahl mehrerer Spektren werden diese Spektren zusammen im Plot-Fenster dargestellt, sodass ein Vergleich der Spektren möglich ist.

Der Spektralbibliotheksbetrachter unterstützt verschiedene ASCII-Formate, das CRISM-Spektralbibliotheksformat (PDS), USGS SpecLib Version 6 und 7 [12], das ASD-Format [19] und Klartext-Zweispaltendaten . Der Spektralbibliotheksbetrachter kann einen gesamten Verzeichnisbaum rekursiv lesen. Siehe Abbildungen 4 und 5, die die Schnittstelle des Spektralbibliotheksbetrachters und ein Plotfenster mit einem Beispielspektrum von Monazit zeigen.

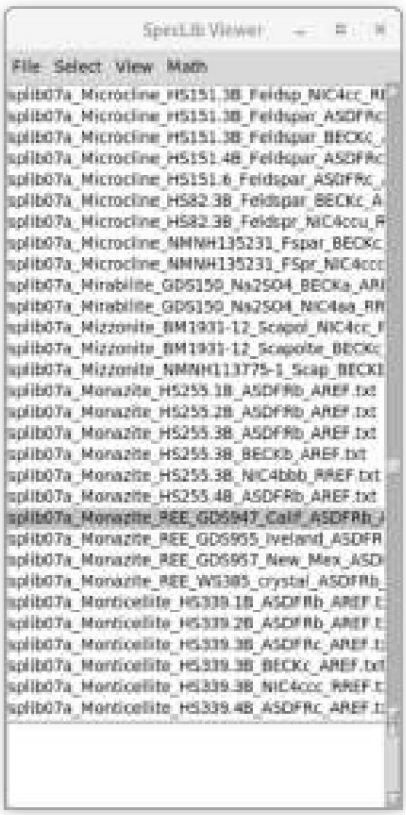


Abbildung 4. Schnittstelle des Spektralbibliotheksbetrachters, in diesem Fall geladen mit der Spektralbibliothek Version 7 des USGS [12].

Darüber hinaus unterstützt der Spektralbibliotheks-Viewer die Manipulation von Spektren mithilfe eines „Spektralmathematik“-Ausdrucks. Eine Teilmenge von Numpy-Funktionen kann auf Spektren angewendet werden. Indizes und Slices können auf Spektren angewendet werden, und es stehen zahlreiche mathematische Operatoren und Funktionen zur Verfügung, um Verarbeitungsvorgänge auf Spektren aufzubauen und diese an (einer Auswahl) einer Spektralbibliothek zu testen.

Spektren mit unterschiedlichen Wellenlängen und Wellenlängenbereichen werden automatisch im laufenden Betrieb neu abgetastet. Alle diese Funktionen sind in der API verfügbar und in der Spektrumklasse von HypPy implementiert (siehe auch Abschnitt 2.4.2 unten). Ergebnisse der spektralen mathematischen Operation im Spektralbibliotheks-Viewer können gespeichert werden, indem zuerst ein Ausgabe

Machine Translated by Google

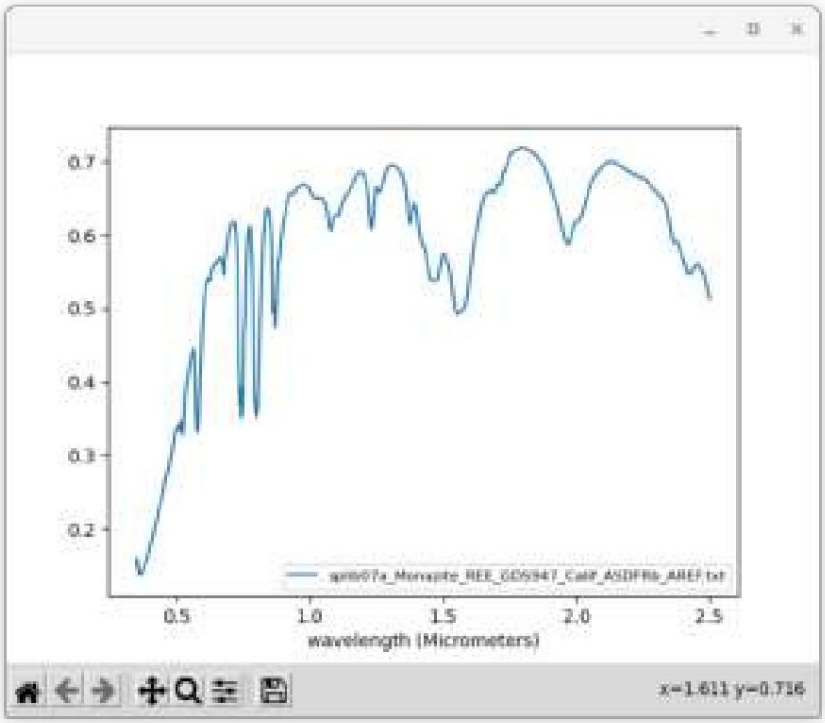


Abbildung 5. Plotfenster mit dem Spektrum Monazite_REE_GDS947_Calif_ASDFRb_AREF. Diese Probe enthält bekanntermaßen Ce, La, Nd und Th. Beachten Sie die Nd-Absorptionsmerkmale bei 740 nm und 800 nm.

2.4. Werkzeuge

Im Folgenden wird eine Teilmenge der Tools von HypPy ausführlicher erläutert.

2.4.1. Band-Mathematik-Tool

Multispektrale und hyperspektrale Bilder können als eine Reihe von Spektralbändern betrachtet werden. Häufig werden Funktionen erstellt, um solche Daten bandweise zu verarbeiten. Das Bandmathematik- Tool implementiert dies.

Ein Beispiel für eine solche Bandmathematikoperation ist die Berechnung des NDVI, bei der die Differenz zwischen einem Nahinfrarotband und einem roten Band normalisiert wird, indem sie durch die Summe des Nahinfrarotbands und des roten Bands dividiert wird. Für den NDVI wird normalerweise für das rote Band ein Bereich von 400 nm bis 700 nm verwendet, und für Infrarot beträgt der B Im folgenden Beispiel wählen wir für Rot ein Band um 675 nm und für Infrarot ein Band um 1000 nm. In HypPy kann der NDVI eines AVIRISNG-Bildes mit einem der folgenden Ausdrücke berechnet werden: (1) oder (2).

(i1[59] ÿ i1[124])/(i1[59] + i1[124]), (i1(675,0) (1)

ÿ i1(1000,0))/(i1(675,0) + i1(1000,0)). (2)

In der Bandmathematik wird die Folge der Bilddateinamen mit den Variablen i1, i2 usw. verknüpft. Im Beispiel des NDVI wird nur ein Eingabebild benötigt, das mit der Variablen i1 verknüpft ist. Die Bilder werden beginnend bei 1 nummeriert und die Bandnummern beginnend bei 0. Im Beispiel ist Band 59 das Band mit einer Mittenwellenlänge von 676,9 nm und Band 124 hat eine Mittenwellenlänge von 1002,4 nm. Bitte beachten Sie, dass in diesem Beispiel die BBL verwendet wird und die ersten sieben Bänder, die als schlecht markiert sind, nicht vorhanden sind.

In der Bandmathematik ist es auch möglich und sogar vorzuziehen, Wellenlängen anstelle von Bandnummern zu verwenden. Das Band mit der nächstgelegenen zentralen Wellenlänge wird automatisch ausgewählt. Beachten Sie, dass in der Bandmathematik, wenn Wellenlängen anstelle von Bandnummern verwendet werden, diese Wellenlängen in runde Klammern und nicht in eckige Klammern eingesch

Machine Translated by Google

Dies wird durch einen „Aufruf“ der Bildklasse implementiert. Python-Objekte können so gestaltet werden, dass sie sich wie Listen oder Iteratoren verhalten und einen Index in eckigen Klammern oder eine Funktion mit einer Parameterliste in runden Klammern haben. Abbildung 6 zeigt ein Beispiel für ein Bandve

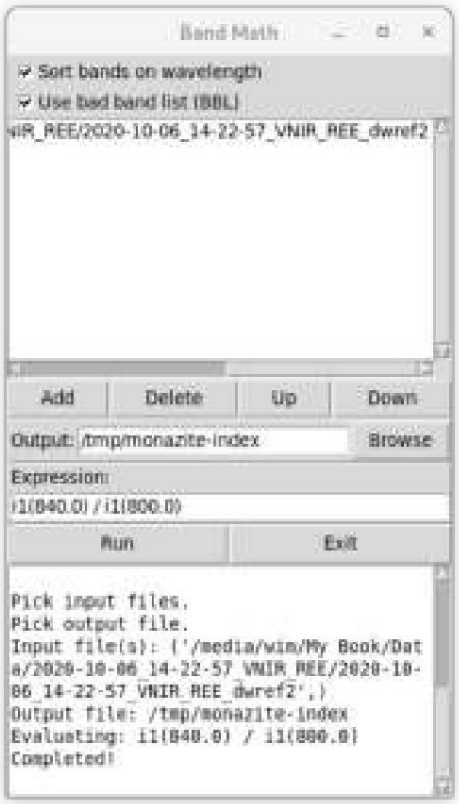


Abbildung 6. Schnittstelle des Bandmathematik-Tools. Die Formel zeigt ein Beispiel für ein Bandverhältnis für einen einfachen Monazitindex basierend auf der Schulter und der Absorptionseigenschaft von Nd bei 840 nm bzw. 800 nm . In diesem Fall gibt es nur eine Eingabedatei i1.

HypPy verwendet die Python-Funktion eval(), um Ausdrücke wie das obige NDVI-Beispiel auszuführen. Bandmathematische Ausdrücke sind im Grunde Python-Ausdrücke. HypPy richtet die Variablen i1, i2 usw. ein. Dies sind HypPy-Bildobjekte, die aufgerufen werden können, um Bänder einer bestimmten Wellenlänge zurückzugeben. Das Verknüpfen der Bildvariablen mit Bilddateien ist ein wesentlicher Schritt bei der Verwendung von Bandmathematischen Ausdrücken. Beachten Sie, dass die in der Bandmathematik verwendete Syntax sich von der in der Spektralmathematik verwendeten unterscheidet. Die Bandmathematik verwendet Bildobjekte, während die Spektralmathematik Spektrumobjekte verwendet; siehe unten.

2.4.2. Spektrumklasse und Spektralmathematik

Spektralmathematik wird verwendet, um mathematische Ausdrücke auf Spektren von Bildern oder Spektralbibliotheken anzuwenden. Bandmathematik verwendet das Bandparadigma; Spektralmathematik verwendet das Spektrumparadigma . In der Spektralmathematik wird das Bild Pixel für Pixel verarbeitet. In der Bandmathematik wird das Bild Band für Band verarbeitet. Spektralmathematik basiert auf der Funktionalität, wie sie in der Spektrumklasse implementiert ist. Die Schnittstelle des Spektralmathematiktools ähnelt der des Bandmathematiktools (siehe Abbildung 6), aber der A

Die Klasse „Spektrum“ wurde zur Manipulation von Spektren erstellt. Die Idee ist, dass ein Spektrumobjekt Informationen über Wellenlängen sowie Werte enthält, sei es Reflexionsgrad, Strahlungsdichte oder andere Daten. Die Klasse „Spektrum“ bietet Zugriff auf die Daten sowie Funktionen zum Bearbeiten der Spektraldaten.

Die folgende Liste fasst einige der Eigenschaften des Spektrum-Containerobjekts zusammen:

- Das Spektrumobjekt enthält Wellenlängen und Werte.
- Das Spektrumobjekt kann einen Namen und eine Beschreibung enthalten.

Machine Translated by Google

• Ein Index kann verwendet werden, um Werte auszuwählen oder Teilmengen des Spektrums zu erstellen. • Eine Ganzzahl als Index bezeichnet eine Bandnummer; eine Gleitkommazahl bezeichnet eine Wellenlänge. • Mathematische Operatoren wie +, ÿ, * und/können auf Spektren angewendet werden. • Das Spektrumobjekt enthält eine Fülle von Operatoren und Methoden.

Derzeit sind etwa 150 Funktionen und 20 Operatoren implementiert (siehe HypPys Handbuch zur Spektralmathematik). • Beim Arbeiten mit Spektren mit unterschiedlichen Wellenlängen werden die Spektren automatis

im laufenden Betrieb neu abgetastet.

Wie bereits gezeigt, können Spektralindizes mithilfe der Bandmathematik implementiert werden. Ein Spektralindex soll einen hohen Wert haben, wenn ein bestimmtes Mineral oder Material vorhanden ist, und einen niedrigen Wert, wenn es fehlt. Ein Beispiel für einen solchen Index wurde im vorherigen Abschnitt gegeben, in dem wir das Beispiel des NDVI vorgestellt haben.

Wenn das Bild jedoch Rauschen enthält, kann auch der resultierende NDVI Rauschen enthalten. In einem solchen Fall kann es vorteilhaft sein, mit Durchschnittswerten in Bandbereichen statt mit einzelnen Bändern zu arbeiten. In der Bandmathematik ist es nicht möglich, Bandbereiche zu verwenden, da die Bänder eines Bereichs einzeln angesprochen werden müssen. In der Spektralmathematik kann jedoch ein Spektralbereich durch einen Ausschnitt des Spektrums ermittelt werden.

Der spektrale mathematische Ausdruck für NDVI könnte nun wie der folgende Ausdruck aussehen (3):

$$(S1[700,0 : 1100,0].\text{Mittelwert}() \text{ ÿ } S1[400,0 : 700,0].\text{Mittelwert}())/$$
$$(S1[700,0 : 1100,0].\text{Mittelwert}() + S1[400,0 : 700,0].\text{Mittelwert}()), \quad (3)$$

wobei S1 die Spektren des ersten Bildes bezeichnet. Die Spektren der nachfolgenden Eingabebilder werden mit S2, S3 usw. nummeriert. In diesem Beispiel wird jedoch nur ein Bild verwendet. Die eckigen Klammern enthalten das, was in Python als Slices bezeichnet wird und durch einen Doppelpunkt ':' g Beachten Sie, dass Gleitkommazahlen als Indizes verwendet werden, die Wellenlängen bezeichnen. Ganzzahlen als Indizes werden als Bandnummern interpretiert. Das Ergebnis eines Ausschnitts eines Spektrums ist ein neues Spektrumobjekt , das eine Teilmenge des ursprünglichen Spektrumobjekts enthält. In diesem Fall enthält die Teilmenge die Spektralwerte ungefähr zwischen 400 nm und 700 nm und zwischen 700 nm un

Bei den Wellenlängen handelt es sich um Näherungswerte, da diese Wellenlängen den **nächstgelegenen Bandnummern zugeordnet werden ,** die möglicherweise nicht die exakt gleiche Wellenlänge wie die zentralen Wellenlängen aufweise HypPy generiert keine Fehler oder Warnungen für die Konvertierung von Wellenlängen in Bänder, um der Geschwindigkeit Vorrang vor Qualitätsprüfungen zu geben, da die Prüfungen viel CPU-Zeit in Anspruch nehmen wü Benutzer und Programmierer können die angeforderten und erhaltenen Wellenlängen jedoch mit den Funktionen wavelength2index() und index2wavelength() überprüfen.

Anschließend wird die Mittelwertfunktion mean() auf diese Teilmengen angewendet. Mit anderen Worten ist das Ergebnis der NDVI, der aus dem durchschnittlichen Spektrum zwischen 400 nm und 700 nm und dem Spektrum zwischen 700 nm und 1100 nm berechnet wird. Auf diese Weise können Spektralindizes erstellt werden, die auf Wellenlängenbereichen statt auf einzelnen Bändern operieren. Der offensichtliche Vorteil ist, dass dadurch das Rauschen in der Ausgabe reduziert wird. Der Nachteil ist, dass solche Ausdrücke langsamer sind als Band-für-Band-Operationen. Darüber hinaus werden im obigen Beispiel die Slices und Mittelwerte zweimal berechnet, da das Spektralmathematik-Tool nur einen einzigen Ausdruck unterstützt. Durch geschickten Einsatz des Walross-Operators ':=', der in Python Version 3.8 [20] eingeführt wurde , kann jedoch eine doppelte Berechnung vermieden werden. Der Ausdruck für den NDVI würde dann wie Ausdruck (4) aussehen:

$$((ir := S1[700,0 : 1100,0].\text{mean}()) \text{ ÿ } (red := S1[400,0 : 700,0].\text{mean}()))/ (ir + red), \quad (4)$$

wobei der Mittelwert des Infrarotbereichs in der Variablen ir und der Mittelwert des Rotbereichs in der Variablen red gespeichert wird und diese Werte später im Ausdruck zur Berechnung des Nenners abgerufen werden.

Bitte beachten Sie, dass wir in den Gleichungen (3) und (4) einen breiten Bereich für Rot und Infrarot verwenden , wobei wir die Chlorophyllabsorption im sichtbaren Teil des Spektrums (400 nm bis 700 nm) als Rot und die Reflexion durch die Blattzellstruktur im Nahinfrarotteil des Spektrums verwenden.

Machine Translated by Google

(700 nm bis 1100 nm) als IR. Die spezifischen Wellenlängenbereiche können je nach verwendetem Fernerkundungssensor oder Technologie variieren. Beispielsweise verwendete Tucker [21] den Bereich 630 nm bis 690 nm für Rot und 750 nm bis 800 nm für IR, die den Bereichen der Landsat MSS- Spektralbänder 5 bzw. 7 entsprechen.

HypPy verwendet die Python-Funktion eval(), um Ausdrücke wie das obige Beispiel zur Mittelwertbildung des NDVI auszuführen. Die spektralen mathematischen Ausdrücke sind Python-Ausdrücke. HypPy richtet jedoch die Variablen S1, S2 usw. ein, die HypPy-Spektrumobjekte sind, die als Python-Containerklasse verwendet werden können. In diesem Sinne ähnelt das Verhalten des Spektrumobjekts einem Python-Listenobjekt, das indiziert, aufgeteilt, durchlaufen usw. werden kann.

In HypPy kann Spektralmathematik im Spektralmathematik-Tool sowie im Spektralbibliotheks -Viewer verwendet werden.

2.4.3. Räumlich-spektrale Filter

Räumlich-spektrale Filter berücksichtigen sowohl spektrale als auch räumliche Informationen . HypPy verfügt über drei Arten von räumlich-spektralen Filtern: Gradientenfilter, Mittelwertfilter und Binning-Filter. Gradientenfilter werden implementiert, indem ein lokaler räumlicher Kernel von Spektren als Vektoren betrachtet und die gewichtete Differenz zwischen diesen Vektoren berechnet wird. Die Mittelwertfilter berechnen die gewichtete Summe dieser Vektoren. Binning-Filter reduzieren die Anzahl der Samples durch Aggregation von Werten.

HypPy verfügt über ein Tool zur hyperspektralen Kantenfilterung. Solche Filter können verwendet werden, um spektrale Variabilität in einer Szene zu erkennen. Bereiche mit einem hohen Wert sind spektral inhomogen und Bereiche mit niedrigen Werten sind spektral homogen. Die Ergebnisse solcher Filter können als erster Schritt bei der Beurteilung der Bildqualität, Klassifizierung oder Segmentierung dienen. Das Ergebnis der Gradientenfilter ist immer ein einbandiges Graustufenbild. Das HypPy-Tool namens „ Hyperspektraler Gradient“ implementiert eine Reihe von Richtungsfiltern (nach oben, unten und zwei Diagonalen) und ungerichteten Filtern (Kanten- und Sobelfilter). Es wurden eine Reihe von Distanzmaßen implementiert, die in Verbindung

Das Distanzmaß des Spektralwinkels (SA) berücksichtigt nur die spektralen Informationen des Eingangsbildes, da der Spektralwinkel nicht auf die Intensität des Eingangsspektrums reagiert . Das Distanzmaß der Intensitätsdifferenz (ID) berücksichtigt dagegen nur die Intensitätsinformationen des Spektrums, nicht aber die spektralen Informationen. Das Distanzmaß der euklidischen Distanz (ED) reagiert sowohl auf die spektralen als auch auf die Intensitätsinformationen der Spektren, obwohl die ED in den meisten Fällen von der Intensität der Eingangsspektren dominiert zu werden scheint. Neben den oben besprochenen Standardmaßen der Spektraldifferenz gibt es zwei weitere Distanzmaße, die in der Hyperspektralanalyse häufig verwendet werden : die Bray-Curtis-Distanz (BC) und die spektrale Informationsdivergenz (SID) [22]. Alle diese Distanzmaße haben ihre eigenen Eigenschaften hinsichtlich der spektralen Ähnlichkeit.

Während die Gradientenfilter einen **zweidimensionalen** 3 × 3-Satz vollständiger Spektren berücksichtigen , berücksichtigen die Mittelwertfilter einen kleineren dreidimensionalen Satz Spektralwerte. Die Mittelwertfilter wenden einen 3 × 3 × 3-Kernel auf den hyperspektralen Datenwürfel an. Der Filter kann alle Werte des Kernels oder einen begrenzten Satz von Kerneln berücksichtigen. Aktuelle Kernels in HypPy berechnen den gewichteten Durchschnitt von 27, 19 oder 7 der Werte des Kernels. Es sind zwei Durchschnittsfunktionen implementiert: der Mittelwert und der Median. Ein Mittelwertfilter ist ein linearer Filter und daher lässt sich seine Wirkung auf das Bild vorhersagen. Ein Medianfilter reduziert die Wirkung von Ausreißern. In dieser Hinsicht kann ein Medianfilter nützlich sein, um Pfeffer-und-Salz-Rauschen herauszufiltern. Die Wirkung eines Medianfilters lässt sich jedoch nicht so einfach vorhersagen wie die eines Mittelwertfilters , was zu unerwarteten Effekten im Ausgabebild führen kann. In bestimmten Fällen können Medianfilter eine mit morphologischen Filtern vergleichbare Wirkung haben, da Bereiche scheinbar schrumpfen oder wachsen, was verwirrend sein und somit die Interpretation des resultierenden Bildes erschweren kann.

Das Ergebnis eines räumlich-spektralen Mittelwertfilters ist immer ein Hyperspektralwürfel mit denselben Abmessungen wie die Eingabe. Wie bei regulären zweidimensionalen räumlichen Filtern müssen jedoch Randeffekte berücksichtigt werden. Der erste und der letzte Wert in jeder der drei Dimensionen (x, y und Band) werden anders behandelt als der Rest des Hyperspektralwürfels.

Machine Translated by Google

Zusätzlich zu diesen Filtern verfügt HypPy über eine Reihe von Binning-Tools (räumliches, spektrales und räumlich-spektrales Binning), um die Anzahl der Samples im Bildwürfel zu reduzieren. Mit diesen Tools können Sie das Datenvolumen verringern, das Rauschen durch Mittelwertbildung reduzieren und nachfolgende Verarbeitungsschritte beschleunigen. Beispielsweise reduziert räumlich-spektrales Binning von xy = 3 × 3 und z = 11 das Datenvolumen um fast den Faktor 100 (3 × 3 × 11 = 99) und verringert die Standardabweichung des Rauschens um fast d

2.4.4. Wellenlängen-Mapper

Das Mapping physikalischer Eigenschaften aus Hyperspektralbildern ist kompliziert und hängt häufig von der Verfügbarkeit von Spektralbibliotheken der Zielmaterialien ab. Viele bestehende Methoden hängen bei der Auswahl der Endelemente und der Nachbearbeitung der Regelbilder von den subjektiven Entscheidungen des Interpreten ab . Wenn keine Vorabinformationen verfügbar sind, sollten Informationen durch Charakterisierung der spektralen Absorptionsmerkmale extrahiert werden. Auf Multispektralbildern können solche Merkmale mithilfe von Indizes gefunden werden, die auf Bandverhältnissen und Bandtiefen basieren; siehe beispielsweise [23]. Hyperspektralbilder zeigen jedoch mehr spektrale Details der Merkmale, was die Mög

Für die Analyse und Interpretation sind Absorptionsmerkmale von Spektralkurven wichtig. Solche Merkmale können als zentrale Wellenlängenposition, Tiefe, Fläche und Asymmetrie parametrisiert werden [24]. Normalerweise sind die wichtigsten Merkmale die zentrale Wellenlängenposition und die Tiefe des Hauptabsorptionsmerkmals. Die Wellenlängenposition eines Absorptionsmerkmals kann zur Identifizierung von Mineralien verwendet werden, da viele Mineralien ihre eigenen einzigartigen Absorptionswellenlängen haben. Die Tiefe eines solchen Merkmals ist ein Hinweis auf die Häufigkeit eines Minerals.

Die zentrale Wellenlänge und Tiefe des tiefsten Merkmals und der darauffolgenden tiefsten Merkmale können automatisch bestimmt werden. Diese Parameter können für alle Pixel (Spektren) eines Hyperspektralbilds berechnet und als Graustufenbilder dargestellt werden, die die Wellenlänge und Tiefe des Merkmals darstellen.

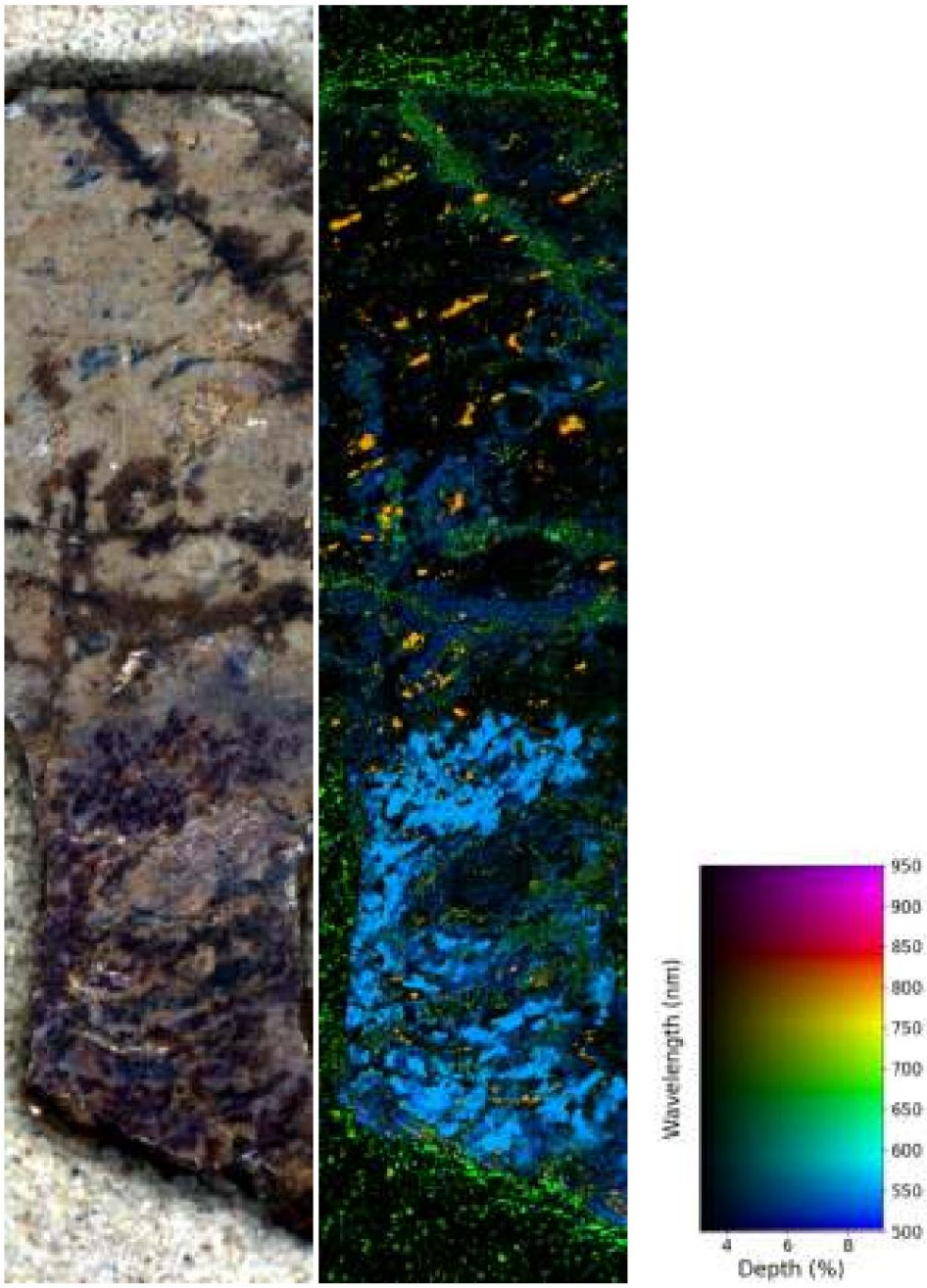
Anschließend können die resultierenden Wellenlängen- und Tiefenbilder mithilfe einer HSI-Transformation (Farbton–Sättigung–Intensität) [17] so fusioniert werden, dass das resultierende Bild die Wellenlänge in Farbe (den Farbton) zeigt. Die Intensität stellt die Tiefe des Merkmals dar. Die Sättigung wird auf 1 gesetzt. Das Wellenlängen-Mapper-Tool von HypPy bestimmt automatisch die Wellenlänge und Tiefe des tiefsten Absorptionsmerkmals aller Pixel eines Hyperspektralwürfels und fusioniert die beiden zu einem Farbbild.

Abbildung 7a zeigt eine Gesteinsprobe in einer natürlichen Farbmischung. Abbildung 7b zeigt das Ergebnis des Wellenlängen-Mappers; Abbildung 7c zeigt die Legende zur Wellenlängenkarte. Der seltenerdhaltige Monazit ist als orangefarbene Körner sichtbar. Die grünen und blauen Körner sind Karbonate. Eine weitere Beschreibung dieser Gesteinsprobe finden Sie in [16].

Der Wellenlängen-Mapper ist unbeaufsichtigt, unkompliziert und wiederholbar; große Bereiche können auf einmal abgebildet werden und Details bleiben im Fokus des gewählten Spektralbereichs erhalten . Das Ergebnis, die sogenannte Wellenlängenkarte, kann von Benutzern mit eingeschränktem Bildverarbeitungshintergrund verstanden werden. Obwohl die Farben auf das Vorhandensein bestimmter Mineralien hinweisen können, ist die resultierende Wellenlängenkarte keine Mineralienkarte. Die Wellenlängenkarte kann als Vorklassifizierungsschritt verwendet werden, um die spektrale Variabilität in der Szene zu verstehen. Das aus der Wellenlängenkarte gewonnene Wissen ist für die zusätzlichen Schritte zur Erstellung einer Mineralienkarte nützlich.

Der Wellenlängen-Mapper benötigt nur zwei Eingabeparameter: den Wellenlängenbereich, auf den fokussiert werden soll, und die Streckung, die für die Tiefe des Merkmals verwendet werden soll. Der Wellenlängen-Mapper wurde erstmals in van Ruitenbeek et al. [25] veröffentlicht, und ein Tutorial zur Methode wurde in [24] veröffentlicht.

Machine Translated by Google



(a) Farbkomposit (b) Wellenlängenkarte c) Legende

Abbildung 7. (a) Natürliche Farbmischung der Gesteinsprobe 15-91-FE (rot 650 nm, grün 550 nm und blau 450 nm). (b) Die Wellenlängenkarte des Bereichs von 500 nm bis 950 nm zeigt die Wellenlänge und Tiefe des tiefsten Merkmals. Monazit ist in Orange (zeigt das 800 nm-Merkmal von Nd) oder Gelbgrün sichtbar, wenn das 740 nm-Merkmal tiefer liegt als das 800 nm-Merkmal. (c) Legende zur Wellenlängenkarte.

2.4.5. Entscheidungsbäume

Ein Entscheidungsbaum ist eine Art Klassifikator, der aus einer Reihe binärer Entscheidungen besteht, mit denen die Klasse für jedes Eingabepixel bestimmt wird. Die Entscheidungen basieren auf den Eigenschaften der Datensätze. Jede Entscheidung unterteilt die Daten in eine von zwei möglichen Klassen oder Klassengruppen. Entscheidungen sind Ausdrücke, die auf ein einzelnes Bild oder eine Reihe von B Jedes dieser Bilder stellt eine Eigenschaft der Eingabepixel dar.

Eine solche Eigenschaft könnte beispielsweise die Wellenlängenposition des tiefsten Absorptionsmerkmals im Spektrum des Eingabepixels sein. Ein Ausdruck könnte erstellt werden, um zu

Machine Translated by Google

ob das Eingabepixel ein Tonmineral mit einem tiefsten Absorptionsmerkmal von etwa 2,2 µm enthalten könnte. Nachfolgende Entscheidungen könnten die Wassermerkmale dieses Pixels überprüfen, um festzustellen, ob das Material einen hohen Kristallinitätsindex aufweist. Nach einer Reihe von Entscheidungen wird dem Pixel eine Klasse zugewiesen.

Entscheidungsbäume sind nützlich, um Klassifikatoren zu erstellen, die auf physikalischen Eigenschaften und nicht auf statistischen Schlussfolgerungen basieren. Dies kann natürlich nur durchgeführt werden, wenn alle für die Entscheidungsbäume erforderlichen Eigenschaften mit ausreichender Sicherheit festgestellt werden können.

Ein Paket wie ENVI unterstützt den Aufbau und die Visualisierung neuer Entscheidungsbäume. ENVI hat jedoch eine Reihe von Nachteilen. Die Art und Weise, wie Entscheidungsbäume in ENVI implementiert werden, ist ziemlich starr, denn sobald ein Knoten im Baum erstellt ist, ist sein Platz im Entscheidungsbaum festgelegt. In einem fertigen ENVI-Entscheidungsbaum ist es unmöglich, Elemente zu verschieben. Wenn Sie in ENVI den Entscheidungsbaum ändern möchten, müssen Sie ihn von Grund auf neu erstellen.

Mit HypPy können ENVI-Entscheidungsbäume ausgeführt werden. Die einzige Voraussetzung ist, dass die für die Entscheidungen verwendeten Ausdrücke in Python-Ausdrücke übersetzt werden können.

Zusätzlich zum ENVI-Entscheidungsbaumformat verfügt HypPy über ein eigenes Entscheidungsbaumformat. Dieses Format ist ein textbasiertes Format, das mit einem Texteditor bearbeitet werden kann.

Zur Visualisierung von Entscheidungsbäumen verwendet HypPy das Graphviz-Paket [26]. Dazu muss der ENVI- oder HypPy-Entscheidungsbaum zunächst in das „dot“-Format konvertiert werden. Die dot-Sprache ist eine umfangreiche Sprache zur Beschreibung von Graphen. Die von ENVI und HypPy verwendeten binären Entscheidungsbäume stellen eine winzige Teilmenge aller möglichen Graphen und Graphentypen dar, die vom Graphviz-Paket unterstützt werden.

2.4.6. Zonenstatistiken

Das Tool „Zonenstatistik“ berechnet Statistiken zu Pixeln eines Eingabebilds innerhalb der durch das Zonenbild definierten Zonen. Das Eingabebild ist ein reguläres Hyperspektral- oder Multispektralbild. Das Zonenbild kann das Ergebnis eines Klassifizierungsbilds oder eines Maskenbilds sein. Für jeden eindeutigen Wert im Zonenbild wird eine Aggregationsfunktion auf dem Eingabebild für alle Positionen ausgeführt, die diesen eindeutigen Wert im Zonenbild aufweisen. Die derzeit implementierten Aggregationsfunktionen sind wie folgt: Mittelwert, Median, Minimum, Maximum, Standardabweichung und Mittelwert plus oder minus zwei Mal die Standardabweichung.

Die Ergebnisse der Zonenstatistik sind hilfreich, um die Eigenschaften von Bereichen im Eingabebild zu bestimmen. Wenn das Eingabebild beispielsweise Rauschen enthält, kann das mittlere Spektrum eines Bereichs mehr spektrale Details offenbaren als das Spektrum eines einzelnen Pixels, in dem spektrale Details von Merkmalen durch Rauschen übertönt werden können.

3. Anwendungsgebiete

Nachfolgend wird die Anwendung von HypPy in verschiedenen Projekten in Bildung, Beratung und Forschung wird ausführlicher besprochen.

3.1. Bildung

HypPy ist eines der Softwarepakete, die in unseren Kursen zur Spektralgeologie an der Universität Twente verwendet werden [27]. Die Übersichtlichkeit der HypPy-Oberfläche macht es zu einem hervorragenden Werkzeug für Studenten, die mit der Spektralgeologie beginnen. Ein Beispiel für veröffentlichte studentische Forschung mit HypPy ist [28]. Ein Tutorial zur Analyse von Absorptionsmerkmalen zur Erzsuche mittels geologischer Fernerkundung finden Sie unter [24].

HypPy kann in den Kurzkursen verwendet werden, um die grundlegenden Konzepte der hyperspektralen Bildgebung zu erlernen und spezielle Operationen zur Verarbeitung und Interpretation hyperspektraler Daten durchzuführen. Beispielsweise wurde HypPy im Kurs „Fernerkundung und Mineralspektroskopie für Explorations- und Bergbaugeologen“ verwendet, der in Istanbul angeboten wurde [29].

Im Dezember 2022 wurde HypPy in einem Spektralworkshop während der jährlichen GRSG-Konferenz [30] verwendet. Der Workshop umfasste einen Überblick über die Theorie und praktische Umsetzung von Spektralmethoden zur Mineralidentifikation, begleitet von einer praktischen Laborübung, bei der die Spektralverarbeitungsanalytik auf einen kürzlich erhaltenen hochauflösenden hyperspektralen Datensatz aus der Luft angewendet wurde, der den sichtbaren und nahen Bereich umfasst.

Machine Translated by Google

Infrarotspektren (VNIR 400 nm bis 1 µm), kurzwelliges Infrarot (SWIR 1 µm bis 2,5 µm) und langwelliges Infrarot (LWIR 8 µm bis 15 µm) Spektralbereiche. Die Laborsitzung verwendete die Hyperspektralverarbeitungssoftware HypPy und ermöglichte ungehinderten Zugriff auf die Probendaten.

3.2. Beratung

Wir untersuchten die Machbarkeit des Einsatzes laborgestützter hyperspektraler Kurzwellen-Infrarot- Bilder (SWIR) für einen kommerziellen Partner, um die mineralogische Zusammensetzung bestimmter Sedimentkernproben aus dem niederländischen Untergrund zu ermitt Die Skriptfunktionen von HypPy erwiesen sich bei diesem Projekt als unverzichtbar.

Unsere Ergebnisse zeigen, dass die Verwendung von hyperspektralen Kurzwellen-Infrarotbildern (SWIR) im Labor mit einer räumlichen Abtastung von 200 µm einen wertvollen zerstörungsfreien Ansatz zur Identifizierung und Kartierung von weißem Glimmer, Ton und Karbonatmineralien in sedimentären Bohrkernen bietet. Gemeinsam mit Partnern an der Universität Groningen gelang es uns mithilfe von HypPy, eine Prototyp-Verarbeitungskette zur Quantifizierung des Mineralgehalts und zur Ableitung einer Reihe geophysikalischer Eigenschaften der Bohrkerne aufzubauen.

3.3. Forschung

Die ersten Arbeiten an HypPy begannen mit dem Mars Planetary Mapping Pilot Project für die Europäische Weltraumorganisation (ESA). Ziel der Pilotstudie war es, die modernen, digitalen Methoden und Arbeitsabläufe der terrestrischen geologischen Kartierung auf die Planetenkartierung zu übertragen. Der Schwerpunkt dieser Studie lag auf den ESA-Datensätzen der hochauflösenden Stereokamera (HRSC) und des Hyperspektralinstruments Visible and Infrared Mineralogical Mapping Spectrometer (OMEGA) an Bord der Mars Express [31]. Der erste Schritt des Arbeitsablaufs bestand darin, die Rohdaten zu verarbeiten, um interessante Merkmale Dazu gehörten digitale Höhenmodelle, die aus den Stereodaten extrahiert wurden, und Mineralienkarten, die mit Hilfe der Hyperspektraldaten erstellt wurden. Diese Werkzeuge wurden verwendet, um die Daten zu interpretieren und eine geologische Karte des Gebiets Nili Fossae zu erstellen . Diese Studie ebnete den Weg für die routinemäßige Anwendung der Wellenlängenkartierungstechnik zur Kartierung der Planeten Mars und Erde [7].

HypPy führt eine Verarbeitungsmethode ein, die darauf ausgelegt ist, OMEGA-Daten von der Strahlungsdichte am Sensor in die Oberflächenreflexion umzuwandeln. Um die besonderen Herausforderungen im Zusammenhang mit OMEGA-Daten zu bewältigen , haben wir innovative Ansätze entwickelt, um ein Transmissionsspektrum pro Szene zu bestimmen , atmosphärische Korrekturen durchzuführen und systematisches und zufälliges Rauschen zu filtern. Siehe beispielsweise van Ruitenbeek et al. [25]. Diese Filter sind jetzt Teil von HypPy. Die Verarbeitungsmethode erzeugt Bilder der Oberflächenreflexion, die die Extraktion von Einzelpixelspektren zur Identifizierung von Oberflächenmerkmalen ermöglichen. In Übereinstimmung mit verschiedenen früheren Studien deuten unsere Ergebnisse unter anderem auf die Existenz von eisenhaltigen Tonen auf der Marsoberfläche hin.

Die Kartierung physikalischer Eigenschaften aus Hyperspektralbildern ist kompliziert und hängt häufig von der Verfügbarkeit von Spektralbibliotheken der Zielmaterialien ab. Viele bestehende Methoden hängen bei der Auswahl von Endelementen und der Nachbearbeitung von Regelbildern von den subjektiven Entscheidungen des Interpreten ab. Wenn keine Vorabinformationen verfügbar sind, sollten Informationen durch die Charakterisierung spektraler Absorptionsmerkmale extrahiert werden. In van Ruitenbeek et al. [25] beschreiben sie eine Methode namens Wellenlängen-Mapper (siehe Abschnitt 2.4.4) , um die Mindestreflexion innerhalb eines Wellenlängenbereichs automatisch zu bestimmen. Wir schlussfolgern, dass sich der Wellenlängen-Mapper als nützliches Werkzeug für die unvoreingenommene Auswahl von Endelementen, die Mineralkartierung und die Interpretation von Bereichen erweist, die die Zonierung von Mineralveränderungen mithilfe von Hyperspektralbildern zeigen. Die Nützlichkeit des Wellenlängen-Mappers wurde bereits in anderen Untersuchungen erkannt. So wurde er beispielsweise verwendet, um Mineralaufschlüsse von lithiumhaltigen Pegmatiten in ei

In einer anderen Studie wurde die Entropiefunktion zur Spektrumklasse hinzugefügt, um die Berechnung eines Index für Sortierprozesse in hydrothermalen Systemen zu ermöglichen. Hydrothermale Prozesse beeinflussen die Zusammensetzung von Gesteinen durch chemische Reaktionen und Flüssigkeitsström

Machine Translated by Google

Die Arbeit zeigte, dass die verbleibenden Muster solcher Prozesse mithilfe der Shannon-Entropie visualisiert werden können. Die Anwendung der Entropie ist nützlich für die Untersuchung hydrothermaler Mineralvorkommen und kann auch bei der Untersuchung früher Lebensumgebungen von Nutzen sein.

Funktionen, die bereits in der Spektrumklasse von HypPy implementiert wurden, könnten verwendet werden, um den Daten Rauschen hinzuzufügen. Dies wurde für die Simulation von luftgestützten SWIR-Hyperspektralsensoren zur Kartierung von Methan (CH4) -Gas verwendet [33]. Die Ergebnisse der Studie zeigten, dass drei derzeit aktive luftgestützte Bildspektrometer die CH4- Fahne nicht erkennen konnten. Der Artikel betonte die Bedeutung sowohl der spektralen Abtastung als auch der radiometrischen Kalibrierung für die erfolgreiche Kartierung von Methan.

Moderne bildgebende Spektrometer sind kompakt, leicht, kostengünstig und im Labor einsetzbar. Präzise hyperspektrale Bilder erfordern minimales Rauschen, präzise Kalibrierung und gleichbleibende Reaktionsgleichmäßigkeit. Aufgrund der Alterung optischer und elektronischer Komponenten sind regelmäßige Kontrollen ratsam. In einer weiteren Studie stellen wir eine Technik zur Quantifizierung von Smile- und Keystone-geometrischen Verzerrungen aus Testbildern vor [34].
Die Messungen werden mithilfe einfacher Modelle für Smile und Keystone sowie bereits vorhandener Tools in HypPy durchgeführt. Zusätzlich werden Funktionen für quadratische Spline-Interpolation, normalisierte Kreuzkorrelation und Optimierung angewendet, um die Smile- und Keystone-Werte zu ermitteln.

Die Ergebnisse zeigen, dass das Lächeln und der Schlussstein eines Laborbildspektrometers kann mit Subpixelpräzision bestimmt werden.

4. Diskussion

Neben HypPy bieten mehrere andere Pakete Funktionen zur Verarbeitung hyperspektraler Bilder.

Um nur einige zu nennen: QUANTools wurde mit der Programmiersprache ENVI/IDL zur automatischen Erkennung mehrerer Absorptionsparameter implementiert [35]. Zur Ausführung von QUANTools wird jedoch eine Laufzeitversion von ENVI/IDL benötigt, und das QUANTools- Paket bietet eingeschränkte Funktionalität, ähnlich einem bestimmten Tool in HypPy, dem sogenannten Wellenlängen-Mapper.

PySptools ist ein Python-Modul, das Spektral- und Hyperspektralalgorithmen implementiert [36]. Allerdings handelt es sich hierbei um ein Paket oder eine Bibliothek, und Sie müssen Code schreiben, bevor Sie es verwenden können. Tatsächlich verwendet HypPy pysptools für seinen linearen Unmixing-Klassifikator. Im Gegensatz zu pysptools bietet HypPy eine Benutzeroberfläche für die Funktionalität.

Ein weiteres erwähnenswertes Toolkit ist Spectral Python (SPy) [37]. SPy ist ein Python-Modul für die Verarbeitung hyperspektraler Bilddaten und bietet Funktionen für Aufgaben wie das Lesen, Anzeigen, Bearbeiten und Klassifizieren hyperspektraler Bilder.
Insbesondere ist SPy im Gegensatz zu HypPy auf die interaktive Nutzung über die Python-Eingabeaufforderung oder über Python-Skripte beschränkt.

Die Softwarepakete Orfeo und EnMAPBox sind beides leistungsstarke Werkzeuge für Fernerkundungs- und Bildverarbeitungsanwendungen [38,39]. Sie werden in der Fernerkundungsgemeinschaft häufig verwendet und bieten umfangreiche Funktionen für die Verarbeitung und Analyse von Satellitenbildern. Während sich diese Pakete auf die Verarbeitung von Satellitenbildern konzentrieren, ist HypPy auch auf die Verarbeitung allgemeiner Hyperspektralbilder ausgerichtet, wie z. B. proximale Bilder, die mit Hyperspektralkameras im Labor aufgenommen wurden.

Genau wie die Sprache Python selbst [6] wurde HypPy eher mit Blick auf Einfachheit als auf Leistung entwickelt [40]. Einige Operationen in HypPy können relativ langsam sein. Die Modularität der Funktionen und Werkzeuge kann jedoch genutzt werden, um experimentelle Programme oder sogar Prototypen von Verarbeitungsketten zu erstellen , die auf beliebiger Hardware laufen.

Was die Geschwindigkeit betrifft, könnten zukünftige Versionen von Python schneller sein. Um ein laufendes Projekt zu nennen: Die Sprache Mojo ist eine Obermenge der Sprache Python mit der Geschwindigkeit der Sprache C [41].

Machine Translated by Google

5. Schlussfolgerungen

In diesem Artikel haben wir das Design, die Implementierung und die Anwendung des Python- Pakets namens Hyperspectral Python (HypPy) beschrieben. Die Idee von HypPy besteht darin, hyperspektrale Bilder mithilfe kostenloser Open-Source-Software verarbeiten zu können. HypPy wurde mithilfe der Python-Sprache entwickelt und stützt sich stark auf die Array-Verarbeitungsfunktionen von Paketen wie NumPy und SciPy. HypPy kann für die frühe Datenanalyse und das Prototyping von Verarbeitungsketten verwendet werden . HypPy wurde mit Blick au Gleichzeitig enthält es mehrere innovative und wirksame Konzepte.

HypPy ist ein Paket in der Entwicklung und für die Entwicklung und hat seine Vorzüge unter Beweis gestellt. Seine Einfachheit und flexible Menüstruktur auf oberster Ebene machen HypPy für den Einsatz in der Ausbildung und für Kurzkurse attraktiv. Neue Methoden für die Forschung können schnell programmiert und getestet werden. Beispiele für solche Forschungsprojekte sind das Pilotprojekt der ESA zur Kartierung des Mars-Planeten und das Forschungsprojekt zur Bestimmung des Trapez- und Smile-Effekts von hyperspektralen Zeilenkameras.

In Projekten wurde es verwendet, um ganze Verarbeitungsketten von Hyperspektraldaten bis hin zu Informationen aufzubauen. Kurz gesagt kann HypPy als Plattform zum Generieren und Testen neuer Ideen für die Verarbeitung von Hyperspektraldaten verwendet werden. Der Kern des Lesens und Schreibens von Hyperspektralbildern , der Wellenlängen-Mapper, Hyperspektralgradientenfilter, räumlich-spektrale Filter, Bandmathematik- und Spektralmathematik-Tools mit eigener Ausdruckssprache, Entscheidungsbaum-Tools, die Verwendung von Zonenstatistiken und dedizierte Bild- und Spektralbibliotheksbetrachter sind gute Beispiele dafür.

Wir erwarten, dass HypPy auch für andere Organisationen und Einzelpersonen in Forschung, Bildung und Projekten, die Hyperspektraldaten verwenden, einen großen Mehrwert darstellt .

Beiträge der Autoren: Konzeptualisierung, WB und FvR; Methodik, WB; Software, WB; Validierung, WB, FvR, HvdW und CH; Ressourcen, AD; Schreiben – Vorbereitung des Originalentwurfs, WB; Schreiben – Überprüfung und Bearbeitung, HvdW; Visualisierung, WB; Überwachung, HvdW und FvdM; Projektverwaltung , WB; Mittelbeschaffung, FvdM und FvR. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu .

Finanzierung: Wir danken der Europäischen Weltraumorganisation (Projekt zur Kartierung des Marsplaneten) und Wijnand van Eindthoven von Deep Atlas für die finanzielle Unterstützung. Die von diesen Organisationen bereitgestellte Finanzierung spielte eine entscheidende Rolle bei der erfolgreichen Durchführung unseres Forschungspr

Datenverfügbarkeitserklärung: HypPy kann von Github und Zenodo heruntergeladen werden, siehe [42]; Name: HypPy (Hyperspectral Python); Lizenz: GPU General Public License Version 3; Kontakt: whbakker@utwente.nl; Hardwareanforderungen: normaler Windows-, Mac- oder Linux-Laptop oder -Desktop; Programmiersprache: Python Version 3; Mindestsoftwareanforderungen: Numpy, Scipy, Pillow, Matplotlib, Pysptools; Programmgröße: 12,2 MB.

Danksagung: Wir danken der Universität Twente für die Bereitstellung der notwendigen Ressourcen und Einrichtungen, die den reibungslosen Ablauf unserer Forschung enorm unterstützt haben. Darüber hinaus danken wir allen Studierenden, die direkt oder indirekt zu dieser Arbeit beigetragen haben, indem sie während ihres Studiums HypPy verwendet haben. Ihre Zusammenarbeit und Ermutigung waren für die Gestaltung von HypPy von

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Abkürzungen

In diesem Manuskript werden die folgenden Abkürzungen verwendet:

API Eine Programmierschnittstelle
BBL Bad Band Liste
BIL Zeilenweise verschachtelte Bänder
BSQ-Band sequentiell
CLI-Befehlszeilenschnittstelle
GUI Grafische Benutzeroberfläche
SWIR Kurzwelliges Infrarot
VNIR Sichtbares und Nahinfrarot

Machine Translated by Google

Verweise

1. NV5 Geospatial Solutions Inc. ENVI – Umgebung zur Visualisierung von Bildern. 2023. Online verfügbar: https://www.nv5 geospatialsoftware.com/docs/using_envi_Home.html (abgerufen am 24. November 2023).
2. Prokakis, E. Freie und Open-Source-Software: Freiheit, Transparenz und Effizienz im Zeitalter der Digitalisierung. J. Politik Ethik Neue Technologien. KI 2022, 1, e31230. [CrossRef]
3. Harris, CR; Millman, KJ; van der Walt, SJ; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, NJ; et al. Array-Programmierung mit NumPy. Nature 2020, 585, 357–362. [CrossRef] [PubMed]
4. Virtanen, P.; Gommers, R.; Oliphant, TE; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Grundlegende Algorithmen für wissenschaftliches Rechnen in Python. Nat. Methods 2020, 17, 261–272. [CrossRef] [PubMed]
5. Hunter, JD Matplotlib: Eine 2D-Grafikumgebung. Comput. Sci. Eng. 2007, 9, 90–95. [CrossRef] 6. van Rossum, G.; Drake, FL Python 3 Reference Manual; CreateSpace: Scotts Valley, CA, USA, 2009. Online verfügbar: https: //dl.acm.org/doi/book/10.5555/1593511 (abgerufen am 31. Juli 2024).
7. Tragheim, DG; Marsh, SH; Pedley, RC; Napier, BR; Bateson, L.; Smith, AG; Marchant, AP; Gunnink, JL; Oosthoek, JHP; Muller, JP The Mars Planetary Mapping Pilot Project; Europäische Weltraumorganisation (ESA): Paris, Frankreich, 2010.
8. Konstantinides, K.; Rasure, JR Die Khoros-Softwareentwicklungsumgebung für Bild- und Signalverarbeitung. IEEE Trans. Bildprozess. 1994, 3, 243–252. [CrossRef] [PubMed]
9. Young, M.; Argiro, D.; Kubica, S. Cantata. ACM SIGGRAPH Comput. Graph. 1995, 29, 22–24. [CrossRef]
10. Ousterhout, JK; Jones, K.; Foster-Johnson, E.; Fellows, D.; Griffin, B.; Welton, D. Tcl und das Tk Toolkit, 2. Auflage; Addision-Wesley Professionelle Computerreihe; Addison-Wesley: Upper Saddle River, NJ, USA, 2009.
11. Roseman, M. Modern Tkinter für vielbeschäftigte Python-Entwickler: Lernen Sie schnell, gut aussehende Benutzeroberflächen für Windows, Mac und Linux mit dem Standard-GUI-Toolkit von Python; Late Afternoon Press: Victoria, BC, Kanada, 2020.
12. Kokaly, RF; Clark, RN; Swayze, GA; Livo, KE; Hoefen, TM; Pearson, NC; Wise, RA; Benzel, WM; Lowers, HA; Driscoll, RL; et al. USGS Spectral Library Version 7; Technischer Bericht; USGS: Reston, VA, USA, 2017. [CrossRef]
13. Bibring, JP; Langevin, Y.; Altieri, F.; Arvidson, R.; Beilud, G.; Berthél, M.; Douté, S.; Drossart, P.; Encrenaz, T.; Vergessen, F.; et al. OMEGA: Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité; Europäische Weltraumorganisation, ESA (Sonderveröffentlichung): Paris, Frankreich, 2009; SP-1291; S. 75–95. Online verfügbar: https://sci.esa.int/s/WnjXOYW (abgerufen am 31. Juli 2024).
14. Vane, G.; Green, RO; Chrien, TG; Enmark, HT; Hansen, EG; Porter, WM Das luftgestützte sichtbare/infrarote Bildgebungsspektrometer (AVIRIS). Remote Sens. Environ. 1993, 44, 127–143. [CrossRef]
15. Jia, G.; Hueni, A.; Tao, D.; Geng, R.; Schaepman, ME; Zhao, H. Spektrale Superauflösung bei der Reflexionsabfrage aus Fernerkundungsdaten von bildgebenden Spektrometern. Opt. Express 2016, 24, 19905–19919. [CrossRef] [PubMed]
16. Marien, C.; Dijkstra, AH; Wilkins, C. Die hydrothermale Veränderung von Karbonatit im Fen Complex, Norwegen: Mineralogie, Geochemie und Auswirkungen auf die Entstehung von Seltenerdressourcen. Mineral. Mag. 2018, 82, S115–S131. [CrossRef]
17. Richards, JA; Jia, X. Remote Sensing Digital Image Analysis; Springer: Berlin/Heidelberg, Deutschland, 1999. [CrossRef]
18. Driscoll, M. Pillow: Bildverarbeitung mit Python; Leanpub: Victoria, BC, Kanada, 2021. Online verfügbar: https://leanpub.com Kissen (abgerufen am 31. Juli 2024).
19. Malvern Panalytical Ltd. ASD-Dateiformat – Version 8. Technischer Bericht, Malvern Panalytical. 2017. Online verfügbar: https: //www.malvernpanalytical.com/ en/learn/knowledge-center/user-manuals/asd-file-format-v8 (abgerufen am 31. Juli 2024).
20. Hettinger, R. Was ist neu in Python 3.8. 2019. Online verfügbar: https://docs.python.org/3/whatsnew/3.8.html (abgerufen am 24. November 2023).
21. Tucker, CJ Rote und fotografische Infrarot-Linearkombinationen zur Überwachung der Vegetation. Remote Sens. Environ. 1979, 8, 127–150. [Querverweis]
22. Chein-I Chang. Spektrale Informationsdivergenz für die hyperspektrale Bildanalyse. In Proceedings of the IEEE 1999 International Geoscience and Remote Sensing Symposium, IGARSS'99 (Kat.-Nr. 99CH36293), Hamburg, Deutschland, 28. Juni–2. Juli 1999; Band 1, S. 509–511. [CrossRef]
23. Hewson, RD; Cudahy, TJ; Mizuhiko, S.; Ueda, K.; Mauger, AJ Nahtlose geologische Kartenerstellung mit ASTER im Broken Hill-Curnamona-Provinz in Australien. Remote Sens. Environ. 2005, 99, 159–172. [CrossRef]
24. Hecker, C.; van Ruitenbeek, FJA; van der Werff, HMA; Bakker, WH; Hewson, RD; van der Meer, FD Spektrale Absorptionsmerkmalsanalyse zur Erzsuche: Ein Tutorial zur Verwendung der Methode in der geologischen Fernerkundung. IEEE Geosci. Fernerkundung Mag. 2019, 7, 51–71. [CrossRef]
25. van Ruitenbeek, FJA; Bakker, WH; van der Werff, HMA; Zegers, TE; Oosthoek, JHP; Omer, ZA; Marsh, SH; van der Meer, FD Kartierung der Wellenlängenposition tiefster Absorptionsmerkmale zur Erkundung der Mineralienvielfalt in hyperspektralen Bildern. Planet. Weltraumforschung 2014, 101, 108–117. [CrossRef]
26. Ellson, J.; Gansner, ER; Koutsofios, E.; North, SC; Woodhull, G. Graphviz und Dynagraph – Statisches und dynamisches Zeichnen von Graphen Werkzeuge. In Graph Drawing Software; Springer Link: New York, NY, USA, 2004; S. 127–148. [CrossRef]
27. ITC. Master of Science-Studiengang Geoinformationswissenschaft und Erdbeobachtung – Studienjahr 2023–2024. 2023. Online verfügbar: https://studyguide.itc.nl/m-geo (abgerufen am 24. November 2023).

Machine Translated by Google

28. Portela, B.; Sepp, MD; van Ruitenbeek, FJA; Hecker, C.; Dilles, JH Verwendung von Hyperspektralbildern zur Identifizierung von Pyrophyllit-Muskovit-Verwachsungen und Alunit in der flachen epithermalen Umgebung des Porphyr-Kupfer- Bezirks Yerington. Ore Geol. Rev. 2021, 131, 104012. [CrossRef]
29. Portela, B. Kurzkurs – Fernerkundung und Mineralspektroskopie für Explorations- und Bergbaugeologen, 7.–11. November 2022, Istanbul. 2022. Online verfügbar: https://www.linkedin.com/posts/brunobvportela_geologicalremotesensing-istanbul- turkey-activity-7001884993555693568-bkXG (abgerufen am 24. November 2023).
30. Geological Remote Sensing Group. GRSG-Jahreskonferenz und -Hauptversammlung 2022: Orbit to Outcrop. 2022. Online verfügbar: https://www.grsg.org.uk/grsg-agm-conference-2022 (abgerufen am 24. November 2023).
31. Chicarro, A.; Martin, PD; Trautner, R. Die Mars Express-Mission: Ein Überblick. In Mars Express: Eine europäische Mission zum Roten Planeten; ESA: Noordwijk, Niederlande, 2004; Band SP-1240. Online verfügbar: https://sci.esa.int/documents/33745/35957/ 1567254632829-ÜbersichtWeb.pdf (abgerufen am 31. Juli 2024).
32. Booysen, R.; Lorenz, S.; Thiele, ST; Fuchsloch, WC; Marais, T.; Nex, PAM; Gloaguen, R. Präzise hyperspektrale Bildgebung von mineralisierten Aufschlüssen: Ein Beispiel von lithiumhaltigen Pegmatiten bei Uis, Namibia. Remote Sens. Environ. 2022, 269, 112790. [Querverweis]
33. Scafutto, RDM; van der Werff, H.; Bakker, WH; van der Meer, F.; de Souza Filho, CR Eine Bewertung von luftgestützten SWIR-Bildspektrometern für die CH4-Kartierung: Auswirkungen von Bandpositionierung, spektraler Abtastung und Rauschen. Int. J. Appl. Earth Obs. Geoinf. 2021, 94, 102233. [CrossRef]
34. Bakker, W.; van der Werff, H.; van der Meer, F. Bestimmung des Lächelns und des Keystones von Hyperspektrallinienkameras im Labor. In Proceedings des 10. Workshops 2019 zu Hyperspektralbildgebung und Signalverarbeitung: Evolution in der Fernerkundung (WHISPERS), Amsterdam, Niederlande, 24.–26. September 2019; S. 1–5. [CrossRef]
35. Kopaÿcková, V.; Koucká, L. Mineralkartierung basierend auf der automatischen Erkennung mehrerer Absorptionsmerkmale. EARSeL EProc. 13, 95–99. [CrossRef]
36. Therien, C. Willkommen bei der PySptools-Dokumentation. 2022. Online verfügbar: https://pysptools.sourceforge.io/ (abgerufen am 24. November 2023).
37. Boggs, T. Spectral Python (SPy) User Guide. 2020. Online verfügbar: https://www.spectralpython.net/user_guide.html (Zugriff am 24. November 2023).
38. Grizonnet, M.; Michel, J.; Poughon, V.; Inglada, J.; Savinaud, M.; Cresson, R. Orfeo ToolBox: Open-Source-Verarbeitung von Remote Sensorbilder. Open Geospat. Data Softw. Stand. 2017, 2, 15. [CrossRef]
39. van der Linden, S.; Rabe, A.; Jakimow, B.; Thiel, F.; Cooper, S.; Okujeni, A.; Hostert, P. Integration von bildgebender Spektroskopie und GIS – Freie und Open Source-Bildanalyse in QGIS mit der EnMAP-Box 3. In Proceedings des OSA Optical Sensors and Sensing Congress 2021 (AIS, FTS, HISE, SENSORS, ES), Washington, DC, USA, 19.–23. Juli 2021; Optica Publishing Group: Washington, DC, USA, 2021; S. HF4E.2. [CrossRef]
40. van Rossum, G.; Fridman, L. Warum Python 3.11 so schnell ist. 2022. Interview auf YouTube. Online verfügbar: https://youtu.be/TLhRuZ9cJWc (abgerufen am 31. Juli 2024).
41. Modular Inc. Mojo – Eine neue Programmiersprache. 2023. Online verfügbar: https://www.modular.com/mojo (aufgerufen am 24. November 2023).
42. Bakker, WH wimhbakker/hyppy: HypPy3. Zenodo 2024. [CrossRef]

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Artikel

Entwurf von Mehrkanal-Spektrum-Intelligenz-Systemen unter Verwendung Ungefährer diskreter Fourier-Transformationsalgorithmus für Antennen Array-basierte Anwendungen zur Spektrumwahrnehmung

Arjuna Madanayake [1], Keththura Lawrance [1,*], Bopage Umesha Kumarasiri [1], Sivakumar Sivasankar [1], Thushara Gunaratne [2], Chamira US Edussooriya [3] und Renato J. Cintra [4]

[1] Fakultät für Elektro- und Computertechnik, Florida International University, Miami, FL 33174, USA; amadanay@fiu.edu (AM); ukuma003@fiu.edu (BUK); ssiva011@fiu.edu (SS)
[2] Herzberg Astronomy and Astrophysics Research Center, Nationaler Forschungsrat Kanada, Penticton, BC V0H 1K0, Kanada; thushara.gunaratne@nrc-cnrc.gc.ca
[3] Fakultät für Elektro- und Telekommunikationstechnik, Universität Moratuwa, Moratuwa 10400, Sri Lanka; chamira@uom.lk
[4] Department of Technology, Universidade Federal de Pernambuco, Caruaru 55015, PE, Brasilien; rjdsc@de.ufpe.br
* Korrespondenz: klawr041@fiu.edu

Zusammenfassung: Das Funkspektrum ist eine knappe und äußerst wertvolle Ressource, die sorgfältige Echtzeitüberwachung und dynamische Ressourcenzuweisung. Dynamic Spectrum Access (DSA) ist eine neue Paradigma für die Verwaltung des Funkspektrums, das KI/ML-gesteuerte Algorithmen für optimale Leistung unter schnell wechselnden Kanalbedingungen und möglichen Cyber-Angriffen im elektromagnetischen Bereich. Schnelles Erfassen in mehrere Richtungen mithilfe von Array-Prozessoren mit anschließender KI/ML-basierte Algorithmen zur Erfassung und Wahrnehmung von Wellenformen, die von der Umgebung ist entscheidend für die Entscheidungsunterstützung in DSA. Als Teil der Richt- und Breitband Spektrumwahrnehmung ist die Fähigkeit zur feinen Kanalisierung breitbandiger Eingänge mithilfe effizienter Fourieranalyse dringend erforderlich. Eine feinkörnige schnelle Fouriertransformation (FFT) über eine große Anzahl von Richtungen ist rechenintensiv und führt zu einer großen Chipfläche und einem hohen Stromverbrauch. Wir gehen dieses Problem an, indem wir die kürzlich vorgeschlagene approximative diskrete Fourier-Transformation nutzen (ADFT), das über eine eigene spärliche Faktorisierung für eine Echtzeitimplementierung mit geringer Komplexität verfügt und Stromverbrauch. Der ADFT wird verwendet, um einen breitbandigen digitalen Mehrstrahl-HF-Beamformer zu erstellen und eine zeitlich spektrumbasierte Aufmerksamkeitseinheit, die 32 diskrete Richtungen über 32 Unterbänder überwacht in Echtzeit mit einem multiplikatorlosen Algorithmus mit geringer Rechenkomplexität. Die Ausgabe von Diese spektrale Aufmerksamkeitseinheit wird als Entscheidungsvariable auf einen intelligenten Empfänger angewendet, der seine Mittenfrequenz und Frequenzauflösung über FFT-Kanalisierer, die speziell für Echtzeit entwickelt wurden Überwachung mit hoher Auflösung. Dieser zweistufige Prozess ermöglicht die Anwendung der Feinverstärkungs-FFT nur auf Richtungen und Bänder von Interesse, bestimmt durch die ADFT-basierte 2D-Raumzeit mit geringer Komplexität Aufmerksamkeitseinheit. Die feinkörnige FFT liefert eine spektrale Signatur, die zukünftige Anwendungsfälle in neuronale Netzwerk-Engines zur Modulationserkennung, IoT-Geräteidentifikation und RFI Identifikation. Beamforming- und spektrale Kanalisierungsalgorithmen, eine digitale Computerarchitektur, und frühe Prototypen mit einem volldigitalen Mehrkanalempfänger mit 32 Elementen und feldprogrammierbarer Es wird ein Gate-Array (FPGA)-basiertes Hochgeschwindigkeits-Software-Defined-Radio (SDR) vorgestellt.

Schlüsselwörter: ungefähre diskrete Fourier-Transformation; Breitband-Spektrum-Wahrnehmung; Breitband Digitales Multibeam-RF-Beamforming; Algorithmen mit geringer Komplexität

check for updates

Zitat: Madanayake, A.; Lawrance, K.; Kumarasiri, BU; Sivasankar, S.; Gunaratne, T.; Edussooriya, CUS; Cintra, RJ Design von Multichannel Spectrum Intelligence-Systeme mit Ungefähre diskrete Fourier Transformationsalgorithmus für Antenne Array-basierte Spektrumwahrnehmung Anwendungen. Algorithmen **2024,** 17, **338.** https://doi.org/10.3390/a17080338

Wissenschaftlicher Redakteur: Frank Werner

Erhalten: 3. Juli 2024
Überarbeitet: 24. Juli 2024
Akzeptiert: 26. Juli 2024
Veröffentlicht: 1. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz. Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennung (CC BY)-Lizenz (https://creativecommons.org/licenses/by/4.0/).

1. Einleitung

In Tolkiens Klassiker Der Herr der Ringe spürt ein allgegenwärtiges „Allsehendes Auge“ und inspiriert bisher die Bewohner von Mittelerde, basierend auf ihrer Nähe zu Objekten zu handeln, die unter dem Einfluss von Lord Sauron stehen [1]. In unserer Welt werden wir dauerha

Machine Translated by Google

Allgegenwärtiges Wissen über das elektromagnetische Spektrum durch Sensorik über Raum, Zeit und Wellenform durch Radiofrequenzwahrnehmung (RF) und ermöglicht so intelligentes Handeln für KI-basierte DSA. Wir brauchen ein RF-Allsehendes Auge für das Spektrum. **Die Fusion von RF-Spektrumsensorik mit KI und maschinellem Lernen (ML) ermöglicht die vollständige Wahrnehmu Allgegenwärtigkeit, also** überall, **und Beständigkeit, also** jederzeit **Informationen darüber zu haben, was in der elektromagnetischen Umgebung möglicherweise geschieht, werden zu unvorstellbar leistungsstarken Funktionen für die intelligenten drahtlosen Systeme von morgen führen. Das Spektrum ist eine immens wertvolle und knappe natürliche Ressource. Das exponentielle Wachstum der drahtlosen Konnektivität führt jedoch unvermeidlich dazu, dass das Spektrum erschöpft ist. Das Problem des Spektrummangels kann theoretisch gemildert werden, indem Kommunikationsnetzwerke in das Millimeterwellenband im Bereich von 30 bis 300 GHz** verlagert **werden. Dennoch musste die Industrie vor kurzem in Bezug auf die Millimeterwellen eine bittere Lektion lernen. Tatsächlich funktioniert die Millimeterwellentechnologie in einer Laborumgebung, es gibt jedoch eine Reihe von Problemen: (i) Es ist schwierig, sie in realen Umgebungen zum Laufen zu bringen, da die scharfen HF-Strahlen, die zur Aufrechterhaltung der Konnektivität notwendig sind, anfällig für Blockaden sind. (ii) Die Strahlsuche ist ein schwer zu lösendes Problem, insbesondere unter Berücksichtigung von Mobilitätsaspekten. (iii) Leistungsverstärker (PAs) sind ineffizient und verschwenden viel Energie. und (iv) die Kanalkohärenzzeiten sind sehr niedrig, was schnellere Aktualisierungen der Kanalzustandsinformationen (CSI) erfordert, was zu einer hohen Rechenlast führt.**

1.1. FR3-Band: Das Goldlöckchen-Spektrum

Einerseits sind die Frequenzbänder FR1/Sub-6 **GHz** – das Legacy-Spektrum – **für erweitertes Wachstum einfach nicht verfügbar. Tatsächlich wurde das 100 -MHz- Subband bei 3,5 GHz** für **unglaubliche 29 Milliarden USD verkauft ,** was die Nachfrage der Mobilfunkbranche deutlich zeigt. Allerdings **kann man mit allem Geld der Welt keine Ressourcen kaufen, die es nicht gibt. Da Millimeterwellen in naher Zukunft wahrscheinlich nicht zu praktischen Systemen führen werden, sucht die Branche nach einem Mittelweg. Dieser spektrale Mittelweg liegt im FR3-Band (6 GHz bis 24 GHz), das von Branchenakteuren liebevoll als** oberes Mittelband bezeichnet wird. In unserer jüngsten Arbeit haben wir **gezeigt, dass das obere Mittelband nicht zu niedrig ist, sodass es genauso überfüllt ist wie das FR1 (Legacy-Band), und nicht zu hoch, sodass der Aufbau einer praktisch skalierbaren Infrastruktur derzeit nicht rentabel ist. Das FR3-Band ist „genau richtig“, da es in einem Bereich liegt, der für skalierbare Systeme günstig ist. Darüber hinaus steht in FR3 Bandbreite für die nächste Generation drahtloser Technologie zur Verfügung, vorausgesetzt, dass es gelingt, sich im Wirrwarr der Frequenzpolitik zurechtzufinden [2].**

1.2. Eine gemeinsame Frequenzökonomie

Bei der Nutzung des oberen Mittelbands für die drahtlose Konnektivität besteht ein dringendes Problem. Der Bereich von 6 GHz bis 24 GHz wurde bereits für unternehmenskritische Anwendungen reserviert und die Lizenzen dafür liegen beim Verteidigungsministerium (DoD) und anderen staatlichen Stellen. Die meisten der unternehmenskritischen Sensoren des DoD, darunter X-Band-Radar, Kommunikationsnetzwerke , wissenschaftliche Experimentalbänder (z. B. Radioastronomie) und Wetterradarsensoren, befinden sich ebenfalls im FR3-Band. Primärnutzer (PUs) werden ihr „wertvolles“ Spektrum wahrscheinlich nie dauerhaft aufgeben ; in Zukunft wird jedoch wahrscheinlich DSA erforderlich sein, bei dem Sekundärnutzer (SUs) wie kommerzielle drahtlose Netzwerke die FR3-Bänder auf eine Weise nutzen können, die für die von PUs verfolgten kritischen Anwendungen kaum oder gar keine Bedrohung darstellt. Obwohl DSA der Gesellschaft enorme Vorteile und Wirtschaftswachstum bringen kann, geht mit einer derart großen Macht auch eine große Verantwortung einher, da die Einhaltung von Vereinbarungen zur gemeinsamen Nutzung mit Nichtabstreitbarkeit und Null-Vertrauen sichergestellt werden muss.

1.3. Die Notwendigkeit einer Echtzeit-KI-Spektralwahrnehmung

„Das Verteidigungsministerium hat festgestellt, dass eine gemeinsame Nutzung möglich ist, wenn bestimmte fortschrittliche Interferenzminderungsfunktionen und ein Koordinierungsrahmen zur Erleichterung der gemeinsamen Nutzung des Spektrums eingeführt werden“, verkündet die Nationale Spektrumstrategie 2023 des Weißen Hauses [3]. Die Früherkennung von Kriegshandlungen steht auf der Prioritätenliste höher als die Servicequalität kommerzieller drahtloser Netzwerke. Obwohl die Wahrscheinlichkeit eines katastrophalen Ereignisses gering ist, ist sie nicht gleich Null. Die einzige Möglichkeit, sicherzustellen, dass SUs nicht in die PU-Spektrur

Machine Translated by Google

und Wellenformdomänen erfolgt durch unabhängige Messungen, wobei Zero Trust angenommen wird, um die Einhaltung von Frequenzfreigabevereinbarungen vollständig zu gewährleisten. Neben echten Fehlern und unbeabsichtigten schädlichen HF-Interferenzen drohen Cyber-Bedrohungen durch bisher unbekannte böswillige Akteure. Die Möglichkeit, kritische drahtlose und Verteidigungsinfrastruktur zu stören , sollte nicht auf die leichte Schulter genommen werden. Zuverlässige, hochempfindliche Messungen mit HF-Sensoren überall und jederzeit ermöglichen eine bessere Widerstandsfähigkeit gegen solche zukünftigen Bedrohungen aus Sicht der Frequenzkriegsführ

1.4. Spektrumsensoren mit allgegenwärtiger Wahrnehmung

Von der HF-Autonomie durch KI-Systeme, die mithilfe von Deep Learning (DL) und Reinforcement Learning (RL) erstellt wurden, wird erwartet, dass sie ein tiefes Wissen – also eine „Wahrnehmung“ – über die erfasste elektromagnetische Umgebung generieren [4]. Die National Spectrum Strategy (The White House, 11/2023) fordert erste Studien in den Bändern 3,1 GHz bis 3,45 GHz, 5,030 GHz bis 5,091 GHz, 7,125 GHz bis 8,400 GHz und 18,1 GHz bis 18,6 GHz in Richtung DSA zwischen kommerziellen und DoD-Einheiten. KI-gestützte Wahrnehmung erfordert HF-Spektrum-Sensorik und die Erkennung von Wellenformparametern (z. B. Modulations- und möglicherweise sogar Kodierungsparametern [5]), Trägeraggregations- und Bandbreitenteilen oder ähnlichen Informationen, Ausbreitungsrichtungen, Leistungspegeln und Kanalverhalten wie Mehrwege-, Doppler- und Intersymbolinterferenzmetriken (ISI) [6]. Typische Spektrumsensoren liefern wahrscheinlich keine verwertbaren Informationen über das Spektrum, außer Nutzungsstatistiken [7–10], bei denen die Erkennung von „Spektrumlöchern“ die möglicherweise für SUs verfügbaren Unterbänder bestimmt [11–17]. Spektrumbetrug ist ein großes Problem [18]. Das Vorhandensein schädlicher HF-Interferenzen (RFI) kann nur durch unabhängiges Messen und Wahrnehmen durch Dritte nachgewiesen werden. Um ein Modell der Spektrallandschaft zu internalisieren, müssen wir die gemessene HF-Umgebung gut verstehen; eine einfache Energieeinteilung mithilfe einer FFT ist für zukünftige Anwendungsfälle zu einschränkend. Die Eindämmun Wir benötigen eine Echtzeitwahrnehmung mit maschineller Intelligenz, da Ereignisse überall und jederzeit mit einer Rate von Tausenden von Ereignissen pro Sekunde auftreten, was jegliche Human-in-the-Loop-Ansätze ausschließt . Die vollständig KI-gestützte Wahrnehmung der HF-Umgebung ist von entscheidender Bedeutung und dient als intelligente Informationsquelle für die automati

1.5. Beitrag des Papiers

In diesem Artikel schlagen wir ein rechnerisch effizientes intelligentes Spektrumsensorsystem vor . Wir integrieren einen DFT-basierten Multibeam-Beamformer, um HF- Signale richtungsabhängig zu erfassen. Um die Rechenkomplexität des Multibeam-Beamformers zu reduzieren, verwenden wir anstelle von DFT die approximative diskrete Fourier-Transformation (ADFT). Das digitale Verarbeitungs-Backend des vorgeschlagenen Multibeam-Beamformers ist auf einem feldprogrammierbaren Gate-Array (FPGA) implementiert. Das digitale Verarbeitungs-Backend, das mit einem Antennenarray aus 32 Patch-Antennen verbunden ist, die mit 5,7 GHz betrieben werden, bestätigt die Echtzeit- Spektrumserfassung. Darüber hinaus demonstriert der Versuchsaufbau die Anwendung von ADFT-basiertem Multibeam-Beamforming in der Spektrumserfassung und bietet einen vielversprechenden Ansatz für eine Echtzeit-, wenig komplexe und effiziente Richtungsspektrumswahr

1.6. Gliederung des Papiers

Der Rest dieses Dokuments ist wie folgt gegliedert. Abschnitt 2 bietet eine umfassende Übersicht, beginnend mit einer Untersuchung des dynamischen Spektrumzugriffs, wobei der Schwerpunkt sowohl auf Signalverarbeitungstechniken als auch auf maschinellen Lern-/Deep-Learning-Ansätzen liegt, gefolgt von einer Diskussion über Radioastronomie und Hochfrequenzstörungen. Abschnitt 3 befasst sich mit der ungefähren diskreten Fourier-Transformation und bietet Einblicke in den theoretischen Hintergrund und die spezifische Implementierung der ADFT. In Abschnitt 4 präsentieren wir einen detaillierten Systemüberblick , der die vorgeschlagene Architektur, das analoge Frontend, den ADFT-Kern und die digitale ADFT-Spektrals Abschnitt 5 beschreibt experimentelle Ergebnisse einschließlich Kalibrierung und Strahlmessungen.

Machine Translated by Google

zeigt die praktische Leistung der vorgeschlagenen Methode. Abschließend werden in Abschnitt 6 die Auswirkungen unserer Erkenntnisse erörtert und mögliche Richtungen für zukünftige Forschungen skizziert.

2. Überprüfung

In diesem Abschnitt überprüfen wir verwandte Arbeiten zum dynamischen Spektrumzugriff. Zunächst betrachten wir Signalverarbeitungstechniken, die für die Spektrumerfassung eingesetzt werden. Im ersten Unterabschnitt betrachten wir außerdem das maschinelle Lernen und kürzlich vorgeschlagene Als Nächstes befassen wir uns mit der Radioastronomie mit Radioteleskopen der nächsten Generation, insbesondere mit der spektralen Erfassung von RFIs.

2.1. Dynamischer Spektrumzugriff: Signalverarbeitung und ML/DL-Ansätze **DSA wurde**

erstmals vor zwei Jahrzehnten auf Grundlage der Messungen der tatsächlichen Spektrumnutzung in Betracht gezogen, die von der Spectrum Policy Task Force der Federal Communications Commission (FCC) durchgeführt wurden **[20]. Im Gegensatz zur Mehrheitsmeinung eines gut ausgelasteten Spektrums zeigten die tatsächlichen Messungen, dass viele Teile des Spektrums zu einer bestimmten Zeit und an einem bestimmten Ort ungenutzt sind, was zu spektralen Löchern führt. Dies ist das Ergebnis der festen Spektrumzugriffsrichtlinie . Unter DSA dürfen sekundäre Benutzer (d. h. nicht lizenzierte Benutzer) spektrale Löcher von primären Benutzern (d. h. lizenzierten Benutzern) nutzen,** ohne den Betrieb **der primären Benutzer durch ein** kognitives Radio **[21–24] zu stören , das ein intelligentes System ist, das auf einem softwaredefinierten Radio basiert. Es gibt jedoch eine Obergrenze für die Kanalkapazität, die mit DSA erreicht werden** kann **[25]. Hier geben wir kurz einen Überblick über** die mit DSA **verbundenen** Spektrumerfassungstechniken , **und der Leser wird auf [26–28] für andere Aspekte von DSA verwiesen, wie z. B. Spektrumverwaltung und Richtlinien zur Spektrumfreigabe.**

Es wurden mehrere Techniken sowohl für die Schmalband- als auch für die Breitband- Spektrumsensierung vorgeschlagen . **Für die Schmalbandsensierung sind die wichtigsten nicht-ML-Techniken die Energieerkennung [29–32], die Erkennung auf Basis angepasster Filter [33–35], die Erkennung zyklostationärer Merkmale [36,37] und die Erkennung auf Basis von Kovarianz [38–40]. Die Ansätze der Energieerkennung und der Erkennung auf Basis von Kovarianz erfordern keine Vorkenntnisse des Signals des primären Benutzers. Allerdings ist erstere bei niedrigen Signal-Rausch-Verhältnissen (SNR) nicht zuverlässig. Die Erkennung auf Basis angepasster Filter bietet optimales Messen und bessere Erkennung bei niedrigen SNR-Werten, aber diese Technik erfordert Vorkenntnisse des Signals des primären Benutzers, die möglicherweise nicht immer verfügbar sind.**
Die zyklostationäre Merkmalserkennung ist im Allgemeinen robust gegenüber Rauschunsicherheiten und funktioniert bei niedrigen SNR-Werten gut; der Nachteil ist jedoch eine lange Erfassungszeit. Bei der Breitbanderfassung kann als einfacher Ansatz ein Breitbandsignal mittels zeitlicher DFT oder einer Filterbank [41–44] in einen Satz Schmalbandsignale aufgeteilt und mithilfe von **Schmalbanderfassungstechniken sequenziell oder parallel erfasst werden .** Die sequenzielle Erfassung erfordert jedoch mehr Zeit, während **die parallele Erfassung mehr Ressourcen erfordert. Da ein Breitbandsignal mit Spektralbereich im Frequenzbereich als spärliches Signal betrachtet werden kann, wurden mehrere auf Komprimierungserfassung basierende Techniken zur Breitbanderfassung vorgeschlagen [45–48]. Diese Ansätze erfordern neben der schnellen Erfassung niedrigere Abtastraten und haben einen geringeren Stromverbrauch . Somit eliminieren auf Komprimierungserfassung basierende Ansätze die meisten Nach**
Im letzten Jahrzehnt wurden eine beträchtliche Anzahl von ML- und DL-Techniken für die Spektrumerfassung vorgeschlagen. ML-Algorithmen wurden hauptsächlich verwendet, um die Kanalbelegungsmuster eines primären Benutzers zu bestimmen und den Kanal als frei oder belegt zu klassifizieren. In [49–53] wurden K-Means-Clustering und Gaußsche Mischmodelle erstmals verwendet, um die Anwesenheit von primären Benutzern zu bestimmen, und dann wird ein Klassifikator wie eine Support Vector Machine und K-Nearest-Neighbor verwendet, um zu bestimmen, ob ein Kanal frei oder belegt ist . DL-Techniken wie Convolutional Neural Networks und Long Short-Term Memory wurden in **[54–57] für die Spektrumerfassung eingesetzt ,** die die herkömmlichen Techniken übertrafen. Kürzlich wurde in **[58] eine transformerbasierte DL-Architektur vorgeschlagen . Dieser DL-Ansatz übertrifft die zuvor vorgeschlagenen Ansätze auf Basis von Convolutional Neural Networks. Ein weiterer DL-Ansatz, der auf kooperative Weise für die Spektru**

Machine Translated by Google

Lernansätze verbessern die Robustheit des Spektrumsensorsystems und ermöglichen es, in dynamischen Umgebungen genauere Entscheidungen zu treffen. Für einen umfassenden Überblick über Spektrumsensortechniken verweisen wir auf [63–65] .

Die räumlich-zeitliche Modellierung kognitiver Funksysteme unter Verwendung mehrdimensionaler Signalverarbeitungskonzepte wurde in [66] vorgestellt. Dabei wurde die Richtungserkennung unter Verwendung dreidimensionaler Filter mit unendlicher Impulsantwort (IIR) als Strahlformer genutzt. Ähnliche Ansätze zur Richtungserkennung unter Verwendung mehrdimensio Darüber hinaus wurden in [70–72] Antennenarraysysteme zur Richtungserkennung eingesetzt . In [73] wurde ein DFT-basierter Multibeamformer zur Echtzeit-Richtungsmessung eingesetzt.

Im Vergleich zu dem in dieser Studie mit ADFT verwendeten Multibeamformer wurde in [73] eine exakte DFT mit 16 gleichzeitigen Strahlen bei 2,4 GHz verwendet. Wir stellen fest, dass sich die Obergrenze der mit DSA [25] erreichbaren Kanalkapazität bei Verwendung der ADFT anstelle der DFT nicht ändert .

2.2. Radioastronomie und RFI

Die Radioastronomie ist auf ein möglichst wenig verunreinigtes Radiospektrum angewiesen, um das Universum zu beobachten [74]. Das Erkennen und Abschätzen schwacher Himmelssignale, die von weit entfernten Galaxien kommen, ist besonders wichtig, um die wichtigsten wissenschaftlichen Ziele der Radioteleskope der nächsten Generation zu erreichen, nämlich des Square Kilometre Array (SKA) (https://www.skao.int/ abgerufen am 24. Juli 2024) [75], das 2030 Atacama Large Millimeter/sub-millimeter Array— Wideband Sensitivity Upgrade (ALMA-WSU) (https://science.nrao.edu/facilities/alma/ Wissenschaft_Nachhaltigkeit/Upgrade der Breitbandempfindlichkeit abgerufen am 24. Juli 2024) [76] und das Very Large Array ngVLA der nächsten Generation) (https://ngvla.nrao.edu/ abgerufen am 24. Juli 2024) [77]. Obwohl sie sich in abgelegenen, dünn besiedelten Regionen der Welt befinden, bedrohen die zunehmende Intensität, Frequenz, Bandbreite und Häufigkeit von RFI die Nützlichkeit der Radioteleskope der nächsten Generation [78]. RFI vom globalen Satellitennavigationssystem (GNSS) [79], terrestrischen und luftgestützten drahtlosen Kommunikationssystemen, Satelliten-Megakonstellationen [80] und anderen Technologien dringen in die Spektren ein, die ausschließlich von Radioastronomen verwendet wurden, und erschweren so die Unterscheidung schwacher Himmelssignale von RFI.

Die Internationale Fernmeldeunion (ITU) (https://www.itu.int/en/Pages/Standardmäßig.aspx abgerufen am 24. Juli 2024) und der FCC (https://www.fcc.gov/ abgerufen am 24. Juli 2024) spielen eine wichtige Rolle beim Schutz des Spektrums für die Radioastronomie und andere passive Nutzer. Die ITU ist über ihren Funkkommunikationssektor (ITU-R) für die Entwicklung internationaler Vorschriften und Empfeh Es erkennt die Bedeutung des Schutzes des Spektrums für passive Dienste wie Radioastronomie, Erderkundungssatellitendienste und Wetterdienste über Satelliten an. Die ITU-R identifiziert spezifische Frequenzbänder für diese passiven Dienste und erlässt regulatorische Bestimmungen, um ihren Schutz vor schädlichen Störungen zu gewährleisten. Diese Bestimmungen umfassen Koordinierungsverfahren, Leistungsgrenzen und Frequenztrennungsanforderungen, um die sensiblen Beobachtungen und Messungen von Radioastronomen und anderen passiven Nutzern zu schützen [81,82]. Ebenso erkennt die FCC in den Vereinigten Staaten die Bedeutung des Schutzes des Spektrums für die Radioastronomi Das Office of Engineering and Technology (OET) der FCC formuliert Richtlinien und Regeln, um schädliche Störungen dieser Dienste zu verhindern. Es legt technische Standards und Lizenzbedingungen fest, die die Bedürfnisse passiver Benutzer, einschließlich Radioastronomen, berücksichtigen . Darüber hinaus arbeitet die FCC eng mit der National Science Foundation (NSF) und anderen relevanten Behörden zusammen, um die Frequenznutzung zu koordinieren und die Integrität wissenschaftlicher Beobachtungen zu schützen. Beide Organisationen stehen jedoch unter unerbittlichem Druck von Industrie und Regierungen, eine stärkere kommerzielle Nutzung des Spektrums zu fordern, das zuvor nichtkommerziellen Benutzern zugewiesen wurde .

Daher wäre es ideal, wenn man die wichtigsten Merkmale von RFI (d. h. Frequenzbereich, Modulationsschema, Quellenstandorte, Intensität usw.) kennt, um Radioteleskopbetreiber bei der Planung und Terminierung von Beobachtungen zu unterstützen. Dies ist jedoch nicht für alle

Machine Translated by Google

Fälle von RFI; daher würden Aufzeichnungen der Radiospektralstärke in Bezug auf Stärke, Richtung, Dauer und Zeitpunkt des Vorfalls den Radioastronomen auch dabei helfen, die kontaminierten Beobachtungen zu kennzeichnen und auszusortieren [83]. In diesem Fall wäre die allgegenwärtige Wahrnehmung **der elektromagnetischen Umgebung ein großer Vorteil in der Nachbearbeitungsphase radioastronomischer Beobachtungen [84].**

3. Näherungs-DFT In

diesem Abschnitt überprüfen wir die Grundlagen der DFT-Näherung, indem wir ihre technische Logik und ihren mathematischen Formalismus anhand eines Optimierungsproblems beschreiben. Als nächstes konzentrieren wir uns auf die hochmoderne 32-Punkt-Näherungs-DFT, die zusammen mit dem zugehörigen schnellen Algorithmus numerisch beschrieben wird. Eine arithmetische Komplexitätsanalyse für die besprochene 32-Punkt-ADFT wird ebenfalls bereitgestellt, mit einem Vergleich m

3.1. Theoretischer Hintergrund

Die diskrete Fourier-Transformation ist ein linearer Operator, der ein N-Punkt-Eingangssignal x = x[0] x[1] . . . x[N **ÿ** 1] **gemäß** **mit dem Ausgangssignal X = X[0] X[1] . . .** X[N **ÿ** 1] in Beziehung setzt. **dem folgenden Ausdruck:**

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot \ddot{y}_{N}^{-nk} \qquad k = 0, 1, \ldots, N-1, \tag{1}$$

wobei ÿN = *exp(2ÿj/N)* die **N-te Einheitswurzel ist** und j ÿ **ÿ1. Per Definition berechnet weist die DFT eine Rechenkomplexität von** O(N2) **auf, die für Echtzeitanwendungen prohibitiv hoch ist.**

In der Praxis wird die DFT mit Hilfe einer von vielen effizienten numerischen Routinen berechnet – schnellen Algorithmen, die zusammen als „FFTs“ bezeichnet werden. Besonders beliebte FFTs sind die auf Radix-2 basierenden Algorithmen wie die Cooley-Tukey-FFT, die **den N-Punkt-DFT-Block rekursiv in zwei N/2-Punkt-DFT-Unterblöcke** zerlegen **kann [85].**

Im Allgemeinen sind FFTs in der Lage, den Rechenaufwand bei DFT- Berechnungen von O(N2) **auf** O(N **log** N) zu reduzieren **[86]. In vielen Fällen [87–92]** wird **durch eine solche Rechenaufwandsreduktion die theoretisch minimale multiplikative Komplexität der DFT erreicht [93]. Infolgedessen ist es unmöglich, die multiplikative Komplexität weiter zu reduzieren. Da es sich hier um ein sehr ausgereiftes Forschungsgebiet handelt, ist die Entwicklung neuer FFTs, die eine erhebliche Reduzierung der Rechenkosten ermöglichen, unwahrscheinlich.**

Unter Verwendung eines fehlertoleranten Paradigmas können jedoch erhebliche Reduzierungen erreicht werden. Wenn bei der DFT-Berechnung eine bestimmte Fehlergrenze zugelassen wird, kann der Rechenaufwand so angepasst werden, dass die niedrigere, aber noch akzeptable Genauigkeit erreicht wird. Als Konsequenz sind für die DFT-Schätzung weniger Rechenoperationen erforderlich [94].

3.2. Optimierungsproblem

Systematisch kann man solche DFT-Schätzer mit geringer Komplexität mithilfe der Matrixapproximationstheorie erhalten. Im Matrixformalismus wird die DFT durch die folgende Operation besc

$$\mathbf{X} = F_N \cdot \mathbf{x}, \tag{2}$$

wobei die N-Punkt-DFT-Matrix, bezeichnet mit FN, **die** N **×** N- **Matrix mit Elementen ist, die durch fk,n** = *ÿ* N , **k,** n = **0, 1, . . . , gegeben** $\ddot{y}_{N}^{-nk}$ **N ÿ 1.**

sind. Im Kontext von Näherungen für diskrete Transformationen ist eine Näherung N qualitativ diskrete Transformation einer Transformation, die exakten und die ungefähren Transformationsdomänensignale sind in einem relevanten Sinne „nahe beieinander“, um quantitativ definiert zu werden. Im Allgemeinen muss eine Näherungstransformation die folgenden Bedingungen erfüllen: (i) Sie kann wichtige Eigenschaften der exakten Transformation weitgehend bewahren; (ii) es besteht eine mathematische Nähe zwischen der exakten und der ungefähren Transformationsmatrix gemäß einer kontextuell relevanten metrischen Funktion – normalerweise einer Leistungskennzahl von

Machine Translated by Google

Verdienst; und (iii) es weist einen deutlich geringeren Rechenaufwand auf, verglichen mit der exakten

Berechnung mittels FFTs. In [95] wird ein Überblick über die Theorie gegeben.

Um die physikalische Interpretation des ungefähren Spektrums zu erhalten,
Es werden ungefähre Transformationsmatrizen gesucht, die im euklidischen Sinne nahe an den
assoziierte exakte Transformationsmatrix. Dies kann erreicht werden durch die Minimierung
der Frobenius-Norm der Differenz zwischen der exakten und der Kandidatenmatrizen für

Näherung, vorbehaltlich der Einschränkung der geringen Komplexität. Daher ist eine
Näherung kann erreicht werden, indem man seine Elemente auf numerische Mengen trivialer

Multiplikanden wie {0, ±1} oder {0, ±1, ±2}.

In Symbolen eine mögliche Formulierung für das Problem der Herleitung einer DFT-Approximation
ist gegeben durch [95]

$$\min_{\hat{f}_{k,n}} \quad \|N \cdot S \cdot \hat{F} - F\|_F^2 \tag{3}$$

$$\text{vorbehaltlich} \quad \hat{f}_{k,n} \in \{0, \pm 1\} \quad , \tag{4}$$

1, . . . , $\hat{}$ N ∈ **1, bezeichnet die Einträge der Näherungsmatrix** $\hat{F}$ **wobei** $\hat{f}_{k,n}$**, wobei k,** n = **0,** **N,**
∥ · ∥F bezeichnet die Matrix-Frobenius-Norm, und S ist eine Normalisierungsmatrix, die verwendet wird, um sicherzustellen,

dass die Basisvektoren von $S \cdot \hat{F}^N$ **haben eine Energie von eins oder nahe eins.**

3.3. 32-Punkt-ADFT

In dieser Arbeit haben wir die 32-Punkte-ADFT übernommen, die mathematisch in [96] eingeführt wurde **und**
theoretisch und in Hardware weiter ausgearbeitet in [97,98]. Die explizite numerische Form der
Die ausgewählte 32-Punkt-ADFT ist in Anhang A angegeben. Abbildung 1 zeigt einen
Vergleich der Amplitudenantworten für die exakte 32-Punkt-DFT und die 32-Punkt-ADFT. Die Phasena
entlang der Maxima, also der Strahlen, bleiben unverändert, und jede Phasenabweichung
tritt in den tiefen Sperrbereichen auf. Daher ist der Einfluss der Näherungsnatur

der ADFT auf die Phasen der strahlgeformten Signale ist vernachlässigbar.

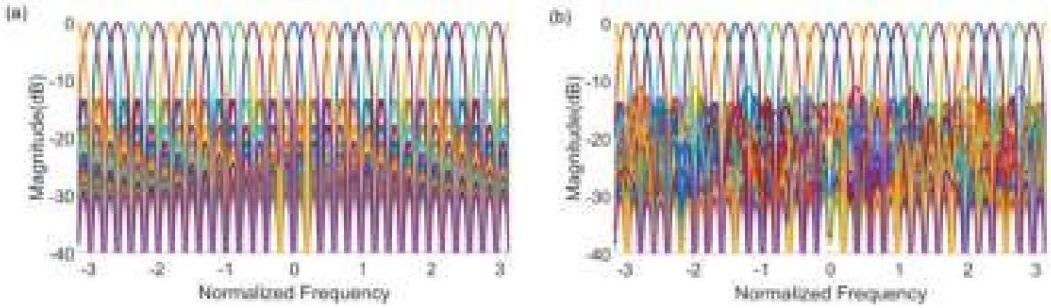


Abbildung 1. Simulierte Amplitudenantwort von (a) idealer 32-Punkt-DFT und (b) niedrigkomplexer 32-Punkt-
ADFT [97,98].

3.4. Schneller Algorithmus und Komplexitätsanalyse

Dargestellt durch die Matrix mit geringer Komplexität $\hat{F}^{32}$**, Die ausgewählte ADFT lässt die**
folgende Faktorisierung:

$$F^{32} = W8 \cdot W7 \cdot W6 \cdot W5 \cdot W4 \cdot W3 \cdot W2 \cdot W1, \tag{5}$$

wobei Wi , i = **1, 2, . . . , 8, dünn besetzte Matrizen mit geringer Komplexität sind. Im Anhang A**

Geben Sie die explizite Definition der Matrizen Wi an , i = **1, 2, ..., 8.**
Zum Vergleich ist in Tabelle 1 die Anzahl der Rechenoperationen für die exakte 32-Punkte-DFT aufgeführt.
berechnet nach seiner Definition und der Radix-2 Cooley–Tukey FFT. Für diese Operation
Beim Zählen nehmen wir an, dass (i) das Eingangssignal rein reellwertig ist; (ii) triviale Multiplikationen
durch 0, ±1 oder ±j **werden nicht gezählt; und (iii) eine Multiplikation zwischen einer reellen Zahl und einer komplexen**
Zahl wird als zwei reelle Multiplikationen ausgeführt – nicht als komplexe Multiplikation [99]. Aufgrund
(i) und (iii), die 704 nichttrivialen Multiplikanden von $\hat{F}^{32}$ **entsprechen 1408 reellen Multiplikationen.**

Machine Translated by Google

Die Auswertung der Anzahl der Zugänge erfolgte als Summe der Anzahl der Real- und Imaginärteil-Additionen. Es wurde auch die arithmetische Komplexität für die diskutierten 32-Punkt-ADFT durch seine Definition und seinen schnellen Algorithmus (Faktorisierung). Die Die Komplexitätsanalyse der Radix-2-Cooley–Tukey-FFT wird detailliert in [99] beschrieben.

Tabelle 1. Leistung der DFT, der Cooley–Tukey FFT und der besprochenen 32-Punkt-ADFT. Trivial Multiplikationen mit ±1 oder ±j werden nicht mitgezählt, es werden nur realwertige Eingangssignale angenommen.

Verfahren	Reelle Multiplikationen	Echte Ergänzungen
Exakte 32-Punkt-DFT (F32)	1408	1666
Radix-2 Cooley–Tukey FFT (N = 32) [99]	88	408
Näherungsweise Dichtefunktionaltheorie ($\hat{F}_{32}$) [96]	0	1282
Schneller Algorithmus für $\hat{F}_{32}$ ($\ddot{y}\prod_{i=1}^{8} W_i$) [97]	0	144

Tabelle 2 zeigt die arithmetische Komplexität für jeden einzelnen Matrixterm der $\hat{F}_{32}^{\text{faktisch}}$ torisierung. Die Matrix W8 stellt keine Rechenkosten dar, da sie nur aus Vorzeichenwechsel- und Datenaustauschoperationen (Vertauschen von Real- und Imaginärteil). Der Datenaustausch ist eine Folge der Multiplikation mit j (einer Rotation um ÿ/2).

Tabelle 2. Leistung der Matrixfaktoren des schnellen ADFT-Algorithmus. Triviale Multiplikationen mit ±1 oder ±j nicht gezählt, es wird nur realwertiges Eingangssignal angenommen.

Betrieb	W1	W2	W3	W4	W5	W6	W7	W8
Reelle Multiplikationen	0	0	0	0	0	0	0	0
Echte Ergänzungen	30	30	14	14	30	14	12	0

4. Systemübersicht

In diesem Abschnitt beschreiben wir das Design und die Realisierung der Antennen, Mikrowellenempfänger , Mixed-Signal-Konverterschaltungen und FPGA-basierten digitalen Realisierungen, die das vorgeschlagene räumlich-zeitliche Spektrumwahrnehmungssystem. Die Antennen und Mikrowellen Systeme wurden intern entworfen und realisiert, während das FPGA-System von die Open-Source-Entwicklergemeinschaft für Radioteleskopinstrumente (CASPER), die sich in verantwortlich für wichtige Radioinstrumente wie das Square Kilometer Array (SKA), Murchison Wide-Field Array, Allen Telescope Array und andere.

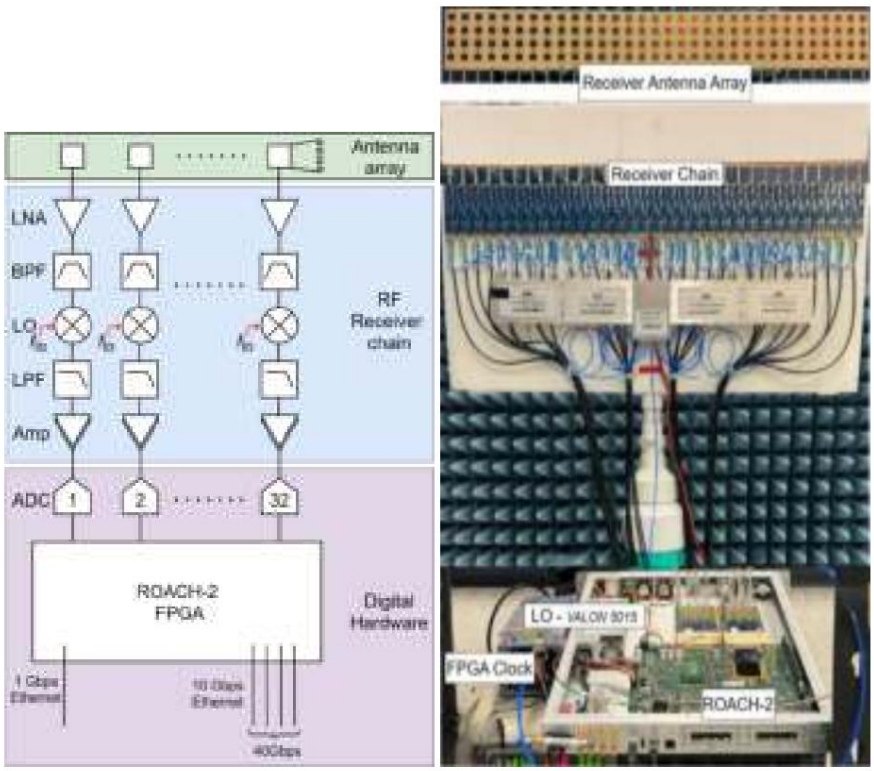
4.1. Vorgeschlagene Architektur

Zur experimentellen Validierung und Messung der von der vorgeschlagenen ADFT erzeugten Strahlen Algorithmen wurde ein 5,7 GHz HF-System mit einem digitalen Verarbeitungs-Backend sorgfältig entworfen und implementiert. Abbildung 2a zeigt die Gesamtarchitektur des Systems basierend auf ein Antennenarray mit 32 Elementen, das wichtige Subsysteme wie das Antennenarray selbst umfasst, die HF-Empfängerkette und die digitale Hardwareeinheit. Abbildung 2b zeigt den Versuchsaufbau der vorgeschlagenen Architektur. Neben den digitalen Kernen, die für die Ausführung der Zur räumlichen DFT der abgetasteten Signale wurden zusätzliche Schaltkreise innerhalb des FPGA entwickelt. Diese zusätzliche Schaltung ermöglicht die Echtzeit-Datenmanipulation und Strahlmessungen über verschiedene Winkel, wodurch die Flexibilität und Leistungsfähigkeit des Systems verbessert wird.

Die auf rekonfigurierbarer Open Architecture Computing Hardware (ROACH-2) basierende digitale Signalverarbeitungsplattform (DSP) ist in Abbildung 3 dargestellt und spielt eine entscheidende Rolle bei dient als Zwischenschicht und ermöglicht die Kommunikation zwischen der FPGA-Hardware (insbesondere sein Speicher) und der Bordcomputer. Die ROACH-2- Karte ist mit dem Haupt-Host-Linux-Server über eine 1- Gbit/s- Ethernet-Verbindung. Eine Python-Routine läuft auf dem Host-Server ist so programmiert, dass er mit der ROACH-2- Plattform interagiert und Steuerungsfunktionen und Abrufen von Daten aus dem FPGA-Speicher beim iterativen Scannen aus verschiedenen Winkeln. Diese einzigartige Systemarchitektur sorgt nicht nur für eine präzise Strahlführung

Machine Translated by Google

Messungen, sondern verbessert auch die Gesamteffizienz und Automatisierung des Experiments und bietet eine robuste Plattform für die Validierung fortschrittlicher Strahlformungsalgorithmen.



(A) **(B)**

Abbildung 2. (a) Die Systemarchitektur; (b) der experimentelle Aufbau des 32-Strahl-Array-Empfängers arbeitet in einem Band von 5,700 GHz bis 5,800 GHz , gefolgt vom digitalen Back-End des **ROACH-2- FPGA-Systems (Xilinx Virtex-6 sx475T FPGA).**



Abbildung 3. ROACH-2-basierte DSP-Plattform basierend auf Xilinx Virtex-6 Sx475 FPGA und 32-Kanal-ADC- Karte.

Wir danken Dr. Dan Werthimer von der UC Berkeley und der CASPER-Community für ihre Beiträge zu den ROACH-2- und CASPER-Tools.

Machine Translated by Google

4.2. Analoge Frontends

Das HF-Frontend des Empfangsmodus-Beamformers wurde aus einem Uniform Linear Array (ULA) mit 32 Elementen und einer Betriebsfrequenz von 5,7 GHz konstruiert, gepaart mit 32 Direktumwandlungs -HF-Empfängerketten auf Leiterplatten. Die Array-Elemente sind im Abstand von 0,6 ÿ angeordnet, was **bei 5,7 GHz** etwa 32 **mm entspricht . Jedes Antennenelement ist als vertikales Subarray aus 4 × 1 Patchantennen konzipiert, was passives Beamforming in der vertikalen Ebene ermöglicht. Dieses analoge Beamforming arbeitet unabhängig von den räumlichen FFT-basierten Beamforming-Algorithmen, die in dieser Arbeit besprochen werden.**

Die Ausgänge der Antennenelemente sind direkt mit 32 Heterodynempfängern verbunden . Diese Empfänger verwenden ein zentrales lokales Oszillatorsystem (LO), das einen rauscharmen Oszillator enthält, der mit einem Leistungsteilernetzwerk mit 32 Ausgängen verbunden ist. Die erste Stufe jedes Empfängers verfügt über einen rauscharmen Verstärker (LNA), der bei 5,8 GHz eine Verstärkung von 16 **dB mit einem Rauschmaß von 2,4 dB liefert.** Nach der Verstärkung werden die Signale durch einen Bandpassfilter **im Bereich von 4,7 GHz bis 6 GHz gefiltert, um Rauschen und Interferenzen außerhalb des Bandes zu eliminieren. Die gefilterten Signale werden dann gemischt und durch Tiefpassfilter geleitet, um herunterkonvertierte Eingänge mit niedriger Zwischenfrequenz (IF) zu erz**

Anschließend werden diese 32 Niederfrequenzsignale um etwa 30 dB verstärkt **und mit zwei ADC16x250-8 Analog-Digital-Wandlerkarten (ADC) digitalisiert. Jede ADC- Karte unterstützt 16 unsymmetrische Eingangskanäle, bietet eine Auflösung von 8 Bit und arbeitet mit Abtastraten von bis zu 250 Msample/s pro Kanal. Dieser Aufbau gewährleistet die hochpräzise Digitalisierung der empfangenen Signale für die weitere Verarbeitung [97].**

4.3. ADFT-Kerne und digitale ADFT-Spektralschätzung **Die**

besprochene 32-Punkt-ADFT wurde mithilfe von Xilinx-Tools implementiert. Die dünn besetzten Matrizen aus der in **(5) gezeigten Faktorisierung wurden zur Optimierung der Leistung als parallele Eingabe-/ Ausgabeblöcke** strukturiert **[97,98]. Die Verifizierung dieser Blöcke wurde mit MATLAB Simulink 8.5 durchgeführt, um die Zuverlässigkeit des Designs vor der Hardwarebereitstellung sicherzustellen. Anschließend wurde das verifizierte ADFT-Design auf das Xilinx Virtex-6 sx475T FPGA abgebildet.**

Um den ADFT-Kern in die **ROACH-2-** FPGA-Plattform **zu integrieren , nutzten wir zwei Bibliotheken der Collaboration for Astronomy Signal Processing and Electronics Research (CASPER)-Community [100,101].** Diese Integration ermöglichte es dem System, 32 gleichzeitige **HF-Strahlen parallel zu erzeugen .** Jeder Strahl entspricht einer diskreten Richtung, die mit der Funktion ÿk = arcsin(k/32) berechnet wird, **wobei** k **zwischen ÿ16 und 15 liegt. Die erfolgreiche Implementierung der ADFT auf dem ROACH-2- FPGA bestätigte nicht nur die Wirksamkeit des Algorithmus , sondern demonstrierte auch die praktische Anwendung fortschrittlicher digitaler Hardware- Designtechniken. Diese Arbeit unterstreicht das Potenzial der ADFT in verschiedenen Anwendungen, die eine effiziente HF-Strahlformung erfordern.**

Um das Problem der hohen Nebenkeulen in der ADFT zu lösen, haben wir eine räumliche Fenstertechnik implementiert. Dabei wurde festgestellt, dass die Nebenkeule im schlimmsten Fall 1,5 dB höher ist als die der (exakten) DFT. Es wurden vier Konfigurationen getestet: drei Fensterfunktionen (Butterworth, Hamming und Hanning) und eine Basislinie ohne Fensterung [85]. Die Auswahl zwischen diesen Fensterfunktionen wurde mithilfe eines Multiplexers gesteuert, der mit einem Steuerauswahlstift ausgestattet war .

Abbildung 4 zeigt die räumliche Fenstereinheit gekoppelt mit den ADFT-Dünnmatrizen. Mathematisch ausgedrückt wird die Fensterung als Diagonalmatrix eingeführt. Die ADFT- basierte Spektralschätzung kann durch die Kombination der Fensterdiagonalmatrix (Dwindowing), des ADFT-Kerns mit geringer Komplexität (Fˆ 32) und einer leistungsnormalisierenden Diagonalmatrix (Dnormalizer) beschrieben werden. Die mathematische Darstellung der vorgeschlagenen ungefähren Spektralschätzung ist gegeben durch

$$\text{ADFTestimator} = \text{Dnormalizer} \cdot F\hat{}\,32 \cdot \text{Dwindowing}. \quad (6)$$

Machine Translated by Google

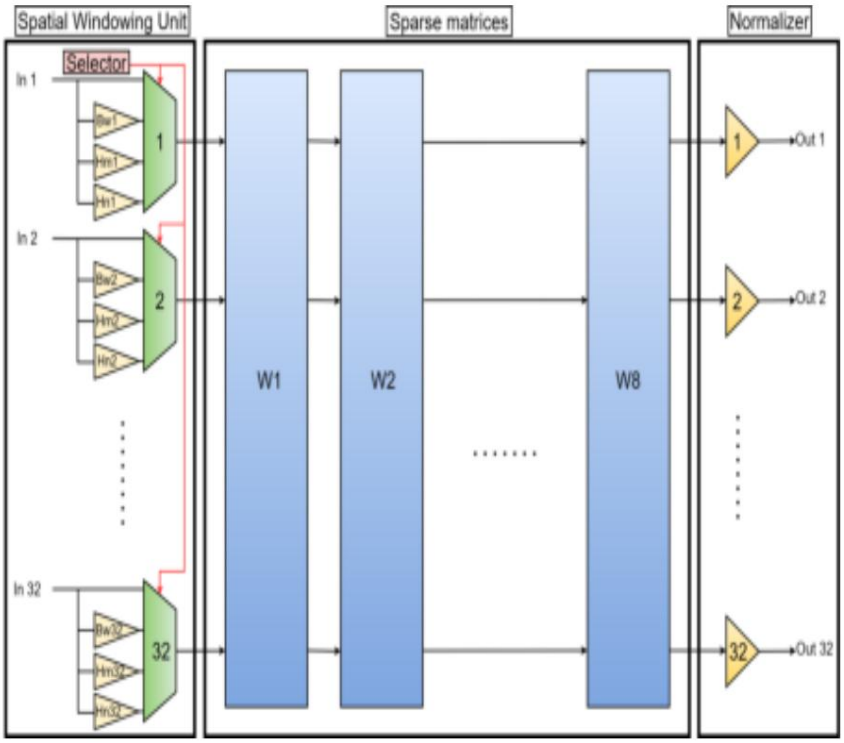


Abbildung 4. ADFT-Architektur mit räumlicher Fensterung und Leistungsnormalisierungseinheiten.

Abbildung 5 veranschaulicht die ADFT-Reaktion und ihre Änderungen unter den Fenstertechniken. Es ist bemerkenswert, dass die Einführung von Fensterfunktionen zu breiteren Hauptkeulen führte, was ein Kompromiss zur Unterdrückung der Nebenkeulen ist. Dieser Kompromiss zwischen Hauptkeulenbreite und Nebenkeulenunterdrückung ist bei der Anwendung von räumlichen Fenstertechniken zu berücksichtigen. Die Implementierung der räumlichen Fensterung erhöht jedoch die Hardwarekomplexität aufgrund der von der Fenstereinheit eingeführten Multiplexer. Daher muss der Benutzer die geeignete Konfiguration basierend auf seinen spezifischen Anforderungen und Einschränkungen auswählen.

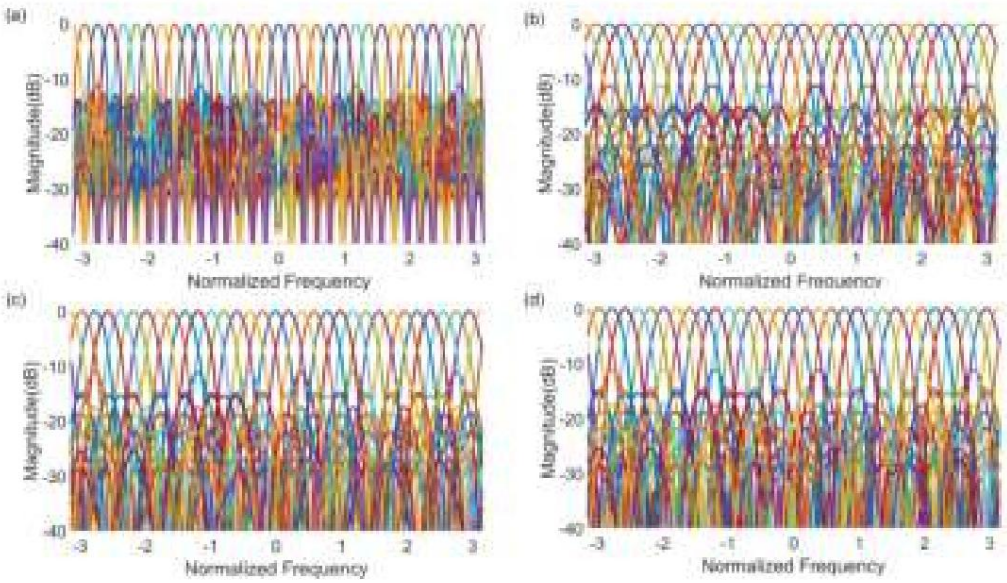


Abbildung 5. Simulierte Magnitudenreaktion der 32-Punkt-ADFT und ihre Modifikationen unter drei verschiedenen Fenstertechniken. Teildiagramm (a) zeigt die Basisreaktion ohne Fenstertechnik. Teildiagramme (b–d) zeigen die ADFT-Reaktionen nach Anwendung der Fenster Butterworth, Humming und Hanning .

Machine Translated by Google

4.4. Digitale FFT-Spektrometer

Das beschriebene System beinhaltet die Implementierung einer räumlichen FFT, die 32 parallele Ausgänge erzeugt, von denen jeder einem bestimmten räumlichen HF-Strahl entspricht. Im Anschluss an diese räumliche FFT wird auf jeden dieser 32 räumlichen HF-Strahlen ein zeitliches FFT-Spektrometer angewendet, wodurch eine Kanalisierung im zeitlichen Frequenzbereich erreicht wird. Jede zeitliche FFT wird mit 1024 Punkten berechnet, was die Verwendung von 32 parallelen FFT-Kernen erforderlich macht. Diese parallele Verarbeitung wird durch die Nutzung real abgetasteter Biplex-FFT-Blöcke erleichtert, die im Xilinx **Virtex-6 SX475T FPGA verfügbar sind .**

Die realwertabgetasteten Biplex-FFT-Blöcke nutzen die Methode der hermiteschen Konjugation und ermöglichen so die Berechnung der realwertabgetasteten FFT. Diese Technik verwendet einen komplexen FFT-Kern, um zwei realwertige Datenströme gleichzeitig zu transformieren. Als Ergebnis erzeugt die FFT-Operation alle Frequenzen, sowohl positive als auch negative, wobei negative Frequenzen das Spiegelbild ihrer positiven Gegenstücke sind **[100].**

Die zeitlichen FFT-Ausgaben für jeden Strahl ergeben 1024 komplexe Werte, die 1024 diskreten Frequenzbereichen entsprechen. Die Leistungsspektraldichte (PSD) wird durch Quadrieren der Größe jedes Frequenzbereichs geschätzt. Diese Schätzung wird durch Mittelung von 1024 zeitlichen FFT-Auswertungen abgeleitet, was alle 1 Million Taktzyklen zu einer neuen PSD-Messung führt . Ein Block-Random-Access-Memory-(BRAM)-Vektorakkumulatorblock sammelt die PSD- Werte und überträgt sie an den Softwareprozess, der vo Abbildung 6 zeigt den gesamten ADFT- und zeitlichen FFT-Fluss, einschließlich Energieintegration und BRAM-Registern. Die Python-Bibliothek casperfpga wird zur Interaktion mit der CASPER-Hardware verwendet und extrahiert Werte aus dem gemeinsam genutzten BRAM.

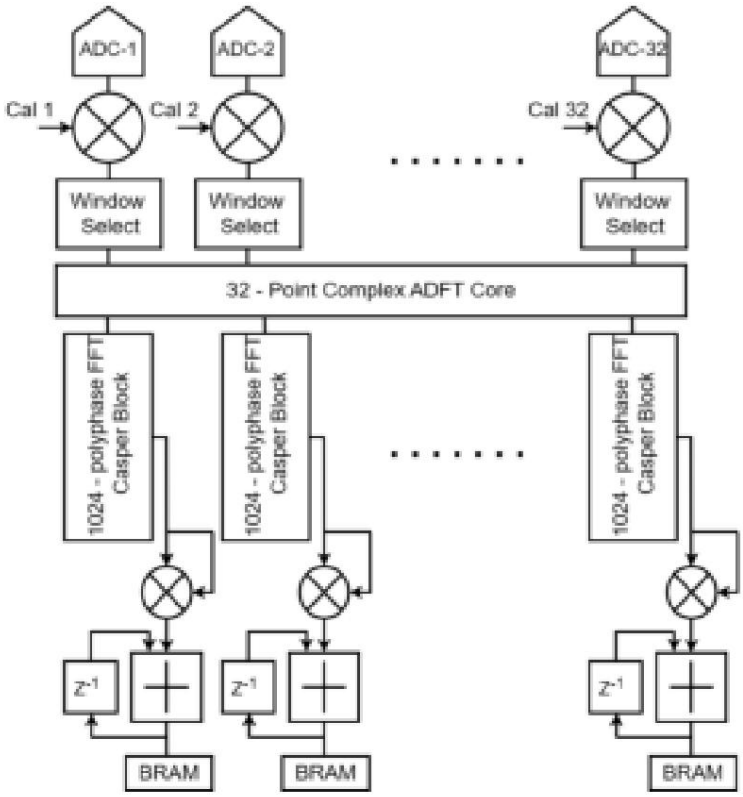


Abbildung 6. ADFT- und FFT-Kalibrierung, Energieintegration und Gesamtsystemblock einschließlich BRAM-Register.

Das FPGA arbeitet mit einer Taktfrequenz von 200 MHz. Bei dieser Konfiguration wird die Systemfrequenzauflösung für jeden Strahl auf ÿF = **100/1024 MHz festgelegt, was ungefähr 100 kHz entspricht. Die Nyquist-Bandbreite des Systems beträgt somit 100 MHz.**
Diese Konfiguration gewährleistet eine präzise Kanalisierung und hochauflösende Frequenzanalyse für jeden der 32 räumlichen HF-Strahlen.

Machine Translated by Google

4.5. Digitale Hochgeschwindigkeitsverbindungen

Die ROACH-2 -Karten sind mit vier SFP+-Ports auf einer einzigen 10- GbE -Mezzanine-Karte ausgestattet, wobei die Ethernet-Schnittstelle von einem integrierten 156,25- MHz- Quarzoszillator gesteuert wird . **Diese Taktfrequenz wird im FPGA intern mit dem Faktor 66 multipliziert, was zu einer effektiven Leitungsgeschwindigkeit von 10,3125 Gbit/s führt.**

Der CASPER 10 GbE Simulink-Kern ermöglicht die Übertragung und den Empfang von UDP- Paketen über IPv4, die in Ethernet-Frames gekapselt sind. Jeder Ethernet-Frame verursacht einen Overhead von 38 Bytes für den Frame-Header, 20 Bytes für den IPv4-Header und 16 Bytes für den UDP-Header, was insgesamt mindestens 74 Bytes Overhead pro Paket ergibt. Um die 64-Bit-Grenzen einzuhalten, können einige Header auch Zero-Padding enthalten, was den effektiven Durchsatz weiter beeinträchtigt. Die maximale Nutzlastgröße für den CASPER 10 GbE- Kern beträgt 8192 Bytes, implementiert im BRAM, wobei weitere 512 Bytes im verteilten RAM für anwendungsspezifische Header verfügbar sind **[100].**

5. Experimentelle Ergebnisse

5.1. Kalibrierung

Die Kalibrierung des RF-Array-Systems ist für eine optimale Strahlformungsleistung unerlässlich und wurde in zwei unterschiedlichen Phasen durchgeführt. Zunächst konzentrierte sich die Kalibrierung auf die ADCs, wobei Open-Source-Routinen verwendet wurden, die zuvor von Mitgliedern der CASPER-Gruppe für diese Hardware entwickelt wurden. Die zweite Phase befasste sich mit der digitalen Korrektur von Fehlanpassungen im Mikrowellen-Frontend. Relative Verstärkungs- und Phasenfehlanpassungen der IF-Ausgänge für jede Kette wurden in Bezug auf eine Referenzkette mit einem 5,71 GHz- Referenzträgereingang bestimmt . **Die relativen Verstärkungsunterschiede wurden mithilfe digitaler Multiplikatoren kompensiert, und Verstärkungsskalierungsfaktoren wurden durch die Messung der durchschnittlichen Leistungswerte jedes ADC- Ausgangs berechnet. Idealerweise sind zur Phasendifferenzkompensation Teilverzögerungsfilter erforderlich, da die RF-Ketten kleinere Verzögerungen als der Abtastperiodenwert aufweisen könnten. Als Zwischenlösung wurden ganzzahlige Verzögerungsfilter entwic**

5.2. Strahlmessungen

Die Leistungsmuster wurden durch Senden eines Dauerstrichsignals (CW) bei fRF = **5,71 GHz erhalten. Die LO-Signalfrequenz fLO bestimmt die ZF als fRF ÿ fLO. Für die Messungen in dieser Studie** wurde **fLO auf 5,7 GHz eingestellt, was eine ZF von 10 MHz ergab , wobei die herunterkonvertierten Ausgänge bei fCLK** = **200 MHz digitalisiert wurden.**

An jedem ADFT-Bin-Ausgang wurden digitale Integratoren verwendet, um die empfangene Energie über eine feste Dauer zu berechnen. Die räumliche Strahlformung erfolgt mit einer Rate von 200 MHz und die PSD wird in Echtzeit berechnet, indem die Größe jedes Frequenzbins quadriert wird. Diese Schätzung wird durch Mittelung von 1024 zeitlichen FFT-Berechnungen bei 200 MHz abgeleitet. Die Geometrie des Messaufbaus beeinflusst den gemessenen Array-Faktor der Strahlen erheblich. Idealerweise sollten Sender und Empfänger weit genug voneinander entfernt sein, sodass die auf das Empfänger-Array auftreffenden Wellen als ebene Wellen angenähert werden können. Gemäß den Standardrichtlinien [102] sollte **der Abstand zwischen Sender** und Empfänger bei 5,7 **GHz 20 m überschreiten , um sicherzustellen, dass sich die Empfängeröffnung im Fernfeld befindet. Die Strahlen wurden auf einem offenen Parkdeck mit einem Sender-Empfänger- Abstand von ungefähr 20 m gemessen.**

Abbildung 7 zeigt die gemessenen Array-Faktormuster für die Strahlen 0 bis 16. Aufgrund der Eingabe von realwertigen IF-Signalen in die ADFT weisen die resultierenden Strahlmuster Symmetrie auf und erzeugen Spiegelbilder für die verbleibenden Bins. Darüber hinaus wird das Vorhandensein von zwei Spiegelspitzen innerhalb der Bins 0 bis 16 beobachtet. Die gemessenen Strahlen entsprechen weitgehend den erwarteten Mustern und weisen einen Abstand von mehr als 10 dB zu

Jeder ADFT-Ausgabebereich entspricht einem HF-Fernfeldstrahl, der in einem diskreten Winkel ausgerichtet ist , der durch ÿk = arcsin(k/N) gegeben ist, **wobei** k = **ÿ16, ÿ15, . . . , 14, 15 und** N = **32 (Anzahl der Array-Elemente). Diese Konfiguration führt zu einem Strahlabstand von 180ÿ/32 = 5,625ÿ und bietet eine Winkelauflösung von ±2,8125ÿ .**

Machine Translated by Google

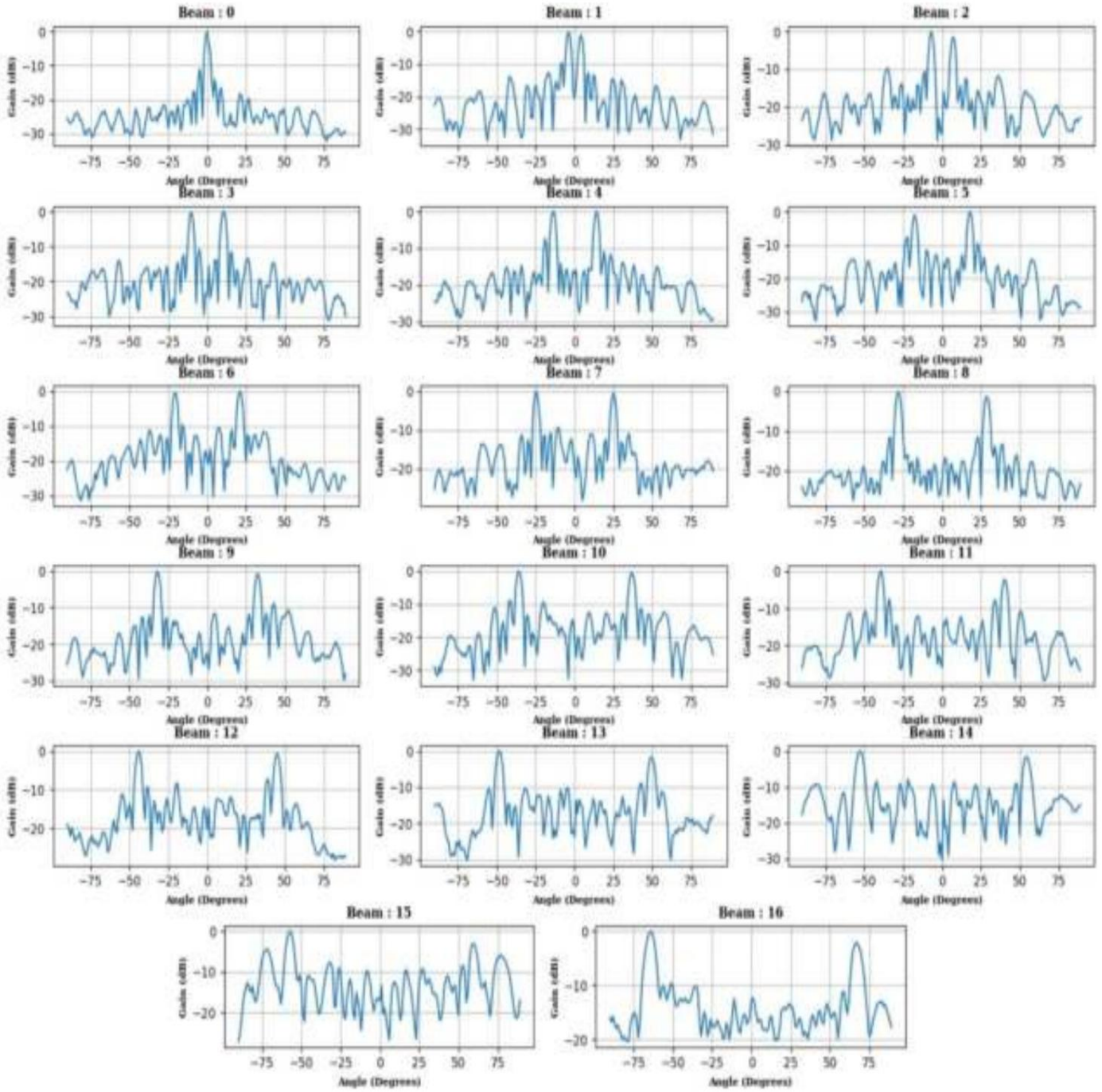


Abbildung 7. Sechzehn Strahlen, gemessen vom 5,7 GHz- Array (vertikale Achse in Dezibel, horizontale Achse der Azimutwinkel [ÿ90° , 90°]). Jede Unterabbildung enthält die gemessenen Array-Faktormuster des Strahls unter Verwendung der realen Eingänge des ADFT-Kerns. Die imaginäre Komponente jedes ADFT- Eingangs wird hier auf Null gesetzt, was erwartungsgemäß zu symmetrischen Hauptkeulen führt.

Abbildung 8 zeigt die gemessene zeitliche PSD-Verteilung der jeweiligen Strahlen bei 35°
mehrere Frequenzen, wenn die Signalquelle in einem Breitseitenwinkel von 15° und, ,
50° positioniert ist . **Der entsprechende Versuchsaufbau mit relativen Quellenpositionen und zu Testzwecken ausgesendeten HF-Frequenzen ist in Abbildung 9 dargestellt.**

Machine Translated by Google

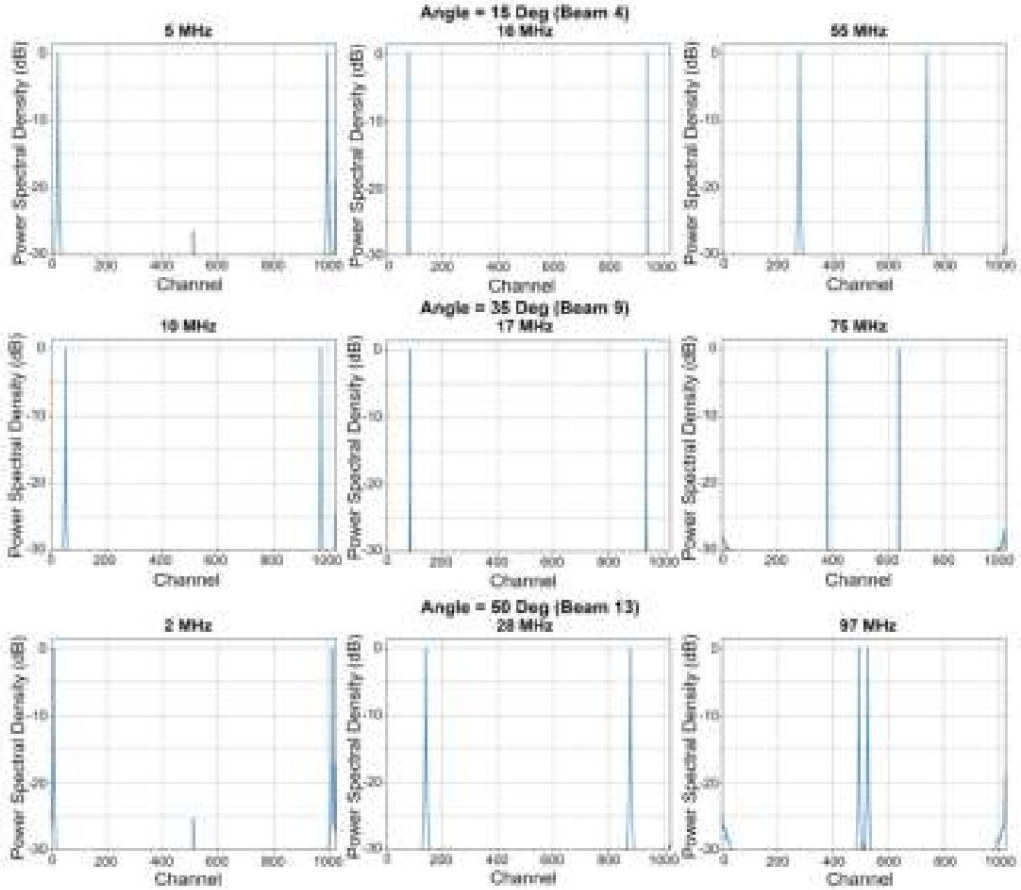


Abbildung 8. Zeitliche PSD über die 1024 diskreten Frequenzbereiche für den jeweiligen HF-Strahl bei mehreren Frequenzen, wenn die Quelle bei 15° platziert ist , 35° , und 50° Breitseiten-Empfängerwinkel.

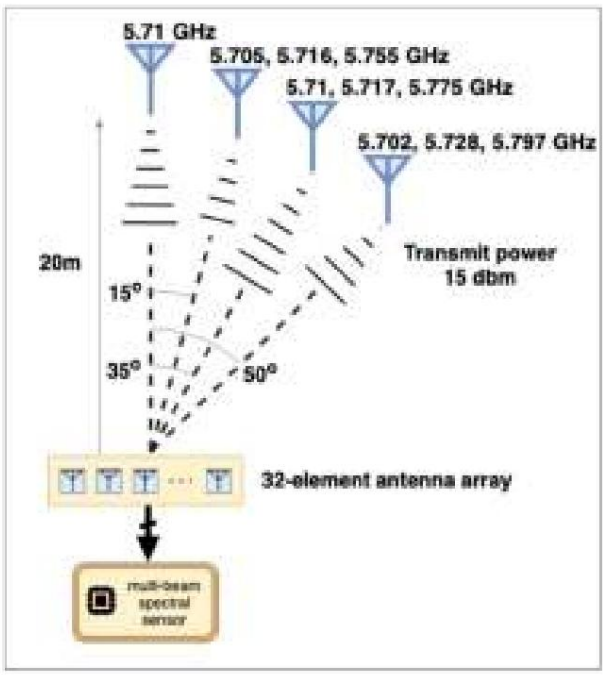


Abbildung 9. Darstellung des Gesamtaufbaus mit Sendeantennen, Empfangsantennen-Array und der Mehrstrahl-Spektralsensor. Der 32-Element-Antennenempfänger erfasst Signale von verschiedene Richtungen, wobei sich die Sender 20 m vom Empfänger entfernt befinden.

Machine Translated by Google

6. Diskussion

In diesem Dokument wird eine digitale Beamforming-Architektur mit geringer Komplexität vorgestellt, die entwickelt wurde, um Erzeugen Sie 32 gleichzeitige RF-Strahlen mit einer 32-Punkt-DFT-Näherung, wodurch Multiplikationsoperationen effektiv eliminiert werden. Wir validierten die vorgeschlagene Methode auf einem v 32-Elemente-Digital- 1-D- Receiver-Array mit 5,7 GHz. Das 1-D- Array verwendet 32 parallele ADCs zum Abtasten der Antennenausgänge und verwendet einen ADFT, der auf einem Xilinx FPGA implementiert ist um 32 HF-Strahlen in Echtzeit zu berechnen. Darüber hinaus wird in diesem Artikel die Verbesserung von unser bestehendes 32-Kanal 5,7 GHz Spektrum-Sensorsystem durch die Implementierung von 32 parallele 1024-bin temporale FFTs auf einer ROACH-2 Hardwareplattform. Dieses experimentelle Die Verifizierung mit Echtzeitmessungen über ein Frequenzband und Einfallswinkel demonstriert die Fähigkeit der Architektur, Informationen aus mehreren Richtungen mit einer Frequenzauflösung von 100 kHz und einer Richtungsauflösung von etwa 6ÿ .

Die Autoren planen, die ADFT auf eine größere Anzahl von Elementen auszuweiten und mehrdimensionale ADFT-Strahlformer zur Strahlsteuerung sowohl in Elevation als auch Azimut Ebenen. Mit einem 32 × 8-Array könnten wir beispielsweise den Strahlensatz auf ein rechteckiges Raster der Größe 32 × 8, wobei sowohl 32-Punkt- als auch 8-Punkt-FFTs durch die entsprechenden ADFT-Algorithmen basierend auf der Faktorisierung dünner Matrizen [103–106].

Autorenbeiträge: Konzeptualisierung und Systemdesign durch AM Theorie und Design von ADFT und schneller Algorithmus von RJC Multidimensionale Signalverarbeitung Review von CUSE Review von Radio Astronomie-Interferenzprobleme von TG Digital Implementierung des ADFT-Kerns von KL Experimental Aufbau, digitales Design der CASPER-Tools, Kalibrierung und Echtzeitexperimente durch SS und BUK Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu.

Finanzierung: Diese Arbeit wurde teilweise von der NSF unterstützt (im Rahmen der Zuschüsse FuSE 2329012, SWIFT 2229471, ICORPS 2243346, SpectrumX), NTIA, CNPq (Brasilien) im Rahmen des Zuschusses 315047/23-2 und Senate Research Ausschuss der Universität Mortuwa im Rahmen des Stipendiums SRC/LT/2020/08.

Datenverfügbarkeitserklärung: Alle Originalbeiträge dieser Studie sind dokumentiert im Artikel. Für weitere Einzelheiten wenden Sie sich bitte an die entsprechenden Autoren.

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Abkürzungen

In diesem Manuskript werden die folgenden Abkürzungen verwendet:

ADC	Analog-Digital-Umsetzer;
ADFT	Ungefähre diskrete Fourier-Transformation;
KI	Künstliche Intelligenz;
ALMA	Atacama großes Millimeter-Array;
BRAM	Block-Direktzugriffsspeicher;
CASPER	Zusammenarbeit für astronomische Signalverarbeitung und Elektronikforschung;
CSI	Informationen zum Kanalstatus;
CW	Kontinuierliche Welle;
DFT	Diskrete Fourier-Transformation;
DL	Tiefes Lernen;
DSA	Dynamischer Spektrumzugriff;
DSP	Digitale Signalverarbeitung;
FCC	Eidgenössische Kommunikationskommission;
FFT	Schnelle Fourier-Transformation;
FPGA	Feldprogrammierbares Gate-Array;
GNSS	Globales Navigationssatellitensystem;
WENN	Zwischenfrequenz;

Machine Translated by Google

Internet der Dinge	**Internet der Dinge;**
ISI	**Intersymbol-Interferenz;**
ITU	**Internationale Fernmeldeunion;**
LNA	**Rauscharmer Verstärker;**
LO	**Lokaler Oszillator;**
ML	**Maschinelles Lernen;**
ngVLA	**Sehr großes Array der nächsten Generation;**
NSF	**Nationale Wissenschaftsstiftung;**
OET	**-Amt für Ingenieurwesen und Technologie;**
Leiterplatte	**Leiterplatte;**
PSD	**Leistungsspektraldichte;**
aus PU	**Hauptbenutzer;**
RF	**Radiofrequenz;**
Informationsmanifold	**Hochfrequenzstörungen;**
RL	**Verstärkendes Lernen;**
ROACH	**Rekonfigurierbare Computerhardware mit offener Architektur;**
SDR	**Softwaredefiniertes Radio;**
SFP	**Steckbarer Kleinformfaktor;**
SKA	**Quadratkilometer-Array;**
SNR	**Signal-Rausch-Verhältnis;**
SO	**Sekundärbenutzer;**
ULA	**Gleichmäßiges lineares Array;**
WSU	**-Breitband-Empfindlichkeitsverbesserung.**

Anhang A

Anhang A.1. Schneller 32-Punkt-ADFT-Algorithmus

Der schnelle Algorithmus für die besprochene 32-Punkt-Approximation DFT besteht aus einer Faktorisierung in niedrigkomplexe, dünn besetzte Matrizen, wie in (5) gezeigt. Die Näherungsmatrix und ihr Die Faktorisierungsterme sind unten explizit angegeben:

$$\mathbf{32} = \qquad \text{, (A1)}$$

Machine Translated by Google

wobei $\ddot{y} = 1 + j$ **und** $\bar{y}$ **bezeichnet komplexe Konjugation:**

$$W1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{(A2)}$$

$$W2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{(A3)}$$

$$W3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{(A4)}$$

Machine Translated by Google

$$W4 = \begin{bmatrix} 1000100000000000000000000000000000000101000000000000000000000000000000000100000000000000000 \\ 0000000000000000010ÿ1000000000000000000000000000000001000ÿ1000000000000000000000000000000000 \\ 0001000000000000000000000000000000000101000000000000000000000000000000001000000000000000000 \\ 0000000000000000000010ÿ1000000000000000000000000000000001000000000000000000000000000000000 \\ 0000010100000000000000000000000000000001000000000000000000000000000010ÿ1000000000 \\ 0000000000000000000000010100000000000000000000000000001000000000000000000000000000000000 \\ 0000010ÿ1000000000000000000000000000000001000000000001000000000000000000001000000000 \\ 0000000000000000000000000100000000000000000000000000001000000000000000000000000000000000 \\ 0000000001000100000000000000000000000000000001000000000000000000000000000000000001000000 \\ 0000000000000000000000001000000000000000000000001000ÿ1000000000000000000000000000000 \\ 0000000000100000000000000000000000000000000001000000000000000000000000000000000010000 \\ 0000000000000000010000000000ÿ1000 \end{bmatrix}, \quad (A5)$$

$$W5 = \begin{bmatrix} 0000000000000000000000000000000100000000000000000000000000000000000010 \\ 0000000000000000000000000000000001 \\ \\ 1010000000000000000000000000000000 \\ 0100000000000000000000000010ÿ1000000000000000000000000011000000000000 \\ 0000000000000000001ÿ1000000000000000000000000000001001000000000000000000000000000 \\ 0000110000000000000000000000000001ÿ1000000000000000000000000000100ÿ1000000000 \\ 0000000000000000000000001001000000000000000000000000011000000000000000000000000000 \\ 000001ÿ1000000000000000000000000000100ÿ1000000000000000000000000011000000 \\ 0000000000000000000000001ÿ1000000000000000000000000001000000000000000000000000 \\ 000000000ÿ1000000000001000000000000000000001000000000000000000000000100000 \\ 1000000000000000000000000010101000000000000000000000000010100000000000000000 \\ 000000001ÿ1000000000000000000000000010ÿ1000000000000000000000001000ÿ100 \\ 0000000000000000000010000ÿ1000000000000000000000000001000000000000000 \\ 0000000000000010101000000000000000000000000010101000000000000000000010 \\ ÿ1000 \end{bmatrix}, \quad (A6)$$

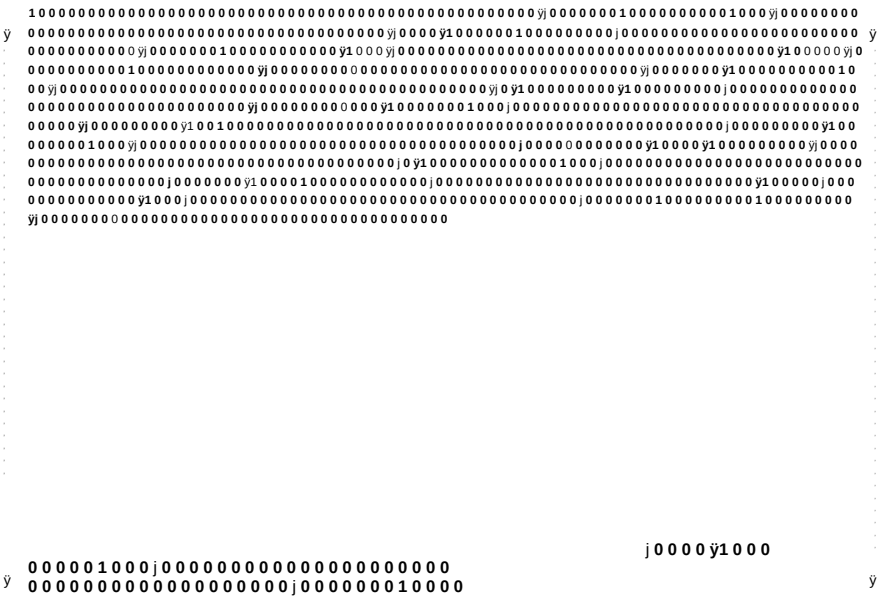
$$W6 = \begin{bmatrix} 0000000000000000000000000010ÿ100 \\ 0000000000000001000000000000001000000000000000000000000001000ÿ1 \\ \\ 1100000000000000000000000000000000 \\ 1ÿ1000000000000000000000000000000 \\ 0010000000000000000000000000000010000000000000000000000000000010000000000000 \\ 0000000000000000001000000000000000000000000000000001000000000000000 \\ 0001000000000000000000000000000010000000000000000000000000000010000000000 \\ 0000000000000000000010000000000000000000000000000001000000000000000 \\ 0001000000000000000000000000000010000000000000000000000000000010000000000 \\ 0000000000000000000010000000000000000000000000000001000000000000000 \\ 0000100010ÿ1000000000000000000000000000010000000000000000000000000011000000000 \\ 0000000000000000000001ÿ1000000000000000000000000001000ÿ1000000000000000000000000 \\ 0000000010000000000000000000000000000010000010000000000000000000000000010000000 \\ 0000000000000000000000000001000010ÿ1 \end{bmatrix}, \quad (A7)$$

$$\begin{bmatrix} 0000000000000000000000000010000000000000000000000000001001000000000000000000 \\ 0000000001000000000000000000000000001000ÿ1000000000000000000000000000100ÿ10000 \\ 0000000000000000000000001000001 \end{bmatrix}$$

Machine Translated by Google

W7 =  , **(A8)**

Und

W8 =  . **(A9)**

Verweise

1. **Tolkien, J.** Der Hobbit und der Herr der Ringe; **Houghton Mifflin Harcourt: Boston, MA, USA, 1954.**
2. **Kang, S.; Mezzavilla, M.; Rangan, S.; Madanayake, A.; Venkatakrishnan, SB; Hellbourg, G.; Ghosh, M.; Rahmani, H.; Dhananjay, A. Mobilfunknetze im oberen Mittelband.** IEEE Open J. Commun. Soc. **2024, 5, 2058–2075. [CrossRef]**
3. **Das Weiße Haus.** Nationale Spektrumstrategie; **Das Weiße Haus: Washington, DC, USA, 2023.**
4. **Restuccia, F.; Melodia, T. Deep Learning auf der physischen Ebene: Systemherausforderungen und Anwendungen für 5G und darüber hinaus.** IEEE Commun . Mag. **2020,** 58, **58–64.**
5. **Triaridis, K.; Doumanidis, C.; Chatzidiamantis, ND; Karagiannidis, GK MM-Net: Ein multimodaler Ansatz zur automatischen Modulationsklassifizierung.** IEEE Commun. Lett. **2023, 28, 328–331. [CrossRef]**
6. **Baldesi, L.; Restuccia, F.; Melodia, T. ChARM: NextG-Spektrum-Sharing durch datengesteuerte dynamische O-RAN- Steuerung in Echtzeit. In Proceedings der IEEE International Conference on Computer Communications (INFOCOM), London, Großbritannien, 2.–5. Mai 2022; IEEE: Piscataway, NJ, USA, 2022; S. 240–249.**
7. **Zhang, L.; Xiao, M.; Wu, G.; Alam, M.; Liang, YC; Li, S. Eine Übersicht über fortschrittliche Techniken zur Frequenzteilung in 5G-Netzwerken.** IEEE Wirel. Commun. **2017,** 24, **44–51. [CrossRef]**
8. **Hu, F.; Chen, B.; Zhu, K. Full Spectrum Sharing in Cognitive Radio Networks toward 5G: Eine Untersuchung.** IEEE Access **2018, 6, 15754–15776. [CrossRef]**
9. **Shokri-Ghadikolaei, H.; Boccardi, F.; Fischione, C.; Fodor, G.; Zorzi, M. Spektrum-Sharing in mmWave-Mobilfunknetzen über Zellassoziation , Koordination und Strahlformung.** IEEE J. Sel. Areas Commun. **2016,** 34, **2902–2917. [CrossRef]**
10. **Lv, L.; Chen, J.; Ni, Q.; Ding, Z.; Jiang, H. Kognitiver nichtorthogonaler Mehrfachzugriff mit kooperativer Weiterleitung: Eine neue drahtlose Grenze für die gemeinsame Nutzung des 5G-Spektrums.** IEEE Commun. Mag. **2018,** 56, **188–195. [CrossRef]**
11. **Agarwal, S.; De, S. eDSA: Energieeffiziente dynamische Spektrumzugriffsprotokolle für kognitive Funknetzwerke.** IEEE Trans. Mob. 2016 , 15, **3057–3071 .**

Machine Translated by Google

12. **Chiwewe, TM; Hancke, GP Schnelle Konvergenz Kooperativer dynamischer Spektrumzugriff für kognitive Funknetzwerke.** IEEE Trans. Ind. Inform. **2017,** 14, **3386–3394.** [CrossRef]
13. **Chang, HH; Song, H.; Yi, Y.; Zhang, J.; He, H.; Liu, L. Distributiver dynamischer Spektrumzugriff durch Deep Reinforcement Learning: Ein auf Reservoir Computing basierender Ansatz.** IEEE Internet Things J. **2018, 6, 1938–1948.** [CrossRef]
14. **Nguyen, HQ; Nguyen, BT; Dong, TQ; Ngo, DT; Nguyen, TA Deep Q-Learning mit Multiband Sensing für dynamischen Spektrumzugriff. In Proceedings des IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Seoul, Republik Korea, 22.–25. Oktober 2018; S. 1–5.** [CrossRef]
15. **Naparstek, O.; Cohen, K. Deep Multi-user Reinforcement Learning für verteilten dynamischen Spektrumzugriff.** IEEE Trans. Wirel. 2019 , 18, **310–323 .**
16. **Wang, S.; Liu, H.; Gomes, PH; Krishnamachari, B. Deep Reinforcement Learning für dynamischen Mehrkanalzugriff in drahtlosen Netzwerke.** IEEE Trans. Cogn. Commun. Netw. **2018, 4, 257–265.** [CrossRef]
17. **Yu, Y.; Wang, T.; Liew, SC Deep-Reinforcement Learning Mehrfachzugriff für heterogene drahtlose Netzwerke.** IEEE J. Sel. Areas Commun. **2019,** 37, **1277–1290.** [CrossRef]
18. **Reston, VA Ligado Networks verklagt die US-Regierung wegen unrechtmäßiger und unentgeltlicher Inbesitznahme von Ligados exklusivem, FCC - lizenziertem Spektrum. 2023. Online verfügbar: https://ligado.com/press/ligado-networks-sues-us-government-for-unlawful- and-uncompensated-taking-of-ligados-exclusive-fcc-licensed-spectrum/ (abgerufen am 28. Mai 2024).**
19. **Barden, S.; Dewdney, P.; Friesen, R.; Murowinski, R.; Savanandam, S.** ATD Strategic R&D Review Report—Signal Processing Portfolio; **Herzberg Astronomy and Astrophysics Research Centre-National Research Council Canada: Victoria, BC, Kanada, 2022.**
20. **FCC Spectrum Policy Task Force.** Bericht der Spectrum Efficiency Working Group, Nov. 2002; **FCC: Washington, DC, USA, 2024.**
21. **Mitola, J.; Maguire, G. Cognitive Radio: Software-Radios persönlicher gestalten.** IEEE Pers. Commun. **1999, 6, 13–18.** [CrossRef]
22. **Haykin, S. Cognitive Radio: Drahtlose Kommunikation mit Hilfe des Gehirns.** IEEE J. Sel. Areas Commun. **2005,** 23, **201–220.** [CrossRef]
23. **Liang, YC; Chen, KC; Li, GY; Mahonen, P. Kognitive Funknetzwerke und -kommunikation: Ein Überblick.** IEEE Trans. Veh. Technol. **2011,** 60, **3386–3407.** [CrossRef]
24. **Wang, B.; Liu, KR Fortschritte bei kognitiven Funknetzwerken: Eine Untersuchung.** IEEE J. Sel. Top. Signal Process. **2011, 5, 5–23.** [CrossRef]
25. **Talebi, SP Primärer Dienstausfall und sekundäre Dienstleistung in kognitiven Funknetzwerken.** Wirel. Commun. Mob. Comput. **2015,** 15, **1982–1990.** [CrossRef]
26. **Zhao, Q.; Sadler, BM Eine Untersuchung zum dynamischen Spektrumzugriff.** IEEE Signal Process. Mag. **2007,** 24, **79–89.** [CrossRef]
27. **Ji, Z.; Liu, KR Dynamic Spectrum Sharing: Ein spieltheoretischer Überblick.** IEEE Commun. Mag. **2007,** 45, **88–94.** [CrossRef]
28. **Bhattarai, S.; Park, JMJ; Gao, B.; Bian, K.; Lehr, W. Ein Überblick über Dynamic Spectrum Sharing: Laufende Initiativen, Herausforderungen, und ein Fahrplan für zukünftige Forschung.** IEEE Trans. Cogn. Commun. Netw. **2016, 2, 110–128.** [CrossRef]
29. **Zeng, Y.; Liang, YC; Zhang, R. Blind kombinierte Energieerkennung für die Spektrumserfassung im kognitiven Radio.** IEEE Signal Prozess. Lett. **2008,** 15, **649–652.** [CrossRef]
30. **Ling, X.; Wu, B.; Wen, H.; Ho, PH; Bao, Z.; Pan, L. Adaptive Schwellenwertsteuerung für energiedetektionsbasierte Spektrumsensorik in Kognitive Radios.** IEEE Wirel. Commun. Lett. **2012, 1, 448–451.** [CrossRef]
31. **Ranjan, A.; Anurag; Singh, B. Design und Analyse der Spektrumerfassung im kognitiven Radio basierend auf Energieerkennung. In Proceedings der International Conference on Signal and Information Processing, Nanded, Indien, 6.–8. Oktober 2016; S. 1–5.**
32. **Arjoune, Y.; Mrabet, ZE; Ghazi, HE; Tamtaoui, A. Spektrumsmessung: Verbesserte Energieerkennungstechnik basierend auf Rauschmessung . In Proceedings of the IEEE Annual Computing and Communication Workshop and Conference, Las Vegas, NV, USA, 8.–10. Januar 2018; S. 828–834.**
33. **Jiang, C.; Li, Y.; Bai, W.; Yang, Y.; Hu, J. Robuste Spektrumerfassung auf Basis statistischer Matched-Filter in Umgebungen mit Rauschunsicherheit . In Proceedings der International Conference on Communication Technology, Chengdu, China, 9.–11. November 2012; S. 1209–1213.**
34. **Zhang, X.; Chai, R.; Gao, F. Auf angepasstem Filter basierende Spektrumerfassung und Leistungspegelerkennung für kognitives Funknetzwerk. In Proceedings der IEEE Global Conference on Signal and Information Processing, Atlanta, GA, USA, 3.–5. Dezember 2014; S. 1267–1270.**
35. **Lv, Q.; Gao, F. Spektrumsmessung und Leistungspegelerkennung mit mehreren Antennen auf Basis angepasster Filter. In Proceedings des IEEE China Summit und der International Conference on Signal and Information Processing, Chengdu, China, 12.–15. Juli 2015; S. 305–309.**
36. **Ilyas, I.; Paul, S.; Rahman, A.; Kundu, RK Vergleichende Bewertung der kognitiven Spektrumerfassung auf Basis der zyklostationären Erkennung. In Proceedings der IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference, New York, NY, USA, 20.–22. Oktober 2016; S. 1–7.**
37. **Cohen, D.; Eldar, YC Komprimierte zyklostationäre Erkennung für Cognitive Radio. In Proceedings der 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5.–9. März 2017; S. 3509–3513.**
38. **Zeng, Y.; Liang, YC Spektrum-Erfassungsalgorithmen für kognitives Radio basierend auf statistischen Kovarianzen.** IEEE Trans. Veh. Technol. **2009,** 58, **1804–1815.** [CrossRef]
39. **Zeng, Y.; Liang, YC Eigenwertbasierte Spektrum-Sensoralgorithmen für kognitives Radio.** IEEE Trans. Commun. **2009,** 57, **1784–1793.** [Querverweis]
40. **Zayen, B.; Hayar, A.; Kansanen, K. Blind Spectrum Sensing für Cognitive Radio basierend auf Signal Space Dimension Estimation. In Proceedings der 2009 IEEE International Conference on Communications, Dresden, Deutschland, 14.–18. Juni 2009; S. 1–5.**

Machine Translated by Google

41. **Farhang-Boroujeny, B. Filterbank-Spektrumsensorik für kognitive Radios.** IEEE Trans. Signal Process. **2008,** 56, **1801–1811.** [Querverweis]
42. **Quan, Z.; Cui, S.; Sayed, AH; Schlecht, HV Optimale Multiband-Joint-Erkennung für die Spektrumerfassung in kognitiven Funknetzwerken.** IEEE Trans. Signal Process. **2009,** 57, **1128–1140.** [CrossRef]
43. **Raghu, I.; Chowdary, SS; Elias, E. Effiziente Spektrumerfassung für Cognitive Radio unter Verwendung von Cosinus-modulierten Filterbänken. In Proceedings der IEEE Region 10-Konferenz, Singapur, 22.–25. November 2016; S. 2086–2089.**
44. **Kumar, A.; Saha, S.; Bhattacharya, R. Verbesserte Kantenerkennung auf Basis der Wavelet-Transformation für die Breitbandspektrumerfassung im Cognitive Radio. In Proceedings of the USNC-URSI Radio Science Meeting, Fajardo, Puerto Rico, 26. Juni–1. Juli 2016; S. 21–22.**
45. **Polo, YL; Wang, Y.; Pandharipande, A.; Leus, G. Compressive wide-band spectrum sensing. In Proceedings der IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19.–24. April 2009; Seiten 2337–2340.**
46. **Arjoune, Y.; Kaabouch, N. Wideband Spectrum Sensing: Ein bayesianischer Compressive Sensing-Ansatz.** Sensors **2018,** 18, **1839.** [Querverweis] [PubMed]
47. **Li, Z.; Chang, B.; Wang, S.; Liu, A.; Zeng, F.; Luo, G. Dynamische kompressive Breitband-Spektrumsensierung basierend auf Kanalenergierekonstruktion im kognitiven Internet der Dinge.** IEEE Trans. Ind. Inform. **2018,** 14, **2598–2607.** [CrossRef]
48. **Yang, L.; Fang, J.; Duan, H.; Li, H. Schnelle Schätzung des komprimierten Leistungsspektrums: Auf dem Weg zu einer praktischen Lösung für Breitband Spektrumsensorik.** IEEE Trans. Wirel. Commun. **2020,** 19, **520–532.** [CrossRef]
49. **Ding, G.; Wu, Q.; Yao, YD; Wang, J.; Chen, Y. Kernelbasiertes Lernen für die statistische Signalverarbeitung in kognitiven Funknetzwerken : Theoretische Grundlagen, Anwendungsbeispiele und zukünftige Richtungen.** IEEE Signal Process. Mag. **2013,** 30, [Querverweis]
50. **Thilina, KM; Choi, KW; Saquib, N.; Hossain, E. Maschinelle Lerntechniken für kooperatives Spectrum Sensing in kognitiven Funknetze.** IEEE J. Sel. Areas Commun. **2013,** 31, **2209–2221.** [CrossRef]
51. **Mikaeil, AM; Guo, B.; Wang, Z. Machine Learning to Data Fusion Approach for Cooperative Spectrum Sensing. In Proceedings der International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, Shanghai, China, 13.–15. Oktober 2014; Seiten 429–434.**
52. **Ghazizadeh, E.; Nikpour, B.; Moghadam, DA; Nezamabadi-pour, H. Eine PSO-basierte Gewichtungsmethode zur Verbesserung maschineller Lerntechniken für kooperatives Spektrum-Sensing in CR-Netzwerken. In Proceedings of the Conference on Swarm Intelligence and Evolutionary Computation, Bam, Iran, 9.–11. März 2016; S. 113–118.**
53. **Khalfi, B.; Zaid, A.; Hamdaoui, B. Wenn maschinelles Lernen auf Compressive Sampling für die Breitband-Spektrumsensierung trifft. In Proceedings der International Wireless Communications and Mobile Computing Conference, Valencia, Spanien, 26.–30. Juni 2017; S. 1120–1125.**
54. **Lee, W.; Kim, M.; Cho, DH Deep Cooperative Sensing: Kooperative Spektrumerfassung basierend auf Convolutional Neural Networks.** IEEE Trans. Veh. Technol. **2019, 68 , 3005–3009.**
55. **Liu, C.; Wang, J.; Liu, X.; Liang, YC Deep CM-CNN für Spectrum Sensing im Cognitive Radio.** IEEE J. Sel. Areas Commun. **2019,** 37, **2306–2321.** [CrossRef]
56. **Soni, B.; Patel, DK; López-Benítez, M. Auf Langzeit-Kurzzeitgedächtnis basierendes Spektrum-Erfassungsschema für kognitives Radio unter Verwendung Primäre Aktivitätsstatistik.** IEEE Access **2020, 8, 97437–97451.** [CrossRef]
57. **Peng, Q.; Gilman, A.; Vasconcelos, N.; Cosman, PC; Milstein, LB Robustes Deep Sensing durch Transferlernen im kognitiven Radio.** IEEE Wirel. Commun. Lett. **2020, 9, 38–41.** [CrossRef]
58. **Zhang, W.; Wang, Y.; Chen, X.; Cai, Z.; Tian, Z. Spectrum Transformer: Ein aufmerksamkeitsbasierter Breitband-Spektrumdetektor.** IEEE Trans. Wirel. Commun. **2024, 1–1.** [CrossRef]
59. **Sarikhani, R.; Keynia, F. Kooperative Spektrumerfassung trifft auf maschinelles Lernen: Deep Reinforcement Learning-Ansatz.** IEEE Commun. Lett. **2020,** 24, **1459–1462.** [CrossRef]
60. **Cai, P.; Zhang, Y.; Pan, C. Koordinationsgraph-basiertes Deep Reinforcement Learning für kooperatives Spectrum Sensing unter Korreliertes Fading.** IEEE Wirel. Commun. Lett. **2020, 9, 1778–1781.** [CrossRef]
61. **Gao, A.; Du, C.; Ng, SX; Liang, W. Ein kooperatives Spectrum Sensing mit Multi-Agent Reinforcement Learning-Ansatz in kognitiven Funknetzwerken.** IEEE Commun. Lett. **2021,** 25, **2604–2608.** [CrossRef]
62. **Ngo, QT; Jayawickrama, BA; He, Y.; Dutkiewicz, E. Multi-Agent DRL-basierte RIS-unterstützte Spektrumerfassung in kognitiven Satelliten-terrestrische Netzwerke.** IEEE Wirel. Commun. Lett. **2023,** 12, **2213–2217.** [CrossRef]
63. **Ali, A.; Hamouda, W. Fortschritte bei der Spektrumerfassung für kognitive Funknetzwerke: Theorie und Anwendungen.** IEEE Commun. Surv. Tutor. **2017,** 19, **1277–1304.** [CrossRef]
64. **Arjoune, Y.; Kaabouch, N. Eine umfassende Untersuchung zur Spektrumerfassung in kognitiven Funknetzwerken: Jüngste Fortschritte, Herausforderungen und zukünftige Forschungsrichtungen.** Sensors **2019,** 19, **126.** [CrossRef]
65. **Zhang, Y.; Luo, Z. Ein Überblick über die Forschung zur Spektrumerfassung basierend auf Deep Learning.** Electronics **2023,** 12, **4514.** [CrossRef]
66. **Wijenayake, C.; Madanayake, A.; Kota, J.; Bruton, L. Raum-Zeit-Spektrale weiße Räume im kognitiven Radio: Theorie, Algorithmen, und Schaltkreise.** IEEE J. Emerg. Sel. Top. Circuits Syst. **2013, 3, 640–653.** [CrossRef]
67. **Madanayake, A.; Wijenayake, C.; Tran, N.; Cooklev, T.; Hum, S.; Bruton, LT Richtungsabhängige Spektrumerfassung mit abstimmbaren mehrdimensionalen raumzeitlichen diskreten Filtern. In Proceedings des IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, San Francisco, CA, USA, 25.–28. Juni 2012; S. 1–6.**

Machine Translated by Google

68. **Wijenayake, C.; Madanayake, A.; Bruton, LT; Devabhaktuni, V. DOA-Schätzung und Quellenlokalisierung in CR-Netzwerken unter Verwendung lenkbarer 2-D-IIR-Strahlfilter. In Proceedings of the IEEE International Symposium on Circuits and Systems, Peking, China, 19.–23. Mai 2013; S. 65–68.**
69. **Gunaratne, TK; Bruton, LT Adaptive 2D FIR Trapezfilter mit komplexen Koeffizienten für Breitband-Beamforming in kognitiven Funksysteme.** Schaltungssysteme. Signalverarbeitung. **2011,** 30, **587–608. [CrossRef]**
70. **Wilcox, D.; Tsakalaki, E.; Kortun, A.; Ratnarajah, T.; Papadias, CB; Sellathurai, M. Über räumliches kognitives Radio unter Verwendung Parasitäre Antennenarrays für einzelne Funkgeräte.** IEEE J. Sel. Areas Commun. **2013,** 31, **571–580. [CrossRef]**
71. **Qian, R.; Sellathurai, M.; Ratnarajah, T. Richtungsabhängige Spektrumerfassung für kognitives Radio unter Verwendung von ESPAR-Arrays mit einer einzigen HF- Kette. In Proceedings der 2014 European Conference on Networks and Communications (EuCNC), Bologna, Italien, 23.–26. Juni 2014; S. 1–5.**
72. **Yazdani, H.; Vosoughi, A. Über kognitive Funksysteme mit Richtantennen und unvollkommener Spektrumerfassung. In Proceedings der 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5.–9. März 2017; S. 3589–3593.**
73. **Pulipati, S.; Ariyarathna, V.; Edussooriya, CUS; Wijenayake, C.; Wang, X.; Madanayake, A. Echtzeit-FPGA-basierte Mehrstrahl- Richtungsmessung von 2,4-GHz-ISM-HF-Quellen. In Proceedings der Moratuwa Engineering Research Conference (MERCon) 2019 , Moratuwa, Sri Lanka, 3.–5. Juli 2019; S. 129–134.**
74. **Richard Thompson, A.; Gergely, TE; Vanden Bout, PA Interferenz und Radioastronomie.** Phys. Today **1991,** 44, **41–49. [CrossRef]**
75. **Wissenschaftliche Ziele – Square Kilometre Array Observatory. Online verfügbar: https://www.skao.int/en/explore/science-goals (Zugri am 27. Mai 2024).**
76. **Carpenter, J.; Brogan, C.; Iono, D.; Mroczkowski, T. Das ALMA2030-Breitband-Empfindlichkeits-Upgrade.** arXiv **2022, arXiv:2211.00195.**
77. **Wissenschaft mit einem Very Large Array der nächsten Generation. Online verfügbar: https://ngvla.nrao.edu/page/science (aufgerufen am 27. Mai 2024).**
78. **Sihlangu, I.; Oozeer, N.; Bassett, BA Natur und Entwicklung von UHF- und L-Band-Radiofrequenzstörungen am MeerKAT- Radioteleskop. In Proceedings des RFI-Workshops 2022, Europäisches Zentrum für mittelfristige Wettervorhersage (ECMRWF), Online, 14.–18. Februar 2022.**
79. **Galt, J. Kontamination durch Satelliten.** Nature **1990,** 345, **483–483. [CrossRef]**
80. **De Pree, CG; Anderson, CR; Zheleva, M. 2023. Die Astronomie ist durch Funkstörungen von Satelliten bedroht – hier erfahren Sie, was dagegen getan werden kann. Online verfügbar: https://www.weforum.org/agenda/2023/03/astronomy-radio-interference-satellites-technology/ (abgerufen am 27. Mai 2024).**
81. **Bevorzugte Frequenzbänder für radioastronomische Messungen unter 1 THz. Empfehlung ITU-R RA.314-11, Internationale Fernmeldeunion – Funkkommunikationssektor. 2023. Online verfügbar: https://www.itu.int/dms_pubrec/itu-r/rec/ ra/R-REC-RA.314-11-202312-I!!PDF-E.pdf (abgerufen am 27. Mai 2024).**
82. **Schutzkriterien für radioastronomische Messungen. Empfehlung ITU-R RA.769-2, Internationale Fernmeldeunion – Funksektor. 2003. Online verfügbar: https://www.itu.int/dms_pubrec/itu-r/rec/ra/R-REC-RA. 769-2-200305-I!!PDF-D.pdf (abgerufen am 27. Mai 2024).**
83. **Fridman, PA; Baan, WA RFI-Minderungsmethoden in der Radioastronomie.** Astron. Astrophys. **2001,** 378, **327–344. [CrossRef]**
84. **Zhang, L.; Jin, R.; Zhang, Q.; Wang, R.; Zhang, H.; Wen, Z. Fusionsmethode zur Erkennung, Lokalisierung und Unterdrückung von RFI durch die Kombination ein- und zweidimensionaler SyntheticAperture-Radiometer.** Remote Sens. **2024,** 16, **667. [CrossRef]**
85. **Proakis, JG; Manolakis, DG** Digitale Signalverarbeitung; **Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2007.**
86. **Oppenheim, AV; Schafer, RW; Buck, J.** Diskrete Signalverarbeitung; **Prentice Hall International Editions; Prentice Hall: Upper Saddle River, NJ, USA, 1999.**
87. **Baghaie, R.; Dimitrov, V. DHT-Algorithmus basierend auf der Kodierung algebraischer Ganzzahlen.** Electron. Lett. **1999,** 35, **1303–1305. [CrossRef]**
88. **Britanak, V.; Yip, P.; Rao, KR** Diskrete Cosinus- und Sinustransformationen; **Academic Press: Cambridge, MA, USA, 2007.**
89. **Feig, E.; Winograd, S. Zur multiplikativen Komplexität diskreter Cosinustransformationen.** IEEE Trans. Inf. Theory **1992,** 38, **1387–1391. [Querverweis]**
90. **Malvar, HS; Hallapuro, A.; Karczewicz, M.; Kerofsky, L. Transformation und Quantisierung mit geringer Komplexität in H.264/AVC.** IEEE Trans. Circuits Syst. Video Technol. **2003,** 13, **598–603. [CrossRef]**
91. **Winograd, S.** Arithmetische Komplexität von Berechnungen; **CBMS-NSF Regionale Konferenzreihe in Angewandter Mathematik; SIAM: Philadelphia, PA, USA, 1980; Band 33.**
92. **Winograd, S. On Computing the Discrete Fourier Transform.** Math. Comput. **1978,** 32, **175–199. [CrossRef]**
93. **Heideman, MT** Multiplikative Komplexität, Faltung und die DFT; **Springer: Heidelberg, Deutschland, 1988.**
94. **Oppenheim, AV; Verghese, GC** Signale, Systeme & Inferenz; **Prentice-Hall Signal Processing Series; Pearson: Upper Saddle River, New Jersey, USA, 2016.**
95. **Cintra, RJ Eine ganzzahlige Näherungsmethode für diskrete Sinustransformationen.** Circuits Syst. Signal Process. **2011,** 30, **1481–1501. [Querverweis]**
96. **Suárez Villagrán, DM Diskrete Fourier-Transformationsnäherungen mit Anwendungen in Erkennung und Schätzung. Master Diplomarbeit, Universidade Federal de Pernambuco, Recife, Brasilien, 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 al. Auf dem Weg zu einem Low-SWaP 1024-Strahl-Digital-Array: Ein 32-Strahl-Subsystem bei 5,8 GHz. IEEE Trans. Antenna **2020,** 68, **900–912. [CrossRef]**

98. Madanayake, A.; Cintra, RJ; Akram, N.; Ariyarathna, V.; Mandal, S.; Coutinho, VA; Bayer, FM; Coelho, D.; Rappaport, TS Schnelle Radix-32-Approximative DFTs für digitales HF-Beamforming mit 1024 Strahlen. IEEE Access **2020, 8, 96613–96627. [CrossRef]**

99. Blahut, RE Schnelle Algorithmen für die Signalverarbeitung; **Cambridge University Press: Cambridge, UK, 2010. [CrossRef]**

100. Die Zusammenarbeit für astronomische Signalverarbeitung und Elektronikforschung. 2024. https://casper.berkeley.edu/ (abgerufen am 23. Mai 2024).

101. Der CASPER-Toolflow. GitHub. 2024. https://github.com/casper-astro/mlib_devel/ (abgerufen am 24. Mai 2024).

102. Stutzman, WL; Thiele, GA Antennentheorie und -design; **Wiley: Hoboken, NJ, USA, 2012.**

103. Kulasekera, S.; Madanayake, A.; Suarez, D.; Cintra, RJ; Bayer, FM Mehrstrahl-Empfängeröffnungen unter Verwendung einer multiplikatorlosen 8-Punkt- Approximations-DFT. In Proceedings der 2015 IEEE Radar Conference (RadarCon), Arlington, VA, USA, 10.–15. Mai 2015; S. 1244–1249. [CrossRef]

104. Zhao, H.; Madanayake, A.; Cintra, RJ; Mandal, S. Analoger Current-Mode 8-Punkt Approximate-DFT Multi-Beamformer mit 4,7 Gbps Kanalkapazität. IEEE Access **2023,** 11, **53716–53735. [CrossRef]**

105. Kulasekera, S.; Madanayake, A.; Wijenayake, C.; Bayer, FM; Suarez, D.; Cintra, RJ Digitale Mehrstrahl-Beamformer mit 8 × 8 RF-Apertur unter Verwendung multiplikatorloser 2-D-FFT-Approximationen. In Proceedings der Moratuwa Engineering Research Conference (MERCon) 2015, Moratuwa, Sri Lanka, 7.–8. April 2015; S. 260–264. [CrossRef]

106. Ariyarathna, V.; Madanayake, A.; Tang, X.; Coelho, D.; Cintra, RJ; Belostotski, L.; Mandal, S.; Rappaport, TS Analoge Approximate-FFT 8/16-Beam-Algorithmen, Architekturen und CMOS-Schaltkreise für 5G Beamforming MIMO-Transceiver. IE Emerg. Sel. Top. Circuits Syst. **2018, 8, 466–479. [CrossRef]**

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Artikel

Vollständige Subhedge-Projektion für schrittweise Hedge-Automaten †

Antonio Al Serhali * und Joachim Niehren *

Inria Center, Universität Lille, 59000 Lille, Frankreich *
Korrespondenz: antonio.al-serhali@inria.fr (AAS); joachim.niehren@inria.fr (JN)

† **Dieser Artikel ist eine erweiterte Version unseres Artikels, der im „International Symposium on Fundamentals of Computation Theory“, Trier, Deutschland, 18.–21. September 2023, veröffentlicht wurde.**

Zusammenfassung: Wir zeigen, wie man schrittweise Hedge-Automaten (SHAs) mit Subhedge-Projektion auswertet **und dabei irrelevante Subhedges vollständig projiziert. Da dies die Weitergabe von Informationen zu endlichen Zuständen von oben nach unten erfordert, führen wir den Begriff der abwärts gerichteten schrittweisen Hedge-Automaten ein. Wir verwenden sie, um In-Memory- und Streaming-Evaluatoren mit vollständiger Subhedge-Projektion für SHAs zu definieren. Anschließend optimieren wir die Evaluatoren so, dass sie zum frühestmöglichen Zeitpunkt über die Mitgliedschaft entscheiden können. Wir wenden unsere Algorithmen auf das Problem der Beantwortung regulärer XPath- Abfragen auf XML- Streams an. Unsere Experimente zeigen, dass die vollständige Subhedge-Projektion von SHAs** die Beantwortung frühester Abfragen auf **XML- Streams tatsächlich beschleunigen kann, sodass sie mit den be**

Schlüsselwörter: Automaten; Projektionsalgorithmus; Streaming-Algorithmus; XML

1. Einleitung

Hedges sind Buchstabenfolgen aus dem Alphabet und Bäumen ÿhÿ , **wobei** h **wiederum ein Hedge ist. Hedges sind abstrahiert von den syntaktischen Details von XML- oder JSON-Dokumenten, können diese aber dennoch darstellen. Die Linearisierung eines Hedges ist ein verschachteltes Wort, das in eine Datei geschrieben werden kann. In diesem Artikel untersuchen wir das Problem des Hedge-Musterabgleichs, d. h., ob ein gegebener Hedge** h **zu einer gegebenen regulären Hedge-Sprache L gehört. Reguläre Hedge- Sprachen können insbesondere in XPath oder JSONPat**

Projektion ist für die Effizienz vieler Algorithmen auf Wörtern, Bäumen, Hecken oder verschachtelten Wörtern notwendig. Intuitiv projiziert ein Algorithmus, wenn er nur ein Fragment der Eingabestruktur besucht: im besten Fall den Teil, der für das betrachtete Problem relevant ist. Die Bedeutung der Projektion für die XML- Verarbeitung wurde bereits von [1–4] erkannt. Saxons In-Memory-Evaluator beispielsweise projiziert XML- Eingabedokumente relativ zu einem XSLT- Programm, das eine Sammlung von XPath- Anfragen enthält, die gleichzeitig beantwortet werden sollen [5]. Das Tool QuiXPath [4] evaluiert XPath- Anfragen im Streaming-Modus mit Teilbaum- und Nachkommenprojektion . Die Projektion während der Auswertung von JSONPath- Anfragen auf JSON- Dokumente im Streaming-Modus wird als Fast-Forwarding bezeichnet [6].

Bei der Projektion der In-Memory-Auswertung wird davon ausgegangen, dass der vollständige Graph des Eingabe-Hedges im Voraus erstellt wurde. Dennoch kann die Projektion Zeit sparen, wenn mehrere Muster auf demselben Eingabe-Hedge abgeglichen werden müssen oder wenn der Graph ohnehin aus unterschiedlichen Gründen erstellt wurde. Im Streaming-Modus ist die Situation ähnlich: Der gesamte Eingabe-Hedge auf dem Stream muss analysiert werden, aber die Auswerter müssen nur Knoten prüfen, die nicht wegprojiziert werden. Da die reine Analyse um zwei oder drei Größenordnungen schneller ist als die Musterauswertung, kann der Zeitgewinn der Projektion beträchtlich s

Ausgangspunkt dieses Artikels ist der Begriff der Subhedge-Irrelevanz von Positionen von Hedge h **in Bezug auf eine Hedge-Sprache** L , **den wir einführen. Dies sind Positionen von** h , bei denen für alle möglichen Fortsetzungen das Einfügen eines Subhedges die Mitgliedschaft **in** L **nicht beeinflusst . Wir liefern einen Algorithmus für Hedge-Muster-Matching mit vollständiger Subhedge-Projektion. Unser Algorithmus entscheidet über Hedge-Muster-Matching, während er die Subhedges, die sich an irrelevanten Subhedge-Positionen befinden, vollständig wegprojiziert. Mit**

check for updates

Zitat: Al Serhali, A.; Niehren, J. Vollständige Subhedge-Projektion für schrittweise Hedge-Automaten. Algorithmen **2024,** 17, **339. https://doi.org/10.3390/a17080339**

Wissenschaftlicher Redakteur: Henning Fernau und Klaus Jansen

Erhalten: 29. Januar 2024
Überarbeitet: 23. Juli 2024
Akzeptiert: 23. Juli 2024
Veröffentlicht: 2. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/4.0/).

Machine Translated by Google

ob ein Hedge h **mit einem Muster** L übereinstimmt , **ohne ein Sub-Hedge von** h zu besuchen , **das sich an einer Position befindet , die in Bezug auf L** irrelevant ist.

Reguläre Hedgesprachen können durch verschachtelte reguläre Ausdrücke dargestellt werden, die aus regulären XPath- Anfragen abgeleitet werden können **[7] , oder durch eine Art Hedgeautomaten, der aus verschachtelten regulären Ausdrücken kompiliert werden kann. Wir verwenden schrittweise Hedgeautomaten (Stepwise Hedge Automata , SHAs) [8], eine neuere Variante der Standard-Hedgeautomaten, die wiederum aus den 1960 SHAs vermischen die Bottom-Up-Verarbeitung von Standard-Baumautomaten mit der Links-nach-Rechts-Verarbeitung von Finite-Wort-Automaten (DFAs). Sie unterstützen weder Top-Down-Verarbeitung noch haben sie einen expliziten Stack, im Gegensatz zu verschachtelten Wortautomaten (NWAs) [11–15]. SHAs haben ein gutes Konzept von Bottom-Up- und Links-nach-Rechts-Determinismus und vermeiden so das problematische Konzept von Determinismus für Standard-Hedge-Automaten und das zu kostspielige Konzept von Determinismus für NWAs. Darüber hinaus sind Compiler von verschachtelten regulären Ausdrücken zu SHAs verfügbar [8], und Determinisierungsalgorithmen [16] erzeugen kleine SHAs für alle regulären XPath- Abfragen in den bestehenden praktischen Benchmarks [7,17].**

Die Motivation für die vorliegende Arbeit besteht darin, einem aktuellen Streaming- Algorithmus für deterministische SHAs , der reguläre XPath- Abfragen so früh wie möglich auswerten kann, eine Subhedge-Projektion hinzuzufügen. Die meisten alternativen Streaming-Ansätze vermeiden die Beantwortung von Anfragen, indem sie Subsprachen regulärer Abfragen ohne Verzögerung berücksichtigen, sodass die Knotenauswahl nur von der Vergangenheit des Knotens und nicht von der Zukunft abhängt **[18,19], oder indem sie e Insbesondere wurde kürzlich gezeigt, dass die früheste Abfragebeantwortung für reguläre monadische Abfragen, die durch deterministische SHAs [21]** definiert sind , **eine geringere Worst-Case-Komplexität aufweist als für deterministische NWAs [22]. Damit wurde die früheste Abfragebeantwortung für reguläre Abfragen zum ersten Mal in der Praxis durchführbar. Andererseits ist sie experimentell immer noch langsamer als die besten existierenden nicht-frühesten Ansätze für allgemeine reguläre monadische Abfragen auf XML- Streams (siehe [23] für einen Vergleich), da letztere Projektion unterstützen.**
Wie Projektionen für **SHAs erstellt werden könnten, wurde vor der Konferenzversion des vorliegenden Artikels** nicht untersucht **[24].**

Betrachten wir das Beispiel des XPath- Filters [self::list/child::item] , der von **einem XML- Dokument erfüllt wird ,** wenn dessen Wurzel ein **Listenelement** ist, das ein untergeordnetes **Element hat . Mit der Kodierung des XML - Dokuments, wie sie in diesem Artikel als Hedges gewählt wird, haben erfüllende Hedges die Form** ÿlist **· h1 ·** ÿitem **· h2ÿ · h3ÿ · h4 und für einige Hedges h1, h2, h3, h4. Wenn dieser Filter auf einem Hedge ausgewertet wird, ist es ausreichend, den ersten Teilbaum auf das Vorhandensein der Wurzelbezeichnung** list **und dann alle seine untergeordneten Elemente zu prüfen, bis eines mit einer Wurzelbezeichnung** item **gefunden wird. Die Unterhedges dieser untergeordneten Elemente, d. h. der Hedge h3 und die Unterhedges der Bäume der obersten Ebene in h1 und h3, können wegprojiziert werden, ebenso wie der Hedge h4: Sie müssen nicht geprüft werden, um diesen Filter auszuwerten. Man muss sich Dieser Informationsstand kann natürlich von oben nach unten aktualisiert werden. Die Evaluatoren von SHAs arbeiten jedoch ausschließlich von unten nach oben und von links nach rechts. Daher müssen Projektionsevaluatoren für SHAs auf allgemeineren Maschinen basieren. Wir schlagen** Downward Stepwise Hedge Automata **(SHAÿ s) vor, eine Variante von SHAs , die Top-down-Verarbeitung unterstützt. Sie sind im Wesentlichen Neumanns und Seidls Pushdown-Forest-Automaten [25], außer dass sie auf unbeschriftete Hedges statt auf beschriftete Wälder angewendet werden. Darüber hinaus ist bekannt, dass NWAs auf verschachtelte Wörter im Wesentlichen auf die gleiche Weise arbeiten [26] , während sie etwas allgemeinere sichtbare Pushdowns ermöglichen. Wir unterscheiden dann Subhedge-Projektionszustände für SHAÿ s und** zeigen, wie man sie verwendet, um **SHAs mit Subhedge-Projektion sowohl im Speicher als auch im Streaming-Modus auszuwerten .**

Als ersten Beitrag präsentieren wir den Safe-No-Change-Compiler von SHAs zu SHAÿ s , der geeignete Subhedge-Projektionszustände unterscheiden kann. Die Idee ist, dass der Safe-No- Change -SHAÿ Kontexte unterscheiden kann, in denen sich die Zustände des SHA sicher nicht ändern. Beispielsweise kann der XPath- Filter [self::list/child::item] durch ein deterministisches **SHA- Operating von unten nach oben definiert werden ,** das unser Compiler auf das **SHAÿ -Operating von oben nach unten abbildet. Die Kontextinformationen, die durch die Top-down-Verarbeitung explizit gemacht werden, beziehen sich auf die Ebenen der Dies erlaubt es uns, Projektionszustände ab Level 2 zu unterscheiden, in denen Subhedges ignoriert werden können. Wir beweisen die Solidität unseres Compilers basierend auf einer nichttrivialen Invariante, die**

Machine Translated by Google

Wir stellen fest (wir weisen darauf hin, dass der Beweis eine Anpassung des ursprünglichen Compilers von FCT [24] erforderte). Wir präsentieren auch ein Gegenbeispiel, das zeigt, dass die Safe-No-Change-Projektion für die Subhedge-Projektion nicht vollständig ist. Es zeigt, dass ein Subhedge irrelevant sein kann, obwohl sich sein Zustand noch ändern kann.

Als zweiten Beitrag schlagen wir den Kongruenzprojektionsalgorithmus vor. Er kompiliert wieder **SHAs zu SHAÿ s, basiert aber auf den Kongruenzbeziehungen von Automatenzuständen. Wir beweisen dann, dass die Kongruenzprojektion nicht nur einen stichhaltigen Algorithmus ergibt, sondern dass dieser Algorithmus auch für die Subhedge-Projektion vollständig ist, d. h. alle stark irrelevanten Subhedges (siehe Definition 9) werden in einen Schleifenzustand ausgewertet. Die Kongruenzprojektion beginnt auf der obersten Ebene der Hedges mit der aus der Automatenminimierung bekannten Myhill–Nerode-Kongruenz .**

Für Sprachen mit Wörtern L identifiziert diese Übereinstimmung Präfixe v , **die dasselbe** (L) = {w **|** v **Restsprache** v ^{ÿ1} · w **ÿ L} haben. Ein Präfix** v **ist dann ein irrelevantes Suffix, wenn sein** (L) **Restsprache** v ^{ÿ1} **auf entweder universell oder leer ist. Im Fall von Hedges, die Wörtern durch Hinzufügen hierarchischer Verschachtelung über einen Baumkonstruktor, die Übereinstimmung muss beim Herabsteigen in Teilbäume angepasst werden.**

Wir zeigen, dass beide Compiler die Größe der Automaten exponentiell vergrößern können, da sie in der Lage sein müssen, Bottom-up-deterministische Automaten auf monadischen Bäumen in Top-down-deterministische Automaten umzuwandeln. Das ist dasselbe wie die Invertierung deterministischer Automaten auf Wörtern, was im schlimmsten Fall bekanntlich zu einer exponentiellen Vergrößerung führt (z. B. für die Sprachfamilie (a + b)· ein (a + b) **ÿ mit (a ÿ b) hat der minimale deterministische Links-nach-links- Automat** 2n+1 **Zustände, während der minimale deterministische Rechts-nach-links-Automat** n + **1 Zustände hat).**

Die exponentielle Explosion kann vermieden werden, wenn man an der Auswertung von Hedges durch Automaten mit Subhedge-Projektion interessiert ist, indem man die kompletten SHAÿ s nicht statisch kons Stattdessen kann bei der Bewertung **einer Absicherung** möglicherweise **nur der erforderliche Teil der SHAÿ s spontan erstellt werden .**

Unser dritter Beitrag ist die Verfeinerung der beiden vorherigen Compiler, so dass sie auch sichere Zustände für die Auswahl an der frühesten Stelle erkennen können. Dazu kombinieren wir diese Compiler mit einer Variante des frühesten Mitgliedschaftstesters von [21] , der im Speicher arbeitet, indem er **SHAÿ s statt** NWAs **kompiliert . Darüber hinaus wird auch ein Mitgliedschaftsfehler an der frühesten Stelle erkannt. Auf diese Weise erhalten wir den frühesten speicherinternen Mitgliedschaftstester für deterministische SHAs.**

Unser vierter Beitrag besteht darin, zu zeigen, wie die bisherigen In-Memory-Evaluatoren im Streaming-Modus ausgeführt werden können, während die Linearisierung des Hedges in ein verschachteltes Wort als Eingabe gelesen wird. Dadurch können wir den jüngsten frühesten Streaming-Mitgliedschaftstester hinter dem AStream-Tool [21] verbessern , **indem wir ihm eine Subhedge-Projektion hinzufüge**

Unser fünfter und letzter Beitrag ist eine Implementierung und experimentelle Evaluierung des Streaming-Algorithmus zur Beantwortung frühester Abfragen für dSHAs durch Einführung einer Subhedge- Projektion in das AStream-Tool [21]. Wir haben sowohl eine Safe-No-Change- als auch eine Kongruenzprojektion implementiert . Dazu haben wir unseren frühesten Mitgliedschaftstester mit Subhedge-Projektion auf einen frühesten Query-Beantwortungsalgorithmus für monadische Abfragen gehoben. Für das Experiment begannen wir mit deterministischen SHAs , die mit dem Compiler aus **[8] für die vorwärts- regulären XPath- Abfragen des XPathMark-Benchmarl Es stellt sich heraus, dass die Kongruenzprojektion bei mindestens der Hälfte der Benchmark-Abfragen viel stärker projiziert als die Safe-No-Change-Projektion. Sie reduziert die Laufzeit für alle regulären XPath- Abfragen erheblich, da große Teile der Eingabe-Hedges wegprojiziert werden können . In unserem Benchmark reicht der projizierte Prozentsatz von 75,7 % bis 100 % des Eingabestroms. Bei XPath-Abfragen, die nur untergeordnete Achsen enthalten, wird der früheste Abfragebeantwortungsalgorithmus von AStream mit Kongruenzprojektion in Bezug auf die Effizienz mit dem besten vorhandenen Streaming-Tool namens QuiXPath [23] konkurrenzfähig, das bei einigen Abfragen nicht immer am frühesten ist . Unsere aktuelle Implementierung der frühesten Kongruenzprojektion in AStream ist bei den Benchmark-Abfragen um einen Faktor von 1,3 bis 2,9 langsamer als QuiXPath. Die Verbesserung** ist bei **XPath-Abfragen mit der untergeordneten Achse geringer ,** bei denen weniger Subhedge-Projektion erforderlich ist.

Machine Translated by Google

möglich. Stattdessen wäre eine Art Nachkommenprojektion erforderlich. Trotzdem ist selbst im schlimmsten Fall in unserem Benchmark die früheste Kongruenzprojektion mit AStream derzeit nur um den Faktor 13,8 langsamer als QuiXPath.

Gliederung. Wir beginnen mit vorbereitenden Bemerkungen zu Hedges und verschachtelten Wörtern in Abschnitt 2. In Abschnitt 3 erinnern wir uns an das Problem der Hedge-Musterübereinstimmung und definieren formal irrelevante Sub-H In Abschnitt 4 erinnern wir uns an die Definition von SHAs, führen SHAÿ s ein und diskutieren ihre In-Memory-Evaluatoren. In Abschnitt 5 führen wir den Begriff der Subhedge-Projektionszustände ein, definieren, wann ein SHAÿ für eine Subhedge-Projektion vollständig ist, und präsentieren einen In-Memory-Evaluator für SHAÿ s mit Subhedge-Projektion. In Abschnitt 6 führen wir die Safe-No-Change-Projektion ein, erklären ihre Solidität und zeigen ihre Unvollständigkeit. In Abschnitt 7 führen wir die Kongruenzprojektion ein und beweisen, dass sie für eine Subhedge- Projektion solide und vollständig ist. Die frühesten In-Memory-Evaluatoren für SHAÿ s mit Subhedge-Projektion folgen in Abschnitt 8. Streaming-Varianten unserer In-Memory-Evaluatoren werden systematisch in Abschnitt 9 abgeleitet. Abschnitt 10 diskutiert, wie wir unsere Algorithmen von Mitgliedschaftstests auf monadische Abfragen übertragen können. Abschnitt 11 diskutiert unsere praktischen Experimente. Abschnitt 12 diskutiert weitere verwandte Arbeiten zu Automatenkonzepten im Zusammenhang mit SHAs und SHAÿ s. Anhang A präsentiert die Neuformulierung des frühesten Abfrage-Antwort-Algorithmus aus [21] basierend auf SHAÿ s und mit Schemabeschränkungen. Anhang B liefert einen sti

Kommentare zur Veröffentlichung. Diese Originalversion des Zeitschriftenartikels wurde auf der FCT-Konferenz [24] veröffentlicht. Im Vergleich zur vorherigen Veröffentlichung wurden die folgenden Be Neu sind:

- **Die Definition von subhedge irrelevanten Präfixen von verschachtelten Wörtern und die Definition von Vollständigkeit für die Subhedge-Projektion (Abschnitt 3).**
- **Die Hinzufügung von Schemabeschränkungen im gesamten Dokument und der Begriff des Schemas Vollständigkeit (Abschnitt 4.3).**
- **Der Begriff der Vollständigkeit für die Subhedge-Projektion (Abschnitt 5.2). • Der Beweis der Solidität des Safe-No-Change-Projektionsalgorithmus in Abschnitt 6.2 wurde verschoben zu Anhang B.**
- **Der Kongruenzprojektionsalgorithmus und der Beweis seiner Solidität und Vollständigkeit für die Subhedge-Projektion (Abschnitt 7). • Eine systematische Methode, um eine Subhedge-Projektion zu einem frühesten SHAÿ** hinzuzufügen **(Abschnitt 8). • Eine systematische Methode, um die Korrektheit von Streaming-Automaten-Evaluatoren auf die entsprechenden In-Memory-Evaluatoren zu reduzieren (Abschnitt 9). • Eine Diskussion darüber, wie man mit monadischen Abfragen umgeht und dabei die Verfügbarkeit Schemabeschränkungen (Abschnitt 10).**
- **Eine Implementierung der Kongruenzprojektion mit experimentellen Ergebnissen für die XPath-Auswertung von XML- Streams (Abschnitt 11).**
- **Eine längere Diskussion verwandter Arbeiten (Abschnitt 12).**
- **In Anhang** A zeigen wir auch, wie man die frühesten **dSHAÿ s aus dSHAs erhält, indem man den Compiler aus [21]** anpasst, **der dSHAs auf die frühesten** dNWAs **abbildet .**

2. Vorbemerkungen

__**Seien** A **und** B **Mengen. Wenn** A **eine Teilmenge von B ist, dann wird das Komplement von** A **in** B durch A = B **** A **bezeichnet, wobei die Menge, relativ zu der das Komplement genommen wird, implizit erhalten bleib Der** Definitionsbereich **der binären Relation** r **ÿ** A **×** B **ist** dom(r) = {a **ÿ** A **|** ÿb **ÿ B.** (a, b) **ÿ r}. Eine** partielle Funktion f : A **ÿÿ** B **ist eine binäre Relation** f **ÿ** A **×** B , **die funktional ist. Eine** totale Funktion f : A **ÿ** B **ist eine partielle Funktion** f : A **ÿÿ** B **mit** dom(f) = **A.**

Wörter. Sei N **die Menge der natürlichen Zahlen, einschließlich 0. Sei das Alphabet ÿ eine Menge. Die Menge der Wörter über ÿ ist ÿ ÿ Ein Wort ist (a1, . . . ,** an) **ÿ ÿ , wobei** n **ÿ** N **geschrieben wird als a1 . . .** an. **Das leere Wort der Länge 0 bezeichnen wir mit** ÿ **ÿ ÿ , und durch v1 · v2 ÿ ÿ Die Verkettung zweier Wörter v1, v2 ÿ ÿ .**

Machine Translated by Google

Wenn v = u · v ÿ · w **ist ein Wort, dann nennen wir** u **ein Präfix von** v **und** ÿ **ein Faktor von** v **und** w **ein von v. Gegeben sei eine beliebige** ÿ **, Suffix, so bezeichnen wir die Menge der Präfixe von Wörtern in** L **mit** prefs(L) **und Teilmenge** L **ÿ ÿ, die Menge der Suffixe der Wörter in** L **durch** suffs(L).

Für jedes Unteralphabet ÿ **ÿ ÿ, die Projektion eines Wortes** v **ÿ ÿ)** ÿ **bis ÿ** ÿ **ist das Wort projÿÿ(v) ÿ in** ÿ **ÿ ,** das man aus v **erhält , indem man alle Buchstaben aus ÿ \ ÿ** entfernt ÿ **·**

(ÿ Hedges. Hedges sind Buchstabenfolgen und Bäume ÿhÿ **mit einem Hedge h. Formaler ausgedrückt hat ein Hedge** h **ÿ Hÿ die folgende abstrakte Syntax:**

h, h ÿ **ÿ Hÿ ::=** *ÿ* **|** ein **|** ÿhÿ **|** h · h **wobei** a **ÿ ÿ**

Wir nehmen an, dass *ÿ* · h = h · *ÿ* = h **und** (h · **h1) · h2** = h · (h1 · h2). **Daher betrachten wir jedes Wort ÿ** aab **in ÿ** ÿ **als Absicherung in Hÿ, d. h.** ÿ **ÿ** = a · a · b **ÿ Hÿ. Jeder Hedge kann als Graph konvertiert oder gespeichert werden. Für die Signatur ÿ** = {Liste, Element, **A, . . . , Z}** ist **der Graph eines Beispiel- Hedges in Hÿ in Abbildung 1 dargestellt . Er kodiert die Sequenz von XML- Dokumenten in Abbildung 2:**

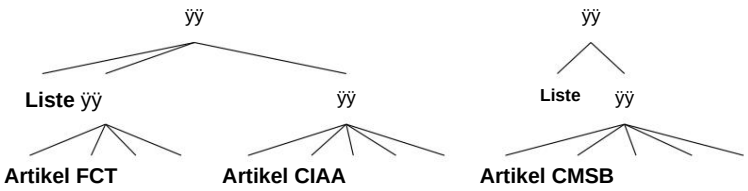


Abbildung 1. Der Graph der Hedge ÿListe · ÿElement · F · C · Tÿ · ÿElement · C · I · A · Aÿÿ · ÿListe · ÿElement · C · M · S · Bÿÿ **ist eine Folge von Bäumen.**

```
<Liste>
   <Element>FCT</
   Element> <Element>ClAA</
Element> </Liste>
<Liste>
   <item>CMSB</item> </
list>
```

Abbildung 2. Die **vom Hedge in Abbildung 1 codierte Folge von XML-** Dokumenten .

Verschachtelte Wörter. Ein verschachteltes Wort über ÿ ist ein Wort über dem Alphabet ÿ^ = **ÿ ÿ** {ÿ, ÿ} , **in dem alle Klammern gut verschachtelt sind. Intuitiv bedeutet dies, dass es für jede öffnende Klammer eine passende schließende Klammer gibt und umgekehrt.**

Die Menge der verschachtelten Wörter über ÿ kann als die Teilmenge der Wörter über **ÿ^ definiert werden, die eine Linearisierung des Hedges, wobei die Linearisierungsfunktion** nw : **Hÿ ÿ (ÿ ÿ** {ÿ, ÿ}) **ÿ wie folgt definiert ist :**

nw(ÿ) = **ÿ,** nw(ÿhÿ) = ÿ · nw(h) · **ÿ,** nw(a) = **a,** nw(h · h ÿ) = nw(h) · nw(h ÿ **).**

Die Menge aller verschachtelten Wörter über ÿ ist also **nw(Hÿ). Sei** hdg **die Inverse der injektiven Funktion** nw **beschränkt auf ihr Bild, so dass die Abbildung von verschachtelten Wörtern auf Hedges mit** hdg(nw(h)) = h **für alle** h **ÿ Hÿ erfolgt.**

Präfixe verschachtelter Wörter. Präfixe, Suffixe und Faktoren verschachtelter Wörter müssen selbst keine verschachtelten Wörter sein. Beispielsweise hat die Absicherung h = a · ÿb · cÿ **die Linearisierung** nw(h) = aÿbcÿ. **Sein Präfix** aÿb **ist nicht gut verschachtelt, da es eine hängende öffnende Klammer hat. Sein Suffix** cÿ **ist ebenfalls nicht gut verschachtelt, da es eine hängende schließende Klammer hat.**

Jeder Algorithmus, der einen Hedge h **von oben nach unten und von links nach rechts durchläuft, überprüft alle Präfixe von** nw(h). **Jedes Präfix** v **von** nw(h) **unterscheidet nicht nur die Position des Hedge** h , **sondern gibt auch den Teil des Hedges an, der sich vor dieser Position in der Linearisierung** nw(h) befindet. **Ein Beispiel ist in Abbildung 3 grafisch dargestellt. Dies gilt insbesondere für Streaming-Algorithmen, die das verschachtelte Wort** nw(h) **als Eingabe in einem Stream** erhalten.

Machine Translated by Google

und kann nur einmal von links nach rechts untersucht werden. Dies gilt jedoch gleichermaßen für In-Memory- Algorithmen, die die Eingabeabsicherung h **als hierarchische Struktur erhalten, deren Graph im Speicher gespeichert ist, und die den Graphen dann von oben nach unten und von links nach rechts durchlaufen. Jeder** ÿÿ- **Knoten wird dann zweimal besucht: einmal beim Lesen einer öffnenden Klammer** ÿ und **beim Abstieg in eine Unterabsicherung und ein weiteres Mal beim Aufstieg zur schließenden Klammer ÿ, nachdem die Unterabsicherung verarbeitet wurde.**

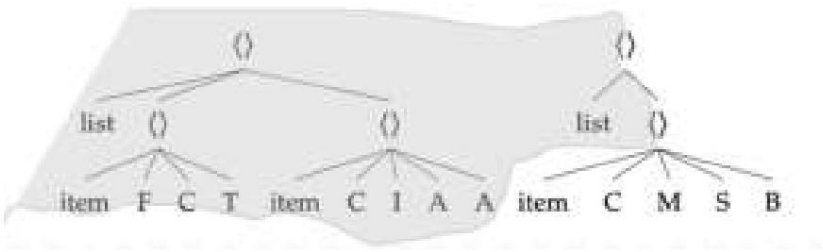


Abbildung 3. Der Teil der Absicherung, der durch das verschachtelte Wortpräfix ÿListe · ÿArtikel · F · C · Tÿ · ÿArtikel · C · I · A · Aÿÿ · ÿListe · **ÿ gekennzeichnet ist.**

3. Problem

Wir definieren formal das Musterabgleichsproblem für Hedges und den Begriff der Sub- Hedge-Irrelevanz. Dies ermöglicht uns, die Teile des Input-Hedges zu charakterisieren, die beim Musterabgleich nicht besucht werden müssen.

3.1. Abgleich von Hedge-Mustern

Wir interessieren uns für Algorithmen, die ein Hedge-Muster mit einem Hedge abgleichen. Wir beginnen mit Algorithmen, die testen, ob eine solche Übereinstimmung vorliegt. Dies wird manchmal als Filterung oder alternativ als boolesche Abfragebeantwortung bezeichnet. In Abschnitt 10 betrachten wir das allgemeinere Problem der Rückgabe des Satzes von Übereinstimmungen. Dieses Problem wird manchmal als monadische Abfragebeantwortung bezeichnet . Unsere Experimente in Abschnitt 11 wenden die monadische Abfragebeantwortung für reguläre XPath-Abfragen an.

Wir beginnen mit Hedge-Mustern, die verschachtelte reguläre Ausdrücke mit Buchstaben in der Signatur ÿ sind. Für diese legen wir eine Menge V **von Rekursionsvariablen fest und nehmen an, dass sie von ÿ disjunkt ist.**

$$\text{ÿ e, e}\quad \textbf{ÿ}\ \text{nRegExpÿ} ::= \ddot{y} \mid \text{ein} \mid z \mid e \cdot e \mid \quad \mid e\,\textbf{ÿ} \mid \text{ÿeÿ} \mid \textbf{µz.e wobei}\ a\ \textbf{ÿ ÿ,}\ z\ \textbf{ÿ}\ V$$

Jeder verschachtelte reguläre Ausdruck e **ÿ** nRegExpÿ **wird von einer Teilmenge von Hedges in** e **ÿ HÿÿV erfüllt. Wir erinnern uns nur informell an die Definition dieser Menge. Der Ausdruck** ÿ wird **nur von dem leeren Wort erfüllt ,** der Ausdruck a **ÿ ÿ nur von der Hedge** a **und der Ausdruck** z **ÿ** V **nur von der einbuchstabigen Hedge z. Ein Ausdruck** e · e **mit dem Verkettungsoperator bzw. e ,e mit Kleenes Sternoperator und** e **ÿ** nRegExpÿ **wird von Wiederholungen erfüllt. Ein Ausdruck** e **aus Wörtern, die e**

ÿ ÿ nRegExpÿ **wird erfüllt durch die Verkettung von Hedges, die** e **und** e **erfüllen. Der** Kleene-Stern **ermöglicht horizontale Rekursion auf Hedges. Der** µ- **Ausdruck µz.e, wobei** z **ÿ** V **und** e **ÿ** nRegExpÿ , wird von **Rekursionsvariable** x **erfüllt, in der** z **wiederholt durch eine Hedge ersetzt wird, die** e erfüllt , **bis sie verschwindet. Wir übernehmen die übliche Einschränkung [28] , dass in jedem verschachtelten regulären Ausdruck µz.e alle Vorkommen von** z **in** e **durch einen Baumkonstruktor** ÿ.ÿ geschützt sind. **Auf diese Weise sorgt der µ-Operator für eine korrekte vertikale Rekursion. Beachten Sie auch, dass wir zur Definition von Untersprachen von Hÿ** nur solche **Ausdrücke in** nRegExpÿ **benötigen, in denen alle Vorkommen von Rekursionsvariablen** z **im Gültigkeitsbereich des Binders** µz gebunden sind.

Es ist bekannt, dass verschachtelte reguläre Ausdrücke in nRegExpÿ alle regulären **Untersprachen von Hedges in Hÿ definieren können , d. h., sie haben die gleiche Ausdrucksstärke wie Hedge-Automaten mit dem Alphabet ÿ. Analoge Ergebnisse für reguläre Baumausdrücke und Baumautomaten sind Volksweisheit [29].**

Machine Translated by Google

Beispiel 1 (Verschachtelte reguläre Ausdrücke von regulären XPath- Filtern). Wir können beispielsweise Teilmengen von Hedge-Kodierungen von **XML-** Dokumenten mit zwei Arten von Elementen in **ÿ** = {Liste, Element} durch Ausdrücke in nRegExpÿ definieren . Für diesen Artikel verwenden wir eine Kodierung des **XML-** Dokuments, die Hedges erzeugt, die den folgenden regulären Ausdruck doc **ÿ** nRegExpÿ erfüllen , wobei z **ÿ** V:

$$\text{treeÿ wobei tree doc} =_{\text{def}} =_{\text{def}} \mu z.\ ÿ(\text{Liste} + \text{Element}) \cdot z \text{ ÿ ÿ}$$

In unserer Anwendung sind die Kodierungen etwas umfangreicher, um verschiedene Typen von **XML-** Knoten wie Elemente, Attribute, Kommentare und Texte unterscheiden zu können. Ein XPath- Filter wie [self::list/child::item] kann dann durch den verschachtelten regulären Ausdruck ausgedrückt werden:

$$\text{filter} =_{\text{def}} ÿ\text{Liste} \cdot \text{doc} \cdot ÿ\text{Element} \cdot \text{docÿ} \cdot \text{docÿ} \cdot \text{doc}$$

Allgemeine Compiler von regulären abwärts gerichteten XPath -Filterabfragen bis hin zu verschachtelten regulären Ausdrücken wurden in [8] vorgeschlagen. Eine Implementierung wurde im Rahmen des AStream-Tools [21] entwickelt.

Wir werden Schemata S **ÿ Hÿ** verwenden , **um die Eingaben unseres Mustervergleichsproblems einzuschränken. Da wir hauptsächlich an regulären Schemata interessiert sind, gehen wir nicht von globaler Regelmäßigkeit aus, da die Safe-No-Change-Projektion nicht darauf angewiesen ist. Das Mustervergleichsproblem für verschachtelte reguläre Ausdrücke in Bezug auf ein Schema S ÿ Hÿ erhält die folgenden Eingaben:**

- **ein verschachtelter regulärer Ausdruck** e **ÿ** nRegExpÿ **und ein**
- **Hedge** h **ÿ S.**

Anschließend gibt es den Wahrheitswert des Urteils h **ÿ e zurück. Der Eingabe-Hedge** h **ÿ S kann entweder durch einen Graphen gegeben sein, der im Speicher liegt, oder durch einen Stream, der die Linearisierung** nw(h) **des Eingabe-Hedges enthält und nur einmal von links nach rechts gelesen werden kann.**

3.2. Irrelevante Subhedges **Wir**

definieren das Konzept irrelevanter Vorkommen von Subhedges in Bezug auf ein gegebenes Hedge-Muster. Was das bedeutet, hängt von der Art des Algorithmus ab, den wir für das Mustervergleichen verwenden. Wir werden Algorithmen verwenden, die auf dem Input-Hedge von oben nach unten, von links nach rechts und von unten nach oben arbeiten. Dies gilt insbesondere für Streaming-Algorithmen.

Intuitiv kann der Mustervergleichsalgorithmus, wenn er von oben nach unten oder von links nach rechts auf einen Knoten trifft, dessen nachfolgender Unter-Hedge irrelevant ist, diesen überspringen, ohne dessen Struktur zu prüfen. Was das Überspringen bedeutet, sollte klar sein, wenn der Hedge durch einen im Speicher abgelegten Graphen gegeben ist. Beachten Sie, dass der vollständige Graph im Voraus erstellt werden muss, auch wenn sich einige Teile davon als irrelevant erweisen könnten. Dennoch kann man viel Zeit gewinnen, wenn man irrelevante Teile überspringt, wenn entweder der Graph bereits aus anderen Gründen erstellt wurde oder wenn viele Mustervergleichsprobleme auf demselben Hedge gelöst werden müssen.

Im Fall von Streams muss das irrelevante Subhedge zwar noch analysiert werden, es wird aber ansonsten nicht analysiert. In den meisten Fällen wird die mögliche Analyse von Automaten durchgeführt, was zwei Größenordnungen mehr Zeit in Anspruch nehmen kann als für die Analyse erforderlich. Daher kann der Streaming-Algorithmus erheblich beschleunigt werden, wenn keine Analyse durchgeführt werden muss.

Definition 1. Sei **S ÿ Hÿ** ein Schema und L **ÿ S** eine Sprache, die dieses Schema erfüllt. Wir definieren diff **ÿ** prefs(Nÿ):
$\overset{M}{\mathbf{S}}$ als am wenigsten symmetrische Relation auf Präfixen verschachtelter Wörter **u,** $u^{ÿ}$

$$\text{Unterschied}_{\overset{M}{\mathbf{S}}}(\text{du}, \text{du}^{ÿ}) \text{ ÿ ÿw ÿ suffs(Nÿ).}\ u \cdot w \text{ ÿ nw(L) ÿ } u^{ÿ} \cdot w \text{ ÿ nw(S \ L).}$$

Ein verschachteltes Wortpräfix u heißt **subhedge relevant** für L mit Schema **S ,** wenn es verschachtelte $ÿ_{v,\,v}$ **ÿ Nÿ** so dass diff $\overset{M}{\mathbf{S}}(u \cdot \mathbf{v}, u \cdot v)^{ÿ}$ Wörter gibt). Andernfalls heißt das Präfix u **subhedge irrelevant** für L mit Schema S.

Also, $\text{Unterschied}_{\overset{M}{\mathbf{S}}}(\text{du}, \text{du}^{ÿ})$ **besagt, dass** u **und** $u^{ÿ}$ **haben Fortsetzungen, die sich in Bezug auf** L **und S unterschiedlich verhalten. Darüber hinaus ist ein Präfix** u **für die Unterhedge irrelevant, wenn die Sprachmitgliedschaft nicht von der Hedge-Einfügung bei** u abhängt , **unter der Bedingung, dass die Schemar**

Machine Translated by Google

garantiert. Der Fall ohne Schemabeschränkungen wird durch die obige Definition subsumiert, indem S = Hÿ **gewählt wird . In diesem Fall ist** das Komplement- Diff die Myhill-Nerode-Kongruenz **Hÿ von L, die in der formalen Sprachtheorie aus der Minimierung von DFAs bekannt ist (siehe jedes Standardlehrbuch zur formalen Sprachtheorie oder Wikipedia (Myhill-Nerode-Theorem: https://en.wikipedia.org/wiki/Myhill-Nerode_theorem, abgerufen am 23. Juli 2024)).**

Diese Übereinstimmung dient als Grundlage für das Gold-artige Lernen regulärer Sprachen aus positiven und negativen Beispielen (siehe [30] oder Wikipedia (Gold-artiges Lernen im Grenzfall https://en.wikipedia.org/wiki/Sprachidentifikation_im_Grenzwert, abgerufen am 23. Juli 2024)).

Für unsere Anwendung auf reguläre XPath-Muster sind Schemabeschränkungen erforderlich.

Beispiel 2 (Reguläre XPath-Filter). Betrachten Sie das Schema **S** = doc für Hedges, **die XML-** Dokumente mit der Signatur **ÿ** = {list, item} darstellen, und die reguläre Hedge-Sprache L = filter, wobei

Filter = ÿListe · Baumÿ · ÿElement · Baumÿ ÿ · Baumÿ ÿ · Baumÿ

ist der verschachtelte reguläre Ausdruck aus Beispiel 1 für den XPath- Filter [self::list/child::item]. Denken Sie daran, dass dieser verschachtelte reguläre Ausdruck nur auf Hedges angewendet werden kann, **die XML-** Dokumente darstellen. Das verschachtelte Wortpräfix u = ÿlist · ÿitem ist für die Sprache L mit dem Schema S als Unterhedge irrelevant. Beachten Sie, dass seine Fortsetzung ÿlist · ÿitemÿÿ mit dem Suffix w = ÿÿ zu L gehört . Daher gilt für jedes h **ÿ Hÿ:** Wenn die Fortsetzung ÿlist · ÿitem · hÿÿ zu **S** = doc gehört, dann gehört sie auch zu L. Dennoch gehört die Fortsetzung ÿlist · ÿitem · h1ÿÿ für den Hedge **h1** = ÿÿ nicht zu L, da sie das Schema **S** = doc nicht erfüllt .

Auch das Präfix ÿitem ist für die Sprache L irrelevant , sogar unabhängig vom Schema. Dies liegt daran, dass L keine Fortsetzung dieses Präfixes zu irgendeinem Hedge enthält. Das Präfix ÿlist ist für die Sprache L bzgl. S nicht irrelevant . Dies kann man am Suffix w = ÿ sehen, da ÿlistÿ nicht zu L gehört , während die Fortsetzung ÿlist · hÿ mit h = ÿitemÿ zu L gehört und beide Fortsetzungen das Schema **S** = doc erfüllen.

Schemabeschränkungen können jedoch auch überraschende Konsequenzen haben, die das Problem noch komplizierter machen.

Beispiel 3 (Überraschende Konsequenzen von Schemabeschränkungen). Betrachten Sie die Signatur **ÿ** = **{a},** das Muster L = {ÿaÿ} und das Schema **S** = L **ÿ** {ÿÿ · **a}.** Das Präfix ÿ ist tatsächlich für L und S subhedge-irrelevant . Um dies zu sehen, betrachten wir alle möglichen schließenden Suffixe einzeln. Das kleinste schließende Präfix ist ÿ. Beachten Sie, dass ein Hedge ÿhÿ **ÿ** L genau dann ist, wenn h = a. Die Mitgliedschaft zu L scheint also vom Subhedge h abzuhängen. Beachten Sie jedoch auch, dass ÿhÿ **ÿ S** genau dann ist, wenn h = a. Wenn also angenommen wird, dass ÿhÿ zum Schema S gehört, muss es auch zu L gehören . Das Muster muss also übereinstimmen, wenn eine Schemamitgliedschaft angenommen wird. Das nächste schließende Suffix ist ÿ · a. Beachten Sie, dass ein Hedge ÿhÿ · a **ÿ S** genau dann ist, wenn h = ÿ. Allerdings ist ÿhÿ · a **ÿÿ** L, sodass das Muster für keinen Subhedge h mit diesem Präfix übereinstimmt, wenn angenommen wird, dass der Eingabe-Hedge das Schema erfüllt. Die Situation ist bei größeren Suffixen ähnlich, sodass die Sprachzugehörigkeit in keinem Fall vom Subhedge am Präfix ÿ abhängt, wenn angenommen wird , dass der vollständige Hedge das Schema S erfüllt.

3.3. Grundlegende

Eigenschaften **Wir zeigen zunächst, dass für Subhedges irrelevante Präfixe auch dann irrelevant bleiben, wenn sie durch das verschachtelte Wort eines beliebigen Hedges erweitert werden. Diese Eigenschaft sollte für jede vernünftige Vorstellung von Subhedges-Irrelevanz zu erwarten sein.**

Lemma 1. Für jedes verschachtelte Wortpräfix u, Sprache **L, S ÿ Hÿ** und verschachteltes Wort v **ÿ Nÿ gilt:** Wenn das Präfix u für L mit Schema S unterheckenirrelevant ist , dann ist das Präfix u · v ebenfalls irrelevant.

Beweis. Sei u ein für L **irrelevantes Subhedge mit Schema S und** v **ÿ Nÿ ein verschachteltes Wort. Wir ÿ Nÿ und ein Wörter** v ÿ , v " **verschachteltes Wortsuffix** w **ÿ** suffs(Nÿ) , **so dass beliebige verschachtelte**

Machine Translated by Google

ÿ u · v · v · w ÿ nw(S) **und** u · v · v · w ÿ **nw(S). Da** v · v ist, ergibt sich ÿ **und** v · v " **sind auch verschachtelte Wörter,** aus **der Subhedge-Irrelevanz von** u **folgendes:**

ÿ u · v · v · w ÿ nw(L) ÿ u · v · v " · w ÿ nw(L).

Daher ist das verschachtelte Wortpräfix u · v für L **mit Schema S unterhemmungsrelevant .** □

Definition 2. Wir nennen eine binäre Relation D auf Präfixen verschachtelter Wörter eine **Differenzrelation auf Präfixen ,** wenn für alle Präfixe verschachtelter Wörter **u,ÿ**prefs(Nÿ) und verschachtelte Wörter v **ÿ Nÿ:**

(u · **v,** u ÿ · v) ÿ D ÿ (u, u ÿ) ÿ **D.**

Lemma 2. Unterschied $\overline{L}^{M}_{S}$ ist eine Differenzrelation.

Beweis. Wir müssen zeigen, dass für alle Präfixe u, ÿ ÿ prefs(Nÿ) **und verschachtelte Wörter** v **ÿ Nÿ , die** u (u · **v,** u ÿ · v) **ÿ** diff **impliziert** (u, u ÿ) ÿ diff **S .** $\overline{M}$ **Seien also u,** u ÿ **Präfixe verschachtelter Wörter sind und** v **ein verschachteltes Wort, so dass** (u · **v,** ÿ uv) ÿ Differenz $\overline{S}^{M}$ **. Dann gibt es ein verschachteltes Suffix** w , **so dass**

u · v · w ÿ nw(L) ÿ u ÿ · v · w ÿ nw(S \ L).

Das Suffix w˜ = v · w **erfüllt dann wie gefordert** u · **w˜** ÿ nw(L) ÿ ÿ **· w˜ ÿ** nw(S \ L). **Daher** (u, u ÿ) ÿ Differenz $\overline{S}^{M}$ u . □

Im schemafreien Fall von Wörtern ist das Komplement diff $\overline{\ÿÿ}^{M}$ **ist eine Äquivalenzrelation, die als Myhill-Nerode-Kongruenz bekannt ist. Im Fall von Schemata ist das Komplement** diff **S** jedoch **möglicherweise nicht transitiv; es handelt sich also möglicherweise nicht einmal um eine Äquivale Dies mag auf den ersten Blick überraschend sein.**

Beispiel 4. Seien **ÿ** = {a, **b},** L = *{ÿ,* **b,** aa} und **S** = L **ÿ** {aab}. Dann gibt es *(ÿ,* a) **ÿ** diff keine Fortsetzung, die $\overline{L_S}$ seit sowohl *ÿ* als auch a in das Schema hinein erweitert . Aus demselben Grund haben wir (a, aa) **ÿ** diff das Komplement diff $\overline{S}^{M}$ **.** Allerdings diff $\overline{L_S}^{M}$ *(ÿ,* aa) , da *ÿ* · b **ÿ** L und aa · b **ÿ S ** L. Daher ist *(ÿ,* aa) **ÿÿ** diff $\overline{S}^{M}$ **,** Also ist nicht transitiv.

Es ist für alle Sprachen L **zu beachten, dass** diff **reflexiv und transitiv ist, also** $\overline{Hÿ}^{M}$ **eine Äquivalenzrelation und damit eine Kongruenz. Für allgemeine Schemata S kann** diff **S** nicht **transitiv sein, wie Beispiel 4 zeigt , und damit nicht kongruent.**

Dies deutet darauf hin, dass Schemata die Entscheidung über die Irrelevanz von Unterabsicherungen erschweren können. Tatsächlich funktioniert die natürliche Art und Weise, wie man Schemata, die auf Komplementen basieren, eliminieren könnte , nicht, wie durch das folgende Lemma (insbesondere das Gegenbeispiel in Beweis 1. ÿÿ 2.) bestätigt wird:

Lemma 3. Für jedes verschachtelte Wortpräfix u **ÿ** suffs(Nÿ) und **L, S ÿ Hÿ** betrachten wir die folgenden beiden Eigenschaften:

1. u ist irrelevant für L mit Schema S. **2.** u ist irrelevant für L **ÿ S** mit Schema $\overline{Hÿ}$ **.**

Dann impliziert Eigenschaft 2 Eigenschaft 1, aber nicht immer umgekehrt.

Beweis. 1. ÿÿ 2.: Hier wird ein Gegenbeispiel präsentiert. Seien L = *{ÿ}* **und S** = *{ÿ,* **a}. Dann ist das Präfix** a **für** L **mit Schema** S irrelevant. Für L **ÿ S** = **Hÿ ** {a} **mit Schema Hÿ** ist das Präfix a **jedoch relevant , da für die verschachtelten Wörter** v = *ÿ* **und** v = a **gilt:** a · v **ÿÿ** L **ÿ S und** a · v $\overline{ÿ}$ **ÿ** L **ÿ S.** $\overline{}$

Machine Translated by Google

ÿ 2.: Als nächstes nehmen wir Eigenschaft 2 an. Betrachten wir beliebige verschachtelte ÿ ÿ Nÿ und 1. ein verschachteltes Wortsuffix w **ÿ** suffs(Nÿ) **so dass** u · v · **w,** u · v **um zu ÿ Wörter v,** v · w **ÿ S. Es genügt zeigen, dass** u · v · w **ÿ S und** u · v **ÿ** · w **ÿ S impliziert** u · v · w **ÿ** L **ÿ** u · v **ÿ** · w **ÿ L. Diese Implikation gilt trivial, wenn** u · v · w **ÿ S oder** u · vu · v · w **ÿ ÿ** · w **ÿ S. Andernfalls haben wir** · w **S und** u · v **ÿ ÿ S. Aus Eigenschaft 2 folgt, wie angenommen,** u · v · w **ÿ** L **ÿ** **ÿ** u · v · w **ÿ L. Also gilt Eigenschaft 1.** □

4. Hedge-Automaten Wir

erinnern an das Konzept der schrittweisen Hedge-Automaten (SHAs), führen ihre Abwärtserweiterung (SHAÿ) ein, diskutieren die Schemavollständigkeit für beide Arten von Automaten, definieren Sub- Hedge-Projektionszustände für SHAÿ s und zeigen, wie man sie zur Auswertung von SHAÿ s mit Subhedge-Projektion verwendet. Wir beschränken uns in diesem Abschnitt auf die Auswertung im Speicher, werden aber in Abschnitt 9 die Streaming-Auswertung besprechen. Die Beziehung zu mehreren eng verwandten Konzepten von Automaten für Hedges, Wälder oder verschachtelte Wörter wird in Abschnitt **12 ausführlicher besprochen .**

4.1. Schrittweise Hedge-Automaten

SHAs sind ein neuer Begriff für Automaten für Hedges [8] , die Bottom-Up-Baum- und Links-nach-rechts-Wortautomaten auf natürliche Weise vermischen. Sie erweitern schrittweise Baumautomaten [31] , indem sie auf Hedges statt auf ungeordneten Bäumen operieren, d. h. auf Folgen von Buchstaben und Bäumen, die Hedges enthalten. Sie unterscheiden sich von verschachtelten Wortautomaten (NWAs) [11,15] dadurch, dass sie Hedges direkt verarbeiten und nicht deren L

Definition 3. Ein schrittweiser Hedge-Automat **(SHA)** ist ein Tupel A = **(ÿ, Q, ÿ, I,** F) , wobei **ÿ** und **Q** endliche Mengen sind, **I,** F **ÿ** Q und **ÿ** = ((a **ÿ)aÿÿ,ÿÿÿ,** @ÿ), wobei a **ÿ Q ×** Q, **ÿÿÿ ÿ** Q **ÿ** und **@ÿ ÿ (Q ×** Q) **×** Q. Ein **SHA** ist deterministisch oder äquivalent ein **dSHA ,** wenn I und **ÿÿÿ** höchstens ein Element enthalten und alle Relationen (a ÿ)aÿÿ und **@ÿ** partielle Funktionen sind.

Die Menge der Zustände Q umfasst eine Teilmenge I **der Anfangszustände, eine Teilmenge** F **der eine Teilmenge ÿÿÿ von Baumanfangszuständen. Die Übergangsregeln in ÿ haben drei Formen: wenn ÿ Endzustände und) ÿ** (q, q, **dann haben wir eine Buchstabenregel, die wir als** q **aÿÿ** q **in ÿ. Wenn ÿ, dann haben wir ein in ÿ. Baumanfangsanwendungsregel, die wir als** q **@p ÿ** q **schreiben, wenn** q **ÿ ÿÿÿ ÿ Q, dann haben wir eine ÿÿ Regel, die wir als ÿÿ** q **in ÿ bezeichnen.**

In Abbildung 4 veranschaulichen wir den Graphen eines dSHA für den XPath- Filter [self::list/child::item]. **Der Graph von SHA wird wie üblich für Zustandsautomaten auf Wörtern erstellt und gezeichnet. Die Zustände sind die Knoten des Graphen; im Beispiel** ist **Q** = **{0, 1, 2, 3, 4, 5}. Die Übergangsregeln definieren die Kanten. Die Kante einer Buchstabenregel** q **aÿÿ** q **in ÿ wird durch den Buchstaben** a **ÿ ÿ annotiert. Im Beispiel** ist **ÿ** = {Liste, Element}. **Jede Regel q@p ÿ** q **gilt mit dem ÿ** p **in ÿ wird als Kante** q **ÿÿ** q gezeichnet **ÿ Zustand** p **ÿ Q. Der Anfangszustand 0 ÿ** I **wird durch einen eingehenden grauen Pfeil angezeigt und der Anfangszustand des Baums 0 ÿ ÿÿÿ durch einen eingehenden grauen Pfeil, der mit dem Symbol ÿÿ annotiert ist . Endzustände wie 4 ÿ F werden durch einen Doppelkreis angezeigt.**

Machine Translated by Google

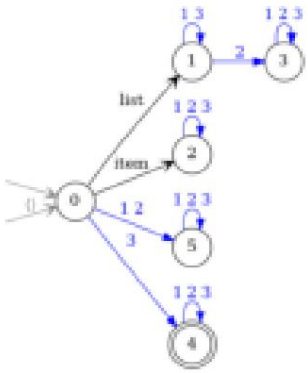


Abbildung 4. Ein dSHA für den XPath- Filter [self::list/child::item], **der schemavollständig für Schema ist** doc.

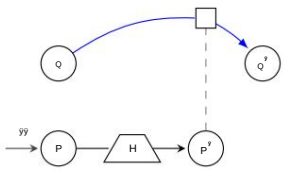
4.1.1. Semantik

Als nächstes definieren wir die Semantik von SHAs, d. h. die Übergänge auf Hedges, die sie definieren, und die Hedge-Sprache, die sie akzeptieren. Für jeden Hedge h **ÿ Hÿ definieren wir die Übergangsschritte ÿ Q, ÿÿ** p **bzgl. ÿ so dass für alle q,** qq **, p,** p a **ÿ ÿ und h,** h **ÿ Hÿ**

$$\frac{q \text{ aÿÿ } \text{ in ÿ}}{qq \text{ aÿÿ } q \text{ bzgl. ÿ}} \qquad \frac{h\ q\ ÿÿ\ q \text{ bzgl. ÿ } q \quad h\ ÿÿ\ q \text{ bzgl. ÿ}}{\text{ bzgl. ÿ}}$$

$$\frac{q \text{ ÿ Q } q}{\text{ÿÿÿ } q \text{ bzgl. ÿ}} \qquad \frac{\text{ÿÿ } h\ ÿÿ\ p \text{ in ÿ } p\ ÿÿ\ p \quad q@p \text{ ÿ } q \text{ in ÿ}}{\text{ÿÿÿ bzgl. ÿ}}$$

Die Übergänge können verwendet werden, um Hedges nichtdeterministisch von unten nach oben und von links nach rechts mithilfe von SHAs auszuwerten. Die ersten drei Regeln geben an, wie Sequenzen von Bäumen und Buchstaben ausgewertet werden, wie etwa ein gewöhnlicher endlicher Automat, wobei angenommen wird, dass die Bäume bereits in Zustände ausgewertet wurden . Die letzte Regel gibt an, wie ein Baum ÿhÿ **von einem Zustand** q **zu einem Z Dies kann wie folgt visualisiert werden:**



Für jeden Anfangszustand p **ÿ ÿÿÿ** eines Baums **muss man die Unterabsicherung** h **nichtdeterministisch auswerten, Für jedes** p **um zu erhalten; dann kehrt man so dass q@p ÿ** q **in ÿ nichtdeterministisch zu einem Zustand** q **irgendeines Zustands** p zurück .

Eine Absicherung wird von A akzeptiert , **wenn sie einen Übergang von einem Anfangszustand zu einem Endzustand erlaubt. Zustand. Die Sprache** L(A) **ist die Menge aller akzeptierten Hedges:**

$$L(A) = \{h \text{ ÿ Hÿ } | \ q \text{ ÿÿ } q \quad \text{bzgl. ÿ, } q \text{ ÿ I, } q \quad \text{ÿ F}\}.$$

4.1.2. Läuft

Läufe können die gesamte Historie einer einzelnen Wahl des nichtdeterministischen Evaluators auf einem Hedge darstellen. Beispielsweise ist der einzige erfolgreiche Lauf des dSHA in Abbildung 4 auf dem Hedge ÿ **Hedge** ÿ Liste · ÿ Artikel ÿ **der Hedge 0 · ÿ0 ·** Liste **· 1 · ÿ0 ·** Artikel · 2ÿ · 3ÿ **· 4 , dessen Berechnung in Abbildung 5 dargestellt ist. Auf der obersten Ebene rechtfertigt der Lauf den Übergang 0 ÿÿ 4 bzgl. ÿ vom Anfangszustand 0 zu einem Endzustand 4. Wenn im Zustand 0 gestartet wird, muss zuerst der Sub-Hedge ausgewertet werden, daher muss der Lauf auf der oberen Ebene ausgesetzt werden, bi**

Machine Translated by Google

ist fertig. Der Subrun auf h ÿ **rechtfertigt schließlich den Übergang 0 ÿÿ** 'H **3 bzgl. ÿ. Zu diesem Zeitpunkt kann der Lauf der oberen Ebene wieder aufgenommen und im Zustand 4 fortgesetzt werden, da 0@3 ÿ 4 in ÿ.**

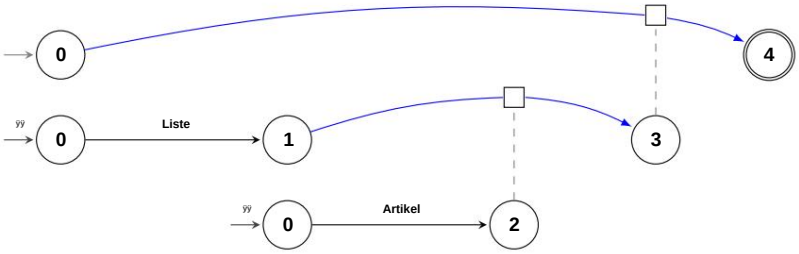


Abbildung 5. Der einzige erfolgreiche Durchlauf des dSHA in Abbildung 4 auf der Absicherung ÿListe · ÿArtikelÿÿ **ist die Absicherung 0 · ÿ0 ·** Liste **· 1 · ÿ0 ·** Artikel · 2ÿ · 3ÿ **· 4, berechnet wie oben dargestellt.**

Beim Zeichnen von Läufen veranschaulichen wir jeden Unterbrechungs-/Wiederaufnahmemechanismus mit einer Box, beispielsweise der Box an der Kante 0 4. Die Box steht für einen zukünftigen Wert, der von der unteren Ebene berechnet wird und die obere Ebene dazu veranlasst, bis zu seinem Eintreffen zu unterbrechen. Schließlich wird die Box durch den Zustand 3 gefüllt, wie durch 3, dargestellt, sodass die Berechnung auf der oberen Ebene fortgesetzt werden kann .

Als nächstes definieren wir Läufe von SHAs auf Hedges formal. Ob ein Hedge mit Buchstaben und Zuständen R **ÿ HÿÿQ ein ÿ-Lauf oder einfach ein Lauf auf einem Hedge** h **ÿ Hÿ ist, geschrieben als** R **ÿ** Laufÿ(h), **wird durch die folgenden Regeln definiert:**

$$\frac{q \text{ aÿÿ } q' \text{ in } ÿ}{q \cdot ein \cdot q' \text{ ÿ Laufÿ(a)}} \qquad \frac{q \cdot R \cdot q' \text{ ÿ Laufÿ(h) } q'' \cdot R' \cdot Q'' \text{ ÿ Laufÿ(h ÿ)}}{q \cdot R \cdot q' \cdot R' \cdot Q'' \text{ ÿ Laufÿ(h} \cdot \text{h } ')}$$

$$\frac{\text{WAHR}}{q \text{ ÿ runÿ(ÿ)} \; q \cdot ÿRÿ \cdot q} \qquad \frac{ÿÿ \; ÿÿ \; p \text{ in ÿ } p \cdot R \cdot p' \text{ ÿ Laufÿ(h) q@p ÿ } ' \text{ ÿ } q' \text{ in ÿ}}{\text{Beachten Sie, dass } ' \text{ Laufÿ(ÿhÿ)}}$$

wenn R **ÿ** runÿ(h), h durch Entfernen aller Zustände aus **R erhalten werden kann , d .** h. projÿ (R) = **h. Jeder Lauf** R **ÿ runÿ(h) kann durch die Abbildung der Präfixe des verschachtelten Wortes** nw(h) **auf Zustände in Q identifiziert werden. Beachten Sie, dass** nw(h) = projÿ (nw(R)). **Für jedes Präfix** v **von** nw(h) **existiert ein eindeutiges Präfix** r **des verschachtelten Wortes** nw(R) , **das mit einem Zustand** q **ÿ Q** endet **und** die Bedingung projÿ (r) = **v erfüllt. Wir nennen dann** q **den Zustand von** R **beim Präfix v. Die Folge 0 · ÿ0 ·** Liste **· 1 · ÿ0 ·** Element **·** 2ÿ **· 4 auf der Hecke** ÿListeÿElementÿÿ **aus Abbildung 5 weist beispielsweise den verschachtelten Wortpräfixen** ÿ **und** ÿ den Zustand 0 **und dem verschachtelten Wortpräfix** ÿ · Liste den Zustan

Ein ÿ-Lauf heißt Lauf von A = **(ÿ, Q, ÿ, I,** F) , **wenn er mit einem Zustand in I beginnt. Ein Lauf von** A ist erfolgreich, wenn es mit einem Zustand in **F endet .**

Lemma 4. h **ÿ** L(A) genau dann, wenn es einen erfolgreichen Lauf R **ÿ runÿ(h)** von A gibt .

Beweis. Unkompliziert mit Induktion über h. □

4.1.3. Determinierung

SHAs können immer mit der aus dem

Determinierung von endlichen Zustandsautomaten auf Wörtern oder von Baumautomaten.

Gegeben sei ein SHA A = **(ÿ, Q, ÿ, I,** F), dann **ist die Zustandsmenge seiner Determinierung** Qdet = **2 Übergangsregeln ÿ** det **Q. Die Zustände der Determinierung sind in Abbildung 6 dargestellt.**

Machine Translated by Google

$$\frac{Q \ddot{y}\ \mathbf{Q}\ a\ \ddot{y}\ \ddot{y}\ Q \qquad \ddot{y} = \{q\ \ddot{y}\ \ddot{y}\ Q \mid \ddot{y}\ \ddot{y}, q\ \ddot{y}\ \mathbf{Q}\}}{Q\ \mathbf{a\ddot{y}\ddot{y}}\ Q\ \ddot{y}\ \ddot{y}\ \ddot{y}^{det}}$$

$$\frac{\ddot{y}\ddot{y}\ddot{y}\ \ddot{y} = \ddot{y} \qquad \mathbf{Q1, Q2}\ \ddot{y}\ \mathbf{Q}\ Q \qquad \ddot{y} = \{q\ \ddot{y}\ \ddot{y}\ \mathbf{Q} \mid \ddot{y}\ \ddot{y}, \mathbf{q1}\ \ddot{y}\ \mathbf{Q1, q2}\ \ddot{y}\ \mathbf{Q2}\}}{\ddot{y}\ddot{y}\ddot{y}\ \ddot{y}\ \ddot{y}^{det} \qquad \mathbf{Q1@Q2}\ \ddot{y}\ \mathbf{Q}\ddot{y}\ \ddot{y}\ \ddot{y}^{det}}$$

Abbildung 6. Die Determinierung ÿ ^{det}**: der Übergangsregeln ÿ eines SHA.**

4.1.4. In-Memory-Mitgliedschaftstests

Das Testen der Mitgliedschaft h **ÿ** L(A) **für einen Hedge h, dessen Graph im Speicher abgelegt ist, kann durchgeführt werden, indem man den eindeutigen Durchlauf der Determinierung von** A **von unten nach oben und von links nach rechts berechnet und testet, ob er erfolgreich ist. Alternativ kann man die eindeutige Menge** P **ÿ Q** berechnen , **sodass** I **ÿÿ** P **in Bezug auf die Determinierung von** A ist , **und testen, ob** P **ÿ** F ÿ= **ÿ. Die Berechnungen sind im Wesentlichen gleich, aber die vollständige Berechnungshistorie geht verloren, wenn nur die Menge** P **der erreichten Zustände und nicht der Durchlauf zurückgegeben wird.**

Zur Determinierung von SHAs können wir auf eine Variante der üblichen Teilmengenkonstruktion zurückgreifen, die von endlichen Automaten auf Wörtern oder Standardbaumautomaten bekannt ist (siehe z. B. [8,16]). Für den Mitgliedschaftstest ist eine Determinierung im laufenden Betrieb ausreichend, die nur den Teil des deterministischen Automaten erze Dieser Teil hat die lineare Größe O(|h|), **während die vollständige Determinierung von** A die exponentielle **Größe** O(2 |A|) **haben kann . Auf diese Weise kann die Mitgliedschaft** h **ÿ** L(A) in der Zeit **O(|h||A|) getestet werden .**

4.1.5. Erreichbarkeit von Hedges

Für jede Menge P **ÿ Q definieren wir die Menge der von** P **aus über einen Hedge erreichbaren Zustände wie folgt:**

$$\text{accÿ (P)} = \{q \quad \ddot{y}\ \text{h}\ \ddot{y}\ \mathbf{Q} \mid \ddot{y}\ q\ \ddot{y}\ \mathbf{P.}\ \ddot{y}\text{h}\ \ddot{y}\ \text{H}\ddot{y}\ . \qquad \ddot{y}\ \mathbf{bzgl.\ \ddot{y}\}.}$$

Wir stellen fest, dass accÿ(P) aus **ÿ und** P **in linearer Zeit in der Größe von** ÿ **berechnet werden kann . Ein Zustand ist von** P **aus genau dann abgesichert erreichbar, wenn er für** P **in der Grafik von ÿ erreichbar ist. Oder durch Berechnung des kleinsten Fixpunkts des folgenden Grund-Datalog-Programms, das von** P **und ÿ abhängt und eine lineare Größe in A hat:**

$$\frac{q\ \ddot{y}}{P\text{- acc(p).}} \qquad \frac{q\ \mathbf{a\ddot{y}\ddot{y}}\ \ddot{y}\ \mathbf{in}\ \ddot{y}}{\text{akk(q}\ \ddot{y}\text{q) :}\ddot{y}\ \text{acc(q).}} \qquad \frac{\mathbf{q@p}\ \ddot{y}\ q \quad \ddot{y}\ \mathbf{in}\ \ddot{y}}{\text{bez.(q}\ \ddot{y}\text{) :}\ddot{y}\ \text{acc(q), acc(p).}}$$

Offensichtlich ist L(A) = **ÿ genau dann, wenn** F **ÿ accÿ(I)** = **ÿ. Daher kann die Leere für SHAs in linearer Zeit entschieden werden. Dies steht im Gegensatz zu der allgemeineren Vorstellung von SHAÿ s, als nächstes vorstellen.**

4.2. Abwärts gerichtete schrittweise Hedge-Automaten

Die Auswertung von SHAs erfolgt von unten nach oben und von links nach rechts, kann aber keine Informationen von oben nach unten weitergeben. Als nächstes schlagen wir eine Erweiterung von SHAs auf SHAÿ s vor, die die Möglichkeit bietet, Informationen mit begrenztem Zustand von oben nach unten weiterzugeben.

SHAÿ s sind im Grunde gleich den Kellerwaldautomaten von Neumann und Seidl [25] , wurden aber von (markierten) Wäldern auf (unmarkierte) Hecken übertragen. **Eine detailliertere Diskussion über die Beziehung von SHAÿ s zu anderen Automatenkonzepten und** insbesondere **zu NWAs** finden Sie **in Abschni groß.**

Definition 4. Ein abwärts gerichteter schrittweiser Hedge-Automat **(SHAÿ)** ist ein Tupel A = **(ÿ, Q, ÿ, I,** F), wobei **ÿ** und **Q** endliche Mengen sind, **I,** F **ÿ** Q und **ÿ** = ((a **ÿ)aÿÿ,ÿÿÿ,** @ÿ). Weiterhin $\ddot{y}$ **ÿ Q ×** Q, gilt a **ÿÿÿ ÿ Q ×** Q und **@ÿ ÿ (Q ×** Q) **×** Q. Ein **SHAÿ** ist deterministisch oder äquivalent ein **dSHAÿ,** wenn I höchstens ein Element enthält und alle Relationen **ÿÿÿ, aÿ** und **@ÿ** partielle Funktionen sind.

Machine Translated by Google

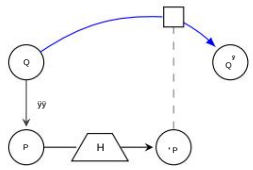
Der einzige Unterschied zu SHAs ist die Form der Baumöffnungsregeln. Wenn (q, q) **ÿ ÿÿ ÿÿÿ ÿ Q, dann haben wir eine Baumanfangsregel, die wir wie folgt bezeichnen:** q **ÿÿ** q , **in ÿ.** Geben **Sie also den** , **wo die Auswertung eines Subhedges beginnt, hängt vom Zustand** q **des übergeordneten Staat** q **an.**

4.2.1. Semantik

Die Definition der Übergangsschritte für SHAÿ s **unterscheidet sich von der für SHAs nur in der Verwendung von Öffnungsregeln in der folgenden Inferenzregel:**

$$\frac{\text{ÿÿ h ÿÿ p in ÿ p ÿÿ pq} \quad \text{bzgl. ÿ q@p} \quad \text{ÿ q} \quad \text{in ÿ}}{\text{ÿÿÿhqÿ ÿ} \quad \text{bzgl. ÿ}}$$

ÿÿ Dies bedeutet, dass die Auswertung des Teil-Hedges h in einem Zustand p **beginnt , sodass** q **ÿÿ** p **in ÿ.**



Der **gewählte Neustartzustand** p kann nun also vom Zustand q darüber abhängen. Auf diese Weise werden Informationen zu endlichen Zuständen von **SHAÿ s von oben nach unten weitergegeben . SHAs arbeiten dagegen rein von unten nach oben und von links nach rechts.**

4.2.2. Läuft

Das Konzept der Läufe kann direkt von **SHAs auf SHAÿ** s **übertragen werden . Im Zustand q genügt es, die Berechnung in Subhedges im Zustand p neu zu starten, so dass**

ÿÿ ÿÿ p **ÿ ÿ. Auf diese Weise werden Informationen über endliche Zustände weitergegeben (während für SHAs ein Baum** q **als Anfangspunkt gewählt werden muss, der unabhängig von q ist). Die einzige Regel der Läufe, die geändert werden muss, ist die folgende:**

$$\frac{\text{ÿÿ ÿÿ p in ÿ p · R · pq} \quad \text{ÿ Laufÿ(h) q@p} \quad \text{ÿ q} \quad \text{in ÿ}}{\text{q · ÿRÿ · q} \quad \text{ÿ Laufÿ(ÿhÿ)}}$$

Ein Beispiel für einen Durchlauf des dSHAÿ in Abbildung 7 ist in Abbildung 8 dargestellt. Hier werden die Informationen auf der Ebene der Knoten weitergegeben. Sie werden durch die Anzahl der Primzahlen dargestellt. Wenn wir beispielsweise den ersten Teilbaum öffnen, der mit item beschriftet ist , **verwenden wir die Übergang dass man vom Zustand 1 auf der ersten Ebene zum Zustand 0 auf der zweiten Ebene gelangt.**

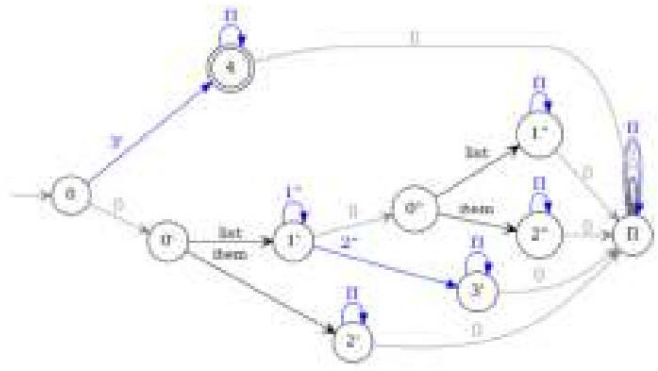


Abbildung 7. Der deterministische SHAÿ mit Subhedge-Projektionszustand ÿ , der durch eine Safe-No-Change-Projektion erreicht wurde.

Machine Translated by Google

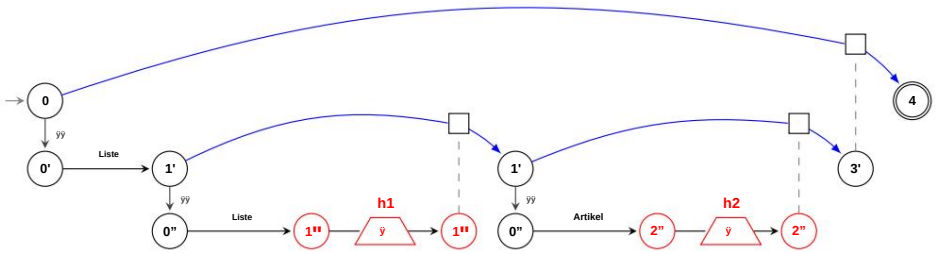


Abbildung 8. Ein erfolgreicher Lauf des SHAÿ in Abbildung 7 auf der Hedge- ÿListe · ÿListe · **h1ÿ** · ÿElement · h2ÿÿ.

Man sieht, dass alle Knoten der Teilbäume h1 und h2 auf den **Projektionszustand ÿ ausgewertet werden .** Man kann also über diese Teilbäume springen, ohne sie zu lesen. Dies ist dadurch gerechtfertigt, dass sie sich auf Ebene 2 befinden, die Information, dass der **Auswerter dieses SHAÿ runter.**

4.2.3. In-Memory-Mitgliedschaftstests

Das Testen der Mitgliedschaft h **ÿ** L(A) **für einen SHAÿ** s im Speicher kann wieder durch **Berechnen des eindeutigen Durchlaufs der Determinierung von** A **durchgeführt werden . Wie zuvor kann der Mitgliedschaftstester auf einer Determinierung während des Vorgangs basieren. Allerdings ist das Determinierungsverfahren für SHAÿ s komplizierter als für SHAs , da SHAÿ s Informationen von oben nach unten und nicht nur von unten nach oben und von links nach rechts weitergeben können. Die Determinierung beruht nicht nur auf der reinen Teilmengenkonstruktion, sondern verwendet auch Zustandspaare in den Teilmengen, im Grunde auf dieselbe Weise wie bei verschachtelten Wortautomaten (NWAs) [14,15]. Dies ist notwendig, um mit dem oben betrachteten Zustandsstapel umgehen zu können . Deshalb kann jeder Determinierungsschritt Zeit O(|A|)** kosten , **wie beispielsweise in [23] angegeben. Die obere Zeitschranke für den Mitgliedschaftstest für SHAÿ s liegt somit in der Zeit O(|h||A|** ^{5}**) und nicht mehr linear kombiniert wie bei SHAs.**

4.2.4. Beziehung zu SHAs Jeder

SHA A = **(ÿ, Q, ÿ, I,** F) kann unter Beibehaltung von Läufen und Determinismus $^{\text{runter}}$ = **(ÿ, Q, ÿ im** $^{\text{runter}}$**, WENN)** $^{\text{runter}}$ auf einen **SHAÿ** A **abgebildet werden . Die einzige Änderung von ÿ wird durch die Vergleich zu ÿ ist folgende Regel beschrieben:**

$$\frac{\text{ÿÿ ÿÿ p in ÿ q ÿ Q}}{\text{ÿ ÿÿ ÿÿ p in}^{\text{runter}}}$$

Unabhängig vom aktuellen Zustand q **ÿ Q ist die SHAÿ** A $^{\text{unten}}$ **können Sie mit der Auswertung der -Subhedge eines Teilbaums in jedem offenen Baumzustand** p **ÿ ÿÿÿ. Beispielsweise ist das** $^{\text{unten}}$ **für dSHAÿ** A **dSHA** A **aus Abbildung 4 der Einführung in Abbildung 9 dargestellt . Umgekehrt kann man um jede Top- dSHAÿ** A **in ein äquivalentes SHA umwandeln, indem man Nichtdeterminismus einführt, down-Verarbeitung zu eliminieren . Die Konstruktion finden Sie in Abschnitt 12.2 .**

Machine Translated by Google

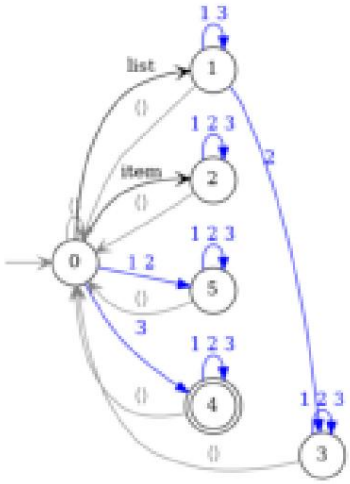


Abbildung 9. Der SHAÿ Ä[Filter] **für den dSHA** A **für den Filter** [self::list/child::item] **in Abbildung 4.**

4.2.5. Hedge-Zugänglichkeit

Die Hedge-Zugänglichkeit kann für **SHAÿ** genauso **definiert werden wie für SHAs. Allerdings kann** die Menge **accÿ(P) im Gegensatz zu SHAs nicht mehr in linearer Zeit in der Größe ÿ berechnet werden. Dies kann immer noch in kubischer Zeit erfolgen, ähnlich wie für NWAs [22]. Das Problem ist, dass die Hedge-Zugänglichkeit für SHAÿ nicht mehr mit dem Zugänglichkeitskonzept des Automatengraph**

Beispielsweise haben wir im **SHAÿ aus Abbildung** 7 accÿ({0}) = **{4}. Beachten Sie, dass 0ÿ ÿÿ in ÿ ist ,** wodurch Knoten **0ÿ nicht eine Übergangsregel 0 ÿÿ 0** gibt , **die von Knoten 0 im Graphen des Automaten aus zugänglich ist, ist dieser Menge gehört, obwohl es Dies liegt daran, dass 0ÿ nicht über irgendeinen Hedge von 0 aus zugänglich ist; das heißt, es gibt keinen Hedge** h , **sodass 0 bzgl. ÿ. Beachten Sie, dass 0 · ÿ · 0 ein Faktor des verschachtelten Wortes einer Folge von ÿ ist. Dies zeigt, dass 0ÿ dennoch von 0 über ein verschachteltes Wortpräfix zugänglich ist .** Für jeden **SHAÿ A kann die Hedge-Erreichbarkeit mit dem folgenden grundlegenden Datalog-Programm kubischer Größe berechnet werden, das von der Anzahl der Zustände von A abhängt:**

$$\frac{\textbf{p@q}\ \ddot{y}\ \ddot{y}\ \ddot{y}\ \text{q}\ \textbf{in}\ \ddot{y}\ \text{p}\ \ddot{y}\ddot{y}\ \text{p}\ \ddot{y}\ \textbf{in}\ \ddot{y}}{\text{acc(p, q) :- acc(p}\ \ddot{y}\ \text{, Q}\ \ddot{y}\ \textbf{).}} \qquad \frac{\text{p}\ \ddot{y}\ \textbf{Q-}}{\text{accm(p, p).}} \qquad \frac{\text{p}\ \textbf{a}\ \ddot{y}\ \ddot{y}\ \text{q}\ \textbf{in}\ \ddot{y}}{\text{acc(p, q).}}$$

$$\frac{\textbf{p, q, r}\ \ddot{y}\ \textbf{Q}}{\text{acc(p, q) :}\ddot{y}\ \text{acc(p, r), acc(r, q).}}$$

Aufgrund der kubischen Zeit der Hedge-Zugänglichkeit für SHAÿ s verwenden wir nur die Hedge-Zugänglichkeit für SHAs , wo es in linearer Zeit erfolgt.

4.3. Schemavollständigkeit

Schemata sind Eingaben des Mitgliedschaftsproblems, das wir betrachten, während Vollständigkeit eine grundlegende Eigenschaft von Automaten ist. Um beide Aspekte angemessen zu kombinieren, schlagen wir den Begriff der Schemavollständigkeit vor, den wir einheitlich sowohl für SHAs als a

Eine Menge von **ÿ- Transitionsregeln nennen wir SHAÿ vollständig, wenn alle ihre Relationen** (a **ÿ)aÿÿ, ÿÿÿ und @ÿ total sind. Für SHAs fordern wir, dass ÿÿÿ eine nichtleere Menge ist (und nicht eine totale Rel**

Ein SHA oder SHAÿ A **heißt vollständig, wenn ÿ vollständig ist und** l ÿ= **ÿ. Wir können jeden SHA oder SHAÿ vervollständigen, indem wir einen neuen Senkenzustand und auch Übergangsregeln in den Senkenzustand für alle linken Seiten hinzufügen, für die keine andere Übergangsregel existiert. Der Senkenzustand wird genau dann initial gemacht, wenn es vorher keinen initialen Zustand gibt. Er ist jedoch nicht final. Wir bezeichnen die Vervollständigung von** A **mit** compl(A).

Definition 5. Ein **partieller Lauf** von A auf einer Absicherung h ist ein Präfix r eines Laufs von compl(A) auf h, so dass

- r endet mit einem Zustand von Q, und

Machine Translated by Google

- r enthält nicht den Senkenzustand von compl(A).

Ein partieller Lauf r von A auf h heißt **blockierend ,** falls es keinen partiellen Lauf r gibt, für ÿ von A auf h, so den r ein striktes Präfix von **rÿ** ist .

Betrachten wir beispielsweise die Absicherung h = ÿListe · ÿArtikelÿÿ. **Der dSHA in Abbildung 4 hat den einzigartigen Lauf** R = **0 ·** ÿ · **0 ·** Liste **· 1 ·** ÿ · **0 ·** Artikel · 2ÿ **· 3 · ÿ · 4 auf** h , **der in Abbildung 5 dargestellt Die Teilläufe von** h **sind daher alle Präfixe von** nw(R) , **die mit einem Zustand enden, also 0, 0 ·** ÿ · **0, 0 ·** ÿ · **0 ·** Liste **· 1 usw. Keiner dieser Teilläufe ist blockierend.**

Definition 6. Ein Schema für einen Automaten A ist eine Hedge-Sprache **S ÿ Hÿ,** sodass L(A) **ÿ** S. Wir nennen den Automaten A schemavollständig für das Schema **S ,** falls kein Hedge h **ÿ S** einen blockierenden Teillauf von A hat.

Schemata S werden normalerweise verwendet, um die Art der Absicherungen einzuschränken, die ein Problem als Eingabe akzeptieren kann. Das dSHA -Mustervergleichsproblem für ein Schema S benötigt zwei Eingaben: eine Absicherung h **ÿ S und einen Automaten A anstelle eines verschachtelten regulären Ausdrucks e. Der Automat** A **wählt dann die Eingabeabsicherungen** h **ÿ S** aus, **die dem Muster entsprechen, d. h. so, dass** h **ÿ** L(A). **Aus diesem Grund nehmen wir an, dass S ein Schema für A ist, d. h.** L(A) **ÿ S. Wir werden oft annehmen, dass** A **für S schemavollständig ist , sodass kein Teildurchlauf von** A **auf einer Eingabeabsicherung** h **ÿ S jemals blockieren kann . Dies kann immer als nächstes basierend auf dem nächsten Lemma 5 erreicht werden .**

Beispiel 5. Der **dSHA** in Abbildung 4 für den XPath-Filter [self::list/child::item] hat die Signatur **ÿ** = {item, list}. Er ist schemavollständig für das Schema doc. Um dies zu erreichen, haben wir diesem Automaten den Zustand **5** hinzugefügt, der von keinem erfolgreichen Lauf verwendet wird. Trotzdem ist dieser **SHA** nicht vollständig. Um ihn zu vervollständigen, müssten wir **5** in Senkenzustände ausführen, indem wir viele Übergänge hinzufügen, wie **0@4 ÿ 5** und **0@5 ÿ** 5, und außerdem für alle anderen Zustände q **ÿ {1, 2, 3, 4, 5}** die Übergangsregeln hinzufügen
Liste Artikel
ÿÿÿ 5, q **ÿÿÿ** 5, **q@4 ÿ** 5 und **q@5 ÿ** 5. q

Wie wir sehen werden, würde die Automatenvervollständigung den Safe-No-Change-Algorithmus für die Subhedge-Projektion für Beispiel 5 unvollständig machen. Ohne Schemavollständigkeit kann die Safe-No-Change-Projektion jedoch unzuverlässig sein: Sie könnte die Ablehnung von Hedges innerhalb des Schemas übersehen . Aus diesem Grund ist es viel besser , **für die Eingabeautomaten der Safe-No-Change-Projektion nur Schemavollständigkeit statt vollständiger Vollständigkeit anzunehmen .** Für die **Kongruenzprojektion wird Schemavollständigkeit implizit angenommen.**

Wir stellen fest, dass die Schemavollständigkeit auch für nicht reguläre Schemata wohldefiniert ist. Für die Safe-No-Change-Projektion müssen wir Schemata tatsächlich nicht auf regulär beschränken, während für die Kongruenzprojektion nur reguläre Schemata berücksichtigt werden.

Wenn A **weiterhin schemavollständig für das Universalschema S** = **Hÿ** ist **und nicht unzugängliche Zustände haben, dann** ist A **vollständig.**

Die Schemavollständigkeit deterministischer Automaten für reguläre Schemata kann immer eingehalten werden, wie wir als nächstes zeigen werden. Dazu definieren wir, dass A mit Schema **S** kompatibel **ist , wenn es für jeden Hedge** h **ÿ S** einen Lauf von A **in** Laufÿ(h) **gibt . Schemavollständigkeit impliziert Kompatibilität, wie das nächste Lemma zeigt. Die Umkehrung gilt nur, wenn Determinismus angenommen wird.**

Lemma 5. Sei A ein deterministischer **SHA** mit Schema S. Dann ist A genau dann mit **S** kompatibel , wenn A für S schemavollständig ist .

Beweis. Angenommen, A **ist schemavollständig für S Wlog., dann** können wir annehmen, dass **S** ÿ = **ÿ. Beachten Sie, dass, wenn** l = ÿ, *ÿ* **ein blockierender Teillauf für jeden Hedge in S \ {ÿ} wäre. Da S** ÿ = **ÿ,** folgt, dass **q0 ÿ** l **existiert . Also** ist **q0 ein Teillauf für jeden Hedge** h **ÿ S \ {ÿ}. Da es aufgrund der Schemavollständigkeit keine blockierenden Teilläufe gibt, kann dieser Lauf schrittweise zu einem Lauf für jedes** h **ÿ S erweitert werden. Daher existiert** für jeden Hedge h **ÿ** S ein Lauf von **A, was Kompatibilität zeigt.**

Machine Translated by Google

Umgekehrt sei A mit **S kompatibel und deterministisch. Wir müssen zeigen, dass** A **schemavollständig für S ist. Sei** v **ein partieller Lauf von** A **auf** h ÿ **S. Aufgrund der Kompatibilität existiert ein Lauf** R ÿ **runÿ(h) , der in einem Anfangszustand von A startet. Aufgrund des Determinismus muss** v ein **Präfix von** nw(R) sein. **Daher** ist v **nicht blockierend.** □

Vorschlag 1. Jeder **SHA** A mit regulärem Schema **S** kann für S schemavollständig gemacht werden .

Beweis. Nach Lemma 5 genügt es, A mit **S kompatibel** zu machen. Sei B **ein dSHA mit S** = L(B). **Automat** B ist mit **S kompatibel .** Da B deterministisch ist, **impliziert Lemma 5 , dass** B **schemavollständig für S ist. Da S ein Schema für** A **mit** L(A) ÿ L(B) ist, **berechnen wir dann die Vervollständigung von** compl(A). **Das Produkt** C = B **×** compl(A) **mit Endzuständen ist schemavollständig für Schema S. Da** L(A) ÿ **S** = L(B), **ist es**
Aus CB F =
F **folgt, dass** A **×** F

L(C) = L(B) ÿ L(A) = L(A). □

Schließlich sei bemerkt, dass das Schema S am gleichen Produkt C = B **× durch die Menge** compl(A) , **außer dass die Menge der Endzustände** F ersetzt wird C **der Schema-Endzustände × (QA ÿ** F **S** = F B {sink}) erkannt werden kann. **Wir haben S** = L(C[F **C/F S]), wobei** C[F **C/F S]** der **dSHA ist S, das aus** C **durch Ersetzen der Menge der Endzustände** F erhalten wird C **mit** F **.**

5. Subhedge-Projektion

Als nächstes führen wir das Konzept der Subhedge-Projektionszustände für SHAÿ s ein, um Präfixe von Hedges zu unterscheiden, die für Subhedge irrelevant sind, und um SHAÿ s mit Subhedge-Projektion zu bewerten.

5.1. Subhedge-Projektionsstaaten

Intuitiv kann ein Subhedge-Projektionszustand nur in sich selbst schleifen, bis ein Hedge geschlossen wird (und dann wird er angewendet). Wenn man in einen Unterbaum hinabsteigt, kann der Subhedge-Projektionszustand immer noch in einen anderen Schleifenzustand geändert werden. Beim Verlassen des Unterbaums muss man jedoch zum ursprünglichen Subhedge-Projektionszustand zurückkehren. Offensichtlich ist jede Senke ein Subhedge- Projektionszustand. Die noch interessanteren Fälle sind jedoch Subhedge-Projektionszustände , die keine Senken sind.

Wir beginnen mit SHAÿ s ohne jegliche Einschränkungen, setzen aber Schema-Vollständigkeit voraus und später Determinismus.

Definition 7. Wir nennen einen Zustand q **ÿ Q** einen **Subhedge-Projektionszustand von ÿ ,** falls ein q existiert, ÿ **ÿ Q** das als Zeuge von q bezeichnet wird , so dass die Menge der Übergangsregeln von **ÿ,** die die ÿ oder q ganz links Position q enthält , in der folgenden Menge enthalten ist:

ÿÿ {q **ÿÿ** q ÿ **, q@q** ÿ **ÿ q,** q ÿ **ÿÿ ÿÿ** q ÿ q **ÿ@q ÿ ÿ** q **ÿ}**
q **ÿ{qÿ aÿÿ** ÿ **,** q **aÿÿ** q **|** a **ÿ ÿ}**

Die Menge aller Subhedge-Projektionszustände von **ÿ** wird mit **Qÿ** shp bezeichnet.

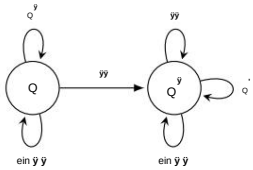
Für jeden vollständigen SHAÿ muss der obige Satz von Übergangsregeln dem Satz an der Position ganz Übergangsregeln von ÿ mit q **oder** q ÿ **links entsprechen. Wenn der SHAÿ** jedoch **nur aus allen schemavollständig für ein bestimmtes Schema S besteht, müssen nicht alle diese Übergänge vorhanden sein. Beachten Sie auch, dass ein Subhedge-Projektionszustand** q seinem Zeugen q ÿ **. Daher jeder Zeuge** Q ÿ eines Subhedge-Projektionszustands **entsprechen kann, der selbst ein Subhedge-Projektionszustand mit** ÿ **.**

Im Beispiel dSHAÿ A = **(ÿ, Q,ÿ, I, F)in Abbildung 7 haben wir Qÿ** = **{ÿ, 4, 2ÿ** shp**, Zeugen** q **3ÿ , 1ÿÿ, 2ÿÿ Der Zeuge aller dieser Subhedge-Projektionszustände ist ÿ. Beachten Sie, dass nicht alle möglichen Übergänge für die Zustände in Qÿ \ {ÿ} vorhanden sind, da dieser Automat nicht vollständig ist (aber immer noch** shp **-schemavollständig Schema** doc**).**
full

Lemma 6. Wenn q **ÿ Qÿ** ein Subhedge-Projektionszustand ist und q **ÿÿ** p bzgl. ÿ ist, dann q = p. shp
H

Machine Translated by Google

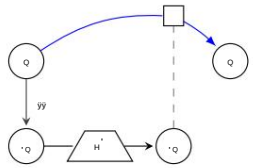
Beweis. Angenommen, q **ÿÿ** p **bzgl. ÿ und** q **ist ein Subhedge-Projektionszustand von ÿ mit Zeugen** Q **. Die einzigen möglichen Übergänge mit** q **oder** q **auf der linken Seite, dann lauten die folgenden:**



Angenommen, h = **t1 · i ÿ · · · · ·** tn **ist eine Folge von Bäumen oder Buchstaben** ti **ÿ ÿHÿÿ ÿ ÿ, wobei 1 ÿ n. Dann gibt es Läufe** Ri **ÿ** runÿ(ti) **und einen Lauf** R **ÿ runÿ(h) , der die Form** h = **q0 · R1 · q1 · ·** Rn **·** qn hat, **wobei q0** = q **und** qn **= p. Wir beweisen für alle 0 ÿ** i **ÿ** n , **dass** qi = q , indem wir es über **i induzieren .** Für i = **0 ist dies trivial. Für** i > **0 zeigt die Induktionshypothese, dass qiÿ1** = **q.**

Fall ti = a **ÿ ÿ. Dann** ist **qiÿ1 aÿÿ** qi **in ÿ. Da qiÿ1** = q **ein Subhedge-Projektionszustand von ÿ ist, folgt, dass** auch qi = **q ist.**

Fall ti = ÿh ÿ **für ein** h **ÿ Hÿ. Da qiÿ1** = q **ein Subhedge-Projektionszustand von ÿ mit dem Subrun Zeuge** q **, von** R ist , **muss** die **Anerkennung** von ti durch das folgende Diagramm gerechtfertigt werden:



Der Teillauf von R **von** h **darf nur den Zeugen** q enthalten **und außerdem** qi = **q.**

□

5.2. Vollständigkeit für die Subhedge-Projektion

Bevor wir definieren, wann ein SHAÿ für eine Subhedge-Projektion vollständig ist, zeigen wir, dass Subhedge-Projektionszustände in Läufen deterministischer SHAs nur bei irrelevanten **Präfixen auftreten können .**

Proposition 2. Sei A = **(ÿ, Q, ÿ, I,** F) ein **dSHAÿ** mit Schema S, R **ÿ runÿ(h)** ist der Run von compl(A) auf einem Hedge h **ÿ** S und q ist ein Subhedge-Projektionszustand für ÿ. Dann ist für jedes Präfix r **·** q von nw(R) das Präfix projÿ (r) von nw(h) für L(A) für Schema S subhedge-irrelevant.

Beweis. Sei R **ÿ runÿ(h) der eindeutige Lauf von** compl(A) **auf einem Hedge** h **ÿ S. Angenommen,** r **·** q **·** s = nw(R) **für einige Subhedge-Projektionszustände** q **von ÿ und seien** v = projÿ (r) **und** w = projÿ (s). **Da** v **·** w = nw(h) **und** h **ÿ S, folgt, dass** v **·** w **ÿ nw(S). Wir legen) ·** w **ÿ** nw(S) **beliebig fest. Sei** ˜h **ÿ Hÿ, so ÿ ÿ Hÿ, so dass** v **·** nw(h **dass ein Hedge** h nw(˜h) = v **·** nw(h) **· w. Für die Subhedge-Irrelevanz von v müssen wir zeigen, dass**

h **ÿ** L(A) **ÿ** ˜h **ÿ** L(A).

„ÿ“ Wir nehmen an, dass h **ÿ** L(A) **und müssen zeigen, dass** ˜h **ÿ** L(A). **Für den partiellen Lauf** r **·** q **von ·** A **auf** ˜h, **es müssen** R **und p existieren, so dass** q **·** R p **ÿ runÿ(h ÿ). Da** q **eine Teilhedge ist) · Projektionszustand von ÿ, es folgt nach Lemma 6 , dass** q = **p. Daher ist** r **·** q **·** nw(R q **·** s **ist ein Lauf von** A **auf** ˜h. **Da** R **für A erfolgreich war ,** muss s **in F enden ; daher ist der obige Lauf auf** ˜h auch für A **erfolgreich . Daher** ist ˜h **ÿ** L(A), **wie wir gezeigt haben.**

„ÿ“ Wir nehmen an, dass ˜h **ÿ** L(A) **und müssen zeigen, dass** h **ÿ** L(A). **Sei R˜ der eindeutige Lauf von** compl(A) **auf** ˜h. **Per Determinismus** muss **R˜ das Präfix** r **· q haben. Es gibt also**

Machine Translated by Google

R ÿ **,** P ÿ ,s, **so dass** p · R ÿ · q ÿ **runÿ(h ÿ) und** r · q · R ÿ · p · s = **R˜. Lemma 6 zeigt, dass** p = **q. Daher** ist r · q · s **der eindeutige Lauf von** compl(A) **auf** h **und dieser Lauf ist akzeptierend. Also** ist h ÿ L(A).

□

Wir stellen fest, dass Proposition 2 ohne die Annahme von Determinismus nicht gelten würde. Um dies zu sehen, können wir jedem dSHA eine Senke **als zweiten Anfangszustand hinzufügen. Man kann dann immer zu dieser Senke gelangen, die ein Subhedge-Projektionszustand ist. Dennoch kann kein Präfix, das zur Senke führt, für das Subhedge irrelevant sein.**

Proposition 2 zeigt, dass Subhedge-Projektionszustände es ermöglichen, Präfixe zu unterscheiden , die für deterministische SHAÿ subhedge-irrelevant sind . **Eine interessante Frage ist, ob alle subhedge-irrelevanten Präfixe auf diese Weise gefunden werden können. Eine eng damit verbundene Frage ist ,** ob es für alle regulären Muster **dSHAs gibt, die alle irrelevanten Subhedges projizieren.**

Beispiel 6. Betrachten wir Beispiel 3 noch einmal mit **ÿ = {a},** einem regulären Muster L = {ÿaÿ} und einem regulären Schema **S** = L **ÿ** {ÿÿ **· a}.** Das Präfix ÿÿ ist für L und S irrelevant , da alle seine Fortsetzungen, die im Schema enden, abgelehnt werden: Es gibt nur eine solche Fortsetzung, nämlich ÿÿ · a. In Abbildung 17 ist jedoch ein **dSHA** für L und **S** mit maximaler Subhedge-Projektion angegeben, das jedoch bei diesem Präfix nicht in einen Subhedge-Projektionszustand übergeht. Der Grund dafür ist, dass es für jedes Präfix in **ÿNÿÿ** in denselben Zustand **{3,** 4}D0 übergeht und das verschachtelte Wort des irrelevanten Subhedges ignoriert. Daher muss dieser Zustand ÿaÿ akzeptieren. Beim Lesen eines a in diesem Zustand geht der **dSHA** jedoch in den Ablehnungszustand {5}D0 über, was zeigt, dass die Absicherung aus dem Schema ÿÿ · a **ÿ** S abgelehnt wird.

Das Beispiel zeigt, dass wir letztendlich mit Mengen von Präfixen argumentieren. Wir heben zunächst die Begriff der Irrelevanz von Subhedge gegenüber Präfixsätzen.

Definition 8. Sei **S ÿ Hÿ** ein Schema und L **ÿ S** eine Sprache, die dieses Schema erfüllt. Eine Menge verschachtelter Wortpräfixe U **ÿ** prefs(Nÿ) heißt **Subhedge relevant** für L mit Schema **S ,** falls es verschachtelte Wörter v **ÿ Nÿ** Präfixe u · v ÿÿ gibt ÿ Andernfalls gilt die Menge ÿ U, so dass die Differenz M S (u ÿ · v · du U als für L bzgl. S **irrelevanter Subhedge** bezeichnet wird .

Um das Problem von Beispiel 6 zu erklären, definieren wir für jedes verschachtelte Wortpräfix w **ÿ** prefs(Nÿ) **eine Klasse** classL (w) **ähnlicher Präfixe, wobei das verschachtelte Wort jedes irrelevanten Unterhedges von links nach rechts** durch **Nÿ** ersetzt wird. Für alle Präfixe w **ÿ** prefs(Nÿ), **verschachtelte Wörter** v **ÿ Nÿ, Buchstaben** a **ÿ ÿ, Teilmengen von Präfixen** U **ÿ** prefs(Nÿ) **und** l **ÿ ÿˆ Buchstaben oder Klammern gilt Folgendes:**

$$\text{KlasseL}_S(w) = \textbf{Klasse_pr\{ÿ\}}\ (w)$$

$$\text{Klasse_prU}(ÿ) = \begin{cases} U \cdot \textbf{Nÿ} & \textbf{wenn U irrelevant für L bzgl. S, sonst} \\ U \end{cases}$$

class_prU(w · ÿ · v) = class_insUÿ(ÿ, v) **wobei Uÿ** = class_prU(w) **class_prU(a ·** v) = class_insUÿ(a, v) **wobei Uÿ** = U **·** a class_prU(ÿv class_insU(l, v) = ÿ ÿ · v) = class_insUÿ(ÿ, v) **wobei Uÿ** = **class_prU(ÿ ·** v) **wenn** U **irrelevant für** L **bzgl. S** U **· Nÿ**

class_prU · l (v) **sonst**

Definition 9. Ein Präfix u **ÿ** prefs(Nÿ) ist **stark subhedge-irrelevant** für L bzgl. **S ,** wenn die Menge classL (u) subhedge-irrelevant für L bzgl. S ist.
S

Beispiel 7. In unserem laufenden Beispiel 6, wo L = {ÿaÿ}, **S** = L **ÿ** {ÿÿ **· a}** und **ÿ = {a},** haben wir classL (ÿ) = ÿ · **Nÿ,** also ist das Präfix ÿ stark subhedge-irrelevant. Die Menge classL (ÿaÿ) = **S ÿNÿÿ** ist subhedge-relevant, also ist das Präfix ÿaÿ nicht stark subhedge-irrelevant. Die Menge classL (ÿÿ · a) = **ÿNÿÿ ·** S a **· Nÿ** ist subhedge-irrelevant, also ist das Präfix ÿÿ · a stark subhedge-irrelevant.

Machine Translated by Google

Definition 10. Ein **dSHAÿ** A = **(ÿ, Q, ÿ, I,** F) ist **vollständig für die Subhedge-Projektion für Schema S,** wenn A schemavollständig für **S** ist und für alle Hedges h **ÿ S** und alle Präfixe v von nw(h), die für ÿ stark Subhedge-irrelevant sind , der eindeutige Lauf von A auf h zu einem Subhedge-Projektionszustand von **ÿ** bei v geht.

Beispiel 8. Der **dSHA** in Abbildung 17 ist für die Subhedge-Projektion für unser laufendes Beispiel 6 vollständig, wobei L = {ÿaÿ} und **S** = L **ÿ** {ÿÿ · **a}.** Er wurde durch Kongruenzprojektion erhalten. Auf allen Präfixen der für Subhedges irrelevanten Klasse classL (ÿ) geht er in den Subhedge-Projektionszustand {0}D1 über. Bei allen Präfixen der Klasse classL (ÿaÿ), die nicht subhedge-irrelevant ist, geht es in den Zustand **{3,** 4}D0, der kein subhedge-Projektionszustand ist. Bei allen Präfixen der subhedge-irrelevanten Klasse classL (ÿÿ · a) geht es in den subhedge-Projektionszustand {5}D0.

Beispiel 9. Das **dSHAÿ** aus Abbildung 7 ist für die Subhedge-Projektion mit Schema doc vollständig. Es kann durch eine Safe-No-Change-Projektion erhalten werden. Im Allgemeinen sind **dSHAÿ ,** die durch eine Safe-No-Change-Projektion erhalten werden, jedoch möglicherweise nicht für die Subhedge-Projektion vollständig, wie wir in Abschnitt 6.3 besprechen werden.

5.3. In-Memory-Evaluator mit Subhedge-Projektion **Als nächstes zeigen**

wir, wie man die Übergänge für SHAÿ s durch Subhedge-Projektion verfeinert. Wir definieren die Übergangsbeziehung mit Subhedge-Projektion ÿÿshpÿ **Q × Q bezüglich ÿ,**
so dass für alle Hedges h, h **ÿ Hÿ und Buchstaben** a **ÿ ÿ,**

$$\frac{q \text{ ÿ } \mathbf{Q\ddot{y}}_{shp} \quad h \text{ ÿ } \mathbf{H\ddot{y}}}{q \;\ddot{y}\ddot{y}shp\; q \;\mathbf{bzgl.}\; \ddot{y}} \qquad \frac{\mathbf{\ddot{y}\ddot{y}\; Q\ddot{y}}\; q shp\; q \;\mathbf{a\ddot{y}\ddot{y}}\; qq \;\mathbf{in}\; \ddot{y}}{a\ddot{y}\ddot{y}shp\; q \;\mathbf{bzgl.}\; \ddot{y}} \qquad \frac{q \;\ddot{y}\ddot{y}\; \mathbf{Q\ddot{y}}_{shp}}{q \;\ddot{y}\ddot{y}\ddot{y}shp\; q \;\mathbf{bzgl.}\; \ddot{y}}$$

$$\frac{q \;\mathbf{\ddot{y}\ddot{y}\; Q\ddot{y}}\; \ddot{y}\ddot{y}shp\; q \; shp\; q \;\mathbf{bzgl.}\; \ddot{y}\; q \;\mathbf{bzgl.}\; \ddot{y}}{\ddot{y}\ddot{y}\ddot{y}shp\; qq \;\mathbf{bzgl.}\; \ddot{y}}$$

$$\frac{\mathbf{\ddot{y}\ddot{y}}\; q \;\mathbf{\ddot{y}\ddot{y}\; Q\ddot{y}}\; \ddot{y}\ddot{y}shp\; p \;\mathbf{in}\; \ddot{y}\; p \; shp\; q \qquad \mathbf{q@p}\; \ddot{y}\; q \;\mathbf{in}\; \ddot{y}}{\ddot{y}h\ddot{y}\; \ddot{y}\ddot{y}\ddot{y}shp \;\mathbf{bzgl.}\; \ddot{y}}$$

qq **Die erste Regel besagt, dass Subhedge-Projektionstransitionen in Subhedge-Projektionszuständen bleiben, bis das Ende des aktuellen Subhedges erreicht ist. Dies ist nach Lemma 6 unter der Bedingung korrekt, dass es keine blockierenden Läufe gibt. Die anderen Regeln besagen, dass sich der Evaluator in anderen Zuständen ohne Subhedge-Projektion verhält.**

Proposition 3. Sei A = **(ÿ, Q, ÿ, I,** F) ein **SHAÿ ,** das schemavollständig für S ist. Dann gilt für alle Hedges h **ÿ S** und Zustände **q,** p **ÿ** Q:

$$q \;\ddot{y}\ddot{y}\; p \;\mathbf{bzgl.}\; \ddot{y} \;\text{genau dann, wenn}\; q \;\ddot{y}\ddot{y}shp\; p \;\mathbf{bzgl.}\; \ddot{y}$$

Beweis. Wir unterscheiden zwei Fälle wie folgt:

Fall q **ÿÿ Qÿ. Dann gilt für jedes** p **ÿ Q,** q **ÿÿ** p **bzgl. ÿ genau dann, wenn** q ÿÿshp p **bzgl. ÿ per Definition der** shp.
projizierenden Übergänge gilt.

Fall q **ÿ Qÿ Dann** ist q ÿÿshp q **bzgl. ÿ. Da** h **ÿ S und** A **schemavollständig für S sind, gibt es** shp.
h **existiert eine Folge von** A **auf h; der Zustand ist also q, so dass** q **ÿÿ** q **dass** q **bzgl . ÿ. Lemma 6 zeigt**
= q, also q **ÿÿ** q **bzgl. ÿ.**

□

Zur Auswertung eines nichtdeterministischen SHAÿ A = **(ÿ, Q, ÿ, I,** F) **mit Subhedge-Projektion auf einen Hedge** h , **dessen Graph im Speicher abgelegt ist, können wir die eindeutige Teilmeng**
P **ÿ Q, so dass** I ÿÿshp P **bezüglich der Übergangsrelation ÿ** Determination der **Determination**

Machine Translated by Google

von A **und testen, ob** P **ÿ** F ÿ= **ÿ. Beim Erreichen eines beliebigen Subhedge-Projektionszustands** P ÿ **während kann der Auswerter direkt zum Ende des aktuellen Subhedges springen, indem er die „Stay“-Regel der Determinisierung von A anwendet:**

$$\frac{P^{ÿ}\ ÿ\ Qÿ\ \text{shp}^{\text{det}}\qquad h\ ÿ\ Hÿ}{P^{ÿ}\ ^{H}_{ÿÿ}\text{shp}\ P^{ÿ}\ \textbf{bzgl. ÿ}\ ^{\text{det}}}$$

Dadurch wird vermieden, dass in einem Unterhedge-Projektionszustand der Determinierung von A ein beliebiges Unterhedge besucht wird. Die Laufzeit hängt daher linear von der Größe des nicht projizierten Teils von h ab , **unter der Annahme, dass die Determinierung von** A **verfügbar ist.**

Wie zuvor muss nur der benötigte Teil der Determinierung von A **zur Auswertung von** h **im laufenden Betrieb erstellt werden. Insgesamt kostet dies eine Polynomzeit in der Größe des benötigten Teils der Determinierung (aber nicht notwendigerweise eine lineare Zeit, da die Determinierung von SHAÿ s kostspieliger ist als die von** SHAs). Beachten Sie auch, dass wir jedes Mal, wenn wir einen neuen Zustand P **Determinierung von** A **entdecken , testen müssen, ob es sich um** ÿ **der einen Subhedge-Projektionszustand handelt. Dazu** ermitteln **und testen wir, ob nur der Übergang mit Zeuge** P muss den Zustand P **'' berechnen , so dass die** ÿ **ÿÿ ÿÿ** P **" bzgl . ÿ** P- **Regeln für Subhedge-Projektionszustände für** P ÿ **" sind verfügbar bzgl. ÿ** det **.**

6. Sichere No-Change-Projektion

Wir wollen das reguläre Mustervergleichsproblem für Hedges mit Subhedge-Projektion lösen . Wir nehmen an, dass das reguläre Muster durch ein **dSHA gegeben ist, während das Schema beliebig sein kann, mit der Ausnahme, dass das dSHA schemavollständig sein muss (siehe Abschnitt 4.3). Anschließend stellen wir den Safe-No-Change-Projektionsalgorithmus vor, der das dSHA in einige dSHAÿ** kompiliert **und dabei Subhedge-Projektionszustände einführt.**

Die Idee der Safe-No-Change-Projektion besteht darin, Informationen von oben nach unten zu übertragen, um Schleifenzustände zu unterscheiden, die sicher nicht mehr geändert werden. Dadurch werden Subhedge-Projektionszustände erzeugt, wie wir anhand von Beispielen veranschaulichen werden. Der Beweis der Solidität der Safe-No-Change-Projektion ist nicht trivial und aufschlussreich für den Beweis der Solidität der Kongruenzprojektion in Abschnitt 7.3. Vollständigkeit kann für die Subhedge-Projektion nicht erwartet werden, da die Safe-No-Change-Projektion nur einfache Schleifen berücksichtigt. Allgemeinere Schleifen können nicht behandelt werden und es ist nicht klar, wie das gemacht werden könnte. Der Kongruenzprojektionsalgorithmus in Abs

Wir halten den Safe-No-Change-Compiler unabhängig vom Schema und müssen daher nicht dessen Regelmäßigkeit annehmen. Wir müssen jedoch die Schemavollständigkeit des Eingabe -dSHA annehmen , **um zu zeigen, dass der Ausgabe- dSHAÿ dieselbe Sprache innerhalb des Schemas erkennt. Das reguläre Mustervergleichsproblem kann dann im Speicher und mit Subhedge-Projektion gelöst werden,** indem der **dSHAÿ auf dem Eingabe-Hedge im Speicher und mit Subhedge-Projektion ausgewertet wird ,** wie in Abschnitt **5.3 erläutert.**

6.1. Algorithmus

Wir beschreiben nun den Safe-No-Change-Projektionsalgorithmus. Sei A = (ÿ, **Q, ÿ, I,** F) **ein SHA mit Schema S. Der Algorithmus nimmt das dSHA** A **als Eingabe und kompiliert es zu einem dSHAÿ** A snc. **Wir werden darlegen, warum** A **für die Subhedge-Projektion unter der Bedingung , dass** A **für Schema S schemavollständig ist, stichhaltig ist . Wir** müssen nicht **einmal annehmen, dass das Schema S regulär ist. Der Beweis ist acht Seiten lang und wird daher in den Anhang delegiert.**

Für jede Menge P **ÿ Q definieren wir die Menge der Zustände, die sicher zu P führen:**

$$\text{sicheres}ÿ\ (P) = \{q\ ÿ\ Q\ |\ \text{acc}ÿ\ (\{q\})\ ÿ\ P\}.$$

Der Zustand q ist **also sicher für** P , **wenn jeder Hedge-Read aus** q , **bei dem der Lauf mit ÿ nicht blockiert, einen Zustand in P erreichen muss. Wir stellen fest, dass** sicheres ÿ (P) **in linearer Zeit berechnet werden kann in d**

Machine Translated by Google

Größe von ÿ unter Verwendung der inversen Hedge-Zugänglichkeit von P. Wir definieren die Menge der Zustände, die sich nicht mehr ändern:

Keine Änderung ÿ = {q **|** q **ÿ** sicheres ÿ **({q})}.**

Also gilt q **ÿ** no-changeÿ **genau dann, wenn** accÿ({q}) **ÿ {q}. Das bedeutet, dass alle Übergangsregeln, die in** q beginnen , **in q schleifen müssen . Im Beispielautomaten aus Abbildung 4 sind die selbstschleifenden Zustände diejenigen in** no-changeÿ = **{2, 3, 4, 5}.**

Lemma 7. Sei A = **(ÿ, Q, ÿ, I,** F) ein **SHA ,** das schemavollständig für das Schema S ist. Für jeden Hedge h **ÿ S** und Zustand q **ÿ** I **ÿ** keine-Änderungÿ gilt dann q **ÿÿ** q bzgl. ÿ.
H

Beweis. Die Schemavollständigkeit von A **für S, angewendet auf** h **ÿ S und** q **ÿ** I, **ergibt die Existenz ÿ** accÿ(q). = **eines Zustandes** q **ÿ Q, so dass** q **ÿÿ** q **bzgl . ÿ. Sei** q **sei ein beliebiger solcher Zustand. Beachten q. Dies**
H ÿ ÿ ÿ
Da q **ÿ** keine-Änderungÿ , **haben wir** q **ÿ** sicheresÿ **({q}), so dass** q **Sie, dass** q **ÿ** accÿ({q}) **impliziert,** dass q
ÿ ÿ
beweist q **ÿÿ** q **bzgl. ÿ.** □
H

Für jeden Zustand q **ÿ Q und jede Teilmenge von Zuständen** Q **ÿ Q definieren wir Folgendes:**

s-downÿ (q, Q) = sicheresÿ **({p ÿ Q | q@ÿp ÿ Q})** s-keine-Änderungÿ (q) = s-downÿ (q, **{q})**

Ein Zustand gehört zu s-downÿ (q, Q) , **wenn alle Zustände** p **ÿ accÿ(q)** die Bedingung **q@ÿp ÿ** Q **erfüllen . Daher** ist p **ein Zustand auf einer niedrigeren Ebene, der sicher in einen Zustand in Q übergehen kann . Ein Zustand** p **gehört zu** s-no-changeÿ (q), **wenn er q sicher nicht ändert , d.ÿh. wenn** {q}@ÿaccÿ({p}) **ÿ {q}.**

Als nächstes kompilieren wir den SHA A **zu einem SHAÿ** A = **(ÿ, , ÿ** snc , snc , F snc). **Dafür sei ÿ** Qsnc **sei ein neues Symbol und betrachten den Zustandssatz wie folgt:**

Qsnc = **{ÿ} ÿ (Q × 2** Q).

In der Praxis beschränken wir den Zustandsraum auf die Zustände, die zugänglich sind oder die dSHAs bereinigen , **und behalten nur die Zustände bei, die bei einem erfolgreichen Durchlauf verwendet werden. Im schlimmsten Fall kann die Konstruktion jedoch tatsächlich exponentiell sein. Beispiel 15 kann angepasst werden,**

Ein Paar (q, P) **bedeutet, dass sich der Zustand** q **im aktuellen Teil-Hedge sicher nicht mehr ändern wird, wenn** q **ÿ** P **ÿ** keine-Änderungÿ **. In diesem Fall kann die Unterabsicherung projiziert werden. Die Sätze von Anfangs- und Endabsicherung Zustände werden wie folgt definiert:**

snc = ich **× {ÿ}** F snc = F **× {ÿ} .**
ich

Wie man die Übergangsregeln von A **aus denen von** A generiert , **wird in Abbildung 10** beschrieben. **Bei Zuständen, die auf der obersten Ebene** (q, P) zugewiesen werden, ist **die Menge** P **leer, sodass nur die Zustände in** keinem Änderungs-ÿ **sicher für keine Änderung sind. Deshalb wird die Definition von** I snc **und** F **abwärts von einem Zustand verwenden Sie** P = **ÿ. Wenn Sie** (q, P) , **für den** q **sicher nicht geändert werden kann, d. h.** q **ÿ** P **ÿ** kein Änderungs-ÿ , **folgende Regel angewendet: dann** gehen,

$$\frac{\text{q ÿ P ÿ keine Änderungÿ}}{\text{ÿÿ (q, P) ÿÿ ÿ in ÿ}}$$

Die Auswertung auf der unteren Ebene führt zum zusätzlichen Zustand ÿ, wo sie dann in einer Schleife ausgeführt wird, bis sie zu q auf der oberen Ebene **zurückkehrt .** Wenn man von einem Zustand (q, P) **abwärts geht , sodass** q **ÿÿ** P **ÿ** keine Änderungÿ ist , **dann wird die folgende Regel angewendet:**

$$\frac{\text{ÿÿ ÿÿ in ÿ q ÿÿ P ÿ keine Änderungÿ}}{\text{P) ÿÿ (q ÿÿ ÿ ,s-keine-Änderungÿ (q)) in ÿ (q,}}$$

Machine Translated by Google

Die Zustände in der Menge s-no-changeÿ (q) **auf der unteren Ebene führen mit Sicherheit nicht dazu, dass** sich q **auf der oberen Ebene für irgendeine nachfolgende Unterabsicherung ändert (wir könnten weitere irrelevante Unterabsicherungen erkennen, indem wir zulassen,** dass sich q **ändert, aber nur auf eine eindeutige Weise. Dies kann mit der alternativen Definition erreicht werden:** s-no-changeÿÿ (q) = ÿrÿQs-downÿ (q, **{r}). Die Berechnung dieser Menge würde eine quadratische Zeit** O(n **|A|) erfordern, während die Berechnung von** s-no-changeÿ (q) **in einer linearen Zeit O(|A|) durchgeführt werden kann).**

$$\frac{\text{q einÿÿ q}^{ÿ}\ \textbf{in ÿ}\ \text{q}\ \textbf{ÿÿ}\ \text{P}\ \textbf{ÿ}\ \text{unverändert ÿ (q, P)}}{\textbf{aÿÿ}\ (\text{q}^{ÿ}, \text{P})\ \textbf{in ÿ}}\ \text{snc} \qquad \frac{\text{a}\ \textbf{ÿ ÿ}\ \text{q}\ \textbf{ÿ}\ \text{P}\ \textbf{ÿ}\ \text{unverändertÿ (q, P)}}{\textbf{aÿÿ}\ (\text{q, P})\ \textbf{in ÿ}}\ \text{snc}$$

$$\frac{\textbf{ÿÿ ÿÿ}^{ÿ}\ \textbf{in ÿ}\ \text{q}\ \textbf{ÿÿ}\ \text{P}\ \textbf{ÿ}\ \text{keine Änderungÿ}}{\text{P)}\ \textbf{ÿÿ}\ (\text{q}^{\textbf{in ÿÿqÿ ,s-keine-Änderungÿ (q))}}\ \textbf{in ÿ}\ (\text{q,}}\ \text{snc} \qquad \frac{\text{q}\ \textbf{ÿ}\ \text{P}\ \textbf{ÿ}\ \text{keine Änderungÿ}}{\textbf{ÿÿ}\ (\text{q, P})\ \textbf{ÿÿ ÿ in ÿ}}\ \text{snc}$$

$$\frac{\textbf{q@p ÿ}\ \text{q}^{ÿ}\ \textbf{ÿÿ}\ \text{P}\ \textbf{ÿ}\ \text{keine-Änderungÿ}}{\text{(q, P)@(p,s-keine-Änderungÿ (q))}\ \textbf{ÿ}\ (\text{q}^{ÿ}, \text{P})\ \textbf{in ÿ}}\ \text{snc} \qquad \frac{\text{q}\ \textbf{ÿ}\ \text{P}\ \textbf{ÿ}\ \text{keine Änderungÿ}}{\text{(q, P)@ÿ ÿ (q, P)}\ \textbf{in ÿ}}\ \text{snc}$$

$$\frac{\text{ein}\ \textbf{ÿ ÿ}}{\textbf{ÿ aÿÿ ÿ in ÿ}}\ \text{snc} \qquad \frac{\text{WAHR}}{\textbf{ÿ@ÿ ÿ ÿ in ÿ}}\ \text{snc} \qquad \frac{\text{WAHR}}{\textbf{ÿÿ ÿ ÿÿ ÿ in ÿ}}\ \text{snc}$$

Abbildung 10. Die Übergangsregeln des SHAÿ A $^{\text{snc}}$ **abgeleitet aus denen des SHA A.**

Bei Anwendung auf den SHA in Abbildung 4 für [self::list/child::item] **ergibt die Konstruktion** den **SHAÿ in Abbildung 11, der tatsächlich dem SHAÿ aus Abbildung 7 bis zur Zustandsumbenennung entspricht. Bei Ausführung auf dem Hedge** ÿlist · ÿlist · **h1ÿ** · ÿitem · h2ÿÿ, **wie in Abbildung 8 gezeigt, müssen die Unterhedges h1 oder h2** nicht besucht werden , **da sie alle ausgehend vom Projektionszustand ÿ erreicht werden.**

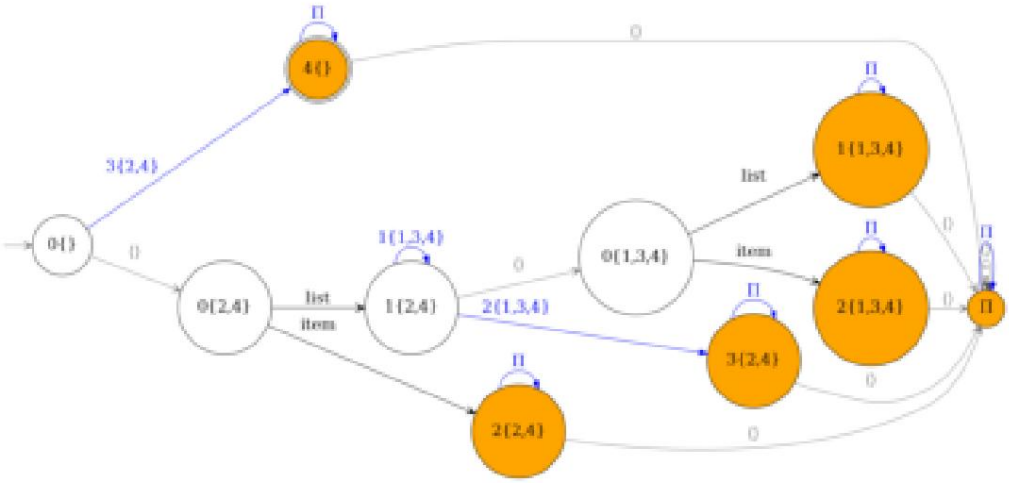


Abbildung 11. Die Safe-No-Change-Projektion dSHAÿ $^{\text{snc}}$ **konstruiert aus dem dSHA** A **für Abfrage** A [self::list/child::item] **ist in Abbildung 4 dargestellt . Die Subhedge-Projektionszustände sind orange eingefärb Unnütze Zustände und Übergänge, die aus dem Schemadokument herausführen,** werden **weggelassen. Zustand ÿ hat einen Übergang, der mit dem Platzhalterbuchstaben „ÿ“ gekennzeichnet ist, der entweder für eine Buchstabenliste** oder **ein** Element steht. **Dieses dSHAÿ ist gleich dem dSHAÿ in Abbildung 7 bis zur** $1^{ÿ}$ **Zustandsumbenennung 0** = **(0,** {}), **0ÿ** = **(0, {2,** 4}), **0ÿÿ** = **(0, {1, 3,** 4}), = **(1, {2,** 4}), **1ÿÿ** = **(1, {1, 3,** 4}), **2ÿ** = **(2, {2,** 4}), **2ÿÿ** = **(2, Denken Sie daran, dass** unverändertes ÿ = **{2, 3, 4, 5}.**

6.2. Solidität

Als nächstes legen wir ein Soliditätsergebnis für eine sichere No-Change-Projektion dar und beweisen es.

Machine Translated by Google

Theorem 1 (Klarheit der Safe-No-Change-Projektion). Wenn ein **SHA** A für ein bestimmtes Schema S schemavollständig ist , dann bewahrt die Safe-No-Change-Projektion für A die Sprache innerhalb dieses Schemas: L(A snc) **ÿ S** = L(A).

Der projizierende In-Memory-Evaluator von A snc **wird effizienter sein als der nicht projizierende Evaluator von A. Beachten Sie die Größe von** A **von A.** snc **kann exponentiell größer sein als das Daher bevorzugen wir für die Auswertung eines dSHA** A **mit Subhedge-Projektion auf einem gegebenen Hedge h möglicherweise nur den benötigten Teil von** A **im laufenden Betrieb. Dieser Teil hat eine Größe** snc **von** O(|h|) **und kann** in der Zeit **O(|A| |h|) berechnet** werden , **sodass die exponentielle Vorverarbeitungszeit vermieden**

6.3. Unvollständigkeit

Die Safe-No-Change-Projektion kann für die Subhedge-Projektion unvollständig sein, sodass nicht alle Präfixe, die stark subhedge-irrelevant sind, auf Subhedge-Projektionszustände abgebildet werden . Dies wird durch das Gegenbeispiel dSHA A **für den XPath- Filter [child::item] in Abbildung 12** gezeigt . **Seine Safe-No-Change-Projektion** A **ÿitem.ÿitem** snc **ist in Abbildung 13** dargestellt. **Beachten Sie, dass das Präfix** u = **hat die Klasse** classL(A) (u) = ÿitem.ÿitem.Nÿ, **die subhedge-irrelevant ist, sodass** u für den **XPath- Filter [child::item] stark subhedge-irrelevant ist . Trotzdem führt sie zum Zustand (2, {1, 3, 5,** 6}), **der kein Subhedge-Projektionszustand ist, da 2 ÿÿ {1, 3, 5, 6}.**
Das Problem besteht darin, dass dieser Zustand immer noch in **(4, {1, 3, 5,** 6}) **geändert werden kann . Dieser Zustand ist in Bezug auf den Filter irgendwie gleichwertig, aber nicht gleich (2, {1, 3, 5,** 6}).

Ein weiteres Unvollständigkeitsproblem sollte erwähnt werden: Die Safe-No-Change-Projektion reagiert empfindlich auf die Vervollständigung von Automaten, wie bereits in Beispiel 5 festgestellt wurde. Dies liegt daran, dass ein Zustand vor der Vervollständigung **zu** No-Changeÿ gehören kann , **aber nicht mehr nach dem Hinzufügen eines Sinks. Trotzdem ändern sich solche Zustände in keinem Baum, der das Schema erfüllt, jemals. Dieses Problem kann beispielsweise auf das Beispiel dSHA in Abbildung 4 für die XPath- Abfrage** [self::list/child::item] angewendet werden. **Daher war es wichtig, für die Safe-No-Change-Projektion nur Schemavollständigkeit anzunehmen und keine vollständige Vollständigkeit zu erzwingen.**

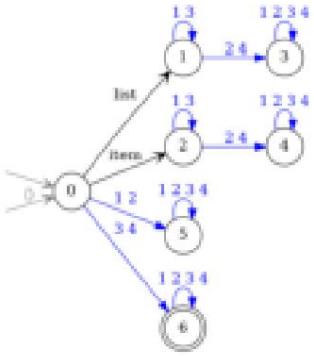


Abbildung 12. Ein schemavollständiger dSHA für den XPath- Filter [child::item]. Es ist ein Gegenbeispiel für die Vollständigkeit der Safe-No-Change-Projektion, siehe Abbildung 13.

Machine Translated by Google

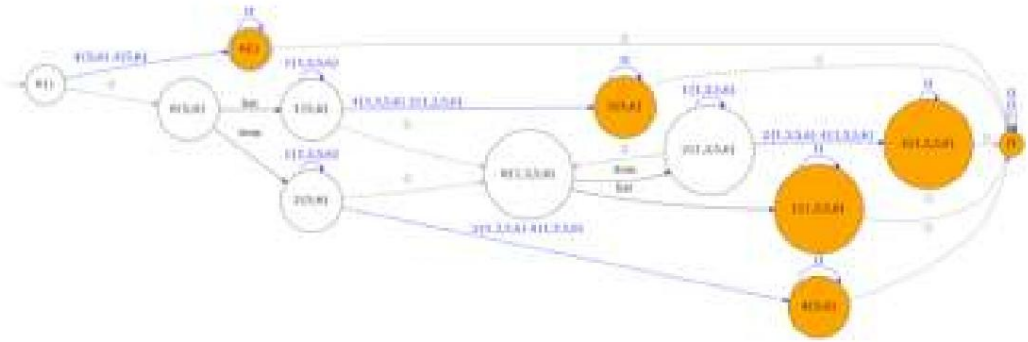


Abbildung 13. Die Safe-No-Change-Projektion. Eine ^{snc} **des dSHA** A **in Abbildung 12. Es ist unvollständig für Subhedge-Projektion mit Schemadokument** im **Zustand {2, {1, 3, 5,** 6}): **Dieser Zustand ist kein Subhedge -Projektionszustand, auch wenn die Präfixe** ÿListe · ÿElement **und** ÿElement · ÿElement, **die zu ihm führen, für die Subhedge irrelevant sind.**

7. Kongruenzprojektion Wir

präsentieren den Kongruenzprojektionsalgorithmus zum Erkennen irrelevanter Subhedges für reguläre Hedgemuster mit regulären Schemabeschränkungen. Wir beweisen, dass die Kongruenzprojektion für die Subhedgeprojektion vollständig ist, wodurch die Unvollständigkeit der Safe-No- Change-Projektion behoben wird. Dafür darf die Kongruenzprojektion das Schema nicht mehr ignorieren, daher müssen wir annehmen, dass das Eingabeschema ebenfalls regulär ist.

7.1. Ideen

Die Grundidee der Kongruenzprojektion besteht darin, das Gegenbeispiel für die Vollständigkeit der Safe-No-Change-Projektion in Abbildung 13 aufzulösen . **Dort wird beim Lesen eines irrelevanten Subhedge-Hedges ein Zustand geändert. Die Zustandsänderung führt jedoch zu einem irgendwie gleichwertigen Zustand, sodass sie nicht wirklich relevant ist. Die Idee besteht daher darin, zu erkennen, wann ein Zustand beim Lesen eines beliebigen Subhedges immer Dies geschieht letztendlich durch die Kongruenzprojektion in Abbildung 14 , die wir erstellen werden .**

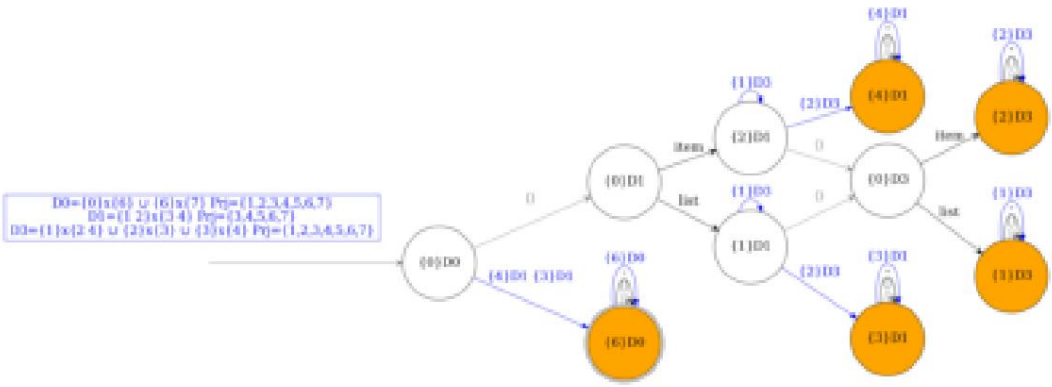


Abbildung 14. Die Kongruenzprojektion dSHAÿ A cgr(doc) **für das Gegenbeispiel der Vollständigkeit der Safe-No-Change-Projektion, d. h. die dSHA** A **für den Filter [child::item] in Abbildung 12 mit den Schema -Endzuständen FS** = **{0, 5, 6}. Beachten Sie, dass der Zustand** {1}D3 , der über das Präfix ÿlist · ÿitem **erreicht wird , und der Zustand** {2}D3 , der über das Präfix ÿitem · ÿitem **erreicht wird , für die Subhedge irrelevant und somit Subhedge- Projektionszustände sind.**

Die offensichtliche Frage ist folgende: Welche Äquivalenzrelation wählen wir? Angenommen, wir möchten testen, ob eine Absicherung, die ein reguläres Schema S erfüllt , **mit einem regulären Muster L übereinstimmt . Im eingeschränkten Fall von Wörtern ohne Schemabeschränkungen könnte die Idee darin bestehen, Myhill-Nerodes Kongruenz $\overline{M}$ zu verwenden, $\overline{M}$ sondern auf Zustände abgebildet. Im allgemeinen Fall jedoch Die Situation wird komplexer, da im allgemeinen Fall** diff **S ist möglicherweise kein**

Machine Translated by Google

Äquivalenzrelation. Es kann also sein, dass es sich nicht um eine Übereinstimmung handelt, wie bereits in Beispiel 4 dargestellt. Darüber hinaus erfordert die Behandlung der Verschachtelung von Hecken eine Aktualisierung der betrachteten Beziehung, wenn Sie sich eine Hecke nach unten bewegen.

7.1.1. Differenzrelationen auf Zuständen

Der Kongruenzprojektionsalgorithmus behält eine Differenzrelation für Zustände von der dSHA , der zu Beginn durch die Differenzrelation auf der Hedge- Differenz **induziert und bei jeder Herabstufung eines Hedges aktualisiert wird.**

Definition 11. Seien **(ÿ, Q, ÿ, _, _)** ein **dSHA.** Eine **Differenzrelation für ÿ** ist eine symmetrische Relation auf Zuständen D **ÿ Q ×** Q, so dass für alle h **ÿ Hÿ**

$$(q \text{ ÿÿ}^{H} q \text{ ÿ bzgl. ÿ ÿ } p \text{ ÿÿ}^{H} p \text{ ÿ bzgl. ÿ ÿ } (q \text{ ÿ}, P^{ÿ}) \text{ ÿ D) ÿ } (q, p) \text{ ÿ D.}$$

Die Menge aller Differenzrelationen für **ÿ** wird mit **Dÿ bezeichnet.**

Wir nennen eine Teilmenge Q **ÿ Q** kompatibel **mit einer Differenzrelation** D **ÿ Dÿ , wenn** $?^{ÿ}$ Q = ÿ. **Dies bedeutet, dass sich keine zwei Zustände von** Q in Bezug auf **D unterscheiden dürfen .**

Definition 12. Seien **(ÿ, Q, ÿ, _, _)** ein **SHA** und D eine Differenzrelation für ÿ. Eine Teilmenge von Zuständen Q **ÿ Q** ist für D bzgl. **ÿ subhedge-irrelevant,** wenn **accÿ(Q)** mit D kompatibel ist . Die Menge aller subhedge-irrelevanten Teilmengen von Zuständen für D bzgl. **ÿ** ist also wie folgt:

$$\text{dPrjÿ (D) = \{Q ÿ Q | accÿ (Q)} \quad ^{2} \text{ ÿ D = ÿ\}.}$$

Ein Zustand q **ÿ Q** ist für D subhedge-irrelevant , falls die Singleton-Zustände {q} für D subhedge-irrelevant sind . Die Menge aller subhedge-irrelevanten Zustände von **Q** wird mit Prjÿ (D) bezeichnet.

Wir betrachten Subhedge-Irrelevanz für Teilmengen von Zuständen, da der Kongruenzprojektionsalgorithmus den Nichtdeterminismus, den er einführt, durch eine Art SHA-Determinierung beseitigen muss. Determinierung ist notwendig, um subhedge-irrelevante Präfixe richtig zu erkennen: Alle einzelnen Zustände in einer Teilmenge können subhedge-irrelevant sein, während die gesamte Teilmenge es nicht ist.

7.1.2. Kleinste Differenzen-Relationen

Für jede binäre Relation R **ÿ Q × Q sei** ldrÿ (R) **die kleinste Differenzrelation auf Zuständen** das **R enthält .**

Lemma 8. (p1, p2) ÿ ldrÿ (R) **ÿ** ÿ(q1, **q2) ÿ R.** ÿh **ÿ Hÿ. p1 ÿÿ q1** bzgl . **ÿ ÿ p2 ÿÿ**^{H} **q2** bzgl . ÿ.^{H}

Beweis. Die Menge {(p1, **p2) ÿ Q2 |** ÿ(q1, **q2) ÿ R.** ÿh **ÿ Hÿ. p1 ÿÿ q2 bzgl. ÿ ÿ**^{H} **p2 ÿÿ**^{H} **Differenzrelation ist, offensichtlich die R enthält, und damit die kleinste derartige Differenzrelation ldrÿ** (R). **Umgekehrt muss jedes Paar in der obigen Menge in jeder Differenzrelation enthalten sein ,** die R **enthält , und damit in** ldrÿ (R).
□

Lemma 9. Für jedes R **ÿ Q ×** Q ist die Differenzrelation ldrÿ (R) der Wert des Prädikats D im kleinsten Fixpunkt des Ground-Datalog-Programms, das durch die folgenden Inferenzregeln generiert wird:

$$\frac{\textbf{p,} \text{ q ÿ Q}}{\text{D(p, q) :- R(p, q).}} \qquad \frac{\textbf{p1 aÿÿ p2} \text{ bzgl . ÿ } \textbf{q1 aÿÿ q2} \text{ bzgl . ÿ}}{\text{D(p1, } \textbf{q1) :ÿ} \text{ D(p2, q2).}}$$

$$\frac{\textbf{p1@p ÿ p2} \text{ bzgl . ÿ } \textbf{q1@q ÿ q2} \text{ bzgl . ÿ D(p1,}}{\textbf{q1) :ÿ} \text{ D(p2, q2).}} \qquad \frac{\textbf{p,} \text{ q ÿ Q}}{\text{D(q, p) :- D(p, q).}}$$

Machine Translated by Google

Beweis. Die erste Inferenzregel garantiert, dass R ÿ **D. Die drei späteren Inferenzregeln charakterisieren Differenzrelationen** D ÿ **Dÿ. Die zweite und dritte Regel besagen, dass Differenzen in** D **über Hedges h rückwärts propagiert werden . Dies geschieht rekursiv, indem Buchstaben- Hedges** h = a **nach der zweiten Regel behandelt werden und Baum-Hedges** h = ÿh **die Symmetrie** ÿ **nach der dritten Regel. Die vierte Regel von Differenzrelationen D ausdrücken. Der kleinste Fixpunkt des durch die obigen Regeln generierten Datalog-Programms enthält** also R **und ist die kleinste Relation, die alle Axiome von Differenzrelationen erfüllt, also ist es** ldrÿ (R). □

Proposition 4. Für jedes R ÿ **Q ×** Q kann die kleinste Differenzenrelation ldrÿ (R) in der Zeit **O(|A|)** berechnet werden. ^{2}).

Beweis: Die Größe des Ground-Datalog-Programms, das ldrÿ (R) **aus Lemma 9** berechnet , **ist höchstens quadratisch in der Größe von A, sodass sein kleinster Fixpunkt in der Zeit O(|A|)** berechnet werden kann. ^{2}) □

7.2. Algorithmus

Wir definieren nun den Kongruenzprojektionsalgorithmus durch einen Compiler aus dSHAs und einer Reihe von Schema-Endzuständen zu SHAÿ s, der, wenn er auf einem Hedge ausgeführt wird, alle für das Subhedge **irrelevanten Präfixe erkennen kann .**

7.2.1. Ein- und Ausgaben Als

Eingaben erhält der Kongruenzprojektionsalgorithmus ein dSHA A = **(ÿ, Q, ÿ, I,** F) **für das reguläre Muster und eine Menge FS mit** F ÿ **FS** ÿ **Q. Das dSHA definiert die reguläre Sprache** L = L(A), **während das reguläre Schema S** durch das **dSHA A[F/FS]** = **(ÿ, Q, ÿ, I,** FS) **erkannt wird .**

Derselbe Basisautomat definiert also die reguläre Sprache und das reguläre Schema, indem er unterschiedliche Sätze von Endzuständen auswählt.

Beispiel 10. Im Beispiel **-dSHA** in Abbildung 4 für den XPath- Filter [self::list/child::item] mit Schemadokument haben wir F = **{4}** und **FS** = **{4, 5}.** Im Automaten für das Zählerbeispiel **[child::item]** für die Safe-No-Change-Projektion in Abbildung 12 haben wir F = **{6}** und **FS** = **{5, 6}.**

Wir stellen fest, dass wir, wenn L **und S** durch unabhängige **SHAs gegeben sind ,** einen **gemeinsamen dSHA** A **und eine Menge FS** wie oben **erhalten können, indem wir das Produkt der dSHAs für** L **und S** berechnen **und vervollständigen. Wie in [16] bemerkt, kann es effizienter sein, den SHA für S in einem ersten Schritt zu bestimmen, das Produkt des SHA für** L **mit dem dSHA für S in einem zweiten Schritt zu bilden und dieses Produkt im dritten Schritt zu bestimmen.**

Die Kongruenzprojektion von A **bzgl. FS wird in ihrer aktuellen Konfiguration eine Teilmenge von Zuständen** Q ÿ **Q und eine Differenzrelation von Zuständen in** D ÿ **Dÿ beibehalten. Dabei kann die Kongruenzprojektion dSHAÿ** immer feststellen, ob das aktuelle Präfix für L **bzgl. Schema S subhedge-irrelevant ist, indem sie prüft, ob die aktuelle Menge von Zuständen** Q **subhedge-irrelevant ist . für die aktuelle Differenzrelation D** irrelevant.

7.2.2. Anfängliche Differenzenrelation

Die anfängliche Differenzrelation auf den Zuständen D init ^{A,FS} **ÿ Dÿ ergibt sich aus der Differenz Beziehung auf Präfixe** diff $^{M}_{S}$ **wie folgt:**

$$D^{A,FS}_{\text{init}} = \{(q_{\text{init}}, q') \mid \ddot{y}(v, v') \ddot{y} \text{diff}^{M}_{S}(q_0, \text{hdg}(v)) \ddot{y} \ddot{y}(q_0, \text{hdg}(v'))\ddot{y}\ddot{y}\}.$$

Das nächste Lemma gibt an, wie D aus A **und FS ohne** $^{A,FS}_{\text{Bezug auf}}$ init **definiert werden kann.**

die Sprachen L = L(A) **und S** = **L(A[F/FS]).**

Lemma 10. $D^{A,FS}_{\text{init}}$ = ldrÿ (F **×** (FS \ F)) **ÿ accÿ(I) 2** .

Machine Translated by Google

Beweis. Für eine Richtung sei (p ∈ A,FS , p ∈ ∈) ∈ D init . **Dann gibt es verschachtelte Wörter** (v ∈ , v ") ∈ diff ∈ hdg(v) **und einem Anfangszustand q0 ∈ I, so dass q0** ∈∈∈ p ∈ **bzgl. ∈ und q0 ,** p ∈ ∈) ∈∈∈ p " **bzgl. ∈. Da** ∈ v , v " **verschachtelte Wörter sind, erfordert die Hedge-** ∈ **folgt. Weiterhin** (v ∈ **, Zugänglichkeit** (p ∈∈ ∈) ∈ diff **S die Existenz eines Hedges** h **∈ H∈, so dass** hdg(v hdg(v ∈∈) · h ∈) · h **∈** L **und ∈ S \ L. Daher gibt es Zustände** q ∈ **∈ F und** q , " **∈ FS \ F, so dass** p ∈ h ∈∈ q ∈ **Und** P " H ∈∈ q ∈∈. **Mit Lemma 8 impliziert dies, dass** (p , p ∈∈) ∈ p ‘‘) **∈** ldr∈ (F × (FS \ F)).

Richtung sei (p . **Mit Lemma 8** zeigt , p∈∈) ∈ ldr∈ **(FS × (FS \ F)) ∈ accï(l) 2 für zu ändere** ∈ **FS \ F und ,** p ∈∈) ∈ **Eigentum** (p ∈ **accï(l) 2** gibt ∈ **∈ F,** q h **∈ H∈, so dass** p ∈ h ∈∈ q ∈ **bzgl. ∈ und** p " h ∈∈ q " **bzgl. ∈. Von** (p ∈ **, es folgt dass es verschachtelte Wörter gibt** v ∈, v " **∈ N∈ und einem Anfangszustand q0 ∈ I, so dass q0** hdg(v ∈) ∈∈∈ ' P **bzgl. ∈ und** q ∈ hdg(v **0 ") ∈∈∈∈∈** p " **bzgl. ∈. Daher** (v ∈ **, v ") ∈** diff ∈ **S , so dass** (q ∈ **, q ") ∈** D A,FS init . □

Als Folge von Lemma 10 und Proposition 4 ergibt sich die anfängliche Differenzenrelation D) **von** A A,FS init kann in der Zeit **O(|A|) berechnet werden** 2 **und FS.**

Proposition 5 (Klarheit der anfänglichen Differenzrelation). Sei A = **(∈, Q, ∈, I,** F) ein vollständiges **dSHA,** F **∈ FS ∈** Q, L = L(A) und **S** = **L(A[F/FS]).** Für jeden Hedge h **∈ H∈** und Zustand **∈∈** q bzgl. **∈** für ein **q0 ∈** I ist das verschachtelte **DA,FS** subhedge- H Wort nw(h) für q **∈** Q subhedge-irrelevant , so dass **q0** L und **S** genau dann sind, wenn q für irrelevant ist init bzgl. ∈.

Beweis. Wir zeigen, dass nw(h) **für** L **und S genau dann subhedge-relevant ist, wenn** q **subhedge-relevant ist. relevant für** D bzgl . **∈.init** A,FS

„∈“ Sei nw(h) ein für L **und** S **relevanter Subhedge . Dann gibt es Hedges** H h ∈ **∈ H∈ und ein verschachteltes Wort** w **∈ N∈, so dass**

$$\mathrm{nw}(h) \cdot \mathrm{nw}(h^{\in}) \cdot w \in L \in \mathrm{nw}(h) \cdot \mathrm{nw}(h \text{ ") } \cdot w \in S \setminus L.$$

Da A **deterministisch ist und q0** ∈∈ q **bzgl. ∈ für das eindeutige q0 ∈ I, folgt, dass es Zustände** p **∈ Q,** q **∈ F und** q gibt ∈ **,** P " **∈ FS \ F, so dass**

q H∈ p ∈ hdg(w) ∈∈∈∈ q ∈ **bzgl. ∈ ∈** q **∈∈** H" p " hdg(w) ∈∈∈∈ q " **bzgl. ∈.**

Von q ∈ **∈ F und** q " **∈ FS \ F, folgt, dass** (q ∈ **,** q **") ∈** D A,FS A,FS init . **Da** D **eine Differenz ist** init **Beziehung,** p hdg(w) **∈∈∈∈** q ∈ **bzgl. ∈ und** p " hdg(w) **∈∈∈∈** q " **bzgl . ∈ bedeutet dies, dass** (p ∈ **, p ") ∈** D A,FS init , **auch. Daher** (p ∈ A,FS , p **") ∈** accï({q}) **2 ∈** D init , **also q, ist relevant für** D A,FS init **bzgl. ∈.**

„∈“ Sei q eine für D **relevante Teilhedge bzgl . ∈. Dann existiert ein Paar** (p A,FS **∈** D A,FS init **∈, p ") ∈** akkï({q}) **2** init . **Es gibt also Hedges** h ∈ **und** h **", so dass**

∈ ∈ h ∈∈ pq „ h bzgl. ∈ ∈ q ∈∈ p " **bzgl. ∈.**

Da D init A,FS = ldr∈ (F × (FS \ F)) **nach Lemma 10 gibt es ein verschachteltes Wort** w **und a) gehört zu Paar** (q ∈ **, q ") ∈ Q2 , so dass entweder** (q ∈ **, q ") oder** (q , Q ∈ F × (FS \ F) **und**

P ∈ hdg(w) ∈∈∈∈ q ∈ **bzgl. ∈ ∈** p " hdg(w) ∈∈∈∈ q " **bzgl. ∈.**

Daher zeigt nw(h) · nw(h) , ∈) · w **in** L **und** nw(h) · nw(h **") ·** w **∈ S ** L **oder umgekehrt. Dies** dass nw(h) für L **und S relevant ist .**

□

Machine Translated by Google

7.2.3. Aktualisierung der Differenzrelation

Für jede Differenzrelation D **ÿ Dÿ und jede Teilmenge von Zuständen** Q **ÿ Q definieren wir eine binäre Relation** downÿ Q(D) **ÿ Q × Q, so dass für alle Zustände p1, p2 ÿ Q gilt:**

(p1, p2) ÿ downÿ Q(D) **genau dann, wenn ÿq1, q2 ÿ Q. (q1 ÿ p1, q2 ÿ p2) ÿ D.**

Für jede Teilmenge von Zuständen Q **ÿ Q und Differenzrelation** D **ÿ Dÿ sei DQ ÿ Dÿ die kleinste Differenzrelation, die** downÿ Q(D) enthält:

DQ = ldrÿ (untenÿ Q(D)).

Lemma 11. Die Differenzrelation **DQ** kann in der Zeit **O(|A|)** berechnet werden. ^{2}) von Q, ÿ und D.

Beweis. Die binäre Relation downÿ Q(D) **kann in der Zeit O(|ÿ|)** berechnet werden . ^{2}) **von Q, D und ÿ. Die kleinste Differenzrelation** ldrÿ (downÿ Q(D)) **kann mit dem Ground-Datalog in der Zeit) aus** downÿ Q(D) **O(|A|** ^{2} **unter Verwendung von Proposition 4 berechnet werden.** □

Beispiel 11. Betrachten Sie den **dSHA** für den XPath- Filter [self::list/child::item] in Abbildung 4 mit dem Satz schemafinaler Zustände **FS** = **{4, 5}.** Da F = **{4},** haben wir **FS ** F = **{5}.** Das A,FS_{init}, das die symmetrische Hülle der **({0} × {1, 4, 5}) ÿ** Unterhedges irrelevanten Zustände sind Prjÿ anfänglichen Differenzrelation ist, ist **D0** = D **{(1,** 4),(4, **5)}.** Die für die **(D0)** = **{1, 2, 3, 4, 5} ,** da nur Zustand **0** auf zwei Zustände in der Differenzrelation **D0 zugreifen kann,** den Endzustand in F = **{4}** und einen nicht-finalen Zustand , der schemafinal in **FS ** F = **{5}** einiger Hedges ist. Die Differenzrelation downÿ **{0}** (D) ist der symmetrische Abschluss von **{1, 2} × {3},** die gleich der Differenzrelation D1 = **D0{0}** = ldrÿ $(\text{down}ÿ^{A,FS}_{init})$ Ist **{0}** (D0)) ist. Daher lauten die Projektionszustände Prjÿ **(D1)** = **{2, 3, 4, 5} ,** da man von den Zuständen **0** und **1** immer noch sowohl **1,** als auch **3** erreichen kann, solange **(1,** 3) **ÿ D1.** Die Differenzrelation D2 = downÿ **{1} (D1)** ist der symmetrische Abschluss von **{(1,** 2),(2, **3)}.** Daher lauten die Projektionszustände Prjÿ **(D2)** = **{1, 2, 3, 4, 5}.**

Vom Zustand 1 aus kann man nur die Zustände **1** und 3 erreichen , die sich für **D2 nicht unterscheiden.**

Projektionszustände für die anfängliche Differenzrelation enthalten alle Zustände, die sicher sind für Auswahl oder sichere Ablehnung in Bezug auf das Schema:

Lemma 12. Alle Zustände in safeÿ (F) und safeÿ (FS **** F) sind subhedge-irrelevant für **DA,FS** $_{init}$.

Auf den Beweis verzichten wir, da das Ergebnis später keine Verwendung findet.

Gegeben sei ein dSHA A = **(ÿ, Q, ÿ, I,** F) **und eine Menge** F **ÿ FS ÿ Q. Wir konstruieren nun die Kongruenzprojektion** A cgr(S) **als folgendes dSHAÿ :**

Ein $_{cgr(S)}$ = **(ÿ,** Q_{cgr} **, ÿ** $_{cgr}$**,** $_{Ich\ cgr(S)}$**,** F cgr(S)),

Dabei sind die Zustände, Anfangszustände und Endzustände wie folgt:

Q_{cgr} **ÿ 2 Q × Dÿ**

$_{Ich\ cgr(S)}$ = {(Ich, D)} $^{A,FS}_{init}$

F cgr(S) = (Q, D) **|** $^{A,FS}_{init}$ Q **ÿÿ** dPrjÿ (D $_{init}$ ^{A,FS}) **ÿ** Q **ÿ** F ÿ= **ÿ ÿ**
Q **ÿ** dPrjÿ (D $^{A,FS}_{init}$) **ÿ accÿ(Q) ÿ** F ÿ= **ÿ**

Die Übergangsregeln in ÿ cgr **werden durch die Inferenzregeln in Abbildung 15** bereitgestellt. **Um die Konstruktion zu veranschaulichen und zu erklären, warum Determinierung erforderlich ist, betrachten wir Beispiel 3 noch einmal, bei dem Schemabeschränkungen Konsequenzen haben, die auf den ersten Blick möglicherweise kontraintuitiv sind.**

Machine Translated by Google

$$\frac{a \text{ ÿ ÿ } Q \text{ ÿ } dPrjÿ\ (D)}{(Q, D) \text{ aÿÿ } (Q, D) \text{ in ÿ } cgr} \qquad \frac{Q \text{ ÿ } dPrjÿ\ (D) \text{ ÿÿ}}{(Q, D) \text{ ÿÿ } (Q, D) \text{ in ÿ } cgr}$$

$$\frac{\text{Q@P ÿ Qÿ in ÿ} \quad ^{det} \quad Q \text{ ÿ } dPrjÿ\ (D)}{(Q, D)@(Q, D) \text{ ÿ } (Q, D) \text{ in ÿ } cgr}$$

$$\frac{Q \text{ aÿÿ Qÿ in ÿ} \quad ^{det} \quad Q \text{ ÿÿ } dPrjÿ\ (D)}{(Q, D) \text{ aÿÿ } (Qÿ\ , D) \text{ in ÿ } cgr} \qquad \frac{Q \text{ ÿÿ } dPrjÿ\ (D)}{(F, D) \quad ^{ÿÿ}\text{ÿÿ (ÿÿÿ, } DQ) \text{ in ÿ } cgr}$$

$$\frac{Q@accÿ(P) \text{ ÿ Qÿ in ÿ} \quad ^{det} \quad Q \text{ ÿÿ } dPrjÿ\ (D)\ P \text{ ÿ } dPrjÿ\ (DQ)}{(Q, D)@(P, DQ) \text{ ÿ } (Qÿ\ , D) \text{ in ÿ } cgr}$$

$$\frac{\text{Q@P ÿ Qÿ in ÿ} \quad ^{det} \quad Q \text{ ÿÿ } dPrjÿ\ (D)\ P \text{ ÿÿ } dPrjÿ\ (DQ)}{(Q, D)@(P, DQ) \text{ ÿ } (Qÿ\ , D) \text{ in ÿ } cgr}$$

Abbildung 15. Die Übergangsregeln ÿ cgr **der Kongruenzprojektion** A cgr(S)

Beispiel 12 (Bestimmtheit während der Kongruenzprojektion ist erforderlich.). Betrachten Sie Beispiel 3 noch einmal mit Signatur **ÿ** = **{a},** Muster L = {ÿaÿ} und Schema **S** = L **ÿ** {ÿÿ **· a}.** Diese Sprache kann durch den **dSHA** A in Abbildung 16 definiert werden. Das Schema **S** kann durch denselben Automaten, aber mit den Schema-Endzuständen **FS** = **{3, 5}** und F = **{3} definiert werden.** Der durch Kongruenzprojektion erzeugte **dSHA** A cgr(S) ist in Abbildung 17 dargestellt . Die eindeutige Absicherung ÿaÿ der Sprache L = L(A) wird im Zustand **({3, 4},** D0) akzeptiert , wobei D0 = **{(3,** 5),(5, **3)}.** Der eindeutige Hedge ÿÿ **·** a in **S ** L wird abgelehnt: Nach dem Lesen des Baums ÿÿ geht der Automat in den Zustand **({3, 4},** D0) , wo er beim Lesen des nächsten Buchstabens a zu einer projizierenden Senke **({5},** D0) geht . Wir stellen fest, dass jeder Hedge in **ÿÿ ÿ ÿ** im Zustand **({3, 4},** D0) von A cgr(S) akzeptiert wird , obwohl nur der einzelne Hedge ÿaÿ zu L gehört . Dies ist sinnvoll , da die Hedges in ÿ(ÿ + a **·** a **·** a **ÿ)ÿ** ohnehin nicht zum Schema **S** gehören . Wir stellen fest, dass **({3, 4},** D0) kein Projektionszustand sein kann, da der Lauf auf Hedge ÿÿ **·** a bis zur Senke **({5},** D0) fortgesetzt werden muss; daher ist Zustand **({3, 4},** D0) unterhedge-relevant.

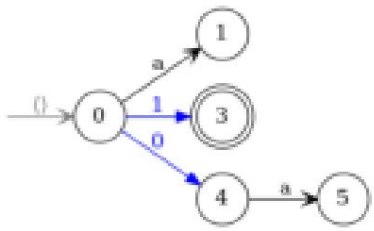


Abbildung 16. Ein dSHA A **für** L(A) = {ÿaÿ} **und Schema-Endzustände FS** = **{3, 5} für das Schema S** = L(A) **ÿ** {ÿÿ **· a}.**

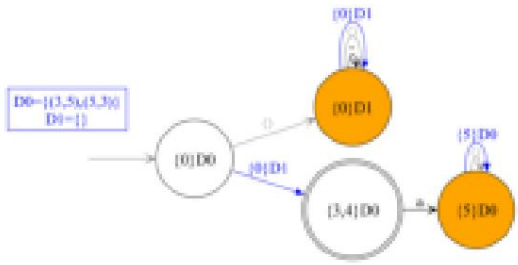


Abbildung 17. Die Kongruenzprojektion A cgr(S) **für das dSHA** A **und die Schema-Endzustände FS in Abbildung 16.**

Machine Translated by Google

Beachten Sie, dass die beiden Einzelzustände **3** und **4** für D0 irrelevant sind, während die Teilmenge mit **ÿ** D0 beide Zustände **{3, 4}** sind für D0 subhedgerelevant, da **(3,** 5) **ÿ** accÿ({3, **4})** die **2** irrelevant ist. Dies zeigt, dass deterministische Konstruktion tatsächlich erforderlich ist, um die Subhedge-Irrelevanz für Fälle wie in diesem Beispiel zu bestimmen. Beachten Sie auch, dass Zustand **0** für D1=ÿ subhedge-irrelevant ist. Das ist in Ordnung, dazu hängt die Akzeptanz von Hedges der ÿ da die by A cgr(S) tatsächlich nicht von h abhängt. Im Gegensatz Form ÿhÿ · h von **hÿ** als dass die Teilmenge der Zustände **{3, 4}** nicht für D0 unterhedge-irrelevant sein kann.

Abschließend wollen wir diskutieren, wie die Übergangsregeln von **dSHAÿ** A cgr(S) hergeleitet werden. Die interessanteste Übergangsregel ist **({0},** D0)@({0}, D1) **ÿ ({3, 4},** D0) in **ÿ** cgr. Sie wird durch die folgende Inferenzregel erzeugt , wobei Q = P = **{0}, accÿ(P)** = **{0, 1, 3, 4, 5},** Q = **{3, 4},** D = D0 und **DQ** = D1:

$$\frac{Q@acc\ddot{y}(P)\ \ddot{y}\ Q\ \ddot{y}\ \text{in}\ \ddot{y}\ \overset{\text{det}}{\ddot{y}}\ Q\ \ddot{y}\ddot{y}\ d\text{Pr}j\ddot{y}\ (D)\ P\ \ddot{y}\ d\text{Pr}j\ddot{y}\ \mathbf{(DQ)}}{(Q, D)@(P, \mathbf{DQ})\ \ddot{y}\ (Q\ \ddot{y}, D)\ \text{in}\ \ddot{y}\ \text{cgr}}$$

Diese Regel besagt, dass, wenn P für **DQ** subhedge-irrelevant ist , alle von P aus erreichbaren Zustände ausprobiert werden müssen , da sie alle von einem anderen Subhedge erreicht werden könnten, das wegprojiziert wurde. Infolgedessen kann man aufgrund der Subhedge-Projektion nicht wissen, in welchen Zustand von Q man gehen soll. Um deterministisch zu bleiben, gehen wir daher in alle möglichen Zustände, d. h. in die ÿ. Im Beispiel Teilmenge Q sind alle Zustände, die erreicht werden können, wenn ein Hedge im Muster {ÿhÿ **|** h **ÿ Hÿ}** gelesen wird , vorausgesetzt , dass die Subhedges h **ÿ ÿ** nicht mit der Subhedge-Projektion untersucht werden.

Beispiel 13 (XPath-Filter [self::list/child::item] **mit Schemadokument).** Die Kongruenzprojektion des Beispiel **-dSHA** in Abbildung 4 ist in Abbildung 18 dargestellt. Dieses **dSHA** ähnelt der Safe- No-Change-Projektion in Abbildung 7, außer dass der nutzlose Zustand **2** entfernt wurde und die vier Projektionszustände nun in sich selbst schleifen, anstatt in einen gemeinsamen Schleifenzustand ÿ zu wechseln. Es sollte auch beachtet werden, dass dort nur Singleton-Zustandssätze verwendet werden, sodass keine Determinierung erforderlich ist. Wie wir sehen werden, ist dies typisch für Experimente mit regulären XPath-Abfragen, bei denen größere Teilmengen selten auftreten.

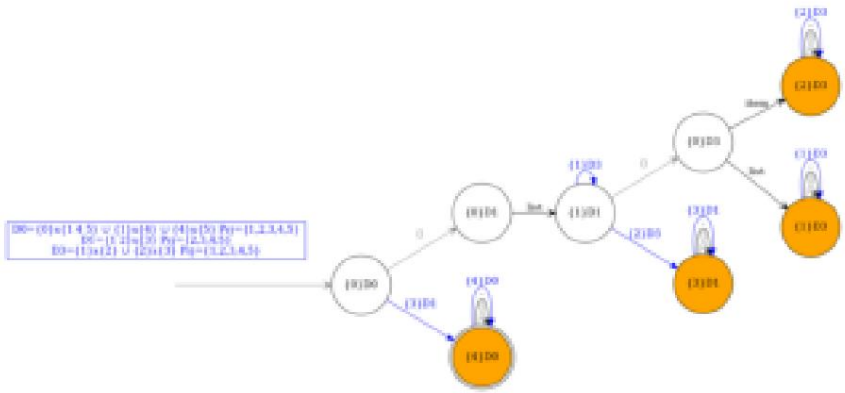


Abbildung 18. Die Kongruenzprojektion A cgr(doc) **des dSHA** A **in Abbildung 4 für den XPath-Filter** [self::list/child::item] **mit Schema** doc. **Subhedge-irrelevante Zustände sind orange eingefärbt. Der Unterstrich steht für eine beliebige Bezeichnung, entweder** Liste **oder** Element.

Beispiel 14 (Gegenbeispiel für die Vollständigkeit der Safe-No-Change-Projektion.). Betrachten wir das Gegenbeispiel für die Vollständigkeit der Safe-No-Change-Projektion noch einmal, d. h. den **dSHA** in Abbildung 12 für die XPath- Abfrage **[child::item]** mit Schema-Endzuständen **FS** = **{0, 5, 6}.** Seine Kongruenzprojektion ist in Abbildung 14 dargestellt . Wir stellen fest, dass das Präfix **ÿitem.ÿitem** zum Zustand {2}D3 führt, der ein Subhedge-Projektionszustand ist, da **2** für D3 subhedge-irrelevant ist. Beachten Sie insbesondere, dass accÿ({2}) = **{2, 4},** sodass keine zwei von **2** aus erreichbaren Zustände in D3 unterschiedlich sind. Dies bedeutet, dass Zustand **2** zwar immer noch in **4** geändert werden kann , dann aber in Bezug auf D3 nicht unterschiedlich wird. Dies löst das Unvollständigkeitsproblem mit der Safe-No-Change-Projektion in diesem Beispiel.

Machine Translated by Google

Die erste **durch die Kongruenzprojektion zugewiesene Eigenschaft für Zustände** (Q, D) besteht darin, dass Q **mit D kompatibel ist . Dies bedeutet, dass sich keine zwei Zustände in** Q in Bezug auf **D unterschei Intuitiv ist jeder Zustand von** Q **so gut wie jeder andere, außer dass er aus dem Schema herausführt. Wenn** also (Q, D) **einem Präfix zugewiesen ist und ein Suffix von einem Zustand in** Q **zu F führt, dann kann es nicht von einem anderen Zustand in Q** zu **FS ** F **führen .**

Lemma 13. Wenn der partielle Lauf von A cgr(S) einem Präfix einen Zustand (Q, D) zuweist , dann ist Q mit D kompatibel .

Wir verzichten auf den Beweis, da dieses aufschlussreiche Ergebnis später nicht direkt verwendet wird. Dennoch wird die Kompatibilität beim Beweis der Korrektheit eine wichtige Rolle spielen.

7.3. Solidität

Als nächstes passen wir das Soliditätsergebnis und den Beweis von der Safe-No-Change-Projektion an die Kongruenzprojektion an.

Theorem 2 (Klarheit der Kongruenzprojektion). Für jedes **dSHA** A = **(ÿ, Q, ÿ, I,** F) und jede Menge F **ÿ FS ÿ** Q bewahrt die Kongruenzprojektion von A in Bezug auf **FS** die Sprache von A innerhalb des Schemas **S** = **L(A[F/ FS]),** d. h.

$$L(A\ cgr(S)\ ) \text{ ÿ } \mathbf{S} = L(A)$$

.

Wir stellen fest, dass S ein Schema für A ist , **da** F **ÿ FS. Darüber hinaus** ist A **schemavollständig für S, da** A **deterministisch ist und S** = **L(A[F/FS]).**

Beweis. Der Beweis der Inklusion L(A) **ÿ** L(A cgr(S)) **ÿ S basiert auf dem folgenden zwei Behauptungen:**

Behauptung 2.1a. Für alle h **ÿ Hÿ,** D **ÿ Dÿ und** Q **ÿ** dPrjÿ (D), (Q, D) **ÿÿ** (Q, D) **bzgl.** ^{H} **ÿ** cgr .

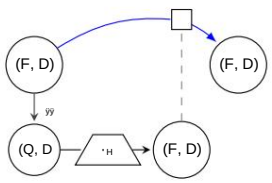
Beweis. Wir beweisen Behauptung 2.1a durch Induktion über die Struktur von h. Angenommen, Q **ÿ** dPrjÿ (D).

Fall h = ÿh ^{ÿ}**ÿ. Die Induktionsvoraussetzung angewendet auf** h^{ÿ} **ergibt** (Q, D) **ÿÿ**^{Hÿ} (Q, D) **bzgl. ÿ** cgr **.**

Wir können daher die folgenden beiden Inferenzregeln anwenden:

$$\frac{Q \text{ ÿ } dPrjÿ\ (D)}{(Q, D) \text{ ÿÿ}^{ÿÿ} (Q, D) \textbf{ in ÿ } cgr} \qquad \frac{Q \text{ ÿ } dPrjÿ\ (D)}{(Q, \textbf{D))@(Q,}\ D) \text{ ÿ } (Q, D) \textbf{ in ÿ } cgr}$$

um das folgende Diagramm bezüglich ÿ cgr zu schließen:



Dies beweist (Q, D) **ÿÿ** ^{H}(Q, D) **bzgl. ÿ** cgr, **wie es die Behauptung verlangt.**

Fall h = **a. Da** q **ÿ** dPrjÿ (D), **können wir die Inferenzregel anwenden:**

$$\frac{a \text{ ÿ } \mathbf{ÿ} \ Q \text{ ÿ } dPrjÿ\ (D)}{(Q, D) \textbf{ aÿÿ } (Q, D) \textbf{ in ÿ } cgr}$$

Dies beweist die Richtigkeit dieser Behauptung.

Machine Translated by Google

Fall h = ÿ. **Wir haben trivial** (Q, D) ÿÿÿ (Q, D) **bzgl. ÿ** cgr .
Fall h = h ÿ · H" **. Nach der Induktionsvoraussetzung, angewendet auf** ÿ **und** h **" erhalten wir Folgendes:**

(Q, D) ÿÿ (Q, D) **und** (Q, D) **ÿ** cgr ÿÿ (Q, D) **bzgl. ÿ** cgr. **Daher** (Q, D) ÿÿÿ (Q, D) **bzgl.** .

Damit ist der Beweis von Behauptung 2.1a beendet. Die nächste Behauptung ist der Schlüssel zum Beweis der Stichhaltigkeit.

Für jede Differenzrelation D definieren wir eine binäre Relation depa (D) **ÿ 2 Q × 2 Q, so dass für zwei beliebige Teilmengen von Zuständen** ÿ **Q" dass ÿ Q,**

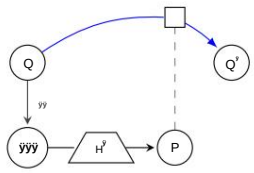
(Q ÿ **, Q ") ÿ** depa (D) **ÿ** Q ÿ **ÿÿ** dPrjÿ (D) **ÿ** Q **ÿ** ÿ **ÿ** Q " ÿ **(1**
Q ÿ dPrjÿ (D) **ÿ** (Q ÿ **ÿ accÿ(Q ") ÿ** Q " a) **ÿ** dPrjÿ (D)) **(2** a)

Beachten Sie außerdem, " **ÿ** dPrjÿ (D) **ÿ** Q ÿ **ÿ accÿ(Q ") impliziert** Q ÿ **ÿ** dPrjÿ (D), **und daher** (Q ÿ **, dass** QQ **") ÿ** depa (D).

Behauptung 2.2a. Sei h **ÿ Hÿ eine Hecke,** ÿ **ÿ Q Teilmengen von Zuständen und** D **ÿ Dÿ eine Differenz bezüg-**
Wenn Q ÿÿ ÿ **Q,** Q **bzgl. ÿ** det, **dann existiert** Q" **ÿ Q, so dass** (Q, D) **ÿÿ** (Q " **,** D) **bzgl. ÿ** cgr
und (Q ÿ **,** QQ **") ÿ** depa (D).

Beweis. Wenn Q **ÿ** dPrjÿ (D), **dann** (Q, D) ÿÿ (Q, D) **bzgl. ÿ** cgr **nach Anspruch 2.1a.** " **= Q. Wir**
und Q **dann** ÿ **und ÿ accÿ(Q ÿÿ)** **Sei** Q **ÿ** dPrjÿ (D) **und damit** Q **ÿ** dPrjÿ (D) , **so dass (1a)**
Q **(2a) von,** (QQ **ÿÿ) ÿ** depa (D) haben. **Es genügt also, die Behauptung unter der Annahme zu**
beweisen, dass Q **ÿÿ** dPrjÿ (D). **Der Beweis erfolgt durch Induktion über die Struktur von h.**

Fall h = ÿh ÿ h ÿ. **Die Annahme** Q ÿÿ Q ÿ det **zeigt, dass es eine Teilmenge von Zuständen bzgl.** ÿ gibt
P **ÿ Q schließt das folgende Diagramm:**



Insbesondere gilt Q@P ÿ Q. **Daraus** ÿ **in ÿ** det. **Sei Dÿ** = **DQ. Da** Q **ÿÿ** dPrjÿ (D) ,
lässt sich Folgendes folgern:

$$\frac{Q \text{ ÿÿ } dPrjÿ\ (D)}{(F, D) \quad ÿÿ\ (ÿÿÿ, Dÿ) \text{ in } ÿ \text{ cgr}}$$

Wir müssen zwei Fälle unterscheiden, je nachdem, ob ÿÿÿ zu dPrjÿ (Dÿ) gehört
oder nicht.

Unterfall ÿÿÿ ÿÿ dPrjÿ (Dÿ). **Die Induktionsvoraussetzung angewendet ergibt die Existenz**
von P ÿ **auf** h **ÿ Q, so dass** (P, P ÿ) **ÿ** depa (Dÿ) **und**

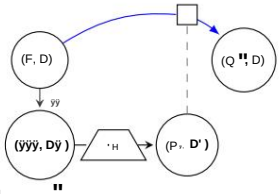
(ÿÿÿ , D ÿ) ÿÿ (P ÿ, D ÿ) **bzgl. ÿ** cgr.

Unterunterfall P **ÿÿ** dPrjÿ (Dÿ). **Da** (P, P **ÿÿ** ÿ) **ÿ** depa (Dÿ), **so folgt, dass** P **ÿ** P **in ÿ** ÿ.
Daher ÿ dPrjÿ (Dÿ). **Sei** Q. " **ÿ Q so, dass Q@P** ÿ **ÿ** F " det. **Wir**
kann P **dann die folgende Inferenzregel anwenden:**

$$\frac{\text{Frage @P}^{ÿ}\ ÿ\ F\ \text{"}\ \text{in } ÿ\ \ \text{det}\ Q\ ÿÿ\ dPrjÿ\ (D)\ P\ ÿ\ ÿÿ\ dPrjÿ\ (Dÿ\ )}{(Q, \textbf{D})@(P\ ÿ, Dÿ\ )\ ÿ\ (Q\ \text{"}, D) \text{ in } ÿ \text{ cgr}}$$

Machine Translated by Google

Daher können wir das Diagramm wie folgt schließen:

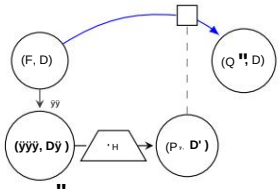


Dies zeigt, dass (Q, D) ÿÿ (Q , D) **bzgl. ÿ** cgr. **Da** P **ÿ** P **in ÿ , Q@P ÿ** Q **, Und Q@P ÿ ÿ** F det, **folgt, dass** Q **ÿ** Q **'', und damit** (Q **,** Q **'') ÿ** depa (D). **Dies zeigt die Behauptung in diesem Fall.**

Unterunterfall P **ÿ** dPrjÿ (Dÿ). **Da** (P, P) **ÿ** depa (Dÿ), **folgt, dass** P **ÿ ÿ** dPrjÿ **accÿ(P ÿ) und** (Dÿ). **Daher können wir die folgende Schlussfolgerung anwenden** P- **Regel für ein** Q **ÿ Hÿ:**

$$\frac{\text{Q@accÿ(P ÿ) ÿ Q}'' \text{ in ÿ det Q ÿÿ dPrjÿ (D) P ÿ ÿ dPrjÿ (Dÿ)}}{\text{(Q, D)@(P ÿ Dÿ) ÿ (Q}'', \text{D) in ÿ cgr}}$$

Daher können wir das Diagramm wie folgt schließen:

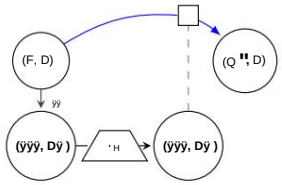


Dies zeigt, dass (Q, D) ÿÿ (Q **und** , D) **bzgl. ÿ** cgr. **Da** P **ÿ accÿ(P ÿ), folgt ÿ Q@P ÿ** QQ **ÿÿ) ÿ** Q@accÿ(P **ÿ) ÿ** Q **aus in ÿ** det **dass** Q Q **ÿÿ. Somit gilt** (Q , depa (D), **also gilt die Behauptung auch in diesem Fall.**

Unterfall ÿÿÿ ÿ Prjÿ (Dÿ). **Behauptung 2.1a zeigt, dass (ÿÿÿ, Dÿ) ÿÿ (ÿÿÿ, Dÿ) bzgl. ÿ** cgr. **Sei** Q **so, dass Q@accÿ(ÿÿÿ) ÿ Qÿÿ in ÿ** det. **Wir können die folgende Schlussfolgerung anwenden -Regel:**

$$\frac{\textbf{Q@accÿ(ÿÿÿ) ÿ Q}'' \text{ in ÿ det Q ÿÿ dPrjÿ (D) ÿÿÿ ÿ Prjÿ (Dÿ)}}{\text{(Q, D)@(ÿÿÿ, Dÿ) ÿ (Q}'', \text{D) in ÿ cgr}}$$

Damit können wir das folgende Diagramm bezüglich ÿ cgr wie folgt schließen:



Dies zeigt, dass (Q, D) ÿÿ (Q , D) **bzgl. ÿ** cgr. **Da** P **ÿ accÿ(ÿÿÿ), folgt Q@P ÿ** Q (QQ **'') ÿ und Q@accÿ(ÿÿÿ) ÿ** Q **aus in ÿ** det , **dass** Q **ÿ** Q **'', und damit** depa (D).

Fall h = **a. Da** Q **ÿÿ** dPrjÿ (D), **können wir die Inferenzregel anwenden:**

$$\frac{\text{Q aÿÿ Q in ÿ det Q ÿÿ dPrjÿ (D)}}{\text{(Q, D) aÿÿ (Q , D) in ÿ cgr}}$$

Daher ist (Q, D) ÿÿ (Q , D) **bzgl. ÿ** cgr. **Mit** Q = Q **, die Behauptung folgt, da** (Q , Q **'') ÿ** depa (D).

Fall h = **ÿ. In diesem Fall gilt** Q = Q. **und** (Q, D) **ÿÿÿ** (Q, D) **bzgl. ÿ** cgr, **also die Behauptung**

Machine Translated by Google

Fall h = **h1 · h2. Da** Q **h2 Q1 ÿÿ** **ÿÿ** Q **h1 bzgl. ÿ** det, **existiert Q1 ÿ Q, so dass** Q **ÿÿ Q1 und bzgl. ÿ** det. Nach der auf Q, **so dass** **h1 angewandten Induktionsvoraussetzung existiert** Q **ÿ Q,** (Q, D) **ÿÿ** (Q **1 ,** D) **bzgl. ÿ** cgr **und (Q1,** Q**) ÿ** depa (D).

Unterfall Q1 ÿ dPrjÿ (D). **Da (Q1,** Q**) ÿ** depa (D), **folgt, dass Q1 ÿ accÿ(Q ÿ** **und** Q **ÿ** dPrjÿ (D). **Weiterhin zeigt** Behauptung **2.1a ,** dass (Q **ÿ** **ÿÿ** (Q **1 ,** D), **und daher** (Q, D) **ÿÿ** (Q **ÿ** **1 ,** D) **bzgl. ÿ** cgr. **Da** Q accÿ(Q1) **und Q1 ÿ accÿ(Q ÿ** **),** **folgt, dass** Q **accÿ(Q ÿ** **). Daher** (Q **ÿ** **ÿ**) **ÿ** depa (D) **und** (Q, D) **ÿÿ** (Q **1 ,** D).

Unterfall Q1 ÿÿ dPrjÿ (D). **In diesem Fall können wir die Induktionsvoraussetzung anwenden auf** **Q1 ÿÿ** Q **bzgl. ÿ** det, was zeigt, dass es Q **gibt** **ÿ Q, so dass** (Q **1 ,** D) **ÿÿ** (Q **,** D) **bzgl. ÿ** cgr **und** **, Q ") ÿ** depa (D). **Daher** (Q, D) **ÿÿ** (Q **,** D) **bzgl. ÿ** cgr **und** (Q **,** (QQ **") ÿ** depa (D).

Damit ist der Beweis von Behauptung 2.2a beendet. □

Beweis der Inklusion L(A) **ÿ** L(A cgr(S)) **ÿS. Da S ein Schema für A ist, haben wir** L(A) **ÿ S, so dass es genügt,** L(A) **ÿ** L(A cgr(S)) **zu zeigen. Sei** h **ÿ** L(A). **Dann existieren q0 ÿ** I **und ÿÿ** q **bzgl. ÿ. Sei** Q = **{q}. Da** A **deterministisch ist, folgt, dass** I q **ÿ F, so dass q0 ÿÿ** Q **bzgl. ÿ** det. **Unter Verwendung** A,FS) **bzgl. ÿ** cgr. **Weiterhin** (I, D) **Damit Behauptung 2.2a impliziert dies die Existenz einer Teilmenge von Zuständen** Q **ÿ Q, so dass** A,FS (Q, Q A,FS) **ÿ** depa (D) **und** (I, D)init **ÿÿ** (Q **,** init) **ÿ** Ich cgr(S)**. Um** h **ÿ** L(A cgr(S)) **zu beweisen, genügt es also zu zeigen, dass** (Q **zwei** **,** D A,FS init) **ÿ** F cgr(S)**. Wir Fälle unterscheiden:**

Fall Q **ÿÿ** dPrjÿ (D **Bedingung (1a) von** (Q , Q **) ÿ** depa (D), **zeigt, dass** Q **ÿ** Q) **ÿ** **so dass** q **ÿ** Q **. Da** q **ÿ F, folgt, dass** Q **ÿ** F ÿ= **ÿ. Somit** (Q **,** D A,FS init F cgr(S) **, so dass** h **ÿ** L(A cgr(S)).

Fall Q **ÿ** dPrjÿ (D **Bedingung (2a) von** (Q, Q A, FS D init **) ÿ** depa (D) **ergibt** Q **ÿ accÿ(Q ÿ). Daher** q **ÿ accÿ(Q ÿ) ÿ** F , **so dass** (Q, **) ÿ** F cgr(S)**, und somit** ist h **ÿ** L(A cgr(S)).

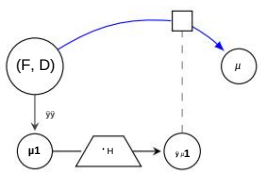
Damit ist der Beweis der ersten Inklusion beendet.

Als nächstes wollen wir die inverse Inklusion L(A cgr(S)) **ÿ S ÿ** L(A) **zeigen . Es wird schließlich ergeben sich aus den nächsten beiden Behauptungen.**

Behauptung 2.1b. Für jede Absicherung h **ÿ Hÿ, Differenzrelation** D **ÿ Dÿ, Projektionszustand** Q **ÿ** dPrjÿ (D) **und Zustand** μ **ÿ** Qcgr gilt: **wenn** (Q, D) **ÿÿ** μ **bzgl. ÿ** cgr, **dann** μ = (Q, D).

Beweis. Mittels Induktion über die Struktur von h **ÿ Hÿ nehmen wir an, dass** (Q, D) **ÿÿ** μ **bzgl. ÿ** cgr:

Fall h = ÿh **ÿ. Es müssen die Zustände µ1,** μ_1 **ÿ** Qcgr **schließt das folgende Diagramm:**



Da Q **ÿ** dPrjÿ (D), **muss die folgende Regel angewendet worden sein, um** (Q, D) **µ1 bzgl. ÿ** cgr zu folgern:

$$\frac{Q \text{ ÿ dPrjÿ (D)}}{(Q, D) \text{ ÿÿ } (Q, D) \text{ in ÿ cgr}}$$

Machine Translated by Google

h **Daher** ist **μ1** = (Q, D). **Die Induktionsvoraussetzung gilt auch für** (Q, D) **bzgl.** ÿ cgr ÿÿ μ **1** = (Q, D) . **Daher** muss μ **durch Anwendung der Regel abgeleitet worden sein: zeigt, dass** μ

ÿ

ÿ

1

$$\frac{\text{Q ÿ dPrjÿ (D)}}{\text{(Q, D)@(Q, D) ÿ (Q, D) in ÿ cgr}}$$

Daher ist μ = (Q, D), **wie erforderlich.**

Fall h = **a. Es gilt folgende Regel:**

$$\frac{\text{a ÿ ÿ Q ÿ dPrjÿ (D)}}{\text{(Q, D) aÿÿ (Q, D) in ÿ cgr}}$$

Daher ist μ = (Q, D).

Fall h = **ÿ. Offensichtlich.**

h1 h2 Fall h = **h1 · h2.** Es muss ein **μ1 geben , so dass** (Q, D) ÿÿ μ **bzgl.** ÿ cgr. **Wenden wir die Induktionsvoraussetzung ÿÿ μ1** auf **h1 an, haben wir μ1** (Q, D).

Wir können also die Induktionsvoraussetzung auf h2 anwenden **und erhalten μ2** = (Q, D).

Damit ist der Beweis von Anspruch 2.1b beendet. □

Als nächstes benötigen wir eine Umkehrung von Anspruch 2.2a. Um Läufe A cgr **wieder mit Läufen von A in Beziehung zu setzen, definieren wir für jede Differenzrelation** D **eine weitere binäre Relation** depb (D) ÿ **2 Q × 2 Q, so dass für zwei beliebige Teilmengen von Zuständen** Q ÿ **Q gilt:**

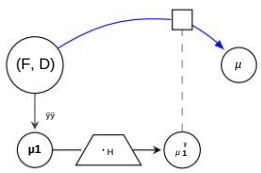
$$(Q \text{ ÿ}, Q \text{ ''}) \text{ ÿ depb (D) ÿ} \quad \begin{array}{ll} \text{ÿ} & (Q \times Q \text{ ÿÿ}) \text{ ÿ D = ÿ ÿ ÿÿ dPrjÿ} \quad (0^{B}) \\ \text{ÿÿ} & Q\text{ÿ (D) ÿ Q ÿ Q ÿ (1 ÿ dPrjÿ (D)ÿÿ Q ÿ} \quad ^{B}) \\ \text{ÿÿ} & Q\text{ÿ accÿ(Q ÿ) (2 ''} \quad ^{B}) \end{array}$$

Behauptung 2.2b. Sei Q ÿ **Q eine Teilmenge von Zuständen, die mit einer Differenzrelation** D ÿ **Dÿ kompatibel ist. Für jede Absicherung** h ÿ **Hÿ und jeden Zustand** μ ÿ Qcgr **mit** (Q, D) ÿÿ μ **bzgl.** ÿ cgr **existiert ein Paar von** ÿ**,** Q **'')** ÿ depb (D), **so dass** μ = (Q ÿ**,** D) **und** Q ÿÿ Q '' **bzgl.** ÿ **.**

Teilmengen von Zuständen (Q **Beweis. Wenn** Q ÿ dPrjÿ (D), dann **zeigt Behauptung** = Q **und** Q **'' Sei die eindeutige Teilmenge der Zustände, so dass** Q ÿÿ Q **2.1b ,** dass μ = (Q, D). **Sei** Q **bzgl.** ÿ det. **Wir müssen zeigen** depb (D). **Da** Q ÿÿ Q **'', haben wir** Q ÿ **accÿ(Q ''), also gilt Bedingung (2b). Bedingung (1b) gilt trivial, da** Q ÿ dPrjÿ (D). **Für Bedingung (0b) ist zu beachten, dass** Q ÿ dPrjÿ (D) **impliziert, dass accÿ(Q) 2** ÿ **accÿ(Q) 2 ×** Q **'')** ÿ D = ÿ**, wenn** ÿ D = ÿ**. Weiterhin** ist Q **erforderlich, so dass** (Q

Wir können also annehmen, dass Q ÿÿ dPrjÿ (D). **Der Beweis erfolgt durch Induktion über die Struktur von** h ÿ **Hÿ. Wir unterscheiden alle möglichen Formen von Hedges** h ÿ **Hÿ:**

Im Fall h = ÿh ÿ**ÿ. Per Definition von** (Q, P) ÿÿ μ **bzgl.** ÿ cgr **muss es μ1,** μ geben ÿ **1** ÿ Qcgr, **so lässt sich folgendes Diagramm schließen:**



Machine Translated by Google

Da Q ÿÿ dPrjÿ (D) ist, **wurde die folgende Inferenzregel angewendet, um** (Q, D) **ÿÿ µ1 bzgl. ÿ** cgr zu folgern, **wobei Dÿ** = **DQ:**

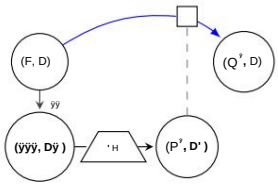
$$\frac{Q\ \ddot{y}\ddot{y}\ dPrj\ddot{y}\ (D)}{(F, D)\quad \ddot{y}\ddot{y}\ (\ddot{y}\ddot{y}\ddot{y}, D\ddot{y})\ \text{in}\ \ddot{y}\ cgr}$$

Daher ist **µ1** = **(ÿÿÿ, Dÿ). Wenn ÿÿÿ ÿÿ** dPrjÿ (D), **dann gilt die Induktionsvoraussetzung für (ÿÿÿ, Dÿ)**, h ÿÿ µ 1 **bzgl. ÿ** cgr **zeigt, dass es** (P , P **") ÿ** depb (D'), **so dass** 1 (P µ , **Dÿ) und ÿÿÿ** = h ÿÿ P **" bzgl. ÿ** det. **Ansonsten kann das gleiche gefolgert werden wie wie eingangs argumentiert. ÿ**

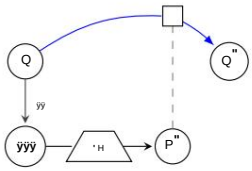
Unterfall P. dPrjÿ (Dÿ). **Die Übergangsregel** (Q, **D)@µ1 ÿ** µ muss **für** Q **wie folgt gefolgert werden ÿ F:**

$$\frac{Q@accÿ(P\ \ddot{y}\ )\ \ddot{y}\ Q\quad \text{in}\ \ddot{y}\quad \text{det}\qquad Q\ \ddot{y}\ddot{y}\ dPrj\ddot{y}\ (D)\ P\quad \ddot{y}\ \ddot{y}\ dPrj\ddot{y}\ (D)}{(Q, \mathbf{D})@(P\ , D\ddot{y}\ )\ \ddot{y}\ (Q\ , D)\ \text{in}\ \ddot{y}\ cgr}$$

Dies zeigt, dass µ = (Q , D). **Wir haben also das folgende Diagramm:**



Lassen " sei die einzige Teilmenge der Zustände, so dass Q@ÿ ÿ F **" bzgl. ÿ** det. **Wir können Sie** Q **dann das folgende Diagramm schließen:**

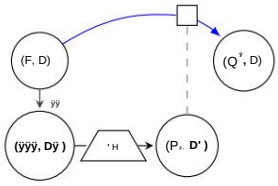


(PD = **ÿ,** , P **ÿÿ) ÿ** depb (Dÿ), **folgt** (P **Aus** **× P ") ÿ D'** = **ÿ, und daher** (QQ **") ÿ** **× Q ") ÿ d. h. Bedingung (0b) von** (Q **tion** , depb (D). **Da** P **ÿ** dPrjÿ (D'), **ist die Bedingung ÿ accÿ(P (2b) von** (PP **ÿÿ), ÿ** depb (Dÿ) **ergibt sich** P **"). Weiterhin ist Q@P in ÿ** det, was **ÿ** F **und** es folgt Q@accÿ(P Q **ergibt , so dass die Bedingungen (1b) und ÿ** QQ **") ÿ** depb (D) **ÿ) ÿ** Q **(2b), von** (QQ **ÿÿ) ÿ** depb (D) . **Daher** (Q ÿh , sind. **Zusammenfassend können wir sagen, Zeigen Sie, dass** µ = , D), Q **ÿÿÿ** Q **" bzgl. ÿ** det, **und** (Q , Q **") ÿ** depb (D), **je nach Bedarf** (Q, **wie**

Unterfall P. **behauptet. ÿÿ** dPrjÿ (Dÿ). **Die Übergangsregel** (Q, **D)@µ1 ÿ** µ muss daher für Q **wie folgt gefolgert werden ÿ F:**

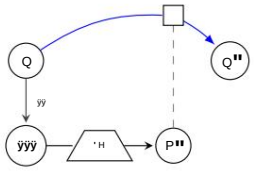
$$\frac{\text{Frage @P}\ \ddot{y}\ \ddot{y}\ F\quad \ddot{y}\ \text{in}\ \ddot{y}\quad \text{det}\qquad Q\ \ddot{y}\ddot{y}\ dPrj\ddot{y}\ (D)\ P\quad \ddot{y}\ \ddot{y}\ddot{y}\ dPrj\ddot{y}\ (D)}{(Q, \mathbf{D})@(P\ , D\ddot{y}\ )\ \ddot{y}\ (Q\ , D)\ \text{in}\ \ddot{y}\ cgr}$$

Dies zeigt, dass µ = (Q , D), **und wir können das folgende Diagramm schließen:**



Machine Translated by Google

Bedingung (0b) von (P, P **") ÿ** depb (D') **ergibt** (P × P **") ÿ D' = ÿ. Sei** Q **bzgl. ÿ sei der eindeutige Teilmenge von Zuständen, so dass ÿ** F det. **Da D' ÿ** downÿ Q(D) **dass × P ") ÿ D' = ÿ, folgt aus Q@P ÿ** Q **ÿ** Q **") ÿ** D = **ÿ. und Q@P ÿ** F **bzgl. ÿ Q@P und Das heißt, Bedingung (0b) von** (QP **") ÿ** depb (D') **, Q") ÿ** depb (D). **Da** P (P det **dass** (Q **ÿÿ** dPrjÿ (Dÿ), **Bedingung (1b) von** (P **ÿ** P **. Aus Q@P ÿ** Q **und Q@P ÿ** F **, daraus folgt, dass** Q **ÿ . Daher sind die Bedingungen (1b) und (2b) von** (Q Q **") ÿ** depb (D) **gelten, so dass** (Q Q **") ÿ** depb (D) **gilt. Außerdem,**



Dies zeigt, dass Q *μ* ÿh ÿ **ÿÿÿ** Q **bzgl. ÿ. Die beiden anderen Voraussetzungen des Anspruchs,** = (Q , D) **und** (Q , Q **") ÿ** depb (D), **wurden bereits früher gezeigt.**

Fall h = **a. Da** Q **ÿÿ** dPrjÿ (D), **muss die folgende Inferenzregel verwendet werden:**

$$\frac{Q \text{ aÿÿ } Q' \text{ in ÿ } Q \text{ ÿÿ dPrjÿ (D)}}{(Q, D) \text{ aÿÿ } (Q', D) \text{ in ÿ cgr}}$$

Also *μ* = (Q , D), Q **aÿÿ** Q **bzgl. ÿ** det. **Da** Q mit **D kompatibel ist** und) **ÿ** D = **ÿ,** D **ist eine Differenzrelation,** Q ist auch mit D **kompatibel . Daher** zeigt (Q **die × Q** (Q **ÿ** Q) **ÿ** depb (D) **gilt ebenfalls. Daher ÿ accÿ(Q ÿ) Bedingung (0b) von** (Q) **ÿ** depb (D). **Trivialerweise erfüllt Q die Bedingungen (1b) und (2b) von , Q (Q , Q so) ÿ** depb (D). (Q **Fall** h = **ÿ. Offensichtlich.**

Fall h = **h1 · h2. Da** (Q, D) **ÿÿ** *μ* **bzgl. ÿ** cgr, **existiert ein μ1 ÿ** Qcgr, **so dass h1 h2** (Q, D) **ÿÿ μ1 und μ1 ÿÿ** *μ* **bzgl. ÿ** cgr. **Wir können die Induktionsvoraussetzung auf h1 anwenden.**

Daher gibt es Teilmengen von Zuständen h1 ÿ Q, so dass μ1 = (Q1, D), Q **ÿÿ** Q **bzgl . Q1,** Q **ÿ und (Q1, Q) ÿ** depb (D). **Wir unterscheiden zwei Fälle:**

h2 Unterfall Q1 ÿ dPrjÿ (D). **Da μ1 = (Q1,** D) **ÿÿ** *μ* **bzgl . ÿ** cgr , **zeigt Behauptung 2.1b ÿÿ (Q1,** D) **bzgl. ÿ** cgr. **dass** *μ* = **(Q1,** D), **so dass (Q1,** D) **Bedingung (2b) von ÿ** accÿ(Q1). **Sei Q2 (Q1,** Q **1) ÿ** depb (D) **und Q1 ÿ** dPrjÿ (D) **implizieren, dass** Q **eindeutige Teilmenge der Zustände, so dass** Q , **ÿÿ Q2 bzgl. ÿ** det. **Dann** ist **Q2 ÿ accÿ(Q ÿ die 1) sodass Q2 ÿ** accÿ(Q1) **und somit Bedingung (2b) von (Q1, Q2) ÿ** depb (D) **gilt. Bedingung (1b) von (Q1, Q2) ÿ** depb (D) **ist trivial, da Q1 ÿ** dPrjÿ (D). **Da Q1 ÿ** dPrjÿ (D) **und Q1 × Q2 ÿ accÿ(Q) 2, folgt, dass Q1 × Q2 ÿ** D = **ÿ , also gilt auch Bedingung (0b) von (Q1, Q2) ÿ** depb (D) . **Daher gilt** Q **(Q1, Q2) ÿ** depb (D). **ÿÿ Q2 bzgl . ÿ** det **und**

Unterfall Q1 ÿÿ dPrjÿ (D). **Da Q1 Relation h1 ÿÿ Q1 und Q1 ist kompatibel mit Differenz ist auch D, folgt, dass** Q **1 kompatibel mit D. Wir können also die h2 -Induktionshypothese auf (Q1,** D) **ÿÿ μ anwenden, was die Existenz von Teilmengen von Zuständen h2) ÿ** depb (D) zeigt, **(Q2,** Q **2 so dass** *μ* = **(Q2,** D) **und Q1 ÿÿ** Q **2 bzgl . ÿ. Wir haben also** Q **ÿÿ** Q **2 bzgl. ÿ und (Q2,** Q **2) ÿ** depb (D), **falls erforderlich.**

Damit ist der Beweis von Anspruch 2.2b beendet. □

Beweis der Inklusion L(A cgr(S)) **ÿ S ÿ** L(A). **Sei** h **ÿ** L(A cgr(S)) **ÿ S. Dann gibt es eine letzte Teilmenge von Zuständen** Q **ÿ** F cgr(S) , **so dass** (I, D) **ÿÿ** (Q, D) **bzgl. ÿ cgr . Da** I

Machine Translated by Google

ein Singleton oder leer, es ist kompatibel mit D init . **Behauptung 2.2b zeigt, dass es a) ÿ** depb (D A,FS
Teilmenge der Zustände Q **ÿ Q, so dass** ich h ÿÿ Q **bzgl. ÿ** det **und** (Q, Q Zustand
(0b) von (Q, Q) **ÿ** depb (D A,FS) **zeigt, dass** (Q, Q A,FS) ÿ D = **ÿ. Da** h **ÿ S ist, folgt, dass es sich**
Q **ÿ FS** ÿ= **ÿ. Der Determinismus von** A **zeigt, dass** Q = {q **ÿ}.** **um ein Singleton handelt. Es gibt also einen Zustand**
Q **dass** Q **ÿ FS, so**

Fall Q ÿÿ dPrjÿ (D A,FS **). Bedingung (1b) zeigt, dass** QÿÿQ , **so dass** q **ÿ Q. Da** Q **ÿ** F cgr **,**
Q **ÿ** F ÿ= ÿ **ist . Da** Q mit D **ÿ Q kompatibel ist** und Q **ÿ** F ÿ= **ÿ,** folgt init , q , **dass** q **ÿ F und damit** h **ÿ** L(A).
ÿÿ FS \ F. Da q **ÿ FS, dies impliziert** q

Fall Q **ÿ** dPrjÿ (D **ÿ** accÿ(Q) A,FS **). Bedingung (2b) zeigt, dass Qÿ ÿ accÿ(Q) , so dass** q
Da Q **ÿ** F cgr, **haben wir accÿ(Q) ÿ** F ÿ= **ÿ. Sei** q **ÿ accÿ(Q) ÿ** F **beliebig. Da** Q **ÿ** dPrjÿ (D
A,FS) **und** (q , q) **ÿ accÿ(Q) 2 , daraus folgt, dass** (q, q A,FS) ÿÿ D . **Außerdem ist ÿ**
q **ÿ** F, **so dass** qh **ÿ ÿÿ FS \ F. In Kombination mit** q **ÿ FS, dies impliziert** q **FS so**
L(A).

Damit ist der Beweis der inversen Inklusion und damit von L(A) = L(A cgr(S)) **ÿ S** beendet.

7.4 Vollständigkeit

Als nächstes zeigen wir die Vollständigkeit der Kongruenzprojektion für die Subhedge-Projektion gemäß Definition 10. Seien A = **(ÿ, Q, ÿ, I,** F) **ein vollständiges dSHA und** F **ÿ FS ÿ Q.**
Automat A **definiert das reguläre Muster** L = L(A) **mit den Schemafinalzuständen im FS, dem regulären Schema S** = **L(A[F/FS]). Wir zeigen zunächst, dass alle subhedge-irrelevanten Zustände von** Differenzrelationen **gemäß Definition** 7 **Subhedge-Projektionszustände** A cgr(S) sind .

Lemma 14. Wenn Q **ÿ** dPrjÿ (D), dann ist (Q, D) ein Subhedge-Projektionszustand von **ÿ** cgr mit Zeuge (Q, D).

Beweis. Wir nehmen an, dass q **ÿ** Prjÿ (D) **und müssen zeigen, dass** (Q, D) **ein Subhedge-Projektionszustand von ÿ** cgr ist. **Wir müssen zeigen, dass alle Übergangsregeln, die mit** (Q, D) beginnen , **für einen Subhedge-Projektionszustand** (Q, D) **mit** (Q, D) **zulässig sind . Sie werden durch die folgenden Regeln erzeugt , die alle Schleifen bilden:**

a **ÿ ÿ** Q **ÿ** dPrjÿ (D)	Q **ÿ** dPrjÿ (D)	Q **ÿ** dPrjÿ (D)
(Q, D) **aÿÿ** (Q, D) **in ÿ** cgr	(Q, D)@(Q, D) **ÿ** (Q, D) **in ÿ** cgr	(Q, D) **ÿÿ** (Q, D) **in ÿ** cgr

□

Wenn also ein partieller Lauf auf einem Präfix u **weist einen Zustand** (P, D) **mit** P **ÿ** dPrjÿ (D) zu. **von** A cgr(S) ist, **dann** ist (P, D) gemäß Lemma **14 ein Subhedge-Projektionszustand und daher gemäß Proposition 2 für die Subhedge irrelevant.**

Lemma 15. Sei A = **A[I/ÿÿÿ],** L = L(A) und **S** = L(EIN **[F/FS]).** Wenn **(ÿÿÿ,** D) **ÿÿÿÿ** (Q, D)
bzgl. Acgr(S) für ein verschachteltes Wort u **ÿ Nÿÿÿÿÿ** und

entweder • Q **ÿÿ** dPrjÿ (D) und q **ÿ** Q,

oder • Q **ÿ** dPrjÿ (D) und q **ÿ** accÿ(Q),

hdg(u, dann existiert ein verschachteltes Wort **uÿ ÿ** classL ÿ(u) und **q0 ÿ ÿÿÿ,** so dass **q0 ÿÿÿÿÿ** q bzgl. ÿ.

Beweis. Nehmen wir **durch Induktion über die Länge von u an, dass (ÿÿÿ,** D) **ÿÿÿÿ** (Q, D) **bzgl.** A cgr(S) .
Im Basisfall haben wir u = **ÿ. Daher ist** Q = **ÿÿÿ.**

Fall Q **ÿÿ** dPrjÿ (D) **und** q **ÿ Q. Dann** ist **ÿÿÿ** = **{q}. Der eindeutige Lauf von** A **auf** u = ÿ **beginnt und endet im Baum-Anfangszustand** q **ÿ ÿÿÿ.**

Machine Translated by Google

Fall Q **ÿ** dPrjÿ (D) **und** q **ÿ** accÿ(Q). **Da** Q **ÿ** dPrjÿ (D), **zeigt Lemma 14 ,** dass (Q, D) **durch** *ÿ(ÿ)* = **Nÿ.** Da q **ÿ ist ein Subhedge-Projektionszustand, so dass** *ÿ* für L **Proposition** 2 **subhedge-irrelevant** ÿ **und S** ÿ accÿ(Q), **ist . Daher hedgen** classL h **und** ÿ **S gibt es außerdem** ein **q0 ÿ I, so dass q0 ÿÿ** q **bzgl. ÿ. Seien** u **ÿ** H classL **in** q **und** u ÿ = nw(h). **Der Lauf von** A **auf** u ÿ **endet** ÿ **S** ÿ(u).

Für den Induktionsschritt unterscheiden wir die möglichen Formen des verschachtelten Wortes u. Es existieren also **u˜,** v **ÿ Nÿ und** a **ÿ ÿ, so dass einer der folgenden Fälle gilt:)** cgr(S)

Fall u = **u˜ · ÿ** · v **· ÿ. Sei** (Q˜ , D) **der Zustand, in dem der Lauf von** (A ÿ **an u˜-Enden.**

Unterfall Q˜ ÿÿ dPrjÿ (D). **Sei Dÿ** = **DQ˜ . Dann existiert** P **ÿ Q, so dass (ÿÿÿ, Dÿ) vÿÿ** (P, **Dÿ) bzgl.** A cgr(S)) cgr(S) **auf** v **in den Zustand** (P, **Dÿ) endet von** (A ÿ

Unterunterfall P **ÿ** dPrjÿ (Dÿ). **Nach Konstruktion von** A cgr(S) **, dann haben wir** Q = **Q˜ @ÿaccÿ(P). Da** q **ÿ Q, gibt es q˜ ÿ Q˜ und** p **ÿ** accÿ(P), **so dass** ÿ(v), **so dass · ÿ q˜@p ÿ** q **in ÿ. Nach Induktionsvoraussetzung existiert** v , **dass der** ÿ **ÿ** KlasseL ÿ **S Lauf von** A ÿ **ÿ auf** v **endet in p. Daher ist der Lauf von** A **auf** u **ÿ** ÿ **= u˜ · ÿ** · v ÿ **endet mit q, und außerdem** mit u ÿ KlasseL (u). **S**

Unterunterfall P **ÿÿ** dPrjÿ (Dÿ). **Nach Konstruktion von** A cgr(S) **, dann haben wir** Q = **Q˜** @ÿP. **Da** q **ÿ Q, existieren q˜ ÿ Q˜ und** p **ÿ P, sodass q˜@p ÿ** q **in ÿ. Nach Induktionsvoraussetzung existiert** v ÿ(v), so dass **die Folge von · ÿ** in **q endet ,** und ÿ **ÿ auf** v **endet in p. Daher ist der Lauf von** A **auf** u **ÿ** ÿ **= u˜ · ÿ** · v ÿ außerdem **u** ÿ KlasseL (u). **S**

Unterfall Q˜ ÿ dPrjÿ (D). **Dann ist** Q = **Q˜ und** q **ÿ accÿ(Q˜). Unter Verwendung der Induktion** hdg(u˜ **(u˜), so Hypothese auf u˜ angewendet, existiert u˜** ÿ **ÿ** KlasseL **S dass ÿÿÿ ÿÿÿÿÿ q. ÿ** classL (u). **Da** u˜ für L **und S irrelevant ist** ÿ**, wir haben u˜** ÿ **S**

Fall u = **u˜ · a. Dieser ist leichter zu erreichen als die vorherigen Fälle und wird daher weggelassen.** □

Lemma 16. Ein partieller Durchlauf von A cgr(S) weist einem Präfix u **ÿ** prefs(Nÿ) einen Zustand (Q, D) zu. Wenn Q **ÿÿ** dPrjÿ (D) und q **ÿ** Q, dann existiert ein Präfix u **ÿ** classL ÿ (u), so dass ein partieller Durchlauf von A **uÿ** den Zustand **S** q zuweist .

Beweis. Durch Induktion über die Länge von u. Im Basisfall haben wir u = **ÿ. Der partielle Lauf von** init . A cgr(S) **auf** u **wird der Zustand** (Q, D) zugewiesen, **wobei** Q = I **und** D = **A FS Angenommen,** Q **ÿÿ** dPrjÿ (D) **und** q **ÿ Q. Dann ist** I = **{q}. Der eindeutige partielle Durchlauf von** A **auf** u = *ÿ* **endet im Anfangszustand** = **A[I/ Induktionsschritt die möglichen Formen** ÿ **ÿÿÿ], L) und S** [F/FS]. **Wir haben q ÿ** I. Lassen Sie A **für den von u unterscheiden. Es existieren also u˜ ÿ** prefs(Nÿ), v **ÿ Nÿ und** a **ÿ ÿ , sodass eine der folgenden Aussagen gilt:**

Fall u = **u˜ · ÿ** · **v. Sei** (Q˜ , **D˜) sei der Zustand, in dem der partielle Lauf von** A cgr(S) **an u˜ -Enden.**

Aus Q **ÿÿ** dPrjÿ (D) **folgt, dass Q˜ ÿÿ** dPrjÿ (D˜). **Weiterhin gilt (ÿÿÿ,** D) **vÿÿ** (Q, D) **bzgl.** A cgr(S)) cgr(S) partielle Lauf von (A **auf** v **geht also** in den Zustand (Q, D). **Unter Verwendung der Induktionshypothese existiert** v ÿ **ÿ** KlasseL ÿ ÿ(v), **so dass der partielle Lauf von** A = **u˜** ÿ **An** ÿ v **endet in q. Daher der partielle Lauf von** A **auf** u **ÿ S' · ÿ** · v in **q endet ,** und außerdem ÿ du KlasseL (u). **S**

Fall u = **u˜ · ÿ** · v **· ÿ. Sei** (Q˜ , **D˜) sei der Zustand, in dem der partielle Lauf von** A cgr(S) **an u˜ -Enden.**

Aus Q **ÿÿ** dPrjÿ (D) **folgt, dass Q˜ ÿÿ** dPrjÿ (D˜). **Weiterhin** gilt **(ÿÿÿ,** D) **vÿÿ** (P, D) **bzgl.** A cgr(S) **für ein** P **ÿ Q.**

Unterfall P **ÿ** dPrjÿ (D). **Dann** ist Q = **Q˜ @ÿaccÿ(P). Also gibt es q˜ ÿ** Q **und** p **ÿ** accÿ(P), **so dass q˜@p ÿ** q **in ÿ. Nach Lemma 15 gibt es** v ÿ(v), ÿ **ÿ** KlasseL ÿ **S**

Machine Translated by Google

so dass ÿÿÿ es hdg(vÿ) **ÿÿÿÿÿÿ** p **bzgl. ÿ. Unter Verwendung der Induktionsvoraussetzung, angewendet einen Anfangszustand q0 ÿ** I **und einen Präfix u˜ Lauf** ÿ **auf u˜, ÿ** KlasseL **(u˜), so dass die von** A **auf u˜** gibt ÿ **geht weiter zum Zustand q˜. Daher der partielle Lauf von** A ÿ **= u˜** ÿ **partielle** ÿ **· ÿ · v · ÿ geht weiter zu** q **und** ÿ u **auf** u **ÿ** KlasseL (u).

Unterfall P **ÿÿ** dPrjÿ (D). **Dann** ist Q = **Q˜** @ÿP. **Also gibt es q˜ ÿ** Q **und** p **ÿ P, so dass ÿ** KlasseL ÿ **dass q˜@p ÿ** q **in ÿ. Mit Lemma 15 existiert** v ÿ (v), **so dass** ÿ **ÿÿ** hdg(vÿ) **ÿÿÿÿÿÿ** p **bzgl. ÿ. Wenn wir die Induktionsvoraussetzung auf u˜ anwenden, gibt es ÿ** KlasseL **ÿ ein Anfangszustand q0 ÿ** I **und ein Präfix u˜** ÿ **(u˜), so dass der partielle Lauf von** A **· ÿ · v · ÿ auf u˜**ÿ **geht weiter zum Zustand q˜. Daher der partielle Lauf von** A ÿ **= u˜** ÿ **weitergeht zu** q **und** u ÿ **auf** u **ÿ** KlasseL (u).

Fall u = **u˜ · a. Dieser ist leichter zu erreichen als der vorherige Fall und wird daher weggelassen.**
□

Das nächste Lemma stellt die Schlüsselinvariante der Kongruenzprojektion dar, die schließlich ihre Vollständigkeit für die Subhedge-Projektion beweist. Dazu definieren wir für jedes F **ÿ FS ÿ** F **die binäre Relation ÿÿF** FS **als symmetrischer Abschluss von** F **×** (FS **** F), **d. h. für alle** (qÿ, q **") ÿ Q2 :**

$$Q\ddot{y}\ \ddot{y}\ddot{y}F_{FS}\ q'' \ddot{y} \quad (q' \ddot{y} F \ddot{y} q \ddot{y} FS \setminus F \ddot{y} q'' \ddot{y} FS \setminus F) \ddot{y}) \quad (q' \ddot{y} F)$$

Lemma 17 (Schlüsselinvariante). Sei A = **(ÿ, Q, ÿ, I,** F) a **dSHA,** F **ÿ FS ÿ** Q, L = L(A) und **S** = **L(A[F/FS]).** Wenn A cgr(S) einen partiellen Lauf auf Präfix u **ÿ** prefs(Nÿ) zum Zustand (P, D) **ÿ 2 Q × Dÿ** hat , dann gilt für alle (pÿ, p **") ÿ** P 2, (Rÿ,r **") ÿ** D, und (vÿ, v **") ÿ N 2** so dass **ÿ** ,

$$P^{\ddot{y}} \ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}^{hdg(v'')} r \quad \text{bzgl. } \ddot{y}\ \ddot{y}\ p'' \ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}^{hdg(v'')} r \quad \text{bzgl . } \ddot{y}$$

es gibt (uÿ, u **ÿÿ) ÿ** KlasseL (u), w **ÿ** suffs(Nÿ), **qÿ ÿÿF FS** q **"** und **q0 ÿ** I, so dass

$$hdg(u \cdot w)^{\ddot{y}} \ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ q \quad \ddot{y}\ \ddot{y}\ q0\ \ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ q \quad hdg(u \cdot w'')\ \text{bzgl.} \quad \text{bzgl . } \ddot{y}$$

Beweis. Induktion über die Anzahl der offenen Klammern von u.

Im Basisfall hat u keine offene Klammer, also **ist** u **ÿ Nÿ** ein A,FS init . **Also hat** A cgr(S) einen partiellen Lauf auf D A,FS **bis** (P, D), **wobei** P **ÿ Q. Sei** (p dem verschachtelten Wort u **ÿ Nÿ verschachteltes Wort. In diesem Fall** ist D = init **ÿ, p ") ÿ** P 2, (Rÿ,r **") ÿ** D A,FS init , **und** (vÿ, v **") ÿ N 2** **so dass ÿ ,**

$$P^{\ddot{y}} \ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}^{hdg(v\ \ddot{y}\ \ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y})} r \quad \textbf{bzgl. } \ddot{y}\ \ddot{y}\ p'' \ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}^{hdg(v'')\ \ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}} r \quad \textbf{bzgl. } \ddot{y}$$

Daraus folgt, dass P **ÿÿ** dPrjÿ (D **es** A,FS init **). Daher können wir Lemma 16** anwenden. **Es zeigt, dass gibt verschachtelte Wörter** (uÿ, u **") ÿ** (classL (u))2 **und q0 ÿ I, so dass**

$$\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}^{hdg(u'')} p\ \textbf{q0} \quad \textbf{bzgl. } \ddot{y}\ \ddot{y}\ \textbf{q0}\ \ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}^{hdg(u'')} p \quad '' \textbf{ bzgl. } \ddot{y}$$

Seit D A,FS init **ÿ** ldrÿ (F **×** (FS **** F)) **nach Lemma 10 gibt es ein verschachteltes Wort** w **ÿ Nÿ und ÿ ÿÿF Zustände** q **FS** q **", so dass**

$$'^{R}\ddot{y}\ddot{y}\ddot{y}\ddot{y}^{hdg(w)} q \quad \textbf{bzgl. } \ddot{y}\ \ddot{y}\ r'' \ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}^{hdg(w)} q \quad '' \textbf{ bzgl. } \ddot{y}$$

Dann haben wir Folgendes:

$$\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}^{hdg(u\ \ddot{y}\cdot v\ \cdot g)} q\ \textbf{q0} \quad \textbf{bzgl. } \ddot{y}\ \ddot{y}\ \textbf{q0}\ \ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}\ddot{y}^{hdg(u''\ v\ \cdot g)} q \quad '' \textbf{ bzgl. } \ddot{y}$$

Machine Translated by Google

Für den Induktionsschritt sei u = **u1 · ÿ · u2 für ein Präfix u1 ÿ** prefs(Nÿ) **und ein verschachteltes Wort u2 ÿ Nÿ, so dass u1 eine offene Klammer weniger hat als u. Lassen Sie** A cgr(S) **einen partiellen Lauf auf dem verschachtelten Wort** u **ÿ Nÿ zu** (P, D) haben. **Lassen Sie (P1, D1) den Zustand sein, den dieser Lauf dem Präfix u1 zuweist. Dann ist Dp= "(ÿ1)P1 , Lassen Sie h SD und** (v , v **") ÿ N 2 so dass ÿ ,**

P hdg(v) **bzgl. ÿ ÿ** p hdg(v ") **bzgl. ÿ**

Da D = (D1)P1 , **existieren** (p **und** p) **ÿ (P1) und Staaten** (r , r) **ÿ D1, so dass** p **1@q** **1@q** **Insbesondere** ist **P1 ÿÿ** dPrjÿ **(D1) , so dass wir Lemma 16** anwenden können. **Es zeigt, dass es gibt** (u) **ÿ** Klasse **und ÿÿÿ ÿÿ** p **". Sei** = ÿ **und** v **bzgl. ÿ Unter Verwendung der** p **auf u1 angewandten Induktionsvoraussetzung , wonach** A cgr(S) **einen partiellen Lauf zum Zustand (P1,** D1) hat,

so dass (p **1 , p 1**) **ÿ (P1)** **und** (r **1 ,r 1**) **ÿ D, es gibt** (u) **ÿ** Klasse **FS** q **" und q0 ÿ I, so dass**

hdg(u **11**w1) **q0 ÿ** hdg(u **11**w1) **bzgl.** **bzgl. ÿ**

Lass uns = u **1** **·** ÿ **·** = u **1** **·** ÿ **·** **, und** w = ÿ **· w1. Das Obige ergibt dann Folgendes:**

hdg(u ·w) **q0 ÿÿ** hdg(u ·w) **bzgl. ÿ** **bzgl. ÿ**

Darüber hinaus (u , u **ÿÿ) ÿ** (Klasse L(u))2 , **also sollte dies gezeigt werden.** □

Satz 6. Sei A = (ÿ, Q, ÿ, I, F) a dSHA, F ÿ FS ÿ Q, L = L(A) und S = L(A[F/FS]).

Wenn A cgr(S) einen partiellen Lauf auf dem Präfix u **ÿ** prefs(Nÿ) zu einem Zustand (P, D) **ÿ 2 Q × Dÿ** hat , so dass P **ÿÿ** dPrjÿ (D), dann ist Klasse L (u) für L bzgl . S subhedge-relevant .

Beweis. Angenommen, A cgr(S) **hat einen partiellen Lauf auf Präfix** u **ÿ** prefs(Nÿ) **zu einem Zustand** (P, D) **ÿ × Dÿ , so dass** P **ÿÿ** dPrjÿ (D). **Da** P **ÿÿ** dPrjÿ (D), **gibt es accÿ(P) 2 2 Q ÿ** D ÿ= emptyset. v **ÿÿ) ÿ N 2**
Es gibt also (p , p **") ÿ** P , (R ,r **") ÿ D, und** (v , **so dass**

P hdg(v ÿ) **bzgl. ÿ ÿ** p hdg(v **") "** **bzgl. ÿ**

Nach Lemma 17 gibt es (u , **so dass** u **ÿÿ) ÿ** Klasse **ÿ ÿÿF** (u), w **ÿ** suffs(Nÿ), q **FS** q **", und q0 ÿ I,**

hdg(u ÿ ·g) hdg(u ·g) **bzgl. ÿ ÿ q0 ÿÿÿÿÿÿÿÿ** **bzgl. ÿ**

Daher ist Klasse (u) ist für L **bzgl. S relevant .** □

Theorem 3 (Vollständigkeit der Kongruenzprojektion). Für jedes vollständige **dSHA (ÿ, Q, ÿ, I,** F) und jede Menge F **ÿ FS ÿ** Q ist die Kongruenzprojektion A cgr(S) für die Subhedge- Projektion für das reguläre Muster L(A) bzgl. des regulären Schemas **S** = **L(A[FS/F])** stichhaltig und vollständig .

Beweis. Die Stichhaltigkeit der Kongruenzprojektion wird in Theorem 2 gezeigt. Um die Vollständigkeit zu beweisen, sei u ÿ prefs(Nÿ) **ein verschachteltes Wortpräfix, das für** L **bzgl. S stark subhedge-irrelevant ist. Per Definition starker Irrelevanz ist die Klasse** classL (u) für L **und** S **irrelevant . Sei** (P, D) der eindeutige Zustand, der **dem Präfix** u **durch den partiellen Lauf von** A cgr(S) zugewiesen wird . **Da** (u) für L **und** S **irrelevant ist , zeigt Proposition 6 , dass** P **ÿ** dPrjÿ (D). **Unter Verwendung von Lemma 14** ist classL **S** (P, D) **dann ein subhedge-Projektionszustand von** A cgr(S)

. □

Machine Translated by Google

7.5. In-Memory-Komplexität **Als**

nächstes diskutieren wir die Komplexität des Mitgliedschaftstests mit vollständiger Subhedge-Projektion basierend auf der In-Memory-Auswertung des Input-Hedges durch die Kongruenzprojektion des Input-Automaten.

Lemma 18. Die Anzahl der Zustände |Qcgr| liegt in O(2 n 2+n), wobei n die Anzahl der Zustände von A ist.

Beweis. Mit der deterministischen Konstruktion sind die Zustände von A cgr(S) **Paare in 2Q × Dÿ. Die maximale Anzahl von Zuständen der Kongruenzprojektion ist also** |Qcgr(S) **|** = **2** |Q| **|Dÿ|** ÿ **2** N **2** 2 n = **2** n 2+n **.** □

Sei diff-rel(Qcgr(S)) **die Menge der Differenzrelationen, die in den Zuständen von** A cgr(S) verwendet werden.

Korollar 1. Die Signatur **ÿ** sei fest. Für jedes vollständige **dSHA** A mit Schema **S** ÿ **Hÿ** kann die Mitgliedschaft in jedem Hedge h **ÿ S** im Speicher in der Zeit **O(|1|)** pro nicht projiziertem Knoten von h getestet werden, nach einer Vorverarbeitungszeit von **O(|A** cgr(S) **|** + n 3d), wobei d die Kardinalität von diff -rel(Qcgr(S)) und n die Anzahl der Zustände von A ist.

Da |Qcgr(S) **| gemäß Lemma 18** in O(2 n 2+n) **liegt , beträgt die Vorverarbeitungszeit O(|A** cgr(S) **|** + n **3d|).** 2n **2+2n** **liegt auch in** O(2).

Beweis. Da A **vollständig ist, zeigt** Kapitel **3 ,** dass A cgr(S) **für die Subhedge-Projektion für Schema S vollständig ist. Der im Speicher befindliche Evaluator von** A cgr(S) **kann auf jedem Eingabe-Hedge** h **ÿ S in der Zeit O(|1|) pro nicht projiziertem Knoten von** h **nach einer Vorverarbeitungszeit von O(|A** cgr(S) **|** + n 3d) ausgeführt werden, **wobei** d **die Kardinalität von** diff-rel(A cgr(S)) **ist. Da die Signatur ÿ fest und** A cgr(S) **deterministisch** ist , liegt |A cgr(S) **| in** O(| Qcgr(S) **|** **2** **). Der dominierende Teil** |A cgr(S) **| liegt in** O(|Qcgr| **2** **).** □

Bei der Anwendung auf die normale Beantwortung von XPath-Abfragen wie in unseren Experimenten in Abschnitt 11 haben wir n **ÿ 80 und** d **ÿ 25, sodass der kubische Teil** O(n 3d) **nicht einschränkend ist. Die Größen der SHAÿ s liegen unter 2600, wenn sowohl Zustände als auch Regeln gezählt werden. Die Obergrenze O(|A** cgr(S) **|** + n 3d) lässt **daher darauf schließen, dass die Vorverarbeitungszeit in der Praxis machbar sein könnte, auch wenn sie im schlimmsten Fall exponentiell ist.**

Hier ist Vorsicht geboten, da die Größen der in Abschnitt 11 angegebenen **SHAÿ nicht zur reinen Kongruenzprojektion** A cgr(S), **sondern zur frühesten Kongruenzprojektion** A , **die wir erst in** cgr(S) **e(S) Abschnitt 8 vorstellen werden. Das Nichterstellen von Übergangsregeln zu Zuständen, die sicher abgelehnt werden können, kann zu einer radikalen Größenreduzierung führen. Alternativ könnte eine wichtige Größenreduzierung in** A cgr(S) **durch das Entfernen aller nutzlosen Zustände und Übergangsregeln erreicht werden.**

Dies würde jedoch die Vorberechnungszeit nicht verringern, im Gegenteil: Es würde die Vorverarbeitungszeit erhöhen, da die Reinigung von Automaten a posteriori ebenfalls Zeit benötigt, die für SHAÿ s nichtlinear sein kann.

Anstatt A cgr(S) statisch **aus den Eingaben** A **und FS** vorzuberechnen , **können wir nur den Teil berechnen, der von** A cgr(S) **zur Auswertung der Eingabeabsicherung** h benötigt **wird. Wir stellen fest, dass die Anzahl der Übergangsregeln des benötigten Teils durch** O(|h|) beschränkt ist , **aber aufgrund der Projektion viel kleiner sein kann . Die Vorverarbeitungszeit sinkt dann auf** O(|A|). **Da der Vollautomat A** cgr (S) im schlimmsten Fall exponentiell in n **sein kann , kann dies die Gesamtverarbeitungszeit reduzieren .**

Positiv ist, dass die On-the-Fly-Auswertung die Gesamtrechenzeit auf O(|h| + n 3d(h)) reduziert, **wobei** d(h) **die Anzahl der Differenzrelationen im Teil von** A cgr(S) ist , **die für die Auswertung von h benötigt wird, also** d(h) **ÿ min(|h|,** d). **Negativ ist, dass** d(h) **Schritte nicht mehr in konstanter, sondern in kubischer Zeit in** O(n) **erfolgen. Es ist auch zu beachten, dass die Gesamtzeit der On-the-Fly-Auswertung der Safe-No-Change-Projektion in O(|A|** + |h| + n s(h)) liegt, **wobei** s(h) **die Anzahl der Teilmengen sicherer Zustände von** A **in Zuständen von** Qcgr ist, **d. h.** s(h) **ÿ |h|. Dies liegt daran, dass jede Teilmenge sicherer Zustände in linearer Zeit in** O(n) berechnet werden kann. **Wie oben argumentiert , beträgt die Gesamtzeit der Kongruenzprojektion, die durch die On-the-Fly-Auswertung erreicht wird, O(|A|** + |h| + n).
3 d(h)). **Dies liegt daran, dass die Berechnung einer Differenzrelation in** O(n 3

Machine Translated by Google

durch Verwendung von kubischen Datalog-Programmen. Der zusätzliche kubische Faktor in n d(h) **anstelle eines linearen Faktors in** n s(h) **ist der Preis, der für die Erzielung einer vollständigen Subhedge-Projektion mit On-the-Fly-Auswertung gezahlt werden muss .** Beachten Sie, dass dieser Preis in Fällen , **in denen** n **ÿ 80,** d(h) **ÿ 15 und** s(h) **ÿ 15 ist, weitgehend akzeptabel ist .**

Beispiel 15 (Exponentieller Anstieg.). Um zu sehen, wie der exponentielle Worst-Case eintreten kann, betrachten wir eine Familie regulärer Sprachen, für die der minimale **DFA** von links nach rechts exponentiell größer ist als der minimale **DFA** von rechts nach links . Die klassischen Beispielsprachen mit dieser Eigenschaft sind Ln = ÿ**ÿ**, wobei n **ÿ** N und **ÿ** = {a, **b}.** Intuitiv gehört ein Wort in **ÿ ÿ** zu Ln , wenn und nur sein n + 1-ter Buchstabe vom Ende gleich **„a“ ist.** Der minimale **DFA** von links nach rechts für Ln hat **2** n + 1 Zustände ,da er sich ein Fenster von n + **1** Buchstaben merken muss . Im Gegensatz dazu hat sein minimaler **DFA** von rechts nach links nur n + **1** Zustände; in dieser Richtung ist es ausreichend, sich die Distanz vom Ende modulo n + 1 zu merken.

Als nächstes betrachten wir das durch µz definierte Schema **S.** ÿ(a + b) · zÿ . Es enthält alle Hedges, die in allen ihren Teilbäumen ein Label von **ÿ** = {a, **b}** als ersten Buchstaben haben und keine weiteren Vorkommen von Buchstaben von **ÿ** an anderer Stelle. Wir betrachten dann die Familie der Hedgesprachen Hn **ÿ Hÿ** mit Schema S, so dass die Folge der Labels eines Wurzel-zu-Blatt-Pfades von hn zu Ln gehört. Beachten Sie, dass Hn von unten nach oben durch den **dSHA** An mit O(n + 1) Zuständen erkannt werden kann, der den minimalen deterministischen **DFA** von Ln auf allen Pfaden des Eingabe-Hedges simuliert . Für einen Evaluator mit Teilhedge- Projektion ist die Situation anders. Bei der Bewegung von oben nach unten muss er sich die Folge s merken, und es gibt **2** n+1
von Bezeichnungen der n + 1-letzten Vorfahren, möglicherweise aufgefüllt mit ÿ b solcher Sequenzen.
Wenn die Sequenz eines Blattes mit einem „a“ beginnt, dann können die folgenden Teilhecken mit den folgenden Blättern wegprojiziert werden. Folglich kann es kein **dSHAÿ** geben, das Hn erkennt , das
projiziert alle irrelevanten Subhedges mit weniger als **2** n+1 Zuständen weg. Insbesondere muss die Größe $_{N}^{cgr(S)}$ von A exponentiell zur Größe von An sein.

8. Früheste Mitgliedschaft mit Subhedge-Projektion Als nächstes

verbessern wir unsere Compiler für die Einführung von Subhedge-Projektionszuständen, so dass sie auch bestimmte Mitgliedschaften oder Nichtmitgliedschaften auf früheste Weise erkennen können. Dazu zeigen wir, wie man den vorherigen frühesten Mitgliedschaftstester für dSHAs aus [21] mit jedem beliebigen Subbaumprojektionsalgorithmus kombinieren kann, so dass er sowohl für die Safe-No-Change-Projektion als auch für die Kongruenzprojektion gilt. Indem wir beide Aspekte orthogonal kombinieren, erhalten wir einen frühesten dSHA mit Subhedge-Projektion, der entweder im Speicher oder im Streaming-Modus ausgewertet werden kann .

8.1. Früheste Mitgliedschaft

Wir beginnen mit einer semantischen Charakterisierung der frühesten Mitgliedschaft, d. h. der Präfixe verschachtelter Wörter von Hedges, deren Suffixe für die Mitgliedschaft irrelevant sind, wenn wir von einer Verarbeitung von oben nach unten und von links nach rechts ausgehen. Beachten Sie, dass die gleiche Charakterisierung bis hin zu verschiedenen Präsentationsoptionen bereits früher im Zusammenhang mit der Stream-Verar

Definition 13. Sei **S ÿ Hÿ** ein Hedge-Schema und L **ÿ S** eine Hedge-Sprache, die dieses Schema erfüllt. Ein geschachteltes Wortpräfix v heißt **sicher für die Zugehörigkeit** zu L mit Schema **S** für alle Suffixe w geschachtelter Wörter, so dass v · w **ÿ** nw(S) gilt, dass v · w **ÿ** nw(L). Es heißt **sicher für die Nichtzugehörigkeit** zu L mit Schema **S** für alle Suffixe w geschachtelter Wörter, so dass v · w **ÿ** nw(S) gilt , dass v · w **ÿÿ** nw(L).

Wir sind nun an SHAÿ interessiert , **die in der Lage sind, bestimmte Mitgliedschaften und Nicht-Mitgliedschaften zu erkennen. Mitgliedschaft syntaktisch beim frühestmöglichen Präfix oder Knoten des Laufs.**

Definition 14. Ein **dSHAÿ** A = **(ÿ, Q, ÿ, I,** F) mit Schema **S ÿ Hÿ** heißt **frühestens für Schema S ,** wenn es einen Zustand sel **ÿ** Q gibt , so dass für alle Hedges h **ÿ** S der eindeutige Lauf R von compl(A) auf h Folgendes erfüllt:

Machine Translated by Google

- es geht zum Zustand sel für genau alle diejenigen Präfixe von nw(h) , die sicher für die Mitgliedschaft in L(A) bzgl. S sind;
- es geht zum Zustand sink für alle Präfixe von nw(h) , die sicher für die Nichtmitgliedschaft in L(A) bzgl . S sind.

Für dSHA A, schemavollständig für Schema S = L(A[F/FS]), können wir ein dSHAÿ Ae(S) konstruieren , **das das früheste für S ist, wie in Anhang A gezeigt. Die Idee besteht darin, den frühesten Mitgliedschaftstester für deterministische SHAs aus [21] so anzupassen, dass er Schemabeschränkungen** berücksichtigt. Darüber hinaus müssen wir die Eingabe **-dSHAs zu dSHAÿ** s **kompilieren und nicht zu dNWAs, wie dort verwendet. Diese Änderung ist unkompliziert, sobald die Beziehung zwischen beiden Automatenmodellen klar wird.**

Da die Konstruktion von Ae(S) **nicht zum Kern des vorliegenden Dokuments gehört, veranschaulichen wir das Ergebnis nur anhand unseres laufenden Beispiels: Für das dSHA** A **mit Schema S** = doc **in Abbildung 4 erhalten wir das SHAÿ** Ae(S) **in Abbildung 19.** Beachten Sie, dass **es beim Präfix** ÿlistÿitem **in den Zustand to** sel wechselt, **was zeigt, dass die Mitgliedschaft sicher ist. Beim Präfix** ÿitem **blockiert es und zeigt damit an, dass die Nichtmitgliedschaft sicher ist. Beim Präfix** ÿlistÿlist wechselt **es** jedoch **nicht in einen Subhedge-Projektionszustand, obwohl dieses Präfix für die Subhedge irrelevant ist.**

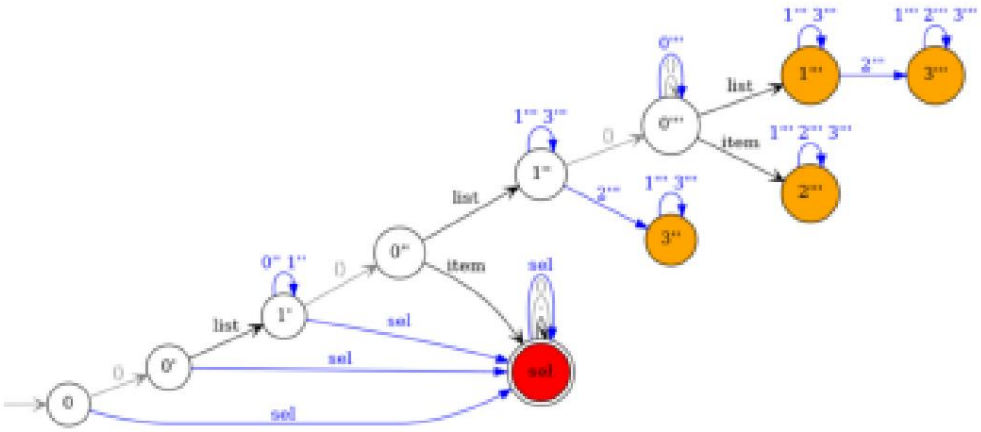


Abbildung 19. Das früheste dSHAÿ Ae(S) **für das dSHA** A **in Abbildung 4 mit Schema S** = doc **für den XPath-Filter** [self::list/child::item].

8.2. Hinzufügen einer Subhedge-Projektion

Wir erweitern jetzt jede früheste Mitgliedschaft für dSHAs mit einer Subhedge-Projektion. Die Idee besteht darin, die Subhedge-Projektion mit der frühesten Mitgliedschaft zu kombinieren. Nehmen wir also an, dass wir ein Schema **S und zwei Automaten,** A ÿ **und** Ae , **mit Schema S haben ,** die **bis zum Schema dieselbe Sprache erkennen ,** sodass L(A ÿ) **ÿ L(S)** = L(Ae) **ÿ L(S). Nehmen wir außerdem an, dass** Ae **einen Auswahlzustand** sel hat, **der eine bestimmte Mitgliedschaft beim frühestmöglichen Präfix anzeigt.**

Um die früheste Mitgliedschaft mit Subhedge-Projektion zu erhalten, kombinieren wir die beiden SHAÿ s zu einem einzigen SHAÿ F e), **wobei beide Automaten grundsätzlich parallel, aber unter gemeinsamer Kontrolle laufen. Die Zustandsmenge von** A $A^{ÿ}_{ÿ}$ = (ÿ, $Q^{ÿ}_{ÿ}$, ÿ, $ÿ^{ÿ}_{ÿ}$, $ÿ^{ÿ}_{ÿ}$) **sind wie folgt:**

$$Q^{ÿ}_{ÿ} = (Qÿ \times (Qe \setminus \{sel\})) \cup \{sel\}$$
$$ÿ^{ÿ}_{ÿ} = (I ÿ \times (Ie \setminus \{sel\})) \cup \{sel \mid sel ÿ Dh\}$$
$$F^{ÿ}_{ÿ} = (F ÿ \times (Fe \setminus \{sel\})) \cup \{sel\}$$

Die Übergangsregeln in ÿ $ÿ^{ÿ}_{ÿ}$ sind in Abbildung 20 **dargestellt . Jeder Lauf von** $A^{ÿ}_{ÿ}$ **synchronisiert parallel , tut dies auch, und wann Läufe von** A ÿ **und** Ae **wie folgt: Immer wenn** Ae zu sel **geht , dann** A $A^{ÿ}_{ÿ}$ **immer es in einen Unterhedge-Projektionszustand geht,** A ÿ **lässt es** Ae **über die nachfolgende Unterhedge springen .**

Machine Translated by Google

$$\frac{q \text{ ein} \ddot{y} \ddot{y} \; q \overset{\ddot{y}}{\to} q' \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}} \qquad r \; a\ddot{y}\ddot{y} \; r \overset{\ddot{y}}{\to} r' \text{ in } \ddot{y}e \; q \; \ddot{y}\ddot{y} \; P \; r\ddot{y} \; \ddot{y} = \text{sel } (q,r)}{a\ddot{y}\ddot{y} \; (q \overset{\ddot{y},R}{\to} q') \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}}} \qquad \frac{q \; a\ddot{y}\ddot{y} \; q' \; a\ddot{y}\ddot{y} \text{ sel in } \ddot{y}e \; (q,r) \; a\ddot{y}\ddot{y} \text{ sel}}{\text{in } \ddot{y} \; {}_{t}^{\ddot{y}}}$$

$$\frac{q \text{ ein} \ddot{y} \ddot{y} \; q \overset{\ddot{y}}{\to} q' \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}} \qquad r \; a\ddot{y}\ddot{y} \; r \overset{\ddot{y}}{\to} r' \text{ in } \ddot{y}e \; q \; \ddot{y} \; P \; r\ddot{y} \; \ddot{y} = \text{sel } (q,r) \; a\ddot{y}\ddot{y}}{(q \overset{\ddot{y}}{\to} q', r) \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}}}$$

$$\frac{q@p \; \ddot{y} \; q \overset{\ddot{y}}{\to} q' \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}} \qquad r@s \; \ddot{y} \; r \overset{\ddot{y}}{\to} r' \text{ in } \ddot{y}e \; r \overset{\ddot{y}}{\to} r' \; \ddot{y} = \text{sel } q \; \ddot{y}\ddot{y} \; P)}{(q,r)@(p,s) \; \ddot{y} \; (q \overset{\ddot{y},R}{\to} q', r') \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}}} \qquad \frac{q@p \; \ddot{y} \; q \overset{\ddot{y}}{\to} q' \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}} \qquad r@s \; \ddot{y} \text{ sel in } \ddot{y}e}{(q,r)@(p,s) \; \ddot{y} \text{ sel in } \ddot{y} \; {}_{t}^{\ddot{y}}}$$

$$\frac{q@p \; \ddot{y} \; q \overset{\ddot{y}}{\to} q' \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}} \qquad r@s \; \ddot{y} \; r \overset{\ddot{y}}{\to} r' \text{ in } \ddot{y}e \; q \; \ddot{y} \; P \; r\ddot{y} \; \ddot{y} = \text{sel } r') \text{ in } \ddot{y}}{(q,r)@(p,s) \; \ddot{y} \; (q \overset{\ddot{y}}{\to} q', r') \; {}_{t}^{\ddot{y}}}$$

$$\frac{\ddot{y}\ddot{y}\ddot{y}\ddot{y}q \overset{\ddot{y}}{\to} q' \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}} \qquad \ddot{y}\ddot{y} \; \ddot{y} \; \ddot{y}\ddot{y} \; rr' \text{ in } \ddot{y}e \; q \; \ddot{y}\ddot{y} \; P \; r\ddot{y} \; \ddot{y} = \text{sel}}{\ddot{y}\ddot{y} \; (q,r) \; \ddot{y}\ddot{y} \; (q \overset{\ddot{y},R}{\to} q') \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}}} \qquad \frac{\ddot{y}\ddot{y}\ddot{y}\ddot{y}q \overset{\ddot{y}}{\to} q' \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}} \qquad R \; \ddot{y}\ddot{y} \; \ddot{y}\ddot{y} \text{ sel in } \ddot{y}e}{\ddot{y}\ddot{y} \; (q,r) \; \ddot{y}\ddot{y} \text{ Auswahl in } \ddot{y} \; {}_{t}^{\ddot{y}}}$$

$$\frac{\ddot{y}\ddot{y}\ddot{y}\ddot{y}q \overset{\ddot{y}}{\to} q' \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}} \qquad \ddot{y}\ddot{y} \; \ddot{y} \; \ddot{y}\ddot{y} \; rr' \text{ in } \ddot{y}e \; q \; \ddot{y} \; P \; r\ddot{y} \; \ddot{y} = \text{sel}}{\ddot{y}\ddot{y} \; (q,r) \; \ddot{y}\ddot{y} \; (q \overset{\ddot{y}}{\to} q', r') \text{ in } \ddot{y} \; {}_{t}^{\ddot{y}}}$$

$$\frac{\text{ein } \ddot{y} \; \ddot{y}}{\text{sel } a\ddot{y}\ddot{y} \text{ sel in } \ddot{y} \; {}_{t}^{\ddot{y}}} \qquad \frac{\mu \; \ddot{y} \; Q\ddot{y} \; {}_{t}}{\text{sel@}\mu \; \ddot{y} \text{ sel in } \ddot{y} \; {}_{t}^{\ddot{y}}} \qquad \frac{\mu \; \ddot{y} \; Q\ddot{y} \; {}_{t}}{\mu\text{@sel } \ddot{y} \text{ sel in } \ddot{y} \; {}_{t}^{\ddot{y}}} \qquad \frac{\text{WAHR}}{\ddot{y}\ddot{y} \text{ sel } \ddot{y}\ddot{y} \text{ sel in } \ddot{y} \; {}_{t}^{\ddot{y}}}$$

Abbildung 20. Die Übergangsregeln $\ddot{y} \; {}_{t}^{\ddot{y}}$ **abgeleitet aus denen der SHA**$\ddot{y}$ **s** A $\ddot{y}$ **mit Projektionszuständen** P **und** Ae **mit Auswahlzustand** sel.

Zur Veranschaulichung betrachten wir den dSHA aus Abbildung 4 noch einmal. Der früheste Teilhedge-Projektion A **e(S)** $^{\text{snc}}$ **dSHA**$\ddot{y}$ **mit ist in Abbildung 21** dargestellt. **Hier kombinieren wir die Safe-No-Cha -Projektion** A $^{\text{snc}}$ **aus Abbildung 11 mit dem frühesten Mitgliedschaftstester** Ae(S) **aus Abbildung 19. Ihre Kombination** A $^{\text{snc}}$ **e(S) stellt das Muster** [self::list/child::item] **mit Schema** doc **auf prägnante Weise dar. Nach dem Präfix** ylistyitem **erkennt es eine bestimmte Mitgliedschaft: für das Präfix** ylistylist **erkennt es, dass die Unterabsicherung irrelevant ist, und nach dem Präfix** yitem erkennt **es eine bestimmte Nichtmitgliedschaft.**

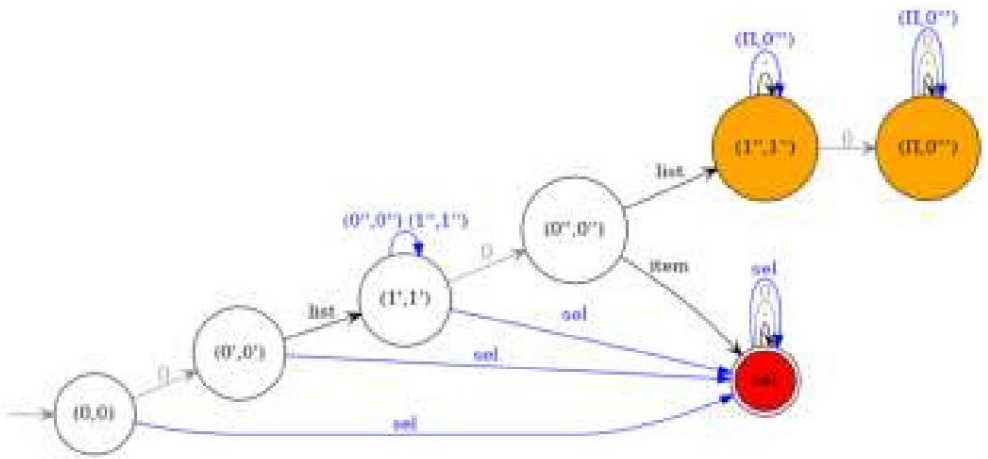


Abbildung 21. Der früheste dSHA$\ddot{y}$ A $^{\text{snc}}$ **mit Safe-No-Change-Subhedge-Projektion für das dSHA** A **in e(S) Abbildung 4 mit Schema S** = doc **für den XPath-Filter** [self::list/child::item].

Machine Translated by Google

8.3. Solidität und Vollständigkeit

Als nächstes zeigen wir, dass die Solidität und Vollständigkeit eines Projektionsalgorithmus erhalten bleiben, wenn er unter der Annahme von Determinismus mit einem frühesten Mitgliedschaftstester kom

Proposition 7. Sei A das *ÿ* und Ae sind **dSHAÿ** s mit demselben Schema **S** und derselben Sprache bis zu Schema, d. h. L(A ÿ) **ÿ L(S)** = L(Ae) **ÿ L(S).** Die Kombination **dSHAÿ** A Sprache bis zu $\ddot{y}_t$ dann hat das gleiche Schema S ebenfalls. Wenn Ae außerdem für S am frühesten ist, dann geht A in einen $\ddot{y}_t$ ist am frühesten für **S** und geht Subhedge-Projektionszustand, wann immer *Aÿ* dies tut.

Wenn q **ein Subhedge-Projektionszustand von** A *ÿ* **und** r **ein Zustand von** Ae , *ÿ* **und** Ae **parallel dann** ist **(q,r) ein Subhedge -Beweis. Input-Hedge ist, wobei** ë . **Der Evaluator für Automat** A $\ddot{y}_t$ **läuft** A **auf dem Projektionszustand von** A **Subhedges übersprungen werden, die in Subhedge-Projektionszuständen von** A **ÿ beginnen. Dazu gehören die Subhedges, die in einem Projektionszustand beginnen, wenn** A **Proposition 2 zu** A **ÿ ausgewertet wird. Solche Subhedges** $\ddot{y}_t$ **zu h. Bewerben sind für** L(A ÿ) **in Bezug auf S irrelevant, da** angenommen wird, dass A deterministisch ist. Da L(A ÿ) **ÿ S** = L(Ae) **ÿ S, hat das Überspringen eines solchen Subhedges keinen Einfluss auf die Akzeptanz von Hedges innerhalb des Schemas** S durch Ae . Daher **ist der Bewerter von** A $\ddot{y}_t$ **ist auch in Bezug auf S am frühesten.** □

Der obige Vorschlag zeigt, dass A **auch für** $\ddot{y}$ **ist für die Subhedge-Projektion vollständig, dann** A $\ddot{y}_t$ **die Subhedge-Projektion vollständig und gleichzeitig am frühesten ist.**

Theorem 4. Für alle vollständigen **dSHAs** A mit Schema S ist das **dSHA** A cgr(S)e(S) das früheste und vollständigste für die Subhedge-Projektion für Schema S.

Beweis. Die Kongruenzprojektion A cgr(S) **ist nach Proposition 2** stichhaltig **und nach Theorem** 3 für **die Subhedge-Projektion bzgl. S** vollständig. Der **in Abschnitt 8.1** besprochene **Automat** Ae(S) **ist der früheste für S. Daher ergibt ihre Kombination** A cgr(S)e(S) **nach Proposition 7 den frühesten Automaten mit vollständiger** Subhedge-Projektion bzgl. S. □

8.4. In-Memory-Komplexität **Als**

nächstes besprechen wir die Komplexität des frühesten Mitgliedschaftstests mit vollständiger Subhedge- Projektion durch einen In-Memory-Evaluator für A cgr(S)e(S) .

Lemma 19. Die Anzahl der Zustände in **Qcgr(S)e(S)** liegt in O(2 n 2+2n+log(n)), wobei n = **|Q|.**

Beweis: Mit Lemma 18 ist **die Anzahl der Zustände von** A cgr(S) **in** O(2 n 2+n), **wobei** n = **|Q|. Die Anzahl der Zustände von** Ae(S) **liegt in** O(n2 n) **und damit in** O(2 n+log(n)). **Die Anzahl der Zustände von Qcgr(S)e(S) liegt damit in** O(2 n 2+n **2** n+log(n)), **was in** O(2 n 2+2n+log(n)) liegt. □

Korollar 2. Lassen Sie die Signatur **ÿ** fest sein. Für jedes vollständige **dSHA** A mit Schema **S ÿ Hÿ** kann die früheste Mitgliedschaft in einem Eingabe-Hedge h **ÿ S** im Speicher in der Zeit **O(|1|)** pro nicht projiziertem Knoten von h nach einer Vorverarbeitungszeit von **O(|A** cgr(S)e(S) **|** + n 3d + n s) getestet werden , wobei d die Anzahl der Differenzrelationen und s die Anzahl der Teilmengen sicherer Zustände in **Qcgr(S)e(S)** ist . Die Vorverarbeitungszeit Zustände beträgt A ist in O(2 n 2+4n+2log(n)), wobei n die Anzahl der

Beweis. Da A **vollständig ist, zeigt** Lemma **4 ,** dass A cgr(S)e(S) **die früheste, stichhaltigste und vollständigste Subhedge-Projektion für Schema S ist. Der im Speicher befindliche Evaluator von** A cgr(S)e(S) **kann auf jedem Eingabe-Hedge** h **ÿ S in Zeit O(|1|) pro nicht projiziertem Knoten von** h **nach einer Vorverarbeitungszeit von O(|A** cgr(S)e(S) **|** + n 3d + n **s. ÿ ist auf die Größe** |A cgr(S)e(S) **| beschränkt durch** O(|Qcgr(S)e(S) ^{2} festgelegt . **Lemma 19 zeigt, dass |Qcgr(S)e(S) | in** O(2 n 2+2n+log(n)) **liegt, also liegt** |A cgr(S)e(S) **| in** O(2 $^{2^{2n\ 2+4n+2log(n)}}$). □

Machine Translated by Google

Wie zuvor **können wir , anstatt** A cgr(S)e(S) **aus den Eingaben dSHA** A **und FS vorab zu berechnen,** nur den Teil berechnen, der für die Auswertung des Eingabe-Hedges h benötigt wird . **Wenn der benötigte Teil von** A cgr(S)e(S) **zur Auswertung** von h **klein ist, verringern sich der Gesamtzeit- und Speicherplatzbedarf erheblich.**

9. Streaming-Algorithmen Wir

zeigen, dass dSHAÿ s auf dem verschachtelten Wort eines Hedges in einer Streaming- Weise ausgewertet werden können, sodass sie dieselben Ergebnisse und das gleiche Projektionsverhalten erzeugen, als würden sie auf Hedges im Speicher in einer Top-down- und von links nach rechts-Weise ausgewertet.

Alle Aspekte im Zusammenhang mit der Subhedge-Projektion bleiben unverändert. Die früheste Abfragebeantwortung im Streaming-Modus bedeutet, dass verschachtelte Wortsuffixe ignoriert werden, wenn sie irrelevant sind, und nicht nur verschachtelte Wörter wie bei der Subhedge-Projektion. Was sich zwischen beiden Auswertungsmodi ändert, ist der Speicherverbrauch. Für die In-Memory-Auswertung muss der gesamte Graph des Eingabe- Hedges gespeichert werden, während im Streaming-Modus bei jedem Ereignis nur ein Stapel und ein Zustand beibehalten werden. Zum Testen der regulären Sprachzugehörigkeit ist der Zustandsraum endlich, während der Stapelraum unendlich wird, da die Tiefe des Stapels von der Tiefe des Eingabe-Hedges abhängt.

Die Eigenschaft, äquivalente Streaming- und In-Memory-Auswerter zu haben, wurde bereits in [26] für Neumanns und Seidls Kellerwaldautomaten [25] und für verschachtelte Wortautomaten [11,15] festgestellt. Wir zeigen hier, wie man dieses Ergebnis auf **dSHAÿ** s **überträgt . Dies erlaubt es uns, die Korrektheit unserer Auswertealgorithmen basierend auf den Eigenschaften der konstruierten Automaten zu zeigen. Es erlaubt auch, Korrektheitseigenschaften von Streaming- Algorithmen zu erhalten, indem man über den In-Memory-Auswerter von SHAÿ s nachdenkt.**

9.1. Streaming-Auswerter für Downward-Hedge-Automaten **Wir**

definieren den Streaming-Auswerter eines SHAÿ A = **(ÿ, Q, ÿ, I,** F) **durch eine sichtbare Kellermaschine . Die Konfigurationsmenge dieser Maschine ist K** = **Q × Qÿ und enthält Zustandspaareund Zustandsstapel. Die sichtbare Kellermaschine liefert für jedes Wort** v **ÿ ÿˆ ÿ eine Streaming-Transitionsrelation vÿÿ ÿ ÿˆ ÿ** *ÿ* **ÿ Qÿ und Wörter v,** v $\overset{\text{str}}{\underset{ÿ}{}}$ **ÿ K × K, so dass für alle Zustände q, ÿ ÿ Q und Stapel** $\overset{ÿ}{}$ **:**

$$\frac{\text{WAHR}}{(q, ÿ)\ \underset{ÿÿ}{\overset{ÿ}{}}\ (q, ÿ)\ \textbf{bzgl. ÿ}} \qquad \frac{(q, ÿ)\ \textbf{vÿÿ}\ (Q^{ÿ}, ÿ) \qquad (Q^{ÿ}, ÿ)\ \underset{ÿÿ}{\overset{ÿ\,v}{}}\ (Q^{''}, ÿ)\ \textbf{bzgl. ÿ}}{(q, ÿ)\ \underset{ÿÿÿÿ}{\overset{ÿ\,v\cdot v}{}}\ (Q^{''}, ÿ)\ \textbf{bzgl. ÿ}}$$

$$\frac{q\ \textbf{einÿÿ}\ q^{ÿ}\ \textbf{in ÿ}}{q\ \textbf{einÿÿ}\ Q^{ÿ}\ \textbf{bzgl. ÿ}} \qquad \frac{\textbf{ÿÿ}\ q\ \textbf{ÿÿ}\ q^{ÿ}\ \textbf{in ÿ}}{(q, ÿ)\ \underset{ÿÿ}{\overset{ÿ}{}}\ (Q^{ÿ}, ÿ \cdot q)\ \textbf{bzgl. ÿ}} \qquad \frac{\textbf{q@p}\ ÿ\ q^{ÿ}\ \textbf{in ÿ}}{(p, ÿ \cdot q)\ \underset{ÿÿ}{\overset{ÿ}{}}\ (Q^{ÿ}, ÿ)\ \textbf{bzgl. ÿ}}$$

Wir stellen fest, dass dieselbe sichtbare Kellermaschine durch Kompilieren des SHAÿ in eine NWA erhalten werden kann, deren Streaming-Evaluator ebenfalls durch eine sichtbare Kellermaschine definiert ist (siehe Abschnitt 12.3 zu verwandten Arbeiten).

Das Konzept der partiellen Läufe des In-Memory-Evaluators ist eng mit Streaming -Läufen verwandt. Um dies zu sehen, stellen wir zunächst fest, dass jeder partielle Lauf v **bzgl. ÿ natürlich einen Stapel wie folgt definiert. Da** v **ein verschachteltes Wortpräfix ist, gibt es eindeutige verschachtelte Wörter v0, . . . ,** vn **ÿ nw(HÿÿQ), sodass**

$$v = \textbf{v0} \cdot ÿ \cdot \textbf{v1} \cdot ÿ \cdot \ \cdots \cdot ÿ \cdot Vn.$$

Diese verschachtelten Wörter müssen ebenfalls Teilfolgen von ÿ sein, sodass Zustände **p0, . . .** pn, **q0, . . . ,** qn **ÿ** Q **existieren , sodass** vi **für alle 0 ÿ** i **ÿ** n von pi **nach** qi **geht . Wir definieren den Stapel** *ÿ* **Sie, dass · · · von** v **durch · qnÿ1 und sagen, dass** v von **p0 nach** qn **geht und den Stapel ÿ hat. Beachten der Stapel** *ÿ* = **q0 · leer ist, wenn** v **ein verschachteltes Wort ist, da in diesem Fall** n = **0 ist, sodass** v = **v0 und** *ÿ* = **ÿ. Beispielsweise** hat **der Teillauf v0** = **0 ·** ÿ · **0 ·** Liste **· 1 ·** ÿ · **0 ·** Element **· 2 den Stapel ÿ0** = **0 · 1, während der Teillauf v1 =** v0 **· ÿ · 3 · ÿ · 4** den Stapel **ÿ1** = ÿ **hat .**

Machine Translated by Google

Lemma 20. Betrachten wir einen **SHA** A = **(ÿ, Q, ÿ, I,** F), der **q,** q als ÿ **ÿ** Q und einem Stapel *ÿ* **ÿ Qÿ .** Sei v mit Präfix eines verschachtelten Wortes in **nw(Hÿ)** angibt . Dann existiert ein partieller Lauf r bzgl . **ÿ** von q nach q ÿ Stapel *ÿ,* so dass projÿ (r) = v genau dann, wenn **(q,ÿ) (Qÿ,** ÿ) bzgl . ÿ.

Beweis. Dies wird durch Induktion über die Länge des Präfixes v durchgeführt . □

Dieses Lemma impliziert, dass jeder im Gedächtnis vorhandene Übergang von A **auf einer Absicherung** h **von** q **nach** q **identifiziert werden mit einem strömenden Übergang von** A **auf** nw(h) **von (q,ÿ) nach** (q ,ÿ).

h **Satz 8.** q ÿÿ q ÿ bzgl. **ÿ ÿ (q,ÿ)** nw(h) (Q ,*ÿ)* bzgl . ÿ.

Beweis. Wenn q **bzgl. ÿ, dann existiert eine Folge** R **von** h **mit ÿ ,** die mit q **beginnt und mit endet.** Q **Betrachten Sie die partielle Folge** v = nw(R) **auf h. Da** v **ein verschachteltes Wort ist, ist sein Stapel leer. Nach Lemma 20 haben wir (q,ÿ)** (Q ,*ÿ)* **bzgl. ÿ und** projÿ (v) = nw(h). **Für die inverse Implikation: Nehmen wir an, dass** (Q ,*ÿ)* **bzgl. ÿ. Sei** u = nw(h). **Nach Lemma 20 gibt es mit (q,ÿ) ein partieller Lauf** v **von ÿ von** q **nach** q **existiert ÿ. Dann gibt es eine Folge** R **von ÿ auf** h , **so dass** h , **dass** v = nw(R). **Daher ist** q ÿÿ q ÿ **bzgl. ÿ.** □

Für jedes deterministische dSHAÿ A, Hedge h und Zustand q kann der Streaming-Übergang auf nw(h) **bezüglich ÿ beginnend mit (q,ÿ)** in der Zeit **O(1) pro Buchstabe berechnet werden,** nach einer **Vorberechnung in der Zeit** O(|A|). **Die Gesamtberechnungszeit beträgt also O(|A|** + **|h|). Der zum Speichern einer Konfiguration benötigte Streaming-Speicher hat die Größe** O(Tiefe(h) + **|A|), da die Größe des sichtbaren Stapels durch die Tiefe des Eingabe-Hedges begre**

9.2. Hinzufügen einer Subhedge-Projektion

Als nächstes erweitern wir die Streaming-Transitionsrelation mit der Subhedge-Projektion, analog zu dem, was wir für die In-Memory-Transitionsrelation getan haben. Wir definieren eine Transitionsrel

Subhedge-Projektion h **ÿ K × K bezüglich ÿ, so dass für alle Wörter v,** v ÿÿ shp ÿ **ÿ ÿˆ ÿ ,**

Buchstaben a **ÿ ÿ, Zustände p, q,** q **ÿ Q und Stapel ÿ,** *ÿ* **ÿ Qÿ :**

$$\frac{q \ \ddot{y}\ Q\ddot{y}_{shp} \qquad v \ \ddot{y}\ nw(H\ddot{y})}{(q, \ddot{y})\ \ddot{v}\ddot{y}\ddot{y}\ (q, \ddot{y})\ \textbf{bzgl. } \ddot{y}\ shp}$$

$$\frac{q \ \ddot{y}\ddot{y}\ Q\ddot{y}_{shp} \qquad q \ \textbf{ein}\ddot{y}\ddot{y}\ q \ \textbf{in } \ddot{y}}{(q, \ddot{y})\ \textbf{a}\ddot{y}\ddot{y}\ (q\ shp, \ddot{y})\ \textbf{bzgl. } \ddot{y}} \qquad \frac{q \ \ddot{y}\ddot{y}\ Q\ddot{y}_{shp}}{(q, \ddot{y})\ \ddot{y}\ (q, \ddot{y})\ \textbf{bzgl. } \ddot{y}\ \ddot{y}\ddot{y}\ shp}$$

$$\frac{q \ \ddot{y}\ddot{y}\ Q\ddot{y}\ shp \qquad (q, \ddot{y})\ \textbf{v}\ddot{y}\ddot{y}\ (q\ shp, \ddot{y}\ \ddot{y}\ )\ \textbf{bzgl. } \ddot{y}\ (q\ \ddot{y}, \ddot{y}\ \ddot{y}\ )\ \ddot{y}\ddot{y}\ (q\ \text{schp}\ \text{"}, \ddot{y}\ \text{"})\ \textbf{bzgl. } \ddot{y}}{(q, \ddot{y})\ \ddot{y}\ddot{y}\ddot{y}\ (q\ \text{schp}\ \text{"}, \ddot{y}\ \text{"})\ \textbf{bzgl. } \ddot{y}}$$

$$\frac{q \ \ddot{y}\ddot{y}\ Q\ddot{y}_{shp} \qquad qq\ \ddot{y}\ddot{y}\ \ddot{y}\ddot{y} \ \textbf{in } \ddot{y}}{(q, \ddot{y})\ \ddot{y}\ddot{y}\ (q\ \text{schp}\ \ddot{y}, \ddot{y} \cdot q)\ \textbf{bzgl. } \ddot{y}} \qquad \frac{p \ \ddot{y}\ddot{y}\ Q\ddot{y}_{shp} \qquad \textbf{q@p } \ddot{y}\ q \ \textbf{in } \ddot{y}}{(p, \ddot{y} \cdot q)\ \ddot{y}\ddot{y}\ (q\ \text{schp}\ \ddot{y}, \ddot{y})\ \textbf{bzgl. } \ddot{y}}$$

Die projizierende Übergangsbeziehung bleibt in einer Konfiguration mit einem Subhedge-Projektionsstatus, bis das Ende des aktuellen Subhedges erreicht ist. Dies ist korrekt, da ein Subhedge-Projektionsstatus ohnehin nicht geändert werden kann.

Machine Translated by Google

Proposition 9. Für jedes **SHAÿ** A = **(ÿ, Q, ÿ, I,** F) gibt es in den Zuständen **ÿ ÿ** Q, verschachteltes Wort v, auf dem **ÿ q,** q keinen blockierenden Teillauf, der von q ausgeht:

(q,ÿ) vÿÿ str (Q ÿ ,ÿ) *bzgl.* ÿ **ÿ (q,ÿ) vÿÿ** (q shp str ÿ ,ÿ) bzgl . **ÿ.**

Beweis. Sei h **ÿ Hÿ die Absicherung, so dass** v = nw(h). **Der Beweis erfolgt durch Induktion über die Größe von v.**

Für die Implikation von links nach rechts nehmen wir an, dass (q,ÿ) vÿÿ Da v = nw(h), **beweist** str (Q ÿ ,ÿ) **bzgl . ÿ.**

Proposition 8 q h ÿÿ str Q ÿ **bzgl. ÿ.**

Fall q **ÿ Qÿ Nach Lemma 6 folgt aus** q shp. **folgende** h ÿÿ q str ÿ **bzgl . ÿ dass** q = q ÿ **. Wir können dann anwenden Regel, da** v **ein verschachteltes Wort ist:**

$$\frac{q \text{ ÿ } Q\ddot{y} \text{ shp} \qquad v = \mathbf{nw(H\ddot{y})}}{\text{str}}$$
$$\mathbf{(q,\ddot{y})\ v\ddot{y}\ddot{y}\ (q,\ddot{y})\ bzgl.\ \ddot{y}}\ \text{shp}$$

Daher ist (q,ÿ) vÿÿ str shp **(q,ÿ) bzgl. ÿ.**

Fall q **ÿÿ Qÿ** schp. **Wir unterscheiden alle möglichen Formen der Absicherung h.**

Unterfall h = h ÿ · H" **Seien** v) ÿ **und** nw(**dabei muss** " = nw(h **") , also** v = v ÿ „ · v = nw(h) . **folgende Regel angewendet worden sein:**

$$\frac{\mathbf{(q,\ddot{y})}\ \ddot{y}\ddot{y}^{\ddot{y}v}\ \text{str}\ (p,\ddot{y})\ \mathbf{bzgl.\ \ddot{y}}\ (p,\ddot{y}) \qquad \ddot{y}\ddot{y}\ddot{y}^{v"}\ \text{str}\ (Q^{\ddot{y}},\ddot{y})\ \mathbf{bzgl.\ \ddot{y}}}{\mathbf{(q,\ddot{y})}\ \ddot{y}\ddot{y}\ddot{y}^{"v\cdot v\ddot{y}}\ \text{str}\ (Q^{\ddot{y}},\ddot{y})\ \mathbf{bzgl.\ \ddot{y}}}$$

Da **v** ÿ **ein verschachteltes Wort ist, folgt, dass** *ÿ* = **ÿ. Nach der Induktionshypothese esis angewendet auf die verschachtelten Wörter v und** v **" erhalten wir (q,ÿ)** ÿ ÿ ÿ v str shp **(p,ÿ) bzgl. ÿ und (p, ÿ)** v ÿ ÿ ÿ " str (q ÿ ,ÿ)

shp **bzgl. ÿ. Daher können wir die folgende Regel anwenden:**

$$\frac{q\ \ddot{y}\ddot{y}\ Q\ddot{y}\ \text{shp} \qquad \mathbf{(q,\ddot{y})}\ \ddot{y}^{\ddot{y}v}\ \text{str}\ \mathbf{(p,\ddot{y})\ bzgl.\ \ddot{y}\ (p,\ddot{y})\ \ddot{y}\ddot{y}}\ \text{shp} \qquad \ddot{y}\ddot{y}\ddot{y}^{v"}\ \text{str}\ (q^{\ddot{y}},\ddot{y})\ \mathbf{bzgl.\ \ddot{y}}}{\mathbf{(q,\ddot{y})}\ \ddot{y}\ddot{y}\ddot{y}^{"v\cdot v\ddot{y}}\ \text{str}\ (q^{\ddot{y}},\ddot{y})\ \mathbf{bzgl.\ \ddot{y}}}\ \text{schp}$$

Dies zeigt (q,ÿ) vÿÿ (q shp str ÿ ÿ , ÿ) **nach Bedarf.**

Unterfälle h = ÿh ÿ **oder** h = a **oder** h = **ÿ. Unkompliziert.**

Für die Implikation von rechts nach links nehmen wir an, dass (q,ÿ) vÿÿ str (q ÿ ,ÿ) **bzgl. ÿ.** schp

Fall q **ÿ Qÿ** schp. **Dann müsste folgende Regel angewendet worden sein mit** q = q ÿ :

$$\frac{q\ \ddot{y}\ Q\ddot{y}\ \text{shp} \qquad v = \mathbf{nw(H\ddot{y})}}{\text{str}}$$
$$\mathbf{(q,\ddot{y})\ v\ddot{y}\ddot{y}\ (q,\ddot{y})\ bzgl.\ \ddot{y}}\ \text{shp}$$

Da wir annehmen, dass v **keine blockierenden partiellen Läufe mit ÿ hat, gibt es einen partiellen Lauf auf** v **mit ÿ ,** der mit **q beginnt .** Da v **ein verschachteltes Wort ist,**

Machine Translated by Google

Run auf v **ist das verschachtelte Wort eines Runs, also existiert ein Run auf dem Hedge** h , **der mit q beginnt . Dies zeigt die Existenz eines Zustands** p , **sodass** q ÿÿ p **bzgl . ÿ. Da** q ÿ **QÿH**

Lemma 6 zeigt q = **p, und damit** q ÿÿ q **bzgl. ÿ. Proposition 8 beweist dann** shp, **dass (q,ÿ)** $\overset{nw(h)}{ÿÿÿÿ}$ $\overset{str}{}$ **(q,ÿ) bzgl. ÿ, was gleich (q,ÿ) vÿÿ** ist $\overset{str}{}$ $(Q^{ÿ},ÿ)$ **bzgl. ÿ, je nach Bedarf.**

Fall q ÿÿ **Qÿ Wir unterscheiden alle möglichen Formen der Hecke h.** shp.

Unterfall h = h $\overset{ÿ}{}$ · H" **Sei** v . $\overset{ÿ}{}$ = nw(h $\overset{ÿ}{}$ **und** v . $\overset{"}{}$ = nw(h **ÿÿ) , also** v = v) $\overset{ÿ}{}$ „ · v = nw(h).

Die folgende Regel muss angewendet worden sein:

$$\frac{\text{q ÿÿ Qÿ shp} \qquad \textbf{(q,ÿ)} \ \overset{ÿ\,v}{} \ \overset{str}{} \ \text{(p, ÿ) bzgl. ÿ (p, ÿ) ÿÿ shp} \qquad \overset{v}{ÿÿÿ}\overset{"}{} \ \overset{str}{} \ \text{shp} \ (q \ \overset{ÿ}{},ÿ) \textbf{ bzgl. ÿ}}{\textbf{(q,ÿ)} \ \overset{ÿ\,v\ „\ v}{ÿÿÿ} \ \text{shp} \ \overset{str}{} \ (q\ ÿ \ \overset{ÿ}{},ÿ) \textbf{ bzgl. ÿ}}$$

Da **v** $\overset{ÿ}{}$ **ein verschachteltes Wort ist, folgt, dass** auch *ÿ = ÿ ist .* **Nach der Induktionshypothese esis angewendet auf die verschachtelten Wörter** $\overset{ÿ}{}$ **und** v **" erhalten wir (q,ÿ)** $\overset{ÿ\,v}{ÿ}$ $\overset{str}{}$ **(p,ÿ) bzgl. ÿ und (p, ÿ)** $\overset{v}{ÿÿÿ}\overset{"}{}$ $\overset{str}{}$ $(Q^{ÿ},ÿ)$ **bzgl. ÿ. Daher können wir folgende Regel anwenden:**

$$\frac{\textbf{(q,ÿ)} \ ÿÿ \ \overset{ÿ\,v}{} \ \overset{str}{} \ \textbf{(p,ÿ) bzgl . ÿ (p,ÿ)} \qquad \overset{v}{ÿÿÿ}\overset{"}{} \ \overset{str}{} \ (Q^{ÿ},ÿ) \textbf{ bzgl. ÿ}}{\textbf{(q,ÿ)} \ \overset{"\,v\,·\,ÿ}{ÿÿÿ} \ \overset{str}{} \ (Q^{ÿ},ÿ) \textbf{ bzgl. ÿ}}$$

Dies zeigt (q,ÿ) vÿÿ $\overset{str}{}$ $(Q^{ÿ},ÿ)$, **je nach Bedarf. ÿ**

Unterfälle h = ÿh $\overset{ÿ}{}$ **oder** h = a **oder** h = **ÿ. Unkompliziert.**

Damit sind die Beweise für beide Richtungen beendet. □

Ein Streaming-Evaluator mit Subhedge-Projektion für deterministische SHAÿ s A **auf Hedges** h **kann daher erhalten werden, indem man die Streaming-Transitionsrelation mit Subhedge-Projektion für** A **von** nw(h) **ausgehend von der Anfangskonfiguration berechnet. Dies kostet höchstens O(1)** Zeit **pro Buchstabe von** nw(h), **also konstante Zeit pro Ereignis des Streams.**

9.3. Streaming-Komplexität der frühesten Mitgliedschaft mit Subhedge-Projektion

Mithilfe von Streaming-Evaluatoren für SHAÿ s können wir die früheste Mitgliedschaft mit Subhedge-Projektion im Streaming-Modus testen, indem wir SHAÿ A cgr(S)e(S) ausführen .

Korollar 3. Für alle vollständigen **dSHAs (ÿ, Q, ÿ, I,** F), F **ÿ FS ÿ** Q, Sprache L = L(A) und Schema **S** = **L(A[F/ FS]),** kann die früheste Mitgliedschaft mit vollständiger Subhedge-Projektion für L bzgl. **S** im Streaming-Modus für jeden Hedge h **ÿ S** in der Zeit **O(|1|)** pro nicht projiziertem Buchstaben von nw(h) nach einer Vorverarbeitungszeit von **O(|A** cgr(S)e(S) **|** + n 3d + n s) getestet werden, wobei d die Anzahl von (S) ist. Der in Differenzrelationen und s ist die Anzahl der sicheren Teilmengen in $\overset{cgr}{}$ Zuständen von **Q e(S)** erforderliche Speicherplatz liegt O(Tiefe(h) + |A cgr(S)e(S) |).

Beweis. Nach Theorem 4 ist **dSHAÿ** A cgr(S)e(S) **das früheste und vollständigste für die Subhedge-Projektion. Nach Proposition 8 und 9 können wir also einen Streaming-Mitgliedschaftstester mit Subhedge-Projektion für einen Hedge** h **ÿ S** erhalten , **indem wir den Streaming-Evaluator mit Subhedge-Projektion von** A cgr(S)e(S) **auf dem verschachtelten Wort** nw(h) ausführen. **Da dieser SHAÿ** deterministisch **ist , kann die Streaming-Evaluierung in der Zeit O(|1|) pro Buchstabe von** nw(h) **nach einer Vorverarbeitungszeit von O(|A** cgr(S)e(S) **|** + n 3d + n s) durchgeführt werden. □

Wie bei der Safe-No-Change-Projektion kann die exponentielle Vorverarbeitungszeit auch hier vermieden werden, indem der Teil des SHAÿ A cgr(S)e(S), **der für die Bewertung der Absicherung benötigt wird, spontan erstellt wird. Die Größe des benötigten Teils liegt in** O(|h|). **Daher kann der Platzbedarf auch**

Machine Translated by Google

auf O(|h|) **reduziert werden ,** was für große Ströme nw(h) **zu viel sein kann . Der benötigte Teil von** A cgr(S)e(S) **kann jedoch viel kleiner als |h| sein. Darüber hinaus muss ein Stapel der Größe** depth(h) **beibehalten werden, aber diese Abhängigkeit von** h **kann sogar für große Ströme nw(h) akzeptabel sein .**

10. Monadische reguläre Abfragen und XPath

Als nächstes betrachten wir das Problem, wie monadische reguläre Abfragen zu Hedges frühestmöglich und mit Subhedge-Projektion beantwortet werden können. Jede monadische Abfrage definiert eine Reihe von Positionen für jeden Eingabe-Hedge. Die XPath- Abfragen self::list[child::item] können **beispielsweise** auf den Hedge ÿlist.ÿlistÿ.ÿitemÿÿ.ÿlist.ÿitemÿÿ **angewendet werden , wo sie die Position des ersten obersten Baums auswählen, der mit** list beschriftet ist, da dieser einen mit item **beschrifteten Kindbaum enthält . Wir identifizieren die Positionen von Hedges durch Ganzzahlen. Beispielsweise wird das ausgewählte** Listenelement **im obigen Hedge durch die Ganzzahl 2 identifiziert. Daher ist die Antwortmenge der obigen Abfrage zum obigen Hedge {2}.** Beachten Sie, dass monadische Abfragen **im Gegensatz zur Subbaumsemantik von XPath im offiziellen** W3C-Standard **keine weiteren Informationen über die ausgewählten Positionen zurückgeben .**

Die Beantwortung dieser Abfrage ist komplizierter als die Überprüfung, ob der XPath- Filter [self::list/child::item] an der Wurzel des ersten Teilbaums des Hedges übereinstimmt, da auch der Antwortsatz erstellt werden muss. Was den Speicherverbrauch betrifft, müssen aktive Antwortkandidaten für die Beantwortung monadischer Abfragen gepuffert werden. Daher kann der Speicherverbrauch von Streaming-Auswertern monadischer Abfragen linear mit der Größe des Streams wachsen, je größer die Anzahl der aktiven Kandidaten ist.

10.1. Monadische reguläre Abfragen

Monadische Abfragen auf Hedges in Hÿ können mit Sprachen identifiziert werden, die annotiert auf Hedges in Hÿÿ{x} ,sind, wobei x ÿÿ ÿ **eine beliebig feste Auswahlvariable ist. Eine monadische Abfrage auf Hedges in Hÿ** kann dann mit der Sprache aller annotierten Hedges in **Hÿÿ{x} identifiziert werden , in denen eine einzelne Position x-annotiert wurde und diese Position durch die Abfrage ausgewählt wird.**

Daher können reguläre monadische Abfragen für Hedges in Hÿ durch verschachtelte reguläre **Ausdrücke in** nRegExpÿÿ{x} **oder durch einen dSHA mit derselben Signatur definiert werden .** Die reguläre **XPath- Abfrage self::list[child::item] kann beispielsweise durch den folgenden verschachtelten regulären Ausdruck definiert werden :**

$$\ddot{y}\text{Liste} \cdot x \cdot \ddot{\mathbf{y}} \cdot \ddot{y} \cdot \text{Element} \cdot \ddot{\mathbf{y}} \cdot \ddot{\mathbf{y}} \cdot \ddot{y}\ddot{y} \cdot \ddot{\mathbf{y}},$$

wobei ÿ wie ein Platzhalter ist, der jede Absicherung in Hÿ ohne x akzeptiert, d. h. für eine Rekursionsvariable z ÿ **V :**

$$\ddot{\mathbf{y}} = \mu z.(+a\ddot{y}\ddot{y}a + \ddot{y}z\ddot{y}) \quad {}^{\ddot{y}} \cdot$$

Beachten Sie, dass der verschachtelte reguläre Ausdruck für den XPath- Filter [self::list/ child::item] in nRegExpÿ **aus dem der XPath- Abfrage self::list[child::item] durch Entfernen des x erhalten werden kann. Darüber hinaus müssen wir für jede monadische Abfrage von Hedges die annotierten Hedges einschränken, die von der Sprache der Abfrage erkannt werden sollen, sodass genau eine Position in jedem Hedge mit x annotiert ist .** Dies kann durch Schnittmenge mit dem **folgenden verschachtelten regulären Ausdruck** onex **erreicht werden :**

$$\mu z.\ (TxT + T.\ddot{y}z\ddot{y}.T)$$

Das bisherige Schema doc **für Mustervergleich muss nun in S** = docx **ÿ** onex geändert werden, **wobei** docx **ÿ** nRegExpÿÿ{x} wie doc **ist , aber über der Signatur ÿ ÿ {x}.**

Für die reguläre XPath-Abfrage self::list[child::item] erhalten wir den dSHA in Abbildung 22. Die Übergangsregeln mit dem Sonderbuchstaben x geben an, wo Positionen an die **Auswahlvariable gebunden werden können :** x **kann nur im Zustand 2 gebunden werden, also in Teilbäumen, die mit einem** Listenelement beginnen . **Derselbe Automat kann verwendet werden, um das Schema S** = docx **ÿ** onex zu definieren , **indem man die Schema-Endzustände FS** = **{14, 20} anstelle der Endzustände** F = **{20} verwendet.** Tatsächlich haben wir den Automaten konstruiert, indem wir den aus dem **XPath-** Ausdruck **erhaltenen Automaten** mit den aus den regulären Ausdrücken docx **und** onex **berech**

Machine Translated by Google

das Schema. Beachten Sie, dass 20 ein Endzustand für die Abfrage (und das Schema) ist, während 14 ein Senkenzustand für die Abfrage (aber nicht für das Schema) ist.

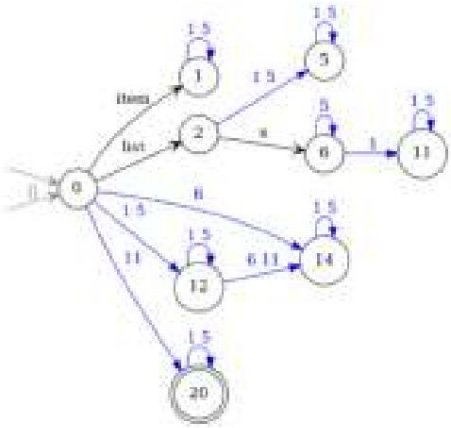


Abbildung 22. Der dSHA für die XPath -Abfrage self::list[child::item]. **Die Auswahlposition wird durch x angezeigt.**

10.2. Früheste Abfragebeantwortung mit Subhedge-Projektion **Ein naiver**

Abfrageauswertungsalgorithmus für einen Hedge-Automaten mit Signatur ÿ ÿ {x} **erhält einen Eingabe-Hedge** h **ÿ Hÿ, errät alle möglichen Positionen, wo** x **in h eingefügt werden kann, und prüft für alle, ob der erhaltene annotierte Hedge von A akzeptiert wird. Weniger naiv kann** x **nur an den Positionen eingefügt werden, wo der aktuelle Zustand einen ausgehenden** x- **Übergang hat. All dies kann verbessert werden, wenn der betrachtete Hedge-Automat Auswahl-, Ablehnungs- oder Projektionszustände unterstützt. In allen Fällen muss der Algorithmus alle Bindungskandidaten zwischenspeichern , die sich noch nicht in einem Auswahl- oder Ablehnungszustand befinden. Im schlimmsten Fall muss er die Bindungen an alle Positionen zwischenspeichern. Daher ist der Zustandsraum des Auswerters monadischer Abfragen nicht mehr endlich.**

Auswahl- und Ablehnungszustände können durch Anpassung des frühesten Abfragebeantwortungsalgorithmus aus [21] identifiziert werden. Dies kann so gemacht werden, dass man einen SHAÿ Ae(S) **mit Signatur ÿ ÿ** {x} erhält , **abhängig vom Schema. Es ist zu beachten, dass die in [21]** verwendeten Automaten **dNWAs statt SHAÿ** s **sind und dass Schemata dort ignoriert werden. Die erforderlichen Anpassungen werden daher in Anhang A erörtert .** Wie man aus Ae(S) **einen effizienten Abfragebeantwortungsalgorithmus im Streaming-Modus** erhält , **ist das Hauptthema von [21]. Dies ist der Basisalgorithmus, den wir für unsere Experimente mit Subhedge-Projektion erweitern werden.**

Um dem Algorithmus zur frühesten Abfragebeantwortung eine Kongruenzprojektion hinzuzufügen, müssen wir den Automaten A cgr (S) ausführen. **Die früheste Kongruenzprojektion für die XPath- Abfrage e(S) self::list[child::item] ist in Abbildung 23** dargestellt. **Sie bindet** x **im Zustand 2D1,** nachdem ein **Listenelement der obersten Ebene gelesen wurde ,** und prüft dann, ob ein untergeordnetes Element item vorhanden ist. Sobald es gefunden wurde, **wechselt sie in den Auswahlzustand 1D3 .** Alle anderen untergeordneten Elemente des Listenelements **der obersten Ebene werden im Zustand 2D3 projiziert . Um der Projektion die früheste Self-No-Change-Abfragebeantwortung hinzuzufügen, müssen wir statt** A cgr den Automaten A **e(S)** ausführen . **Wir geben keine Details an, behaupten aber** für die vorliegende Abfrage **e(S) , dass die Projektionsleistung**

Machine Translated by Google

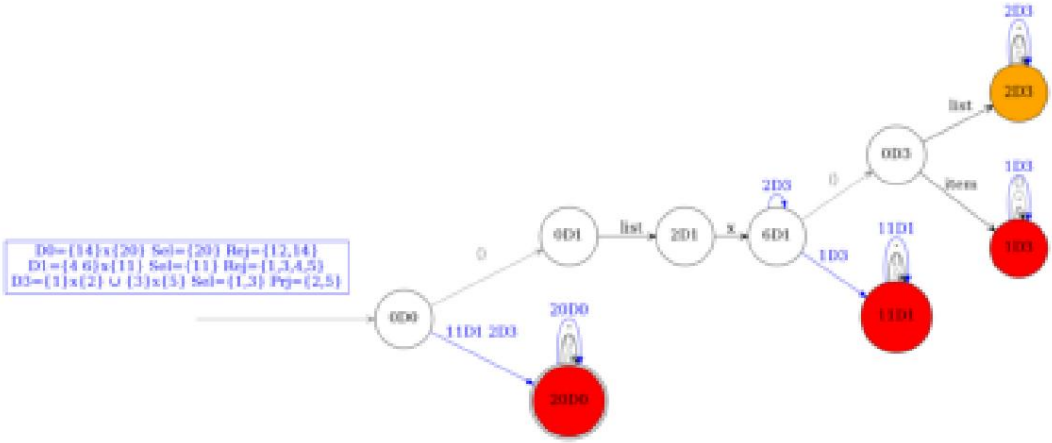


Abbildung 23. Die früheste Kongruenzprojektion A cgr(doc ÿ onex)e(docÿonex) **für den dSHA** A **in Abbildung 22 mit FS = {14, 20}. Sie definiert die Abfrage des XPath** self::list[child::item].

Glücklicherweise beeinflussen die Soundness- und Vollständigkeitssätze für den frühesten Mitgliedschaftstest mit Subhedge-Projektion nur die Konstruktion dieser SHAÿ s, nicht aber, wie sie ausgewertet werden sollen. Es gibt jedoch ein zusätzliches Problem: Verschachtelte Wortpräfixe können nicht als irrelevant für das nachfolgende Subhedge betrachtet werden, wenn die Auswahlvariable x dort gebunden **werden kann , selbst wenn die Akzeptanz nicht davon abhängt, wo sie gebunden wird. Dies liegt daran, dass die Position der Bindung von jedem Abfragebeantwortungsalg**

Definition 15. Eine Bindung des Präfixes v nennen wir **irrelevant** für L und **S ,** sofern es keine Absicherung h gibt, die x und das Suffix w enthält, so dass v · w **ÿ** nw(S) und v · nw(h) · w **ÿ** nw(L).

Unser Subhedge-Projektions-Evaluator kann nur auf subhedge-irrelevante Präfixe projizieren, die auch bindungsirrelevant sind. Gegeben sei ein dSHA A **und ein Satz von Schema-Endzuständen FS. Man kann entscheiden, ob ein Präfix subhedge-irrelevant ist, wenn der aktuelle Zustand** q **Hedges akzeptiert ,** die x enthalten , **aber auf einen anderen Zustand zugreifen kann, der nicht sicher für die Ablehnung ist und x-Bindungen zulässt . Im bodennahen Datalog können wir bindungsirrelevante Zustände** q unterscheiden , **indem wir die folgenden Regeln für alle q,** q auferlegen

binding_irrelev(q) :- state(q), nicht binding_relev(q). binding_relev(q) :- nicht bind_x(q), acc(q,q'), bind_x(q'), nicht rej(q').

Beachten Sie in unserem Beispiel der XPath -Abfrage self::list[child::item], dass der Zustand 2D3 von dSHAÿ A cgr(doc **ÿ** onex)e(docÿonex) **in Abbildung 23 für die Bindung irrelevant ist, da** x an das Listenelement **der obersten Ebene im Zustand 2D1** gebunden sein muss und daher nicht an ein darunter liegendes Listenelement **gebunden werden kann .**

11. Experimente mit Streaming-XPath-Auswertung Wir

vergleichen experimentell drei Streaming-Auswerter für dSHA, für die früheste monadische Abfragebeantwortung, ohne Subhedge-Projektion, mit Safe-No-Change-Projektion und mit Kongruenzprojektion.

11.1. Benchmark **dSHA** für XPath-Abfragen

Wir beginnen mit dSHA und definieren monadische Abfragen für die regulären XPath- Abfragen A1-A8 aus dem XPathMark-Benchmark [7] , die in Tabelle 1 aufgeführt sind . Diese XPath- Abfragen zeigen die meisten Merkmale des regulären XPath-Fragments. In Tabelle 2 haben wir 14 weitere XPath-Pfadabfragen hinzugefügt, die wir ebenfalls für Tests als nützlich erachtet haben.

Machine Translated by Google

Tabelle 1. Normale XPath-Abfragen A1-A8 aus der XPathMark-Sammlung.

Abfrage-ID	XPath-Ausdruck
A1	/site/geschlossene_auktionen/geschlossene_auktion/anmerkung/beschreibung/ Text/Schlüsselwort
A2	//geschlossene_Auktion//Schlüsselwort
A3	/site/geschlossene_auktionen/geschlossene_auktion//Schlüsselwort
A4	/site/geschlossene_auktionen/geschlossene_auktion[Anmerkung/Beschreibung/ Text/Schlüsselwort]/Datum
A5	/site/geschlossene_auktionen/geschlossene_auktion [Nachkomme::Schlüsselwort]/Datum
A6	/site/people/person[Profil/Geschlecht und Profil/Alter]/Name
A7	/site/people/person[Telefon oder Homepage]/Name
A8	/site/people/person[Adresse und (Telefon oder Homepage) und (Kreditkarte oder Profil)]/Name

Tabelle 2. Zusätzliche reguläre XPath-Abfragen für die XPathMark-Dokumente.

A0	/Website
A1_0a	/Website/*
A1_0b	/Website/@*
A1_0c	/Website//@*
A1_1a	//Bieter/Personref[beginnt mit(@Person,'Person0')]
A1_1d	//Bieter/Personenreferenz[@Person='Person0']
A1_2	//Person
A1_3	/site/regionen/afrika//@*
A1_4	/site/regionen/afrika/*
A1_5	/site/regionen/*
A1_6	//geschlossene_Auktion/Annotation//Schlüsselwort
A2_1	//geschlossene_Auktion[Nachkomme::Schlüsselwort]
A4_0	/site/geschlossene_auktionen/geschlossene_auktion[Anmerkung]/Datum
A4_1	/site[offene_auktionen]/geschlossene_auktionen

Deterministische SHAs für A1-A8 wurden bereits in [21] beschrieben. Für die anderen XPath-
Abfragen haben wir **über verschachtelte reguläre Ausdrücke zu dSHAs** kompiliert.
Um die Eingabe -dSHAs für unsere Evaluatoren zu erzeugen, schneiden wir diese Automaten
mit einem dSHA für das Schema- Dokumentÿ $_{x}$ **ÿ** onex **wobei** docÿ $_{x}$ **ist das Schema, das wir für Hedge verwenden**
Kodierungen von realen XML- Dokumenten mit x- **Annotationen. Dadurch konnten wir identifizieren**
die Schema-Endzustände FS. Wir haben das Ergebnis minimiert und vervollständigt. Tabelle 3 gibt die Größe an der Eingabe -dSHAs für unsere Evaluatoren, die mit dem obigen Verfahren erhalten wurden. Für jedes dSHA A, Die Größe wird durch zwei Zahlen size(n) angegeben, die erste für die Gesamtgröße und die zweite für die Anzahl der Staaten n.

Für den Eingabe -dSHA A **jeder Abfrage haben wir statisch den gesamten dSHAÿ s** berechnet
A cgr(S)e(S) **unter Verwendung der notwendigen Teile des Determinationsalgorithmus aus [16]. Die**
Die Größe der erhaltenen **dSHAÿ s und die Anzahl** d **der Differenzrelationen sind in Tabelle 3** angegeben.
auch. Die größte Größe ist 2091(504) für den Eingangs- dSHA für A8. Die größte Anzahl von Differenz
Die Relationen d = 24 werden auch für diese Abfrage erhalten. Die Größe dieser Automaten ist also tatsächlich viel

Machine Translated by Google

kleiner als man es von einer Konstruktion erwarten würde, die im schlimmsten Fall hoch exponentiell ist
Fall. Die einzige Abfrage, für die wir keine entsprechende deterministische Abfrage erhalten konnten,
früheste Kongruenzprojektion A cgr(S)e(S) **ist A5.**

Die Berechnungszeit der frühesten Kongruenzprojektion betrug zwischen 2,3 und 26 s.

Tabelle 3. Gesamtgröße und Anzahl n **der Zustände des Eingangs- dSHA** A **für die Abfrage mit schema-final Zustände FS, die Gesamtgröße und Anzahl der Zustände des SHAÿ** A cgr(S)e(S), **erhalten durch früheste Kongruenz Projektion und die Zahl** d **ihre Differenzrelationen.**

Abfrage dSHA A Größe(n)	Größe(#Staat)	Rel. d ID	dSHAÿ #Diff- ID	Abfrage dSHA A	dSHAÿ #Diff-Größe(n)	Größe(#Status)	Rel. d
A1 482(68)	1296(324)	16	A1_0c	230(43)	238(62)		6
A2 224(42)	316(82)	7	A1_1a	305(54)	384(101)		8
A3 320(53)	662(156)	10	A1_1d	305(54)	382(101)		8
A4 629(74)	1651(404)	18	A1_2	194(39)	152(42)		5
A5 438(63)	1226(269)	13	A1_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)	5	A2_1	232(42)	295 (78)		7
A1_0a 224(42)	145(44)	6	A4_0	392(59)	719(184)		12
A1_0b 203(39)	68(23)	5	A4_1	292(49)	287(78)		8

Wir weisen darauf hin, dass wir die vollständigen **dSHAÿ s für die Safe-No- Change-Projektion** A snc,
die reine Kongruenzprojektion A cgr(S) **und die früheste Abfragebeantwortung noch nicht statisch berechnet habe**
Ae(S) . **Wir glauben, dass diese Automaten größer werden könnten als** A cgr(S)e(S) , **den wir in einem einzigen Versuch konstruiert haben.**

11.2. Streaming-Auswertungstool: AStream

Wir verwenden und entwickeln das AStream-Tool zur Beantwortung monadischer dSHA- Abfragen auf XML- Streams mit Schemabeschränkungen. Version 1.01 von AStream wurde in [21] vorgestellt.
unterstützt die früheste Abfragebeantwortung ohne Projektion. Gegeben sei ein dSHA A, **das eine monadische**
Abfrage, eine Reihe von Schema-Endzuständen FS und ein XML- Stream w, konstruiert es on-the-fly die
benötigte einen Teil des SHAÿ Ae(S) **beim Parsen von** w **und Auswerten** von A **auf der Hedge-Kodierung w.**

Für die FCT-Konferenzversion des vorliegenden Papiers [24] haben wir AStream erweitert mit
Safe-No-Change-Projektion. Dies führt zur Version AStream 2.01. Sie konstruiert die früheste Safe- No-Change-
Projektion dSHAÿ $A^{snc}_{e(S)}$ **on the fly während der Auswertung der definierten monadischen Abfrage**
durch A **auf der Hedge-Kodierung des XML- Streams w. Subhedge-Projektion kann eingeschaltet werden oder**
aus, um beide Versionen ohne weitere Implementierungsunterschiede vergleichen zu können.

Für die vorliegende Zeitschriftenversion haben wir die früheste Kongruenzprojektion hinzugefügt und in
AStream integriert, was zu AStream 3.0 führte.

AStream 3.0 unterscheidet sich von den beiden vorherigen Versionen dadurch, dass der dSHAÿ $A^{cgr}_{e(S)}$ **Ist**
statisch und vollständig unabhängig von der Eingangsabsicherung aufgebaut. Wir verwenden dann eine generische
frühester Streaming-Evaluator für SHAÿ s, der in Ablehnungszuständen ablehnt und in Auswahlzuständen auswählt
Staaten und Projekte in allen Subhedge-Projektionsstaaten. Dieser Evaluator könnte auch mit
$A^{snc}_{e(S)}$ **oder** Ae(S) , **solange diese nicht zu groß werden.**

Es ist zu beachten, dass $A^{cgr}_{e(S)}$ **erwies sich für unseren gesamten Benchmark als angenehm klein,**
mit den anderen dSHAÿ s $A^{snc}_{e(S)}$ **und** Ae(S) **riskieren, größer zu werden. Wie bereits erwähnt, haben wir nicht**
konstruieren diese dSHAÿ s bisher für unsere Benchmark-Abfragen. Aus diesem Grund haben wir weiterhin

Machine Translated by Google

das früheste Streaming mit Safe-No-Change-Projektion mit AStream 2.01, während wir AStream 3.0 für das früheste Streaming mit Kongruenzprojektion verwendet haben.

Alle AStream-Versionen verlassen sich auf Javas abc-Datalog, um die kleinsten Fixpunkte von Datalog-Programmen von unten nach oben zu berechnen. Datalog-Programme werden für alle logischen Schlussfolgerungen benötigt: zum Berechnen von Teilmengen von Zuständen, die auf allen Ebenen sicher für Ablehnung, Auswahl oder Subhedge- Projektion sind. Auch die Differenzrelationen werden basierend auf Datalog berechnet. Beachten Sie, dass die früheste Abfrageauswertung im laufenden Betrieb die Ausführung von Datalog während der Abfrageauswertung erfordert, während bei einem statischen Ansatz zum vollständigen Erstellen von dSHAÿ s Datalog nur zur **Vorverarbeitungszeit während der Automatenkonstruktion b**

11.3. Bewertungsmaßnahmen

Wir möchten die Zeit messen, die für die Ausführung der drei Streaming-Abfrage-Evaluatoren für alle Aus den XPath-Ausdrücken in der Sammlung in den Tabellen 1 und 2 erhaltene **dSHAs .**

Ein Vorteil des XPathMark-Benchmarks ist, dass er mit einem Generator für XML- Dokumente geliefert wird, auf die die Abfragen angewendet werden können und deren Größe skaliert werden kann. Wir haben ein XML- Dokument mit einer Größe von 1,1 GB erstellt, das groß genug ist, um Streaming obligatoris Unsere Experimente zeigen, dass die Effizienz der Abfrageauswertung linear mit der Größe des nicht projizierten Teils des XML- Dokuments wächst. Dies gilt für jeden unserer drei Ev Daher würde die Messung der Zeit des Evaluators bei XML- Dokumenten anderer Größen keine neuen Erkenntnisse bringen, außer dass auch der Speicherverbrauch moderat bleibt.

Der Zeitgewinn durch Projektion für eine Abfrage Q **ist der prozentuale Zeitgewinn durch Projektion bei der Abfrageauswertung, wenn man die reine Dokumentanalysezeit** tparse **Sekunden außer Acht**

$$\text{ZeitgewinnQ} = \mathbf{100\ ÿ\ (1\ ÿ}\ (\text{tnoprojectQ}\ \mathbf{ÿ}\ \text{tparse})/(\text{tprojectQ}\ \mathbf{ÿ}\ \text{tparse})).$$

Wir können den Zeitgewinn für die früheste Safe-No-Change-Projektion von AStream 2.01 und für die früheste Kongruenzprojektion von AStream 3.0 messen, da die Projektion in beiden abgeschaltet werden kann. Der Nachteil des Zeitgewinns besteht darin, dass er von den Details der Implementierung abhängt .

Ein besseres Maß für die Projektionsleistung ist der Ereignisgewinn. Der Parser sendet einen der benötigten Ereignisse an den SHAÿ , Datenstrom und ignoriert dabei Teile des Eingabedokuments, die wegprojiziert werden können. Dazu muss er vom Automaten immer darüber informiert werden, welche Art von Ereignissen im aktuellen Zustand projiziert werden können. Der Ereignisgewinn der Projektion event-gainQ ist dann der Prozentsatz der Ereignisse, die als irrelevant erkannt und daher von einem **Projektionsauswerter für die Abfrage** Q **nicht mehr erstellt werden . Man könnte erwarten, dass der Zeitgewinn immer kleiner ist als der Ereignisgewinn.**

$$\text{time-gainQ}\ \mathbf{ÿ\ event\text{-}gainQ}$$

Tatsächlich wird dies durch unsere Experimente bestätigt. Wir glauben, dass diese Diskrepanzen zwischen diesen beiden Maßen anzeigen, wie viel Spielraum für die Optimierung der Implementierung eines Evaluators bleibt. Auf diese Weise können wir feststellen, dass bei AStream 3.0 noch Raum für weitere Optimierungen bleibt.

11.4. Experimente

Wir haben Javas XMLStreamReader- **Schnittstelle v1.0 aus dem Paket** javax.xml.stream verwendet, um **XML- Dateien in Scala zu parsen und zu streamen. Das Parsen eines 1,1 GB großen XML- Dokuments erforderte** tparse = 15 s, während die Abfrage ohne Projektion im Durchschnitt tnoprojectavg = 752 s **dauerte , wobei die Zeit zwischen 600 und 900 s variierte . Die durchschnittliche Verarbeitungszeit für 1 MB betrug somit 0,72 s. Nachdem wir dies wussten, haben wir die erwartete Auswertungszeit anhand der Größe des nicht projizierten Teils des XML - Dokuments vorhergesagt. Sie lautet wie folgt:**

$$\mathbf{EreignisgewinnQ\ ÿ\ 1100\ ÿ\ 0{,}72\ s} + \text{tparse}\ \mathbf{s.}$$

Ohne Projektion generiert der Parser 1.012.018.728 Ereignisse für das 1,1 GB große XML- Dokument. Dies ist die Basislinie zur Berechnung des Event-GainQ für alle Abfragen Q.

Machine Translated by Google

Wir haben AStream mit Scala v2.13.3 auf einer Maschine mit dem Betriebssystem Ubuntu ausgeführt 20.04.06 LTS (64-bit). Bei dem Rechner handelt es sich um einen Dell Precision-7750, ausgestattet mit einem Intel® Core™ i7-10875H CPU @ 2,30 GHz × 16 und 32 GB RAM.

11.4.1. Früheste Kongruenzprojektion

Wir präsentieren die Maßnahmen unseres Bewerters mit der frühesten Kongruenzprojektion in Tabelle 4. Der Ereignisgewinn ist hoch und variiert zwischen 75,7 % für A1_0c und 100 % für A0, A1_0a, A1_0b, A1_4, A1_5 und A4_1. Der Ereignisgewinn für Abfragen ohne Nachkommenachse liegt über 98,2 %. Bei diesen Abfragen können alle Teilbäume, die nicht auf dem Hauptpfad der Abfrage liegen, wegprojiziert werden. Insbesondere müssen nur Teilbäume bis zu einer festen Tiefe untersucht werden. Diese sind die Abfragen A1, A4, A6-A8 und A4_0 und die oben aufgeführten Abfragen mit Ereignisgewinn 100 %. Der Zeitgewinn für all diese Abfragen ist etwas geringer als der Ereignisgewinn, aber immer noch darüber 98,2 %.

Tabelle 4. Zeitgewinn und Ereignisgewinn durch früheste Übereinstimmungsprojektion mit AStream3.01 auf einer 1,1 GB XPathMark-Stream.

Abfrage	#Ans- Wer	Zeit in s	Zeit Gain in s	Ereignis Gewinnen	Abfrage	#Ans- Wer	Zeit in s	Zeit Gain in s	Ereignis Gewinnen
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 %

Als nächstes betrachten wir den zweiten Typ von Abfragen mit der absteigenden Achse, nämlich A2. A1_0c, A1_1a, A1_1d, A1_2, A1_6 und A2_1. Intuitiv wird ein geringerer Ereignisgewinn erwartet in diesen Fällen, da die Subhedge-Projektion mit der absteigenden Achse nicht direkt ausschließen kann ein vollständiger Teilbaum unter einem bestimmten Element; alle untergeordneten Elemente müssen überprüft werden. Beispielsweise muss die Abfrage A1_0c gleich /site//@* **alle Attribute aller Elemente auswählen unter der Stammelementsite , was die Überprüfung jedes XML- Elements im Dokument erfordert, für die Attributpräsenz, mit Ausnahme der** Site des Stammelements.

Der für die vorgenannten Abfragen gemeldete geringere Ereignisgewinn stimmt mit den obigen Erwartung, die Prozentsätze von 75,7 % bis 81,1 % zeigt. Außerdem zeigen diese Abfragen eine größere Lücke zwischen Zeitgewinn und Ereignisgewinn, die zwischen 2,2 % und 5,7 % liegt und Gegenstand zukünftiger Optimierungen sein. Es ist anzumerken, dass einige dieser Abfragen mehrere Features gleichzeitig, wie die Abfragen A1_1a und A1_1d, die eine absteigende Achse haben, Filter mit Attributen und Stringvergleich.

Die einzige Ausnahme von diesem Trend für die Abfragen mit absteigender Achse sind A3 und A1_3, mit entsprechenden Ereignisgewinnen von 97,8 % und 99,8 %. Der Grund für diesen hohen Gewinn ist dass diese Abfragen von der Wurzel aus beginnen, mit einem untergeordneten Achsenpfad der Tiefe 3, bevor die Nachkommenachse. Dieser anfängliche Pfad ermöglicht den Ausschluss und die Projektion der meisten Elemente Untersite , die nicht mit ihren gesamten Teilbäumen in den angegebenen Pfad passen.

Machine Translated by Google

11.4.2. Früheste sichere Projektion ohne Änderungen

Wir vergleichen die früheste Safe-No-Change-Projektion mit der frühesten Kongruenzprojektion in Tabelle 5. Wir beschränken uns dabei auf die Abfragen A1–A8 aus dem XPathMark in Tabelle 1.

Die XPath- Abfragen A1, A4, A6, A7 und A8 haben keine untergeordneten Achsen. Die Zeit Der Gewinn für die sichere No-Change-Projektion bei diesen Abfragen liegt zwischen 92,3 und 98,9 %. Für die frühesten Kongruenzprojektion war der Zeitgewinn für A1, A4 und A6 mit mindestens 96,6 % besser. Für A7 und A7 liegt der Zeitgewinn der Safe-No-Change-Projektion nahe am Ereignisgewinn von Kongruenzprojektion, aber der Zeitgewinn der Kongruenzprojektion ist 1,5% geringer. Wir hoffen, dass eine optimierte Version der Kongruenzprojektion die Safe-No-Change-Projektion übertreffen könnte. Projektion in allen Fällen.

Die XPath- Abfragen mit Nachkommenachse sind A2, A3 und A5. Für A2 und A3 in Insbesondere ist der Zeitgewinn der Safe-No-Change-Projektion mit nicht einmal 10,9% sehr gering, während Kongruenzprojektion ergibt mindestens 77,9%. Wir glauben, dass dies ein Fehler in unserem aktuellen Implementierung der Safe-No-Change-Projektion, die möglicherweise die Projektion von Attribute und Textknoten für diese Abfragen (aber nicht für die anderen). Kongruenzprojektion auf A2 gewinnt 77,9 % der Zeit und 81,1 % der Ereignisse. Das ist viel besser, aber immer noch weit entfernt von der Projektionskraft für Abfragen ohne Nachkommenachsen. Für A3 gewinnt die Kongruenzprojektion 97,4 % der Zeit. Dies ist viel besser als bei A2, da die absteigende Achse in A3 am Ende liegt, während sie für A2 am Anfang des Pfades liegt. Aber auch für A2 ist die Kongruenzprojektion sehr hilfreich.

Tabelle 5. Zeitgewinn für die früheste sichere Nichtänderungs- und früheste Übereinstimmungsprojektion sowie der Ereignisgewinn der frühesten Übereinstimmungsprojektion.

Abfrage AUSWEIS	Zeit (s) Snc	Zeit (s) Kongress	Zeitgewinn Snc	Zeitgewinn Kongress	Ereignisgewinn Kongress
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. Vergleich mit externen Tools

QuiXPath erwies sich als das effizienteste externe Tool mit großer Reichweite für regelmäßige XPath- Auswertung auf XML- Streams [4]. Daher vergleichen wir die Effizienz der Kongruenz Projektion auf QuiXPath in Tabelle 6 und damit durch Transitivität auf die vielen alternativen Tools. Wir stellen fest, dass QuiXPath eine frühe Abfragebeantwortung durch Kompilierung durchführt, um möglicherweise nichtdeterministische NWAs mit Auswahlzuständen, ohne immer die frühesten zu sein. Abgesehen von Dabei basiert seine Effizienz sowohl auf der Teilbaumprojektion als auch auf der Nachkommenprojektion.

Machine Translated by Google

Tabelle 6. Vergleich der parsefreien Zeitangaben in Sekunden für XPathMark-Abfragen mit QuiXPath und früheste Kongruenzprojektion.

Abfrage, Parse-frei, Parse-freier Bruch. AUSWEIS	Zeit (en) Zeit (en) QuiXPath-Kongr.-Prj.		Abfrage, Parse-frei, Parse-freier Bruch. AUSWEIS	Zeit (s) Zeit (s) QuiXPath-Kongr.-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		

Die Experimente mit QuiXPath in [4] nutzen die parsefreie Zeit. Daher das gleiche gilt für die Kongruenzprojektion durch Reduzierung der gemessenen Gesamtzeit um 15 s Analysezeit.

Verglichen mit den parsingfreien Zeiten für XPath- Abfragen ohne Nachkommenachse, Die früheste Kongruenzprojektion zeigt deutliche Verbesserungen. Im Durchschnitt unsere Projektion ist nur um den Faktor 1,3–2,9 langsamer als QuiXPath. Das zeigt, dass unsere Die Implementierung ist bereits fast konkurrenzfähig mit bestehenden XML- Streaming-Tools, und ist gleichzeitig der einzige, der eine schnellstmögliche Beantwortung der Anfrage garantiert.

Bei Abfrage A2 mit einer absteigenden Achse ist die Kongruenzprojektion um den Faktor 13,8 langsamer als QuiXPath. Der Grund dafür ist, dass QuiXPath die Nachkommenprojektion unterstützt, während die Kongruenzprojektion nur die Unterhedge betrifft. Für Abfrage A3 mit einer Nachkommenachs Die Kongruenzprojektion liegt nur um den Faktor 1,6 zurück, so dass die Nachkommenprojektion weniger hier relevant.

12. Verwandte Automatenmodelle

Eine große Anzahl alternativer Automatenkonzepte für Bäume, Wälder und Hecken wurden in der Literatur vorgeschlagen. Wir vergleichen sie mit SHAs und SHAÿ s im folgenden Unterabschnitt und diskutieren Sie die Auswirkungen.

12.1. **SHAs** im Vergleich zu Baum-, Wald- und Hedge-Automaten

Standard-Hedge-Automaten. Standard-Hedge-Automaten [9,10,32,33] arbeiten auf beschrifteten Hedges in einer Bottom-up-Manier, ähnlich wie SHAs. Sie haben auch die gleiche Ausdruckskraft, horizontale Sprachen werden jedoch unterschiedlich spezifiziert, was zu einem problematischen Begriff von Determinismus [34]. Aus diesem Grund schlägt die eindeutige Minimierung für deterministische Standard Hedge-Automaten. Dennoch haben sie die gleiche Ausdruckskraft wie SHAs , wenn sie auf beschriftete Hecken.

Standard-Baumautomaten und SHA-Minimierung. Syntaktisch ist jeder SHA A ein Standard Baumautomat über die folgende Rangsignatur:

- ein unäres Symbol für jeden Buchstaben a ÿ ÿ;
- ein Binärsymbol @;
- eine Konstante ÿÿ.

Semantisch gibt es keine perfekte allgemeine Entsprechung. Nur für Unterklassen gibt es eine perfekte Die Korrespondenz kann über eine binäre Kodierung erfolgen. Dies wurde in [8] für die Unterklasse von dSHAs mit I = ÿÿÿ. In [35] konnte auch für die Unterklasse der Multimodul- dSHAs. Dies führt zu eindeutigen Minimierungsergebnissen für beide Unterklassen deterministischer SHAs.

Bojanczyks Waldautomaten. Standard-Hedgeautomaten haben ebenfalls die gleiche Ausdruckskraft wie Bojanczyks Waldautomaten (siehe Abschnitt 3.3 von [36]). Diese basieren auf der Idee von Übergangsmonoide, statt auf der Idee von Übergangsregeln wie SHAs oder finiten

Machine Translated by Google

Zustandsautomaten oder Waldautomaten von Neuman und Seidl. Als Konsequenz gibt es einen exponentiellen Unterschied in der Prägnanz zwischen SHAs und Bojanczyks Waldautomaten

12.2. SHAÿ s im Vergleich zu SHAs

Wir beginnen mit der Erklärung, warum SHAÿ die gleiche Ausdruckskraft wie SHAs haben . Im Grunde ist es der gleiche Grund, warum NWAs die gleiche Ausdruckskraft wie SHAs haben **(siehe z. B. [8]). Für die einfachere Richtung haben wir bereits argumentiert, dass jeder SHA** A = **(ÿ, Q, ÿ, I,** F) in einen **SHAÿ**

Elimination. Umgekehrt können wir jeden SHAÿ A = (ÿ, Q, ÿ, I, F) in einen äquivalenten SHA Downward und dabei möglicherweise Nichtdeterminismus wie folgt einführen:

$$\text{eliminierenÿ} = \text{ich} \times Q \qquad \frac{\text{ÿÿÿÿq} \quad \text{ÿ in ÿ}}{\text{ÿÿ (q, q} \quad \text{ÿelimÿ) in ÿ}}$$

$$F^{\text{eliminiereÿ}} = F \times Q$$

$$\frac{q\ \text{aÿÿ} \quad \text{ÿ in ÿ r ÿ } Q}{q\ (r, q)\ \text{aÿÿ (r, q} \quad \text{ÿelimÿ) in ÿ}} \qquad \frac{\text{q@p ÿ q} \quad \text{ÿ in ÿ r ÿ } Q}{(r, q)@(q, p)\ \text{ÿ (r, q} \quad \text{ÿelimÿ) in ÿ}}$$

Satz 10. L(A) = L(elimÿ (A)).

Die Konstruktion ist analog zur Umwandlung von NWAs in SHAs [8] oder in Hedge Automaten [26]. Die Korrektheitsbeweise für diese Compiler sind Standard.

Eindeutige Minimierung. Die eindeutige Minimierung schlägt für dSHAÿ s fehl, wie es bei deterministischen Mehrwegautomaten üblich ist. Dies passiert sogar bei deterministischen Zweiwegautomaten auf Wörtern. Im Gegensatz dazu genießt die Unterklasse der dSHAs mit der Einschränkung, dass der Anfangszustand und der Anfangszustand des Baumes gleich sind, eindeutige Minimie

Pushdown-Forest-Automaten von Neumann und Seidl. Es ist auch bekannt, dass Standard-Hedge-Automaten die gleiche Ausdrucksstärke aufweisen wie die Pushdown-Forest- Automaten von Neumann und Seidl **[25]. Letztere können mit SHAÿ** s für beschriftete Hedges identifiziert werden. **Um Pushdown-Forest-Automaten auf Standard-Hedge-Automaten abzubilden, ist eine Abwärtseliminierung erforderlich , ebenso wie für die Abbildung von SHAÿ s auf SHAs.**

12.3. SHAÿ s im Vergleich zu NWAs

Der Streaming-Evaluator für SHAÿ s über eine sichtbare Kellermaschine kann auch durch Kompilieren eines SHAÿ zu einem verschachtelten Wortautomaten (NWA) [11] erhalten werden , **dessen Streaming- Evaluator durch eine sichtbare Kellermaschine [12] gegeben ist, die früher als** eingabegesteuerte Automaten **[13–15] bezeichnet wurde.**

Definition 16. Ein verschachtelter Wortautomat **(NWA)** ist ein Tupel **(ÿ, Q, ÿ, ÿ, I,** F), wobei ÿ, ÿ und **Q** Mengen sind, **I,** F **ÿ** Q und **ÿ** = ((a **ÿ)aÿÿ,** ÿ ÿ enthält die folgenden Relationen: a **Q**
ÿ ÿ ÿ Q × (ÿ × Q) und ÿ **, ÿ ÿ Q × ÿ ×** Q. Ein **NWA** ist deterministisch oder äquivalent ein **dNWA ,** wenn I ÿ enthält höchstens ein Element und alle oben genannten Beziehungen sind partielle Funktionen.

Die Elemente von ÿ heißen Stapelsymbole. Die Übergangsregeln in ÿ haben wiederum drei Formen: interne Regeln q **aÿÿ** q ÿ**ÿÿ und** **ÿÿÿÿ in ÿ wie bei SHAs, Eröffnungsregeln** q **ÿÿÿ** q **in ÿ wenn** (q) = (q **,** ÿ), **Abschlussregeln** q **ÿÿÿ** q **in ÿ wenn** (q, ÿ) = q **.**

Streaming-Auswerter für NWAs. Der Streaming-Auswerter für NWAs kann als Kellermaschine zur Auswertung der verschachtelten Wörter von Hedges in Streaming-Manier betrachtet werden. Eine Konfiguration der Kellermaschine ist ein Paar in K = **Q × ÿ , das einen Zustand und einen Stapel enthält.**

Machine Translated by Google

Für jedes Wort v ÿ ÿ^ ÿ **, definieren wir eine Streaming-Transitionsrelation** str **ÿ K × K, so dass für alle q,** q^ÿ **vÿÿ ÿ Q und** ÿ ÿ ÿ **, gilt folgendes:**

$$\frac{\text{WAHR}}{(q, ÿ) \overset{ÿ}{ÿÿ}_{\text{str}} (q, ÿ) \textbf{ bzgl. ÿ}} \qquad \frac{(q, ÿ) \overset{ÿ}{ÿÿ}_{\text{str}} (Q^{ÿ}, ÿ) \qquad (Q^{ÿ}, ÿ) \overset{ÿ}{{}_{v\,ÿÿ}}_{\text{str}} (Q^{"}, ÿ) \textbf{ bzgl. ÿ}}{(q, ÿ) \overset{ÿ\,v\cdot v}{ÿÿÿÿ}_{\text{str}} (Q^{"}, ÿ) \textbf{ bzgl. ÿ}}$$

$$\frac{q \textbf{ einÿÿ } q^{ÿ} \textbf{ in ÿ}}{q \textbf{ einÿÿ } q^{ÿ}_{\text{str}} \textbf{ bzgl. ÿ}} \qquad \frac{\textbf{ÿÿÿÿÿÿ ÿ}^{ÿ} \textbf{ in ÿ}}{(q, ÿ) \overset{ÿ}{ÿÿ}_{\text{str}} (Q^{ÿ}, ÿ) \textbf{ bzgl. ÿ}} \qquad \frac{\textbf{ÿ}}{\textbf{ÿÿÿÿÿÿ}^{ÿ} \textbf{ in ÿ}}{(q, ÿ \cdot ÿ) \overset{ÿ}{ÿÿ}_{\text{str}} (Q^{ÿ}, ÿ) \textbf{ bzgl. ÿ}}$$

Von SHAÿ s zu NWAs. Für jedes SHAÿ A = **(ÿ, Q, ÿ, I,** F) können wir das **NWA** A **(ÿ, Q, ÿ,** ÿ) nwa = **definieren ,** nein , ICH nein , F nwa) **unter Beibehaltung des Determinismus, so dass ÿ** = **Q, während ÿ das für alle** a **ÿ ÿ und q,** p **ÿ Q Übergangsregeln enthält:**

$$\frac{q \textbf{ aÿÿ}^{ÿ} \textbf{ in ÿ}}{qq \textbf{ aÿÿ } q^{ÿ} \textbf{in ÿ}}_{\text{nwa}} \qquad \frac{ÿÿ\, q \textbf{ ÿÿ}^{ÿ} \textbf{ in ÿ}}{\substack{q \\ \textbf{ÿÿq}\, q \textbf{ ÿÿÿ}^{ÿ} \textbf{in ÿ}}}_{\text{nwa}} \qquad \frac{\textbf{q@p ÿ } q\, ÿ^{ÿ} \textbf{ in ÿ}}{\textbf{ÿq}\; p \textbf{ ÿÿÿ } q \;{}^{ÿ} \textbf{ in ÿ}}_{\text{nwa}}$$

Zum Beispiel die dNWA A nwa **des dSHAÿ** A **in Abbildung 9, erhalten aus dem dSHA für Das** [self::list/child::item] **aus der Einleitung ist in Abbildung 24** dargestellt. **Das dNWA** A **des dSHAÿ** nwa A **in Abbildung 21 ist in Abbildung 25 dargestellt.**

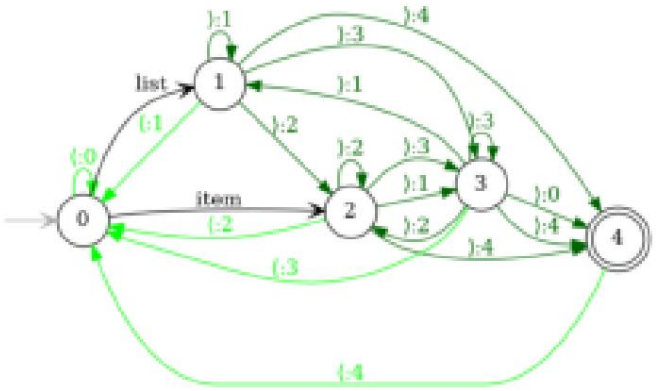
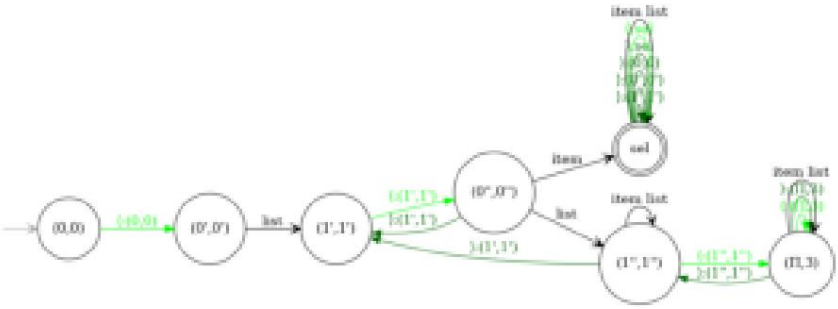


Abbildung 24. Der dNWA A **der** nwa **für das dSHAÿ** A **in Abbildung 9 , das durch Anwendung des** Down- **Operators auf dSHA in Abbildung 4 für den Filter** [self::list/child::item].



dNWA (A snc) nwa e(doc) **für das früheste dSHAÿ mit Safe-No-Change-Projektion in Abbildung 25. Die Abbildung 21.**

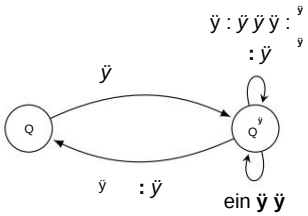
Machine Translated by Google

Lemma 21. L(A nwa) = L(A).

Die Läufe von SHAÿ s A **und NWAs** A nwa **identifiziert werden können.**

Projektion für NWAs. Projektions-Evaluatoren für NWAs wurden im Rahmen der Projektion von **NWAs [4] vorgeschlagen. Sie basieren auf der folgenden Vorstellung von Zuständen irrelevanter Teilbäume für NWAs.**

Definition 17 (Variante der Definition 3 von [4]). Wir nennen einen Zustand q eines **NWA** einen **Zustand irrelevanter Teilbäume ,** wenn es zwei verschiedene Stapelsymbole **ÿ,** *ÿ* gibt, so dass die rechts gezeigten Übergänge existieren, aber keine weiteren öffnenden Übergänge mit ÿ, keine weiteren Übergänge mit *ÿ* und keine weiteren schließenden Übergänge in q, die ÿ aufpoppen. In diesem Fall **,** schreiben wir q **ÿ** i-Baumÿ\ÿ.



Lemma 22. Jeder Subhedge-Projektionszustand eines vollständigen **SHAÿ** A ist ein Zustand irrelevanter Teilbäume von Anwa

Anschließend wurde in [4] gezeigt , **wie man eine NWA mit einer Teilmenge Qÿ** abbildet **von Zuständen irrelevanter** shp- **Teilbäume zu einem projizierenden NWA, dessen Streaming-Semantik einen Streaming-Evaluator mit Subhedge-Projektion ergibt. Die Darstellung der Subhedge-Projektion für SHAÿ s ohne Umweg über NWAs führt auf direkterer Weise zum gleichen Ergebnis.**

Es sollte auch beachtet werden, dass die Projektion von NWAs neben der Subhedge-Projektion auch die Nachkommenprojektion unterstützt . Bei SHAs bleibt dies künftiger Forschung überlassen.

13. Schlussfolgerungen und zukünftige Arbeiten

Wir haben das Konzept irrelevanter Subhedges für die Mitgliedschaft in Hedgesprachen unter Schemabeschränkungen eingeführt. Wir haben Algorithmen mit Subhedge-Projektion entwickelt, die über irrelevante Subhedges springen, um die Mitgliedschaft in den **durch dSHAs definierten regulären Hedgesprachen zu testen . Alle unsere Algorithmen können im Speichermodus und im Streamingmodus ausgeführt werden. Die Schwierigkeit bestand darin, die benötigten Informationen zu endlichen Zuständen für die Subbaumprojektion von oben nach unten zu übertragen, da SHAs von unten nach oben arbeiten. Wir haben dies auf der Grundlage von Compilern von SHAs zu Downward -SHAs gelöst.**

Dieser erste Compiler gibt Sicherheitsinformationen zu unveränderlichen Zuständen weiter. Der zweite Compiler, dessen Vollständigkeit wir für die Subhedge-Projektion beweisen, propagiert Differenzrelationen. Anschließend haben wir die Subhedge-Projektion mit dem frühesten Mitgliedschaftstest kombiniert und sie mit der frühesten monadischen Abfragebeantwortung kombiniert. Wir haben sowohl Safe-No-Change als auch vollständige Kongruenz in unser AStream-Tool integriert und ihre Nützlichkeit experimentell bestätigt : Wir haben gezeigt, dass es das AStream-Tool bei der Beantwortung **regulärer XPath-Abfragen auf XML-** Streams **tatsächlich beschleunigen kann . Darüber hinaus zeigte die Kongruenzprojektion viel höhere Ereignisprojektionsgewinnraten als Safe-No-Change, die zwischen 75,7 % und 100 % für die Klasse von Abfragen ohne Nachkommenachse lagen, und wurde mit der Zeit mit den besten vorhandenen Streaming-Tools für reguläre XPath-Abfragen ohne rekursive Achse konkurrenzfähig.**

Eine noch offene theoretische und praktische Frage ist, wie man eine Nachkommenprojektion **für SHAs erhält , wie sie zur Beschleunigung der Auswertung von XPath- Abfragen mit einer Nachkommenachse** erforderlich ist . **Auf der Implementierungs- und experimentellen Seite wäre es schön, auch die In-Memory-Version unserer Subhedge-Algorithmen zu untersuchen . Schließlich wä**

Machine Translated by Google

Streaming-Tools zur Beantwortung frühester Abfragen werden weiter ausgebaut, bis sie die besten Streaming-Tools in allen Fällen übertreffen und gleichzeitig für Abfragen besser sind, bei denen die anderen Tools nicht die frühesten si

Beiträge der Autoren: Schreiben – Originalentwurf, AAS und JN. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu.

Finanzierung: Diese Forschung erhielt keine externe Finanzierung.

Datenverfügbarkeitserklärung: Datensatz auf Anfrage bei den Autoren erhältlich.

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Anhang A. Früheste Mitgliedschaft Wir

passen den frühesten Mitgliedschaftstester für dSHAs aus [21] so an, dass er zu **SHAÿ s statt zu NWAs kompiliert und Schemata berücksichtigt.**

Anhang A.1. Schemasicherheit ist ein Inklusionsproblem **Die**

Idee zum Testen bestimmter Mitgliedschaften und bestimmter Nichtmitgliedschaften besteht darin, zu berücksichtigen, dass Zustände sicher sind, um Endzustände zu erreichen, außer wenn man das Schema verlässt. Sei A = **(ÿ, Q, ÿ, I,** F) **ein SHA mit Schema S, sodass** F **ÿ FS ÿ Q mit** S = **L(A[F/FS]) existiert. Beachten Sie insbesondere, dass** L(A) **ÿ** S **ÿ Hÿ .** Die Menge der Zustände, die sicher sind, um eine Teilmenge P **ÿ Q für alle Hedges zu erreichen, die das Schema S nicht verlassen, ist wie folgt:**

$$\text{sicher}_{S}ÿ\,(P) = \text{sicher}ÿ\,(P\ ÿ\ (Q \setminus \text{FS})).$$

Wenn A **deterministisch und schemavollständig für S** ist, **dann ist die Sicherheit eines Zustands von** A **in Bezug auf Schema S ein Inklusionsproblem.**

Lemma A1. Sei S = **L(A[F/FS]).** Wenn A deterministisch und schemavollständig für S ist, dann gilt für alle Zustände q **ÿ** Q:

$$\mathbf{L(A[I/\{q\}, F/FS])\ ÿ\ L(A[I/\{q\}])\ ÿ}\ q\ ÿ\ \text{sicheres}ÿ\,(F).$$

Beweis. „ÿ“. Angenommen, q **ÿÿ** safe$_{S}$ÿ(F). **Die Definition der schemabasierten Sicherheit ergibt** accÿ({q}) **ÿ** (FS **** F) ÿ= **ÿ. Es existiert also eine Absicherung** h **ÿ** H**ÿ, sodass** q **ÿÿ FS ** F **bzgl. ÿ. Insbesondere gilt** q **ÿ L(A[I/{q}, F/FS]). Deterministisch gesehen gibt es auf** h keinen anderen Übergang **von q, und daher** gilt q **ÿÿ L(A[I/{q}]). Daher gilt L(A[I/{q}, F/FS]) ÿÿ L(A[I/{q}]. „ÿ“. Angenommen,** q **ÿ** safeÿ (F). **Sei** h **ÿ L(A[I/{q}, F/FS]).** ^{ÿ} **ÿ FS,** h , **so dass** q **ÿÿ** q **ÿ F.** ^{ÿ} **bzgl . ÿ.** Daher ist q ^{ÿ} **Dann existiert** q **ÿ** accÿ({q}), **sodass** accÿ({q}) **ÿ** F **ÿ (Q \ F** ^{ÿ} **Daher** ist h **ÿ L(A[I/{q}]). impliziert** q □

Lemma A1 mit FS = **Q zeigt Lemma 5 von [21], das für alle vollständigen dSHAs** besagt A **dass Sicherheit ein Universalitätsproblem ist:**

$$\mathbf{L(A[I/\{q\}])} = \mathbf{Hÿ}\ ÿ\ q\ ÿ\ \text{sicheres}ÿ\,(F).$$

Dies liegt daran, dass die Schemavollständigkeit von A **für S** = **Hÿ** die Vollständigkeit von **A impliziert** und **somit L(A[I/{q}, F/Q])** = **Hÿ.**

Anhang A.2. Algorithmus

Wir präsentieren jetzt unseren ersten Mitgliedschaftstester für dSHAs mit Schemabeschränkungen. Gegeben sei ein SHA A, so berechnen wir einen SHAÿ Ae(S) = **(ÿ,** Qe , ÿe ,I **e(S) ,** Fe(S)) **wie folgt. Sei** sel **ein**

Machine Translated by Google

frisches Symbol. Wir beginnen mit einer Menge von Zuständen, die sicher für die Auswahl S0 = safeÿ (F) **bzw. sicher für die Ablehnung R0** = safeÿ **(Q** \ F) sind. **Die Zustandsmengen von** Ae(S) **sind dann wie**

$$Qe = (Q \times 2\ Q \times 2\ Q)\ \ddot{y}\ \{sel\} =$$
$$_{\text{Ich e(S)}}\ \{(q, \mathbf{R0, S0)} \mid q\ \ddot{y}\ \mathbf{I,}\ q\ \ddot{y}\ddot{y}\ \mathbf{R0}\ \ddot{y}\ \mathbf{S0\}}\ \ddot{y}\ \{sel \mid I\ \ddot{y}\ \mathbf{S0}\ \ddot{y} = \ddot{y}\}$$
$$Fe(S) = \{(q, \mathbf{R0, S0)} \mid q\ \ddot{y}\ \mathbf{F\}}\ \ddot{y}\ \{sel\}$$

Zur Veranschaulichung betrachten wir noch einmal das dSHA aus Abbildung 4. Dieses dSHA ist nicht vollständig, aber schemavollständig für das Schemadokument . Wir erhalten s-downÿ **(0, {4})** = **{3, 4}. Es mag kontraintuitiv erscheinen, dass nicht nur Zustand 3, sondern auch Zustand 4 down sicher für die Auswahl bleibt. Dies spiegelt die Tatsache wider, dass kein richtiges Subhedge eines Hedges, das das beabsichtigte Schema erfüllt, jemals in Zustand 4 gelangen kann .** Wenn wir die früheste Konstruktion mit sicherer No-Change-Projektion auf **dSHA in Abbildung 4 anwenden, erhalten wir das dSHAÿ in Abbildung A1. Die Übergangsregeln in** ÿe sind in **Abbildung** A2 **angegeben .**

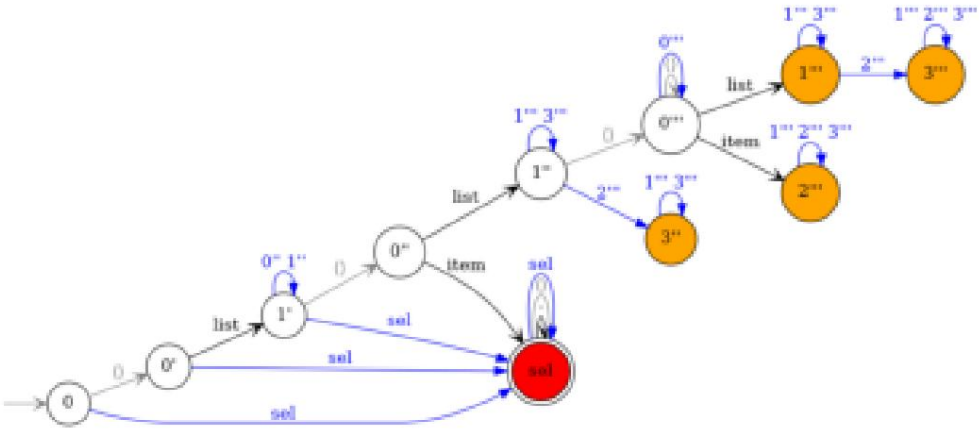


Abbildung A1. Das früheste dSHAÿ Ae **für das dSHA** A **in Abbildung 19 mit** Schemadokument **für den XPath- Filter** [self::list/child::item]. **Die Zustände von** Ae **entsprechen den folgenden Tupeln: 0** = **(0, {1, 2, 3, 5}, {4}), 0ÿ** = **(0, {2, 4, 5}, {3,** 4}), **1ÿ** = **(1, {2, 4, 5}, {3,** 4}), **0ÿÿ** = **(0, ÿ,** $_{1}$" **{2, 4,** 5}), = **(1, ÿ, {2, 4,** 5}), **0ÿÿÿ** = **(0, ÿ,** ÿ), **1ÿÿÿ** = **(1, ÿ,** ÿ), **2ÿÿÿ** = **(1, ÿ,** ÿ) **und 3ÿÿÿ** = **(1, ÿ,** ÿ).

Machine Translated by Google

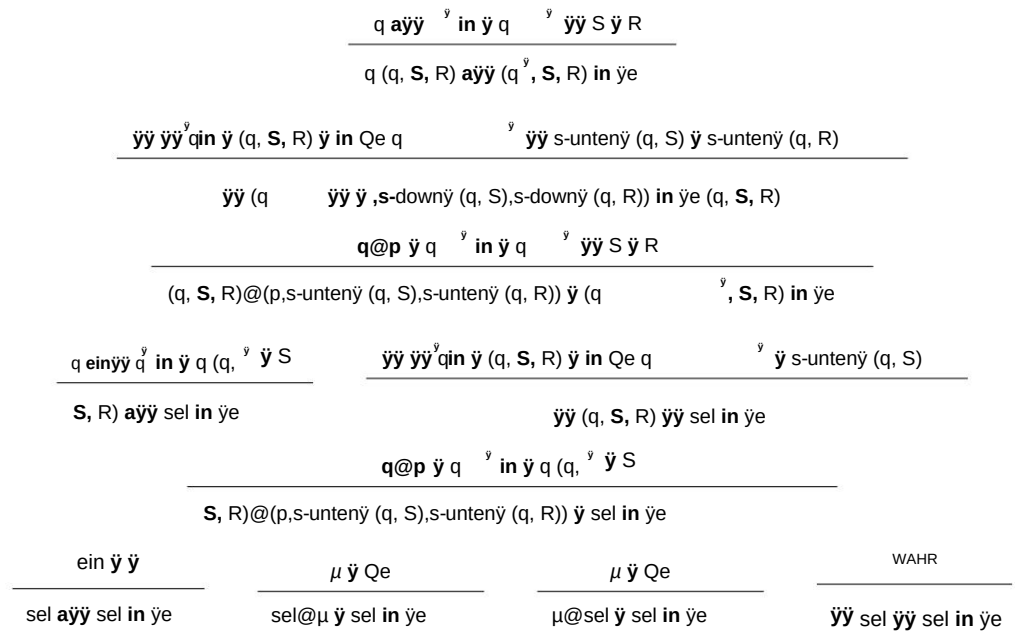


Abbildung A2. Die frühesten Übergangsregeln ÿe, **abgeleitet aus den Übergangsregeln ÿ einiger SHA.**

Anhang A.3. Solidität und Vollständigkeit

Als Nächstes stellen wir Ergebnisse zur Solidität und Vollständigkeit für unseren frühesten In-Memory-Mitgliedschaftstester mit Projektion bereit.

Proposition A1 (Korrektheit). Für jeden **SHA** A mit Schema **S** = **L(A[F/FS]),** Hedge h **ÿ** S und partiellem Lauf v von Ae(S) auf h, der im Zustand sel endet, dann gilt h **ÿ** L(A).

Wenn also der Teildurchlauf den Zustand sel erreicht, **dann ist die Mitgliedschaft von** projÿ (v) in L(A) **bezüglich S sicher. Die Auswertung von a SHAÿ** s Ae(S) **ergibt den frühesten im Speicher befindlichen Mitgliedschaftstester für dSHA A. Eine einfache Anpassung ist wie folgt: Immer wenn** sel **erreicht wird, kann die Auswertung von vorne angehalten werden und die Eingabeabsicherung akzeptieren.**

Theorem A1 (Vollständigkeit). Sei A = **(ÿ, Q, ÿ, I,** F) ein **dSHA ,** das schemavollständig ist für Schema **S** = **L(A[F/FS]).** Für jeden Hedge h **ÿ S** und Präfix v von nw(h): Wenn v sicher für die

- Mitgliedschaft in L(A) mit S ist, dann existiert ein partieller Lauf w für Ae(S) auf h, so dass v = projÿ (w) und w mit dem Zustand sel endet.
- Wenn für v die Nichtmitgliedschaft in L(A) mit S sicher ist , dann existiert ein blockierender partieller Lauf w für Ae(S) auf h, so dass projÿ (w) ein Präfix von v ist.

Dies bedeutet, dass die Sicherheit bezüglich Mitgliedschaft und Nichtmitgliedschaft beim frühesten Präfix erkannt wird, wenn der Evaluator für Ae(S) ausgeführt wird .

Beweis. Wir stellen fest, dass dieser Algorithmus im Grunde derselbe ist wie der früheste Mitgliedschaftstester aus Proposition 6 von [21], außer dass er auch auf sichere Ablehnung prüft und Schemata berücksichtigt. Die Tatsache, dass hier **NWAs statt SHAÿ s verwendet werden, ist nicht wesentlich. Wir können also den Beweis der Solidität und Vollständigkeit problemlos von dort übernehmen.** □

Anhang B. Beweise für Abschnitt 6 (Safe-No-Change-Projektion)

Beweis. Wir müssen beweisen, dass kein sich ändernder Zustand q **ÿ** P **ÿ** keine-Änderungÿ mehr **stichhaltig ist.**

Wenn q **ÿ** keine Änderung ÿ , **folgt dies aus der Schemavollständigkeit von ÿ, so dass man weder eine Absicherung aus dem Schema blockieren noch den Zustand ändern kann. Im Fall** q **ÿ P ist die**

Machine Translated by Google

Die Intuition besagt, dass sich weder der Zustand auf der Ebene darüber, beispielsweise r, ändern kann, da P = s-keine-Änderungÿ (r), **noch kann der Automat aufgrund der Schemavollständigkeit irgendeine Absicherung durch das Schema bl**

Wir beweisen zunächst die Inklusion L(A) ÿ L(A snc). **Da** L(A) ÿ **S per Definition von Schemata ist, impliziert dies** L(A) ÿ L(A snc) **ÿ S. Der Beweis basiert auf den folgenden drei Behauptungen A1.1a, A1.2a und A1.3a. Beachten Sie, dass Schemavollständigkeit für diese Richtung nicht erforderlich ist.**

h **Behauptung A1.1a. ÿ ÿÿ ÿ bzgl. ÿ** snc **für alle Hedges** h **ÿ Hÿ.**

Der Beweis ist unkompliziert durch Induktion über die Struktur von h. Dabei werden die letzten drei Übergangsregeln von ÿ in Abbildung 10 verwendet, snc sodass **für jede folgende Absicherung immer in ÿ** verbleiben kann .

Behauptung A1.2a. Für alle h **ÿ Hÿ,** q **ÿ Q und** P **ÿ Q, so dass** q **ÿ** P **ÿ** keine Änderungÿ ,

(q, P) **ÿÿ** H (q, P) **bzgl. ÿ** snc .

Wir beweisen Behauptung A1.2a durch Induktion über die Struktur von h.

Fall h = ÿh ÿ h ÿ. **In diesem Fall können wir Behauptung** A1.1a verwenden, **um zu zeigen, dass ÿ ÿÿ ÿ bzgl. ÿ** snc **und die Inferenzregeln** ÿ

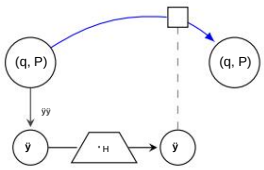
q **ÿ** P **ÿ** keine Änderungÿ

(q, P) **ÿÿ ÿÿ ÿ in ÿ** snc

q **ÿ** P **ÿ** keine Änderungÿ

(q, **P)@ÿ ÿ** (q, P) **in ÿ** snc

um das folgende Diagramm bezüglich ÿ snc zu schließen:



Damit ist (q, P) **ÿÿ** (q, P) **bzgl. ÿ** H snc bewiesen, **wie es die Behauptung verlangt.**

Fall h = **a. Da** q **ÿ** P **ÿ** keine Änderung ÿ, **können wir die Inferenzregel anwenden:**

a **ÿ ÿ** q **ÿ** P **ÿ** unverändertÿ (q, P)

aÿÿ (q, P) **in ÿ** snc

Dies beweist die Richtigkeit dieser Behauptung.

Fall h = **ÿ. Wir haben trivial** (q, P) **ÿÿÿ** (q, P) **bzgl. ÿ** snc .

Fall h = h ÿ · H **". Nach der Induktionshypothese angewendet auf** h ÿ **und** h **" wir haben:** (q, P) **ÿÿ** (q, H ÿ P) **und** (q, P) **ÿÿ** H" (q, P) **bzgl. ÿ** snc. **Daher gilt** (q, P) **ÿÿÿ** (q, P) **bzgl. ÿ** H ÿ · H" snc .

Damit ist der Beweis von Behauptung A1.2a beendet. Die nächste Behauptung, bei der der Induktionsschritt etwas mühsamer zu beweisen ist, ist der Schlüssel zum Beweis der Stichhaltigkeit. Wir definieren das folgende Prädikat für alle q, Q ÿ Q und P ÿ Q:

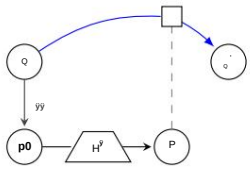
Q ÿ **"** ÿP q genau dann, wenn (q = q ÿ **"** ÿ q , Q ÿ **"** ÿ P)

Behauptung A1.3a. Sei h **ÿ Hÿ eine Absicherung, q,** q h **dass accÿ(P) ÿ** P **und** q **ÿÿ** P **ÿ ÿ Q Zustände und** P **ÿ Q eine Teilmenge von Zuständen, wie** keine-Änderungÿ . **Wenn** q **ÿÿ** q **bzgl. ÿ, dann existiert** q **", so dass** (q, P) **ÿÿ** H (q **" ,** P) **bzgl. ÿ** snc **und** q ÿ **"** ÿP q **" .**

Machine Translated by Google

Beweis. Induktion über die Struktur von h.

Fall h = ÿh ÿ h ÿ. **Die Annahme** q ÿÿ qp ÿ **Q schließt** ÿ **bzgl. ÿ zeigt, dass es Zustände p0 ÿ ÿÿÿ** gibt **und das folgende Diagramm:**

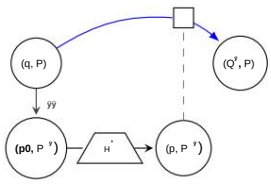


Sei P ÿ = s-keine-Änderungÿ (q) **und beachten Sie, dass aus accÿ(P ÿ) ÿ** ÿ **. Da** q **ÿÿ** P **ÿ** keine Änderungÿ **ist,** P **Folgendes gefolgert werden kann:**

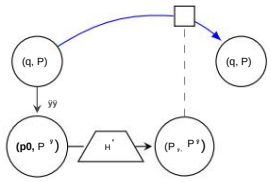
$$\frac{\text{ÿÿ p0 in ÿ q ÿÿ P ÿ keine Änderungÿ}}{\text{ÿÿ (q, P) ÿÿ (p0, P ÿ) in ÿ}}\ \text{snc} \qquad \frac{\text{q@p ÿ q ÿ in ÿ q ÿÿ P ÿ keine Änderungÿ}}{\text{(q, P)@(p, P ÿ) ÿ (q ÿ, P) in ÿ}}\ \text{snc}$$

Unterfall p0 ÿÿ P ÿ **ÿ** keine Änderung ÿ . **Die Induktionsvoraussetzung gilt für** h , **so** ÿ **zeigt, dass es gibt** p **ÿÿ** (p) **, dass (p0,** P ÿ, H ÿ P ÿ) **bzgl. ÿ** snc **und** p ÿP p ÿ **Wir unterscheiden die letztgenannte Prädikat rechtfertigen:**

Das Unterunterfall ÿ = **p. Daher (p0,** P ÿ) **ÿÿ** (p, P ÿ) **bzgl. ÿ** snc, **so dass wir die** -p- **Diagramm lautet wie folgt:**



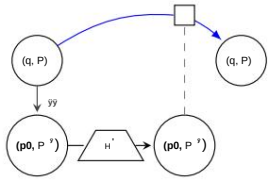
Dies zeigt, dass (q, P) H ÿÿ (q . ÿ, P) **bzgl. ÿ** snc, **und damit die Behauptung, da** ÿ ÿP q ÿ ÿ **. Unterunterfall p,** pq ÿ **ÿ** P ÿ **Da** p ÿ **ÿ** P ÿ **und** P ÿ q = s-keine-Änderungÿ (q), **haben wir q@p in ÿ. Daher können wir das Diagramm wie folgt schließen:**



Da p **ÿ** P s-keine- ÿ **und q@p ÿ** q ÿ **in ÿ haben wir** q = q ÿ **per Definition von** P ÿ = Änderungÿ (q). **Dies zeigt, dass** (q, P) H **ÿÿ** (q, P) **bzgl. ÿ** snc. **Da** q ÿP **q, ist die Anspruch folgt.**

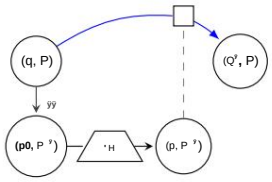
Unterfall p0 ÿ P ÿ **ÿ** keine Änderung ÿ . **Behauptung A1.2a zeigt dann, dass (p0,** P ÿ, H ÿ P ÿ**) ÿÿ (p0,** P ÿ**) bzgl . ÿ** snc **.**

Unterunterfall p0 ÿ P ÿ **. Da** p **ÿ** accÿ(p0) **und accÿ(P ÿ) ÿ** P = s-keine- ÿ **, daraus folgt, dass ÿ** P ÿ, **auch. Per Definition** P p ÿ Änderungÿ (q) **und die Vollständigkeit von ÿ** können **die Mitgliedschaften** ÿ **und** p **ÿ** P ÿ **impliziert, dass q@ÿp0** = {q} = **q@ÿp. Wir p0 ÿ** P **nun das folgende Diagramm wie folgt schließen:**



Unterunterfall p0 ÿ keine Änderungÿ .**In diesem Fall** ist **p0** = p , **so dass** ÿ **ÿ** q@ÿp0. **Daher,**

Machine Translated by Google



Fall h = **a. Da** q **ÿÿ** P **ÿ** keine Änderungÿ , **können wir die Inferenzregel anwenden:**

$$\frac{\text{q einÿÿ q' in ÿ q ÿÿ P ÿ unverändert ÿ (q, P) aÿÿ (q', P) in ÿ}}{\text{snc}}$$

Dies zeigt, dass (q, P) **ÿÿ** (q', P) **bestätigt die Behauptung, da** q ÿP q'.

Fall h = **ÿ. In diesem Fall haben wir** q = q' **und** (q, P) **ÿÿÿ** (q, P) , **also gilt die Behauptung.**

h **Fall** h = **h1 · h2. Da** q **ÿÿ** q' **h1 bzgl. ÿ, es existiert q1 ÿ Q, so dass** q **ÿÿ q1 bzgl. ÿ und**

h2 q1 ÿÿ q' **bzgl. ÿ. Da** q **ÿÿ** P **ÿ** keine Änderung ÿ , **wenden wir die Induktionshypothese auf h1 an. so Dies impliziert, dass es** q gibt **1 dass**

h1 (q, P) **ÿÿ** (q' **1 ,** P) **bzgl. ÿ** snc **und q1** ÿP q **1** .

Wir unterscheiden die beiden Fälle q1 ÿP q **: 1**

Unterfall q1 = q **1 . Wir unterscheiden auch hier zwei Unterfälle: Unterunterfall**

q1 ÿÿ P. Die auf **h2 angewandte Induktionsvoraussetzung ergibt folgendes:**

ing:

ÿq" **. (q1,** P) **ÿÿ** (q" **h2** , P) **bzgl. ÿ** snc **ÿ** q **ÿ** ÿP q" **,**

Somit

ÿq" **.** (q, P) **ÿÿ** (q" **H** , P) **bzgl. ÿ** snc **ÿ** q **ÿ** ÿP q"

.

Unterunterfall q1 ÿ P. Nach Behauptung A1.2a haben wir (q1, P) **ÿÿ (q1,** P). **Wir haben auch** Q' **ÿ** acc({q1}), **und da wir** acc(P) **ÿ P annehmen, impliziert dies** q **ÿ P. Daher gilt** (q, P) **ÿÿ (q1,** P) **und** q **, q1 ÿ P, was bedeutet, dass die Behauptung** q **ist** ÿPq1 **.**

Unterfall q1, q **1 ÿ P. Da** q **1 ÿ P, Behauptung A1.2a impliziert** (q, P) **h2 ÿÿ 1 ,** P) **bzgl . ÿ** snc **.**

ÿ accÿ(P) ÿ P. **ÿÿ** (q **1 ,** P) **bzgl. ÿ** snc. **Da** q (q **ÿ accÿ({q1}) und q1 ÿ P, folgt, dass Also** (q, P) Q' **Hier haben wir wie im vorherigen Unterunterfall angenommen, dass** acc(P) **ÿ** P **wird durch die Behauptung vorausgesetzt.** = q **1 . Dann haben wir** (q, P) **ÿÿ** (q" **H** , P) **bzgl . ÿ** snc **Seien** q **und** Q' q **ÿ** P , **was die Behauptung zeigt.**

Damit ist der Beweis der Behauptung A1.3a beendet. □

Beweis der Inklusion L(A) **ÿ** L(A snc). **Sei** h **ÿ** L(A). **Dann gibt es q0 ÿ** I **und** q **ÿ F, so dass q0 ÿÿ q. Wir unterscheiden zwei Fälle:** H

Fall q0 ÿ keine Änderung ÿ . **Per Definition von** keine Änderung ÿ **und da** q **ÿ** accÿ(q0) **haben wir q0** = **q.**

Behauptung A1.2a zeigt, dass (q0, ÿ) **ÿÿ (q0,** ÿ) **bzgl. ÿ** snc **und somit (q0,** ÿ) **ÿÿ** (q, ÿ), **so dass** h **ÿ** L(A snc).

Fall q0 ÿÿ keine Änderungÿ . **Behauptung A1.3a mit** P = **ÿ zeigt, dass (q0,** ÿ) **ÿÿ** (q, ÿ) **bzgl. ÿ** snc **und daher** h **ÿ** L(A snc).

Machine Translated by Google

Damit ist der Beweis der ersten Inklusion beendet. Als nächstes wollen wir die inverse Inklusion L(A snc) **ÿ S ÿ** L(A) zeigen. **Sie ergibt sich schließlich aus den folgenden drei Behauptungen A1.1b, A1.2b und A1.3b.**

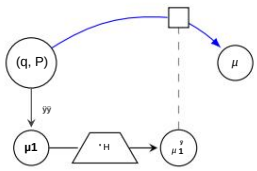
Behauptung A1.1b. Für jede Absicherung h **und jeden Zustand** *μ* **ÿ** Qsnc, **falls ÿ ÿÿ** *μ* **bzgl. ÿ** $\overset{H}{\text{snc}}$ **dann ist** *μ* = **ÿ.**

Der Beweis ist unkompliziert durch Induktion über die Struktur von h: Die einzigen Übergangsregeln von ÿ snc **mit ÿ auf der linken Seite ergeben sich aus den letzten drei Regeln in Abbildung 10. Diese erfordern, in ÿ zu bleiben, egal, welche Absicherung folgt.**

Behauptung A1.2b. Für jede Absicherung h gilt P **ÿ Q, Zustand** q **ÿ** P **ÿ** keine Änderung ÿ **und Zustand** *μ* **ÿ** Qsnc: **wenn** $(q, P) \overset{H}{\text{ÿÿ}}$ *μ* **bzgl. ÿ** snc, **dann** ist *μ* = (q, P).

Beweis. Mittels Induktion über die Struktur von h. Angenommen, (q, P) $\overset{H}{\text{ÿÿ}}$ *μ* **bzgl. ÿ** snc:

Fall h = ÿh **ÿ. Es müssen die Zustände µ1,** *μ* $\overset{ÿ}{\mu_1}$ **ÿ** Qsnc **schließt das folgende Diagramm:**



Da q **ÿ** P **ÿ** keine Änderung ÿ , **muss die folgende Regel angewendet worden sein, um** (q, P) **µ1 bzgl.** **ÿÿ ÿÿ ÿ** snc zu folgern:

$$\frac{\text{q ÿ P ÿ keine Änderung ÿ}}{\text{(q, P)} \quad \text{ÿÿ ÿÿ ÿ in ÿ}} \text{ snc}$$

Daher ist **µ1 = ÿ. Behauptung A1.1b zeigt, dass** *μ* **sich** $\overset{ÿ}{\mu_1}$ **= ÿ. Also muss** *μ* **aus der Anwendung der Regel ergibt:**

$$\frac{\text{q ÿ P ÿ keine Änderung ÿ}}{\text{(q, P)@ÿ ÿ (q, P) in ÿ}} \text{ snc}.$$

Also ist *μ* = (q, P), **wie erforderlich.**

Fall h = **a. Es muss folgende Regel angewendet worden sein:**

$$\frac{\text{q ÿ P ÿ unverändert ÿ}}{\text{(q, P) aÿÿ (q, P) in ÿ}} \text{ snc}$$

Daher ist *μ* = (q, P).

Fall h = **ÿ. Offensichtlich.**

h = **h1 · h2. Fall** Es muss **µ1 geben , sodass** (q, P) **ÿÿ** *μ* **bzgl. ÿ** snc. **Durch Induktion ÿÿ µ1- Hypothese auf h1 angewendet, haben wir µ1** = (q, P). **Wir können also die Induktionshypothese auf h2** anwenden , **um µ2** = (q, P) zu erhalten.

Damit ist der Beweis von Behauptung A1.2b beendet. Als nächstes benötigen wir eine Umkehrung von Behauptung A1.3a. □

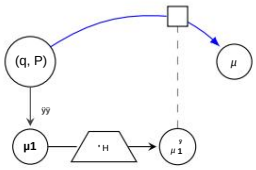
Behauptung A1.3b. Seien q **ÿ Q,** P **ÿ Q, so dass** q **ÿÿ** P **ÿ** keine Änderung ÿ **und accÿ(P) ÿ P. Für jedes** h **ÿ Hÿ, so dass** (q, P) **ÿÿ** *μ* **bzgl. ÿ jeder blockierende** $\overset{H}{\text{snc}}$ **für ein** *μ* **ÿ** Qsnc **und so dass ÿ nicht hat partielle Lauf auf** h **ausgehend von q, existiert** q $\overset{ÿ}{}$**, q '', so dass**

$$\mu = (q \overset{ÿ}{}, P),\ q \overset{H}{\text{ÿÿ}} q'' \text{ bzgl. ÿ und } q \overset{ÿ}{} \text{ ÿP } q''.$$

Machine Translated by Google

Beweis: Mittels Induktion über die Struktur von h **ÿ Hÿ können wir Fälle für alle möglichen Formen von h unterscheiden.**

Fall h = ÿh **ÿÿ** μ **ÿ**[ÿ] **bzgl. ÿ. Definitionsgemäß müssen** (q, P)μ existieren[H] , **damit folgendes Diagramm** [ÿ]**1 ÿ** Qsnc, **wie geschlossen werden kann:**



Da q **ÿÿ** P **ÿ** keine-Änderungÿ , **wurde die folgende Regel angewendet, um** (q, P) **ÿÿ µ1** zu [ÿÿ]folgern , **wobei** P[ÿ] = s-keine-Änderungÿ (q):

$$\frac{\text{[ÿÿ]}\ \text{ÿÿ p in ÿ q ÿÿ P ÿ keine Änderungÿ}}{\text{ÿÿ (q, P) ÿÿ (p, P}^{\text{ÿ}}\text{) in ÿ}}\ \text{snc}$$

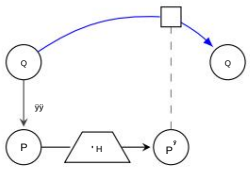
Daher existiert p **ÿ ÿÿÿ, so dass µ1** = (p, P[ÿ]**). Wir legen einen solchen Zustand** p **beliebig fest.**

Da ÿ keine blockierenden Läufe auf h **von** q aus **hat, existiert** p[ÿ]**, Q**[ÿ]**" ÿ Q, so dass** p [Hÿ] **ÿÿ** p **ÿ** q**, und q@p end von" bzgl. ÿ. Außerdem** hat **ÿ keine blockierenden Teillauf auf** h[ÿ] **p.**

Unterfall p **ÿ** P[ÿ]**·. In diesem Fall können wir Behauptung A1.2b auf** (p, P[ÿ]) [h ÿ]μ 1**bzgl. ÿ in der Reihenfolge ÿÿ zeigen, dass** μ , [ÿ]**1** = (p, P[ÿ]**anwenden. Da** p **ÿ** accÿ({p}) **und** p **ÿ** P = s-keine-Änderungÿ (q), **um zu folgt, dass q@ÿp** = **{q}. Daher wurde die folgende Regel angewendet, um** (q, P) **ÿÿ µ zu folgern:**

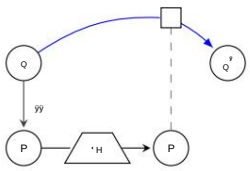
$$\frac{\text{q@p ÿ q in ÿ q ÿÿ P ÿ unverändert ÿ (q,}}{\text{P)@(p, P}^{\text{ÿ}}\text{) ÿ (q, P) in ÿ}}\ \text{snc}\ ,$$

Dies zeigt, dass μ = (q, P). **Sei** [ÿ] = q , **so dass** μ = (q[ÿ]**,** P). **Da** p **bzgl. ÿ**[ÿ]**ist per Definition habe** p[ÿ] q **ÿ** accÿ({p}), **so dass q@p von** s-keine-Änderungÿ **({q}) ÿ** q **bzgl. ÿ .**
Daher können wir das folgende Diagramm schließen:



dass q**" von der** [ÿ] = q **'. Dann gilt** q . **Sei** q = **q, so** [ÿ] ÿP q **"** ÿh ÿ **und** q [ÿ]**ÿÿÿ** q[ÿ] **bzgl . ÿ, als Behauptung gefordert wird.**

Unterfall p **ÿ** keine Änderungÿ . [h ÿ] **In diesem Fall** kann p[ÿ] **bzgl. ÿ impliziert, dass** p[ÿ] **= p. Daher ÿÿ** p **das folgende Diagramm schließen:**



Dies zeigt, dass q **ÿÿÿ** q[ÿ]ÿH ÿ [ÿ] **bzgl. ÿ.**

Machine Translated by Google

Unterfall p ÿÿ P ÿ, ÿ ÿ keine Änderung ÿ . **Nach der Induktionshypothese angewendet auf** (p ÿ P ÿ ÿ ÿ

Machine Translated by Google

□

Beweis der Inklusion L(A snc) **ÿ S ÿ** L(A). **Sei** h **ÿ** L(A snc) **ÿ S. Da** h **ÿ** L(A snc), **existiert q0 ÿ** I **und** q **ÿ F, so dass (q0,** ÿ) **ÿÿ** (q, ÿ) **bzgl. ÿ** snc. **Mit Lemma 7 haben wir ÿÿ** q bzgl . **ÿ. q0 Wir unterscheiden zwei Fälle:**

Fall q0 ÿ keine Änderungÿ . **Behauptung A1.2b zeigt, dass** q = **q0. Da** A **schemavollständig für S,** h **ÿ S und q0 ÿ** I **ÿ** keine Änderungÿ ist, **zeigt Lemma 7 ,** dass **q0 ÿÿ q0 bzgl. ÿ. Da** q = **q0, ergibt dies** h **ÿ** L(A).

Fall q0 ÿÿ keine Änderungÿ . **Da** A **schemavollständig für S und** h **ÿ S** ist , **gibt es keine blockierenden Läufe auf** h , die in **q0 starten .** Daher können wir Behauptung **A1.3b mit** P = **ÿ auf (q0,** ÿ) **ÿÿ** (q, ÿ) **bzgl. ÿ** snc **anwenden . Dies zeigt, dass q0 ÿÿ** q **bzgl. ÿ und daher** h **ÿ** L(A).

Damit ist der Beweis der inversen Inklusion und damit von L(A) = L(A snc) abgeschlossen. □

Verweise

1. **Marian, A.; Siméon, J. Projecting XML Documents. In** Proceedings of the VLDB; **Morgan Kaufmann: San Francisco, CA, USA, 2003; S. 213–224.**
2. **Frisch, A. Reguläre Baumsprachenerkennung mit statischen Informationen. In Proceedings of the Exploring New Frontiers of Theoretical Informatics, 18. Weltcomputerkongress der IFIP, 3. Internationale Konferenz für Theoretische Informatik der TCS, Toulouse, Frankreich, 22.–27. August 2004; S. 661–674.**
3. **Maneth, S.; Nguyen, K. XPath Whole Query Optimization.** VLPB J. **2010, 3, 882–893.**
4. **Sebastian, T.; Niehren, J. Projektion für verschachtelte Wortautomaten beschleunigt XPath-Auswertung von XML-Streams. In Proceedings der International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM), Harrachov, Tschechische Republik, 23.–28. Januar 2016.**
5. **Kay, M. The Saxon XSLT and XQuery Processor. 2004. Online verfügbar: https://www.saxonica.com (abgerufen am 29. Juli 2024).**
6. **Gienieczko, M.; Murlak, F.; Paperman, C. Unterstützung von Nachkommen in SIMD-beschleunigtem JSONPath. In Proceedings der 29. ACM International Conference on Architectural Support for Programming Languages and Operating Systems, La Jolla, CA, USA, 27. April–1. Mai 2024; Band 1, S. 15–29.**
7. **Al Serhali, A.; Niehren, J. Eine Benchmark-Sammlung deterministischer Automaten für XPath-Abfragen. In Proceedings of the XML Prague 2022, Prag, Tschechische Republik, 9.–11. Juni 2022.**
8. **Niehren, J.; Sakho, M. Determinisierung und Minimierung von Automaten für verschachtelte Wörter erneut betrachtet.** Algorithms **2021,** 14, **68. https://doi.org/10.3390/a14030068.**
9. **Comon, H.; Dauchet, M.; Gilleron, R.; Löding, C.; Jacquemard, F.; Lugiez, D.; Tison, S.; Tommasi, M. Baumautomatentechniken und Anwendungen. Online verfügbar: http://tata.gforge.inria.fr (abgerufen am 29. Juli 2024).**
10. **Thatcher, JW Charakterisierung von Ableitungsbäumen kontextfreier Grammatiken durch eine Verallgemeinerung der Automatentheorie.** J. Comput Syst. Sci. **1967, 1, 317–322. https://doi.org/10.1016/S0022-0000(67)80022-9.**
11. **Alur, R. Die Verbindung von Wörtern und Bäumen. In Proceedings des 26. ACM SIGMOD-SIGACT-SIGART Symposiums über Prinzipien der Database Systems, Peking, China, 11.–14. Juni 2007; ACM-Press: New York, NY, USA, 2007, S. 233–242.**
12. **Alur, R.; Madhusudan, P. Sichtbar Pushdown-Sprachen. In Proceedings des 36. jährlichen ACM-Symposiums zur Theorie des Rechnens, Chicago, IL, USA, 13.–16. Juni 2004; Babai, L., Hrsg.; ACM: New York, NY, USA, 2004; S. 202–211. https: //doi.org/10.1145/1007352.1007390.**
13. **Mehlhorn, K. Pebbling Mountain Ranges und seine Anwendung der DCFL-Erkennung. In Proceedings of the Automata, Languages and Programming, 7. Kolloquium, Noordweijkerhout, Niederlande, 14.–18. Juli 1980; Proceedings; de Bakker, JW, van Leeuwen, J., Hrsg.; Springer: Berlin/Heidelberg, Deutschland, 1980; Band 85, S. 422–435. https://doi.org/10.1007/3-540-10003- 2_89.**
14. **von Braunmühl, B.; Verbeek, R. Input Driven Languages are Recognized in log n Space. In** Topics in the Theory of Computation; **North-Holland Mathematics Studies; Karpinski, M., van Leeuwen, J., Hrsg.; North-Holland: 1985; Band 102, S. 1–19. https://doi.org/10.1016/S0304-0208(08)73072-X.**
15. **Okhotin, A.; Salomaa, K. Komplexität von eingabegesteuerten Kellerautomaten.** SIGACT News **2014,** 45, **47–67. https://doi.org/10.1145/2636805.2636821.**
16. **Niehren, J.; Sakho, M.; Al Serhali, A. Schemabasierte Automatenbestimmung. In Proceedings des 13. Internationalen Symposiums für Spiele, Automaten, Logik und formale Verifizierung, GandALF 2022, Madrid, Spanien, 21.–23. September 2022; Ganty, P., Monica, DD, Hrsg.; EPTCS: 2022; Band 370, S. 49–65. https://doi.org/10.4204/EPTCS.370.4.**
17. **Lick, A.; Schmitz, S. XPath Benchmark. Online verfügbar: https://archive.softwareheritage.org/browse/directory/1ea68cf5bb3f9f3f2fe8c7995f1802ebadf17fb5 (abgerufen am 29. Juli 2024).**

Machine Translated by Google

18. **Mozafari, B.; Zeng, K.; Zaniolo, C. Leistungsstarke komplexe Ereignisverarbeitung über XML-Streams. In** Proceedings der SIGMOD- Konferenz; **Candan, KS, Chen, Y., Snodgrass, RT, Gravano, L., Fuxman, A., Candan, KS, Chen, Y., Snodgrass, RT, Gravano, L., Fuxman, A., Hrsg.; ACM: New York, NY, USA, 2012; S. 253–264. https://doi.org/10.1145/2213836.2213866.**
19. **Grez, A.; Riveros, C.; Ugarte, M. Ein formaler Rahmen für die Verarbeitung komplexer Ereignisse. In Proceedings der 22. Internationalen Konferenz zur Datenbanktheorie (ICDT 2019), Lissabon, Portugal, 26.–28. März 2019; S. 5:1–5:18. https://doi.org/10.4230/LIPIcs. ICDT.2019.5.**
20. **Muñoz, M.; Riveros, C. Streaming Enumeration on Nested Documents. In Proceedings of the 25th International Conference on Database Theory, ICDT 2022, Edinburgh, Großbritannien, 29. März–1. April 2022; Olteanu, D., Vortmeier, N., Hrsg.; Schloss Dagstuhl - Leibniz-Zentrum für Informatik: 2022; Band 220, S. 19:1–19:18. https://doi.org/10.4230/LIPIcs.ICDT.2022.19.**
21. **Al Serhali, A.; Niehren, J. Früheste Abfragebeantwortung für deterministische schrittweise Hedge-Automaten. In Proceedings of the Implementation and Application of Automata—27. Internationale Konferenz, CIAA 2023, Famagusta, Nordzypern, 19.–22. September 2023; Proceedings; Nagy, B., Hrsg.; Springer: Berlin/Heidelberg, Deutschland, 2023; Band 14151, S. 53–65. https: //doi.org/10.1007/978-3-031-40247-0_3.**
22. **Gauwin, O.; Niehren, J.; Tison, S. Earliest Query Answering for Deterministic Nested Word Automata. In Proceedings des 17. Internationalen Symposiums zu Grundlagen der Computertheorie, Wroclaw, Polen, 2.–4. September 2009; Springer: Berlin/Heidelberg, Deutschland, 2009; Band 5699, S. 121–132.**
23. **Debarbieux, D.; Gauwin, O.; Niehren, J.; Sebastian, T.; Zergaoui, M. Frühe verschachtelte Wortautomaten für die Beantwortung von XPath-Abfragen in XML-Streams.Theor.** Comput. Sci. **2015,** 578, **100–125. https://doi.org/10.1016/j.tcs.2015.01.017.**
24. **Al Serhali, A.; Niehren, J. Subhedge-Projektion für schrittweise Hedge-Automaten. In Proceedings of the Fundamentals of Computation Theory – 24. Internationales Symposium, FCT 2023, Trier, Deutschland, 18.–21. September 2023; Proceedings; Fernau, H., Jansen, K., Hrsg.; Springer: Berlin/Heidelberg, Deutschland, 2023; Band 14292, S. 16–31. https://doi.org/10.1007/978-3-031-43587-4_2.**
25. **Neumann, A.; Seidl, H. Locating Matches of Tree Patterns in Forests. In Proceedings of the Foundations of Software Technology and Theoretical Computer Science, Chennai, Indien, 17.–19. Dezember 1998; Springer: Berlin/Heidelberg, Deutschland, 1998; Band 1530, Seiten 134–145.**
26. **Gauwin, O.; Niehren, J.; Roos, Y. Streaming Tree Automata.** Inf. Process. Lett. **2008,** 109, **13–17. https://doi.org/10.1016/j.ipl.2008 .08.002.**
27. **Franceschet, M. XPathMark Performance Test. Online verfügbar: https://users.dimi.uniud.it/~massimo.franceschet/xpathmark/ PTbench.html (abgerufen am 25. Oktober 2020).**
28. **Hosoya, H.; Pierce, BC XDuce: Eine statisch typisierte XML-Verarbeitungssprache.** ACM Trans. Internet Technol. **2003, 3, 117–148. https://doi.org/10.1145/767193.767195.**
29. **Gécseg, F.; Steinby, M.** Baumautomaten; **Akadémiai Kiadó: Budapest, Ungarn, 1984.**
30. **Gold, EM-Sprachidentifikation im Grenzfall.** Inf. Control. **1967,** 10, **447–474. https://doi.org/10.1016/S0019-9958(67)91165-5.**
31. **Carme, J.; Niehren, J.; Tommasi, M. Abfragen nicht eingestufter Bäume mit schrittweisen Baumautomaten. In Proceedings of the Rewriting Techniques and Applications, 15. Internationale Konferenz, RTA 2004, Aachen, Deutschland, 3.–5. Juni 2004; Proceedings; van Oostrom, V., Hrsg.; Springer: Berlin/Heidelberg, Deutschland, 2004; Band 3091, S. 105–118. https://doi.org/10.1007/978-3-540-25 979-4_8.**
32. **Paar, C.; Quéré, A. Definition und Etüde des zweisprachigen Regelwerks.** Inf. Control **1968,** 13, **565–593.**
33. **Murata, M. Hedge Automata: Ein formales Modell für XML-Schemata. 2000. Online verfügbar: http://www.xml.gr.jp/relax/hedge_ nett.html (abgerufen am 29. Juli 2024).**
34. **Martens, W.; Niehren, J. Zur Minimierung von XML-Schemas und Baumautomaten für nicht geordnete Bäume.** J. Comput. Syst. Sci. **2007,** 73, **550–583.**
35. **Niehren, J. Stepwise Hedge Automata sind exponentiell prägnanter als Forest Automata. 2024,** in Vorbereitung.
36. **Bojanczyk, M.; Walukiewicz, I. Forest-Algebren. In** Proceedings of the Logic and Automata: History and Perspectives [in Honor of Wolfgang Thomas]; **Flum, J., Grädel, E., Wilke, T., Hrsg.; Amsterdam University Press: Amsterdam, Niederlande, 2008; Band 2, S. 107–132.**

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Artikel

Hierarchisches Optimierungsframework für das Layoutdesign von Stern-Baum-Gassammelpipeline-Netzwerk in diskreten Räumen

Yu Lin [1], Yanhua Qiu [1], Hao Chen [2], Jun Zhou 3,*, Jiayi He[3], Penghua Du [3] und Dafan Liu [4]

1 Forschungsinstitut für Sammel- und Transporttechnik, PetroChina Southwest Oil & Gasfield Company, Chengdu 610041, China; linyu223@petrochina.com.cn (YL); qiuyan0320@petrochina.com.cn (YQ)
2 Sichuan Shale Gas Exploration und Entwicklung Co., Ltd., Chengdu 610051, China; chenhao05@petrochina.com.cn
3 Schule für Erdöltechnik, Southwest Petroleum University, Chengdu 610500, China; hjywyyx@163.com (JH); scdupenghua@163.com (PD)
4 Abteilung für Schiefergasexploration und -entwicklung, CNPC Chuanqing Drilling Engineering Co., Ltd., Chengdu 610066, China; liudf_cqyyq@cnpc.com.cn
* Korrespondenz: zhoujunswpu@163.com

Zusammenfassung: Das Gassammelleitungsnetz ist eine kritische Infrastruktur für die Sammlung und Beförderung von Erdgas von der Förderstelle zur Verarbeitungsanlage. In diesem Dokument wird ein Designoptimierungsmodell für ein sternförmig-baumförmiges Gassammelleitungsnetz in einem diskreten Raum vorgestellt, das darauf abzielt, die optimale Konfiguration dieser Infrastruktur zu bestimmen. Ziel ist es, die für den Aufbau des Netzes erforderlichen Investitionen zu reduzieren. Zu den wichtigsten Entscheidungsvariablen gehören die Standorte der Stationen, der Standort der Anlage, die Verbindungen zwischen Bohrlöchern und Stationen sowie die Verbindungen zwischen den Stationen. Es werden mehrere Gleichheits- und Ungleichheitsbeschränkungen formuliert, die sich hauptsächlich auf die Verbindung zwischen Bohrlöchern und Stationen, den Übertragungsradius und die Kapazität der Stationen beziehen. Das Design eines sternförmig-baumförmigen Pipelinenetzes stellt ein komplexes, nichtdeterministisches, polynomisches (NP) und schwieriges kombinatorisches Optimierungsproblem dar . Um diese Herausforderung zu bewältigen, wird ein hierarchisches Optimierungsframework in Verbindung mit einem verbesserten genetischen Algorithmus (IGA) vorgeschlagen. Die Wirksamkeit des genetischen Algorithmus wird durch Tests und Vergleiche mit anderen herkömmlichen Algorithmen validiert. Anschließend werden das Optimierungsmodell und die Lösungsmethodik auf die Layoutgestaltung eines Pipeline-Netzwerks angewendet. Die Ergebnisse zeigen, dass die optimierte Netzwerkkonfiguration die Investitionskosten im Vergleich zum ursprünglichen Entwurf um 16 % senkt. Darüber hinaus wird beim Vergleich des optimalen Layouts unter einer Stern-Stern-Topologie festgestellt, dass die für die Stern-Stern-Topologie erforderlichen Investitionen 4 % höher sind als die für die **Stern-Baum-Topologie erforderlichen .**

Schlüsselwörter: Topologische Struktur; Layout-Design; hierarchische Optimierung; genetischer Algorithmus

check for updates

Zitat: Lin, Y.; Qiu, Y.; Chen, H.; Zhou, J.; Er, J.; Du, P.; Liu, D. Hierarchisches Optimierungsframework für die Layoutgestaltung eines Stern-Baum-Gassammelpipeline-Netzwerks in diskreten Räumen. Algorithms **2024,** 17, **340. https://doi.org/10.3390/a17080340**

Wissenschaftliche Herausgeber: Frank Werner, Ming-Feng Ge und Chen Liu

Erhalten: 3. Juli 2024
Überarbeitet: 27. Juli 2024
Akzeptiert: 3. August 2024
Veröffentlicht: 5. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/ 4.0/).

1. Einleitung

1.1. Motivation

Derzeit befindet sich die Weltgemeinschaft im Übergang in ein Zeitalter, das von sauberer und effizienter Energie geprägt ist, wobei Erdgas sich als vorherrschende Energiequelle herauskristallisiert. Dem Global Energy Outlook zufolge wird der weltweite Erdgasverbrauch bis 2050 voraussichtlich kontinuierlich steigen, mit einer durchschnittlichen jährlichen Wachstumsrate von über 1,4 % (Richard, 2020). Der deutliche Anstieg des Erdgasbedarfs erfordert verstärkte Anstrengungen beim Bau von oberflächlichen Gaspipelinenetzen. Die für Gaspipelinenetze erforderlichen Investitionen sind beträchtlich und machen oft über 30 % der gesamten Entwicklungskosten eines Gasfeldes aus [1].
Folglich **kann eine rationale Optimierung des Designs dieser Pipelinenetze** erhebliche wirtschaftliche Vo

Das System umfasst in der Regel Brunnen, Gassammelstationen, zentrale Aufbereitungsanlagen und Pipelines (Abbildung 1). Zunächst wird das aus den Brunnen geförde

Machine Translated by Google

Die Gassammelstation wird über Produktionspipelines versorgt. Nach der Erfassung in der Gassammelstation wird das Gas dann durch Sammelpipelines zur zentralen Aufbereitungsanlage transportiert, wo es gereinigt wird und den für die externe Verteilung erforderlichen Standards entspricht Basierend auf der Netzwerktopologie können Gassammelpipeline-Netzwerke in drei grundlegende Typen eingeteilt werden: Stern-, Baum- und zyklische Topologien [2]. In einer Sterntopologie sind mehrere Bohrlöcher direkt mit einer einzigen Station verbunden, wie in Abbildung 2a dargestellt. In einer Baumtopologie sind Bohrlöcher über benachbarte Bohrlöcher mit der Station verbunden, wie in Abbildung 2b dargestellt. Bei einer zyklischen Topologie bilden miteinander verbundene Bohrlöcher eine Schleife, wie in Abbil

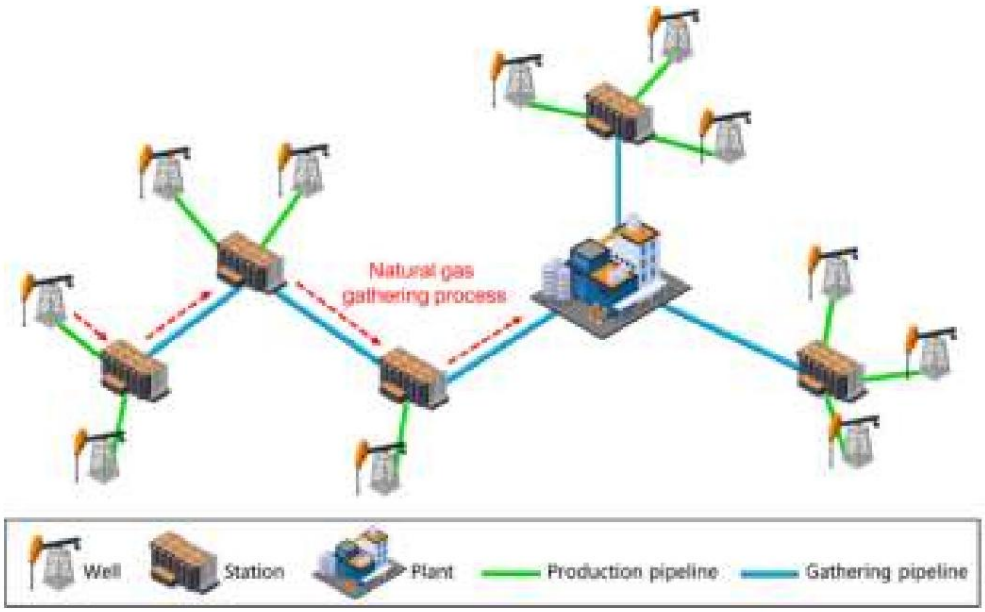


Abbildung 1. Typisches Gassammelpipelinenetz.

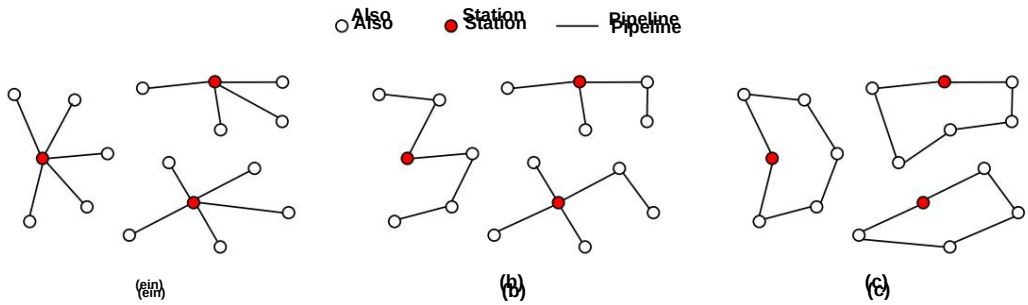


Abbildung 2. Struktur des Pipeline-Netzwerks: (a) Sterntopologie, (b) Baumtopologie, (c) zyklische Topologie.

In der Praxis werden in groß angelegten Gassammelpipeline-Netzwerken häufig verschiedene topologische Kombinationen beobachtet. Derzeit werden bei der tatsächlichen Anlage von Gasfeldern basierend auf einer Vielzahl von Faktoren, einschließlich Gasquellenverteilung, Baukosten, Verwaltungskomfort, Betriebseffizienz usw., häufig Netzwerkstrukturen mit Stern-Stern-Topologie und Stern-Baum-Topologie bei der Entwicklung von Gasfeldern verwendet. Die Netzwerkstruktur mit Stern-Stern-Topologie aus mehreren Gasquellen, die direkt mit einer zentralen Gassammelstelle verbunden sind, macht das Layout einfach und leicht zu verwalten und zu kontrollieren. Ihre Vorteile bestehen darin, dass die Gesamtlänge der Pipeline reduziert werden kann, die anfänglichen Baukosten gesenkt werden können und das von den Gasfeldern produzierte Erdgas bequem zentral verarbeitet und transportiert werden kann , was für Gasfelder mit einem kleinen gasführenden Gebiet, einer großen Anzahl von Bohrlochköpfen und einer geringen Leistung einer einzelnen Quelle geeignet ist. Die Netzwerkstruktur mit Stern-Baum-Topologie kombiniert die Eigenschaften der Stern- und dendritischen Struktur. In dieser Struktur sind jede Gasquelle und ihre Gassammelstationen sternförmig

Machine Translated by Google

gleicht die Kosten und die Effizienz der Langstreckensammelleitung aus und bietet gleichzeitig eine gewisse Flexibilität zur Anpassung und Erweiterung je nach Entwicklungsstadium des Feldes. Abbildung 3 zeigt ein typisches Stern-Baum-Gassammelpipeline-Netzwerk. Diese Netzwerkstruktur eignet sich besonders für Szenarien, in denen die Bohrlöcher in lokalen Blöcken konzentriert sind, diese Blöcke jedoch weit verstreut sind und erhebliche Entfernungen zwisc

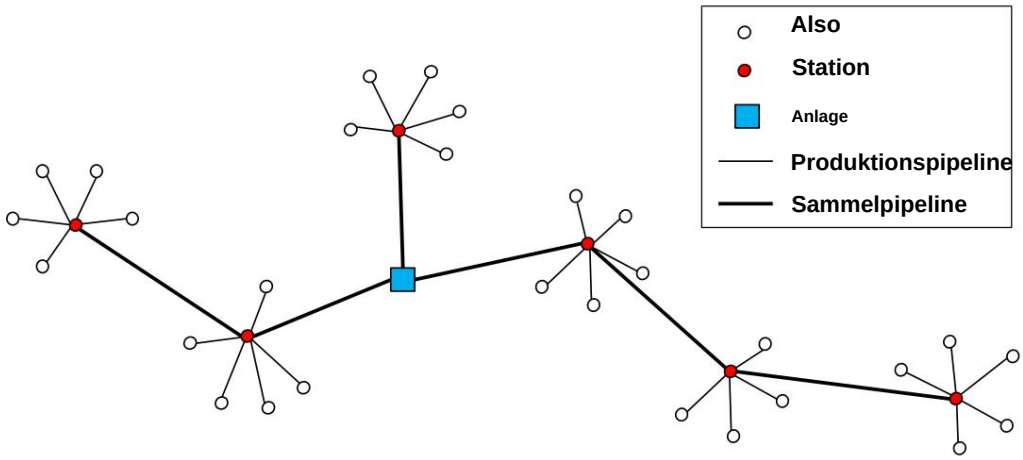


Abbildung 3. Stern-Baum-Pipeline-Netzwerk.

Im Allgemeinen kann das Optimierungsproblem in Entwurfsoptimierung und Betriebsoptimierung unterteilt werden **[3]. Bei der Entwurfsoptimierung geht es darum, die am besten geeignete Layoutkonfiguration, Stationsplatzierungen, Rohrgrößen und Gerätegrößen für ein neues Pipeline- Netzwerk auszuwählen, um optimale Bauinvestitionskosten zu erzielen. Bei der Betriebsoptimierung geht es dagegen um die Verfeinerung der Betriebsstrategien des bestehenden Pipeline- Netzwerks, was die Optimierung der Pipeline-Durchflüsse und Knotendruckparameter zur Erreichung technischer oder w**

Die Herausforderung der Layoutoptimierung von Gassammelpipeline-Netzwerken umfasst mehrere Teilprozesse, wie die Gruppierung von Bohrlöchern, die Standortbestimmung von Stationen und Anlagen sowie die Herstellung der Konnektivität zwischen den Anlagen [4]. Abbildung 4 zeigt den Layoutoptimierungsprozess. Darüber hinaus ist die Bestimmung des Stations- (oder Anlagen-)Standorts in zwei Szenarien unterteilt. Das erste Szenario besteht darin, dass die Station an jedem beliebigen Ort innerhalb des Gasfeldes gebaut werden kann . Das zweite Szenario besteht darin, dass die Station gemeinsam mit dem Bohrloch gebaut werden muss, um Bauinvestitionskosten zu sparen. Mit anderen Worten kann der Stationsstandort nur aus den vorhandenen Bohrlochstandorten ausgewählt werden, was als Auswahl des Stationsstandorts in einem diskreten Raum definiert ist. Da dieses Szenario in der Ingenieurpraxis weiter verbreitet ist, konzentriert sich dieser Artikel auf dieses Problem.

1.2. Literaturübersicht

Aus Literaturrecherchen geht hervor, dass die wissenschaftliche Gemeinschaft umfassende Studien zur Optimierung des Designs von Gaspipelinenetzen durchgeführt hat. Diese Studien konzentrieren sich hauptsächlich auf die Modellierung und Lösung von Optimierungsproblemen im Zusammenhang mit dem Netzwerklayout . Abhängig von der Topologie des Pipelinenetzes – etwa Stern-, Baum- oder zyklische Topologie – treten unterschiedliche Optimierungsprobleme auf. Sanaye und Mahmoudimehr [5] entwickelten ein Optimierungsmodell für Gaspipelinenetze mit Baumtopologie, das Parameter wie Netzwerklayout, Rohrdurchmesser und Knotendrücke berücksichtigt. Rodriguez et al. [6] untersuchten das Design mehrstufiger Sternnetze und befassten sich dabei mit optimalen Konfigurationen und Anlagenplatzierungen. Fu [7] erkundete das mehrzielige Optimierungsdesign ringförmiger Ölsammelpipelinenetze. Wang et al. [8] formulierten ein mehrperiodisches MILP- Modell (Mixed Integer Linear Programming) zur Optimierung von Netzwerken mit Baumtopologie, das Gelände und Hindernisse berücksichtigt. Zhou et al. [9] etablierte ein MINLP (Mixed Integer Nonlin-

Machine Translated by Google

ear Programming) Modell zur Optimierung von Pipeline-Netzwerken mit zweistufiger Sterntopologie für unterirdische Erdgasspeicher.

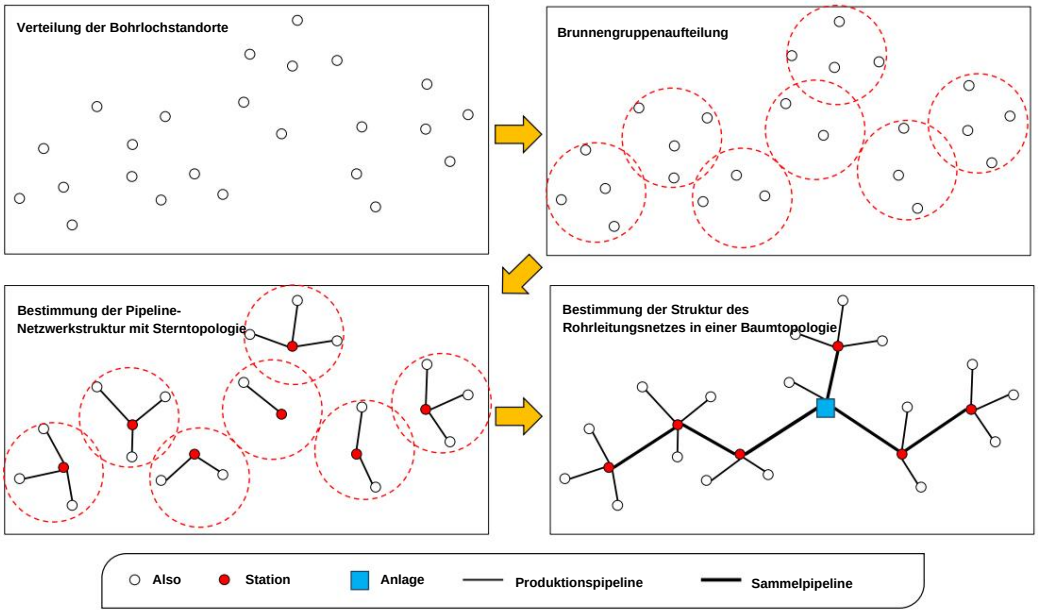


Abbildung 4. Hierarchisches Optimierungsdiagramm.

In Bezug auf die Optimierungslösungsmethoden wurden in der frühen Forschung vorwiegend Operations Research-Techniken, insbesondere heuristische Algorithmen, eingesetzt, um die Gasansammlung zu a Pipeline-Netzwerk-Optimierungsmodelle. Aus einer mathematischen Optimierung und Operation Aus wissenschaftlicher Sicht ist die Planung optimaler Gassammelleitungsnetze eine NP-schweres kombinatorisches Optimierungsproblem, das erhebliche Herausforderungen mit sich bringt und Anwendung in verschiedenen Industriebereichen [10,11]. Balinski [12] formulierte ein Optimierungsmodell für die Platzierung von Standorten in diskreten Räumen mit unbegrenzten Standortkapazitäten, das hauptsächlich mit heuristischen Algorithmen gelöst. Gängige exakte Algorithmen sind der Verzweigungsalgorithmus und Begrenzungsverfahren, Schnittebenenalgorithmus und Spaltengenerierungsalgorithmus [13,14]. Traditionelle Optimierungstheorien haben jedoch oft Schwierigkeiten, Probleme effizient zu lösen mit Erdgasleitungsnetzmodelle aufgrund langer Rechenzeiten und gelegentlicher Diskrepanz zwischen theoretischen Ergebnissen und praktischen Anwendungen.

In den letzten Jahren sind kommerzielle Solver und heuristische Algorithmen allmählich die häufigste Art, mit der Optimierung des Pipeline-Netzwerklayouts umzugehen. Ein kommerzieller Solver kann die mühsame Algorithmus-Programmierarbeit eliminieren, aber die Lösungsmethode ist begrenzt, und es ist schwierig, die eingebauten Parameter des Algorithmus zu ändern. Die heuristische Algorithmus hat einen großen Programmieraufwand, aber der Algorithmus hat eine gute Skalierbarkeit und können sogar die Vorteile mehrerer Algorithmen kombinieren, um Optimierungsergebnisse zu erzielen. Uester und Dilaveroglu [15] lösten ein Problem der Planung und Optimierung eines Gaspipeline-Netzwerks mithilfe des Online-Bonmin-Solvers. Zhang et al. [16] entwickelten eine umfassende MILP-Layoutoptimierungsmodell, das Gelände- und Hindernisüberlegungen berücksichtigt. Es gibt auch Wissenschaftler, die Methoden der Graphentheorie und des maschinellen Lernens verwenden, um Optimierungsprobleme zu lösen [17,18]. Sie verwendeten den Gurobi-Solver, um die Modell. Rosa et al. [19] stellten einen praktischen und effizienten Optimierungsansatz für Unterwasser-Produktionsnetzwerke mit dem GAMS-Solver. Liu et al. [20] schlugen eine verbesserte Partikelschwarm-Optimierungsalgorithmus mit erweiterten globalen Suchfunktionen zur Bewältigung das MINLP-Modell zur Optimierung von Pipeline-Netzwerklayouts. Hong et al. [21] schlugen ein Modell zur Optimierung des Rohrleitungsnetzes, das hydraulische Eigenschaften berücksichtigt und verwendete zur Lösung einen Ameisenkolonie-Algorithmus.

Machine Translated by Google

1.3. Beiträge

Investitionen in oberflächliche Gasfelder machen etwa 30 bis 40 % der gesamten Entwicklungsinvestitionskosten aus , und der Anteil der Investitionen in den Bau oberflächlicher Sammel- und Transportpipelinenetze macht oft 60 bis 70 % der oberflächlichen Investitionskosten aus. Es ist ersichtlich, dass ein Erdgassammel- und -transportpipelinenetz eine wichtige Position beim Bau von unterirdischen Gasfeldern einnimmt und seine Designoptimierung die wirtschaftlichen

Zurzeit reichen einige traditionelle Optimierungsmethoden nicht aus, um mehrstufige Strukturen von Pipeline-Netzwerken zu untersuchen und eine effektive Lösung zu finden. Daher schlägt dieser Artikel eine Reihe von Optimierungsmethoden für das komplexere Stern-Baum- Pipeline-Netzwerk zur Gassammlung in diskreten Räumen vor.

Basierend auf der untersuchten Literatur haben Wissenschaftler bedeutende Fortschritte bei der Optimierung des Entwurfs von Gaspipeline-Netzwerken gemacht. Ihr Schwerpunkt lag jedoch hauptsächlich auf der Optimierung von Einzelnetztopologien und vernachlässigte die Komplexität, die durch mehrere topologische Kombinationen entsteht, insbesondere die in Öl- und Gasfeldern vorherrschenden Stern-Baum-Mehrebenenkonfigurationen . Die Einbeziehung unterschiedlicher Netzwerktypen erhöht die Komplexität der Lösung dieser

Modelle erheblich. Wir gehen dieses Problem wie folgt an: (1) Wir entwickeln ein MINLP-Modell für den opt Raum.

(2) Es wird ein hierarchisches Optimierungskonzept vorgeschlagen. Im ersten Schritt liegt der Schwerpunkt auf der Optimierung des sternförmigen Rohrleitungsnetzes, das die Bohrlöcher mit den Stationen verbindet. In der zweiten Phase wird die Gestaltung des Rohrleitungsnetzes mit Baumtopologie optimiert, das die Stationen mit den Anlagen verbindet.

(3) Ein verbesserter genetischer Algorithmus (IGA) wird vorgeschlagen und der Kruskal-Algorithmus wird verwendet um optimale Layoutlösungen für das Baumnetzwerk zu erreichen.

Abschnitt 2 beschreibt das mathematische Modell. Abschnitt 3 führt eine hierarchische

Optimierungsrahmen. Abschnitt 4 stellt einen praktischen Fall vor und untersucht das Optimierungsergebnis . Abschließend bietet Abschnitt 5 eine Zusammenfassung.

2. Mathematisches Modell

Innerhalb der Planung des Erdgassammel- und -transportpipelinenetzes gibt es zwei Hauptmodelltypen: ein MIP-Modell und ein MINLP-Modell. Im MIP-Modell werden verschiedene Punkt-zu-Punkt-Entfernungen im Voraus berechnet. Bei der Erschließung von Gasfeldern ist der Maßstab der Anzahl der Bohrlöcher sehr groß, und die Berechnung der Entfernung im Voraus ist sehr umfangreich; daher sind im Modell dieses Dokuments die Pipelinelänge von den Bohrlöchern zu den Stationen Lij **und die Pipelinelä**

Entfernungen, die nichtlineare Gleichungen sind, bzw. $Lij = \sqrt{xi \ddot{y} xj^2 + yi \ddot{y} yj^2}^{0,5}$

und $Ljj\ddot{y} = \sqrt{xj \ddot{y} xj \ddot{y}^2 + yj \ddot{y} yj \ddot{y}^2}^{0,5}$ **, was die Lösungseffizienz verbessern kann. In**

Außerdem sind die Stationsstandort-Entscheidungsvariable *(ÿj),* **die Werksstandort-Entscheidungsvariable** *(ÿk)* **und die Pipeline-Anschluss-Entscheidungsvariable** *(ÿij,ÿjjÿ)* **binäre Variablen. Daher handelt es sich bei dem Modell in diesem Dokument um ein MINLP-Modell. Das Modell ist gut skalierbar, sodass die Anzahl der Knoten in Zukunft erhöht werden kann.**

2.1. Zielfunktion Es wird

ein MINLP-Modell vorgeschlagen. Als **Zielfunktion des Modells wird die Mindestinvestition verwendet ,** wie in Gleichung (1) dargestellt.

$$\textbf{Min .}\ F = FPP + FGP + FS + FC \quad \textbf{(1)}$$

wobei F **die Investition in das Pipeline-Netzwerk in CNY ist.** FPP **die Investition in die Produktionspipeline in CNY ist.** FGP **die Investition in die Sammelpipeline in CNY ist.** FS **die Investition in die Station in CNY ist.** FC **die Investition in die Anlage in CNY ist.**

Machine Translated by Google

(1) Investitionen in die Produktionspipeline.

Die Investition in Produktionspipelines entspricht den Kosten für die Verbindung aller Bohrlöcher und Stationen im Netzwerk, wie in Gleichung (2) gezeigt.

$$FPP = \ddot{y}\ \ddot{y}\ \ddot{y}ijcijLij \quad (2)$$

wobei ÿij **die Verbindungsentscheidungsvariable zwischen Brunnen und Stationen ist, 0 oder 1.** cij **ist der Preis der Pipeline zwischen Brunnen und Stationen, CNY/m.** Lij **ist die Pipelinelänge von Brunnen zu Stationen, m.** Lij = xi ÿ xj (xj , yj) **ist die** ^{2} + yi ÿ yj ^{2} ^{0,5} **,** (xi , yi) **ist die Brunnenkoordinate, und Stationskoordinate.**

(2) Sammeln von Pipeline-Investitionen.

Die Investition in die Sammelleitungen entspricht den Kosten für die Verbindung der Leitungen zwischen allen Stationen im Netzwerk, wie in Gleichung (3) dargestellt.

$$FGP = \ddot{y}\ \ddot{y}\ \ddot{y}jjÿ\ cjjÿ\ Ljjÿ \quad (3)$$

wobei ÿjjÿ **die Verbindungsentscheidungsvariable zwischen den Stationen ist, 0 oder 1.** cjjÿ **der Preis der Pipeline zwischen den Stationen ist, CNY/m.** Ljjÿ **ist die Pipelinelänge zwischen den Stationen, M.** Ljjÿ = xj ÿ xjÿ - **Koordinate.** ^{2} + yj ÿ yj ÿ ^{2} ^{0,5} **,** (xj , yj) **und** (xj ÿ , yj ÿ) **sind die Stationen oder Anlagen**

(3) Bahnhofsinvestitionen.

Die Stationsinvestition ist die Summe der Investitionskosten aller Stationen im Netz, wie in Gleichung (4) dargestellt.

$$FS = \ddot{y}\ \ddot{y}j\ fS \quad (4)$$

wobei ÿj **die Entscheidungsvariable für den Stationsstandort ist, 0 oder 1.** fS **sind die Kosten der Station, CNY.**

(4) Anlageinvestitionen.

Die Anlageninvestition ergibt sich aus der Summe der Investitionskosten aller Anlagen im Leitungsnetz , wie in Gleichung (5) dargestellt.

$$FC = \ddot{y}\ \ddot{y}k\ fP \quad (5)$$

wobei ÿk **die Entscheidungsvariable für den Werksstandort ist, 0 oder 1.** fP **sind die Kosten des Werks, CNY.**

2.2. Einschränkungen

(1) Eindeutigkeitsbeschränkung von Brunnenverbindungen.

Jeder Brunnen ist ausschließlich einer Station zugeordnet, wie in Gleichung (6) gezeigt.

$$\ddot{y}\ \ddot{y}ij = 1\ddot{y}i\ \ddot{y}\ l \quad (6)$$

(2) Anlagenanzahlbeschränkung.

Innerhalb des Rohrleitungsnetzes existiert eine einzelne Anlage, wie in Gleichung (7) gezeigt.

$$\ddot{y}\ \ddot{y}k = 1 \quad (7)$$

(3) Brunnen- und Stationszugehörigkeitsbeschränkung.

Nun kann i **der Station** j zugewiesen werden , **wenn Station** j **existiert, wie in Gleichung (8) gezeigt.**

$$\ddot{y}ij\ \ddot{y}\ \ddot{y}j\ \ddot{y}i\ \ddot{y}\ l,\ \ddot{y}j\ \ddot{y}\ J \quad (8)$$

Machine Translated by Google

(4) Einschränkung des Übertragungsradius.

Der Gasdruck kann während des Transports aufgrund von Rohrreibung abnehmen. Um den Druckabfall in einem vernünftigen Rahmen zu halten, sollte der Abstand zwischen der Bohrung und der Station kleiner als ein bestimmter Wert sein, der als Übertragungsradius bezeichnet wird und in Abbildung 5 dargestellt ist.

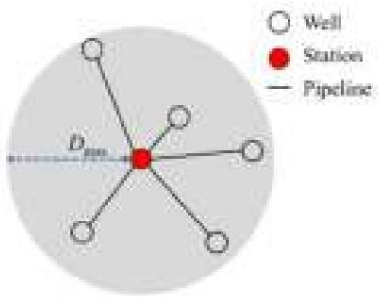


Abbildung 5. Schematische Darstellung des Übertragungsradius.

Die Übertragungsradiusbeschränkung wird wie in Gleichung (9) gezeigt berechnet.

$$\text{ÿijLij ÿ } \mathbf{Dmax}_{\text{ÿiÿ}} \mathbf{I,} \text{ ÿj ÿ J} \quad \text{(9)}$$

wobei Dmax die maximal zulässige Rohrleitungslänge von der Bohrung zur Station in m ist.

max

(5) Kapazitätsbeschränkung der Stationen.

Im Allgemeinen wird die Einschränkung des Sammelradius nicht berücksichtigt, wenn die Einschränkung der Stationskapazität behandelt wird, die unter das Problem der kapazitätsbegrenzten Standortwahl (CFLP) fällt. In diesem Dokument werden sowohl die Radiusbeschränkung als auch die Kapazitätsbes Wenn die Sammelradiusbeschränkung eingehalten wird, die Kapazitätsbeschränkung jedoch nicht, wird das Layout des Rohrleitungsnetzes neu gestaltet (siehe Abbildung 6).

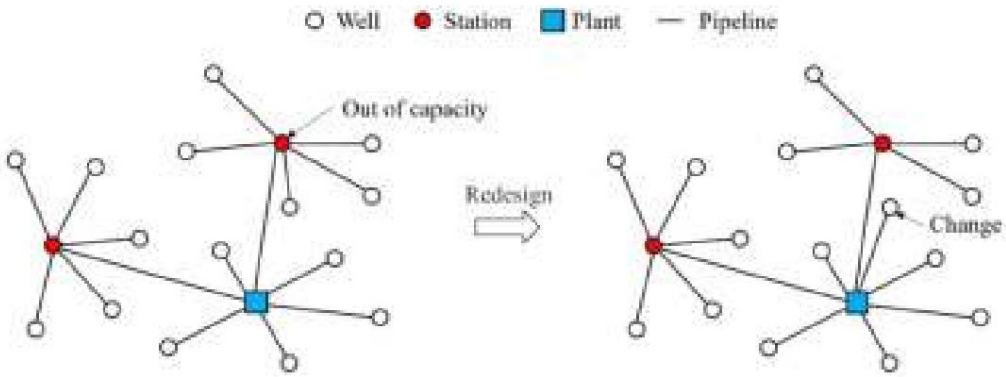


Abbildung 6. Neugestaltung des Pipeline-Netzwerks unter Verletzung von Kapazitätsbeschränkungen.

Daher darf die Kapazität einer Station den Wert aus Gleichung (10) nicht überschreiten.

$$\text{ÿjqj ÿ } \mathbf{qmax} \text{ ÿj ÿ J} \quad \text{(10)}$$

max

wobei qj der Knotengasfluss der Station j in m3/d ist. qmax ist die Kapazität der Station in m3/d.

(6) Einschränkung der Baumstruktur. **max**

Die Baumverbindungsbeschränkung stellt sicher, dass die generierte Topologie die kürzeste und schleifenfreie ist, wie in Gleichung (11) gezeigt.

$$\text{ÿ ÿ ÿjj = 2ÿ ÿj} \quad \text{(11)}$$

$$\text{jÿ}_{\text{ÿ}}\text{ÿ}_{\text{ÿ}}\text{jÿ}_{\text{ÿ}} \qquad \text{jÿJ}$$

ÿÿ ÿ2

ÿ ÿ 0

Machine Translated by Google

(7) Knotenflussbeschränkung.

Die Knotenflussbeschränkung stellt sicher, dass das Gesamtgasvolumen konstant bleibt, wie in Gleichung (12) gezeigt.

$$\ddot{y}\ qi\ \ddot{y}\ \ddot{y}\ \underset{ich\ddot{y}ich}{} \quad \underset{k\ddot{y}K}{} \ \ddot{y}kqk = \mathbf{0} \tag{12}$$

wobei qi **der Knotengasfluss von Bohrloch i in m3/d ist.** qk **ist der Knotenfluss von Anlage k in m3/d.**

2.3. Entscheidungsvariablen

Zu den Entscheidungsvariablen gehören die Stationsstandort-Entscheidungsvariable *(ÿj),* **die Werksstandort- Entscheidungsvariable** *(ÿk)* **und die Pipeline-Anschluss-Entscheidungsvariable** *(ÿij,ÿjjÿ).* **Alle diese Entscheidungsvariablen sind binäre Variablen, wie in Gleichung (13) gezeigt.**

$$\ddot{y}j\ ,\ \ddot{y}k\ ,\ \ddot{y}ij,\ \ddot{y}jj\ddot{y} = \mathbf{\{0,}\ 1\}\ i\ \ddot{y}\ \mathbf{I,}\ j\ \ddot{y}\ \mathbf{J,}\ j \qquad \overset{\ddot{y}}{} \ \ddot{y}\ \mathbf{J,}\ k\ \ddot{y}\ K \tag{13}$$

3. Lösungsmethode

Die Entwurfsoptimierung eines Stern-Baum-Pipeline-Netzwerks gilt als NP-schwer, was die Suche nach einer umfassenden Lösung erschwert [22]. Um dieses Problem zu lösen, wird in dieser Studie ein hierarchisches Optimierungsframework vorgeschlagen. Das Framework umfasst zwei Phasen: In der ersten Phase werden die Aufteilung der Bohrlochgruppen und die Stationsstandorte innerhalb des Sterntopologienetzwerks bestimmt, während in der zweiten Phase die Stationsverbindungen und Anlagenstandorte im Baumtopologienetzwerk festgelegt werden. Insbesondere die Optimierung des Sterntopologienetzwerks in der ersten Phase dient als grundlegende und zentrale Komponente des gesamten Optimierungsproblems. Seine Komplexität und sein Umfang übertreffen die der Baumtopologienetzwerkoptimierung der zweiten Phase. Daher konzentriert sich dieses Dokument auf die Verbesserung der Entwurfsoptimierung und stellt einen IGA vor, der auf diese Herausforderung zugeschnitten ist. Zusätzlich wird der Kruskal-Algorithmus, der den Minimal-Spannbaum-Ansatz nutzt, für die Entwurfsoptimierung des Baumtopologie-Pipeline-Netzwerks der zweiten Phase eingesetzt .

3.1. Grundlegender genetischer

Algorithmus **Ein GA ist eine heuristische Suchmethode, die sich an den Prinzipien der natürlichen Selektion und der Genetik orientiert. Ziel ist es, die Zuteilung von Einrichtungen durch Simulation des Evolutionsprozesses biologischer Chromosomen zu optimieren. Der grundlegende Lösungsprozess des GA [23] läuft wie folgt ab:**

(1) Kodierung: Die Informationen über die Verteilung der Brunnen und Stationen werden in der Chromosomensequenz kodiert.

(2) Generierung der Anfangspopulation: Eine feste Größe von Chromosomen (Individuen) wird zufällig generiert, um die Anfangspopulation P(t) zu bilden **(in diesem Dokument werden 200 Gruppen g Dabei** bezeichnet t **den Evolutionsgenerationszähler und** T **stellt die für den Algorithmus festgelegte maximale Anzahl von Evolutionszyklen dar.**

(3) Fitnessfunktion: Jedes Individuum in der Population wird anhand einer Fitnessfunktion bewertet. Als Bewertungskriterium wird typischerweise der Gesamtinvestitionswert verwendet.

(4) Selektion und Vererbung: Durch Simulation **des Prozesses der biologischen Evolution wird die Ausgangspopulation selektiert und gekreuzt ,** und **durch Schritt 3** wird die Population der nächsten Generation **P(t + 1)** gewonnen.

(5) Beurteilung der Abbruchbedingung: Wenn t ÿ **T, dann** ersetzt t **+ 1 t und wir gehen zu Schritt 2 über; wenn** t **> T, betrachten wir das Individuum mit optimaler Fitness durch den Evolutionsprozess als optimale Lösung und beenden die Operation [24].**

Der obige Prozess wird auch durch das in Abbildung 7 dargestellte Flussdiagramm beschrieben .

Machine Translated by Google

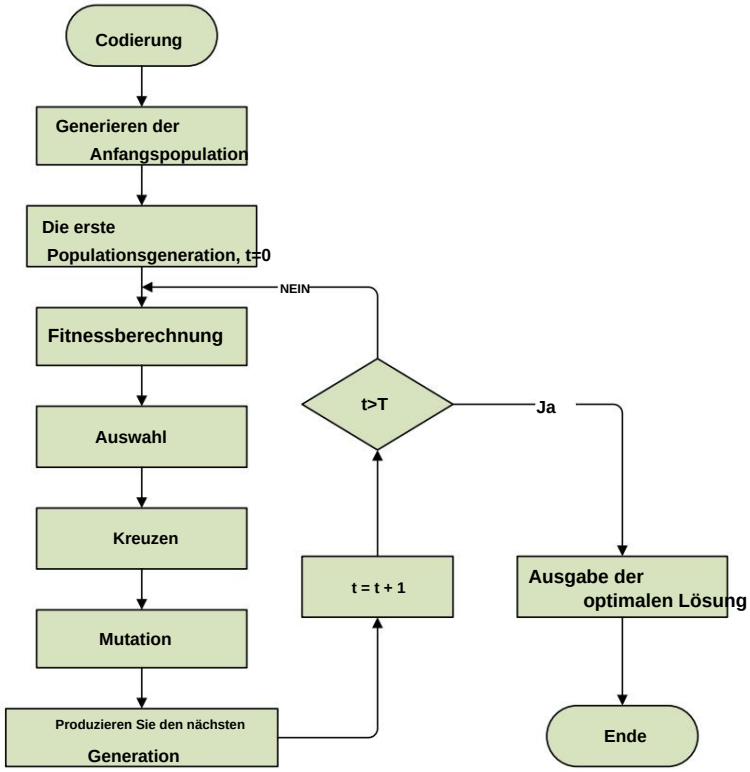


Abbildung 7. Flussdiagramm des grundlegenden GA.

3.2. Verbesserung des Algorithmus

Angesichts der diskreten Natur des Raums und der spezifischen Struktur von Gasfeld-Pipeline-Netzwerken stimmen der traditionelle GA und seine zugehörigen Parameter möglicherweise nicht vollständig mit dem in diesem Dokument vorgeschlagenen mathematischen Modell überein. Daher ist es wichtig, den grundlegenden GA zu verbessern, um diesen Eigenschaften besser gerecht zu werden und das Pipeline-Netzwerkdesign effektiv zu optimieren. Die Verbesserung umfasst die Gesamtkodierung der Bohrloch-Station-Korrespondenz, die Formulierung der Fitnessfunktion der Netzwerklösung und die Neugestaltung des Crossover-Operators und des Mutationsoperators.

(1) Kodierungsdesign.

Unter Berücksichtigung der Eigenschaften der Einschränkungen und unter Verwendung der Verbindungsbeziehung zwischen Brunnen und Stationen als Kodierungsgrundlage wird eine Realzahlkodierung verwendet, um die Effizienz des GA zu verbessern. Die chromosomenkodierende Sequenz ist so strukturiert, dass sie sich direkt auf die Zielfunktion auswirkt, die in diesem Fall die Gesamtinv Die Länge jedes Chromosoms entspricht der Gesamtzahl der Bohrlöcher, und innerhalb der Sequenz wird die Kennung jedes Bohrlochs systematisch in aufsteigender Reihenfolge dargestellt. Diese Kodierungsmethode stellt sicher, dass das GA die Konstruktionsbeschränkungen optimal berücksichtigt und effektiv auf die Optimierung der Konfiguration des Gaspipeline-Netzwerks abzielt. Die entsprechende Stationsnummer jedes Bohrlochs wird an jeder Position innerhalb des Chromosoms stochastisch generiert, um die Verbindung zwischen dem Bohrloch und seiner jeweiligen Station anzuzeigen. Im vorgeschlagenen Chromosomenkodierungsschema stellt jede Ziffer in einem Chromosom = (C1, C2, ... Cm) **die Nummer der Station dar, die jedem Bohrloch**

(2) Entwurf eines Crossover-Operators.

Für den GA sind die Crossover-Operationsmethode und die Crossover-Wahrscheinlichkeit äußerst wichtige Parameter, die die Optimierungsqualität des Algorithmus direkt beeinflussen. Unter Verwendung von CAP101-Daten basierend auf OR-LIBRARY (50 Brunnen, 25 Stationen) [25] werden **die Crossover-Operationsmethode und die Crossover-Wahrscheinlichkeit auf ihren Einfluss auf die Lösung getestet und der beste Crossover-Modus und die beste Crossover-Wahrscheinlichkeit bestimmt.**

Jede der Crossover-Techniken wurde 50 Mal getestet und ihre Ergebnisse wurden zum Vergleich gemittelt. Mit zunehmender Anzahl iterativer Generationen wurde der Algorithmus zunehmend

Machine Translated by Google

Die Multipoint-Crossover-Methode zeigte eine überlegene Leistung, was zu ihrer Auswahl führte. Abbildung 8 zeigt die durchschnittlichen Testergebnisse.

Die Crossover-Wahrscheinlichkeit variiert zwischen 0,5 und 0,9. Ein Vergleich der Auswirkungen dieser unterschiedlichen Crossover-Wahrscheinlichkeiten auf die Ergebnisse wurde durch Mittelung der Ausgestammt aus 50 Tests. Wenn die Crossover-Wahrscheinlichkeit zwischen 0,7 und 0,9 liegt, sind die Ergebnisse ähnlich. Eine angemessene Erhöhung der Crossover-Wahrscheinlichkeit kann die Konvergenz der Lösung verbessern, daher sollte die Crossover-Wahrscheinlichkeit im Design auf einem hohen Wert gehalten werden. Die Abbildung näherte sich der optimalen Lösung an. Unter den verschiedenen untersuchten Cro Point-Crossover-Verfahren. Das Multi-Point-Crossover-Verfahren zeigte eine überlegene Leistung, was zu seiner Auswahl. Abbildung 8 veranschaulicht die durchschnittlichen Testergebnisse.

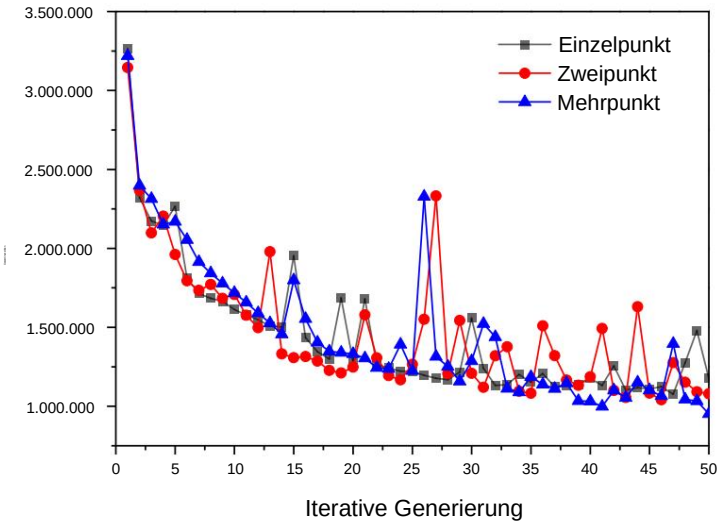


Abbildung 8. Konvergenzkurve verschiedener Crossover-Betriebsmethoden.
Abbildung 8. Konvergenzkurve verschiedener Crossover-Betriebsmethoden.

Die Crossover-Wahrscheinlichkeit variiert zwischen 0,5 und 0,9. Ein Vergleich der Effekte
Die Bewertung dieser unterschiedlichen Crossover-Wahrscheinlichkeiten auf die Ergebnisse erfolgte durch die Mittelung der Ergebnisse aus 50 Tests. Wenn die Crossover-Wahrscheinlichkeit zwischen 0,7 und 0,9 liegt, sind die Ergebnisse sind ähnlich. Eine angemessene Erhöhung der Crossover-Wahrscheinlichkeit kann die Konvergenz verbessern der Lösung, daher sollte die Crossover-Wahrscheinlichkeit im Entwurf auf einem hohen Wert gehalten werden. Abbildung 9 zeigt die Auswirkung unterschiedlicher Crossover-Wahrscheinlichkeiten auf die Ergebnisse im Mehrpunkt-Kreuzungsmethode.

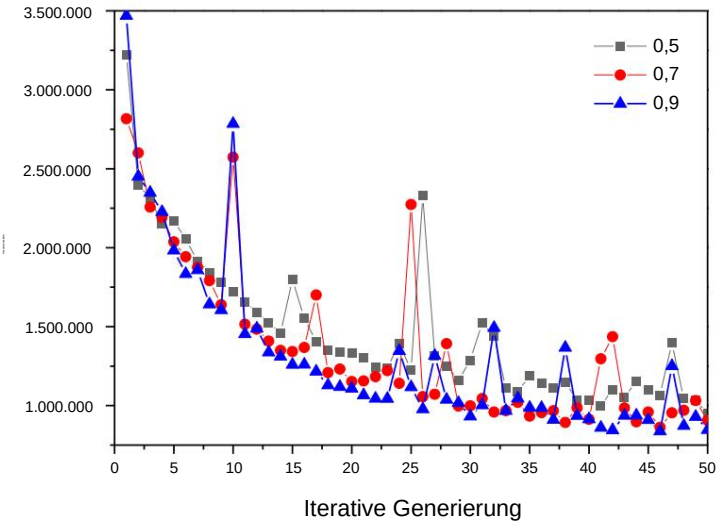


Abbildung 9. Konvergenzkurve verschiedener Crossover-Wahrscheinlichkeiten.

(8) Entwurf von Mutationsoperatoren

Der Mutationsoperator ist eine Maßnahme, die verhindert, dass der GA in das lokale Optimum fällt, indem er die genetischen und evolutionären Werte des Rupturallels simuliert Optimum durch Simulation der genetischen und evolutionären Werte des Rupturallels [26]. Im Prozess Beim Debuggen des Mutationsoperators wird eine Variationsfunktion für den GA getestet, um das Risiko zu verringern, in späteren Phasen in lokale Optima zu fallen. Die Mutationsoperatoren verwenden einen adaptiven Ansatz, wie in Formel (14) angegeben.

Der Mutationsoperator ist eine Maßnahme, die verhindert, dass der GA in den lokalen lokale Optimum fällt, indem er die genetischen und evolutionären Werte des Rupturallels simuli das Debuggens des Mutationsoperators wird eine Variationsfunktion für den GA getestet, um

ÿ ÿ ÿ ÿ ÿ + ÿ (14)
ÿ ÿ
= ÿ ÿ +

Machine Translated by Google

das Risiko verringern, in späteren Phasen in lokale Optima zu fallen. Die Mutationsoperatoren verwenden ein adaptiver Ansatz, wie in Formel (14) angegeben.

$$MP = \ddot{y} \cdot \frac{\text{Fitmax-Fit}}{\text{FitmaxÿFitmin}} \quad \text{Passform } \ddot{y} \; \frac{\text{Fitmax+Fitmin}}{2}$$
$$MP = \ddot{y} \quad \text{Passform} < \frac{\text{Fitmax+Fitmin}}{2} \tag{14}$$

wobei MP **die Mutationswahrscheinlichkeit ist;** Fit **der Fitnesswert; Fitmax und Fitmin die maximaler und minimaler Fitnesswert jeder Generation; und** $\ddot{y}$ **ist der Mutationskoeffizient.**

Durch Tests wurde festgestellt, dass bei einem Mutationskoeffizienten unter 0,003 die Die iterative Stabilität und Konvergenz des gesamten Lösungsprozesses ist besser und die Lösung Zeit ist sinnvoller; wenn der Mutationskoeffizient jedoch einen Wert von mehr als 0,1 hat, Die ermittelte Lösung kann weit von der optimalen Lösung entfernt sein, was die Stabilität beeinträchtigt der Iteration und der Konvergenzgeschwindigkeit und anderer Probleme. Daher ist die Mutation Der Koeffizient wurde so gewählt, dass er unter 0,003 liegt.

Die Auswirkungen der Mutationswahrscheinlichkeit auf Investitionen sind in Tabelle 1 dargestellt. Diese Experimente mit CAP101 zeigen, dass bei einer Mutationswahrscheinlichkeit von einem bestimmten Wert, es gewährleistet die Bevölkerungsvielfalt, ohne die Konvergenz der GA zu beeinträchtigen zur globalen Optimallösung. Der Hauptgrund hierfür ist, dass bei reellen Zahlen Codierung verwendet wird, enthält jede Position auf jedem Chromosom mehr Informationen als wenn andere Kodierungsmethoden verwendet werden, und wenn die Mutationswahrscheinlichkeit hoch ist, ist es einfach, das dominante Individuum direkt zerstören.

Tabelle 1. Auswirkungen unterschiedlicher Mutationswahrscheinlichkeiten auf die GA-Leistung.

Problemgröße	Gesamtzahl der Berechnungen	$\ddot{y}$	Durchschnittliches optimales Ergebnis
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. IGA-Tests

Um die optimale Gestaltung von Stern-Baum-Pipeline-Netzwerken in diskreten Räumen zu verbessern, In diesem Dokument werden Verbesserungen am Kodierungsschema, am Crossover-Operator und am Mutationsoperator des grundlegenden GA hervorgehoben. Die Lösungsschritte des IGA sind in Abbildung 10 darg

Zur Validierung der Wirksamkeit des modifizierten genetischen Algorithmus bei der Lösung der Pipeline Netzwerk-Layout-Design-Problem, wurden eine Reihe von Algorithmen-Tests durchgeführt. Alle Algorithmen wurden in der Software MATLAB 2022 implementiert. Die numerischen Experimente wurden durchgeführt auf einem Computer mit einer Intel(R) Core (TM) i5-2450M CPU @ 2,50 GHz, 4 GB RAM und das Betriebssystem Windows 7.

Zum Vergleich mit dem verbesserten GA wurden zwei klassische Algorithmen ausgewählt, nämlich Partikelschwarmoptimierung mit Einschnürungsfaktor (CPSO) und Tabu-Suche (TABU). Mehrere Probleme in der OR-LIBRARY wurden für algorithmische Tests verwendet, und die Lösung Ergebnisse sind in Tabelle 2 dargestellt. LOCAL ist die tatsächlich optimale Lösung. n **steht für gut Nummer.** m **steht für Stationsnummer. Für kleinere Anzahlen von Brunnen und Stationen, wie z. B. Im Fall von CAP101 sind die Vorteile der Verwendung des IGA nicht ohne weiteres ersichtlich. Tatsächlich IGA benötigt möglicherweise mehr Rechenzeit als andere Algorithmen. Da jedoch die Wenn die Anzahl der** n **(Brunnen) und** m **(Stationen) zunimmt, werden die Vorteile der Effizienz der IGA in**

Machine Translated by Google

tt

Lösungsprobleme werden immer offensichtlicher. Es tendiert dazu, zur optimalen Lösung zu konvergieren schneller als andere Algorithmen, wenn n **und** m **größer werden.**

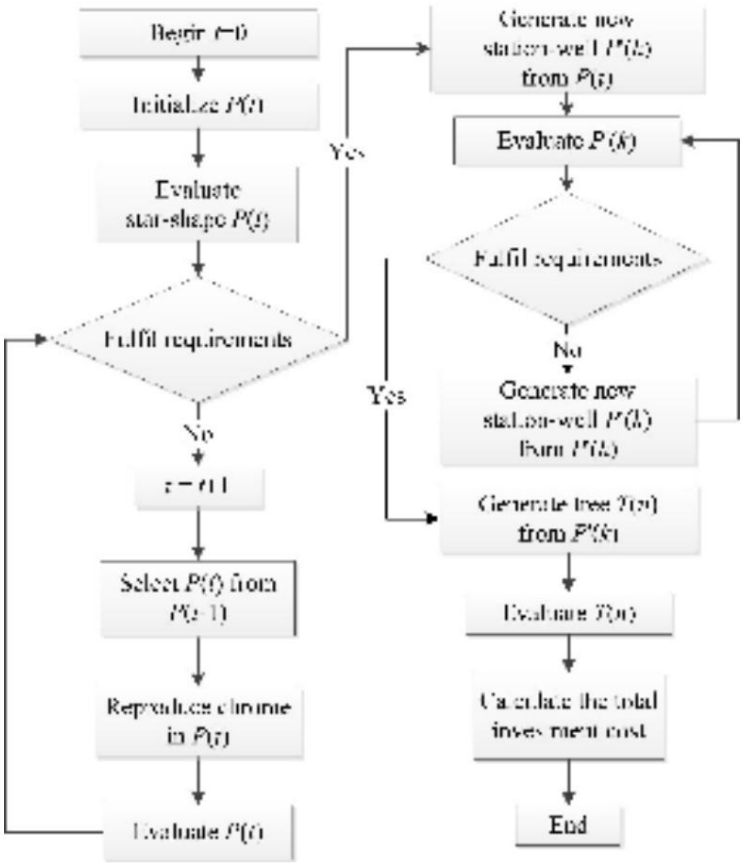


Abbildung 10. Flussdiagramm des IGA.

ff

Tabelle 2. Vergleich mit anderen Algorithmen.

Problem	Größe (n × m)	Algorithmus	Optimale Lösung (Durchschnitt)	Zeit (s)
CAP101	50 × 25	LOKAL	796.648,437	-
		TABU	797.508,70	0,19
		CPSO	796.648,44	0,88
		IGA	796.668,16	0,37
CAP103	50 × 25	LOKAL	893.782,11	-
		TABU	894.008,10	0,24
		CPSO	893.782,11	0,99
		IGA	895.192,66	0,17
CAP104	50 × 25	LOKAL	928.941,75	-
		TABU	928.941,80	0,15
		CPSO	928.941,75	0,60
		IGA	928.946,48	0,17

Machine Translated by Google

Tabelle 2. Fortsetzung

Problem	Größe (n × m)	Algorithmus	Optimale Lösung (Durchschnitt)	Zeit (s)
CAP131	50 × 50	LOKAL	793.439,56	-
		TABU	794.299,90	0,20
		CPSO	793.439,56	3.62
		IGA	793.902,48	0,37
CAP133	50 × 50	LOKAL	893.076,71	-
		TABU	893.782,10	0,30
		CPSO	893.076,71	3,78
		IGA	893.994,01	0,22
CAPB	1000 × 100	LOKAL	13.198.556,43	-
		TABU	13.214.718,1	3,91
		IGA	13.211.519,82	3.26

3.4. Kruskal-Algorithmus

Bei der Anbindung der einzelnen Stationen in einer Baumstruktur besteht die grundlegende Aufgabe darin, das Problem des minimalen Spannbaums. Zu den häufig verwendeten Algorithmen beim Entwurf des minimalen Spannbaums gehören der Prim-Algorithmus und der Kruskal-Algorithmus, der basiert auf dem Greedy-Prinzip. Durch den Vergleich der beiden Algorithmen wurde festgestellt, dass die Der Kruskal-Algorithmus ist dem Prim-Algorithmus in der Lösungszeit immer überlegen, daher haben wir gewählt den Kruskal-Algorithmus als Algorithmus für den Entwurf von Baumtopologie-Netzwerken zu verwenden. Der Kruskal Algorithmus ist ein Spannbaumalgorithmus, der auf gewichteten Kanten basiert. Im Lösungsprozess die kleinste gewichtete Kante wird nacheinander ausgewählt und dem Baumsatz hinzugefügt. Wenn die hinzugefügte Kante erzeugt eine zyklische Topologie, die gewichtete Kante wird entfernt, bis alle Kanten ausgewählt [27].

Der Lösungsprozess des Kruskal-Algorithmus ist in Abbildung 11 dargestellt. **Alle Knoten sind in den Verbindungsknotensatz** U **und den unverbundenen Streuknotensatz V** unterteilt. V = **{v1, v2, v3, v4, v5} und** U = {ÿ}. **Zunächst werden die Knoten, die die kürzeste Kante bilden, ausgewählt und zum Join-Knotensatz hinzugefügt, d. h.** U = **{v4, v5} und** V = **{v1, v2,** v3}. **Dann in Auf die gleiche Weise werden die Kanten v1, v2, v1, v3, v4, v2 und v4, v3 für den Verbindungsknotensatz ausgewählt. Wenn ein Es entsteht eine zyklische Typologie, die Kante, die sie erzeugt hat, wird entfernt und die Lösung ist abgeschlossen. Schließlich** ist V = *{ÿ}* **und** U = **{v1, v2, v3, v4,** v5}.

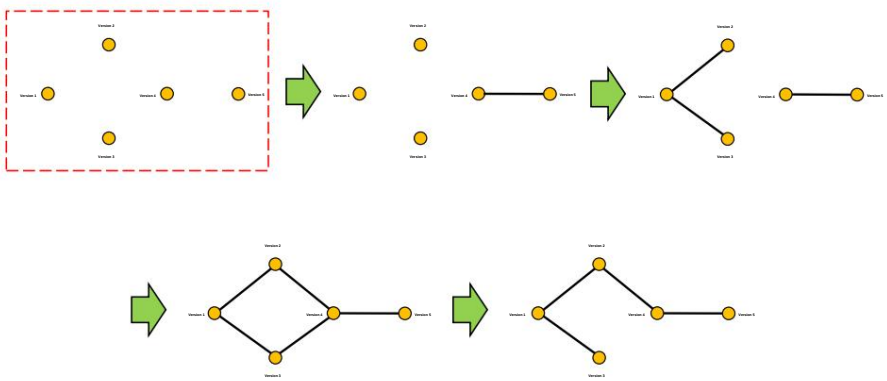


Abbildung 11. Schematische Darstellung des Lösungsprozesses des Kruskal-Algorithmus.

4. Ergebnisse und Diskussion

Für ein Erdgasfeld-Sammel- und Transportleitungsnetzsystem ist der Aufbau Design ist ein NP-Problem, das der Berechnung nicht förderlich ist. Um die Effizienz der Lösung werden folgende vier Annahmen getroffen: (1) Für die externe

Machine Translated by Google

Berechnungsbedingungen bei der Layoutgestaltung, die Standardbedingungen werden als Vorgabe für die Berechnung verwendet (Druck 0,1 MPa, Temperatur 20 °C). (2) Die Durchflussrate in der Gaspipeline ist instabil. Um dieses Problem der Einfachheit halber zu lösen, wird der Gasfluss in der Pipe als stabiler Fluss vereinfacht. (3) Bei der Pipe-Strömung ändert sich die Temperatur der vergrabenen Pipe normalerweise mit der Länge der Pipe, und wenn die Pipe zu lang ist, nähert sich die Gastemperatur schließlich der Umgebungstemperatur an. Um die Berechnungen zu vereinfachen und da sich das Sammelpipeline-Netzwerk normalerweise in derselben Gegend befindet, wird in diesem Dokument angenommen, dass die Temperatur der Pipe-Pipeline konstant ist. (4) Für die Fließrichtung des Gases im Netzwerk wird festgelegt, dass jeder Knoten das Gas nur von einer Richt

Zur Analyse und **Lösungssuche wurde ein reales Erdgaspipelinenetz in China ausgewählt [28]. Der ursprüngliche Aufbau des Pipelinenetzes im Gasfeld ist in Abbildung 12** dargestellt. **Dieses Gasfeld umfasst 93 Bohrungen, von denen jede zwischen 2000 und 5000 m3/d Gas produziert, mit Bohrlochdrücken zwischen 0,2 und 1,0 MPa und einer durchschnittlichen Produktionstemperatur von 20 °C. Die Spezifikation der primären Produktionspipelines zwischen Bohrungen und Stationen war De160 × 9,5 mm, während die Spezifikation der sekundären Sammelpipelines zwischen Stationen (und der Anlage) De225 × 12,8 mm war. Detaillierte Kosten sind in Tabelle 3 aufgeführt.**

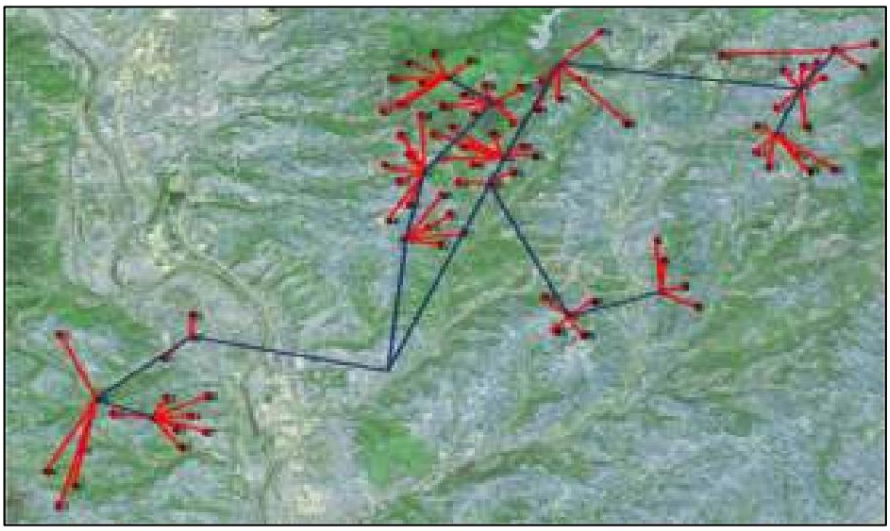


Abbildung 12. Skizze des ursprünglichen Layouts des Rohrleitungsnetzes.

Tabelle 3. Voraussichtliche Kosten des Pipeline-Netzwerks.

Projekt	Kosten
Station	CNY 52,17 × 104
Anlage	CNY 1725.93 × 104
Rohrleitungskosten (De160 × 9,5 mm)	CNY/km 15,636 × 104 CNY/
Rohrleitungskosten (De225 × 12,8 mm)	km 25,406 × 104

Die maximale Radiusbeschränkung betrug 1,5 km und die Verarbeitungskapazität einer einzelnen Station betrug 20.000 m3/Tag. Da die Daten aus einem Niederdruckgasfeld stammten, wurden Druckbeschränkungen in diesem Fall nicht berücksichtigt. Im ursprünglichen Algorithmus (GA) dauerte die Berechnung ffi jeweils mehr als 2 Stunden. Nach der Optimierung der Algorithmusstruktur (IGA) wurde die Berechnungszeit auf 15–30 Minuten verkürzt, wodurch die Berechnungseffizienz um über 400 % gesteigert wurde. Der IGA wurde verwendet, um das Layout des Pipeline-Netzwerks zu lösen.

Unter Verwendung des etablierten hierarchischen Optimierungsrahmens haben wir ein optimiertes Layoutschema abgeleitet. Das anfängliche Sterntopologielayout in der ersten Phase ist in dargestellt

Machine Translated by Google

Abbildung 13. Abbildung 14 zeigt die Ergebnisse der Gruppierung der Bohrlöcher und der Bestimmung der Stationsstandorte . Die endgültige optimierte Anordnung ist in Abbildung 15 dargestellt. Dieses Netzwerk besteht aus vierzehn Gassammelstationen und einer Aufbereitungsanlage, also einer Station weniger als im ursprünglichen Gasfeldentwurf.



Abbildung 13. Verbindungsbeziehung zwischen Brunnen und Stationen.

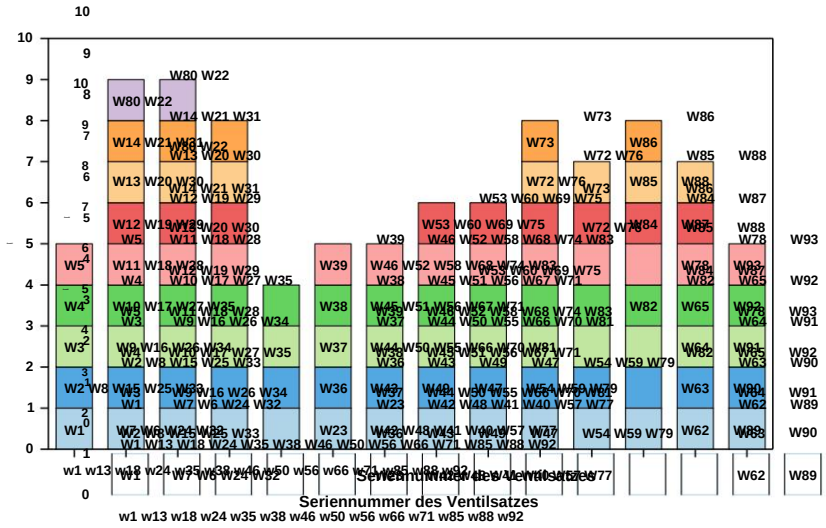


Abbildung 14. Ergebnisse für Brunnenzuteilung, Stationsstandort und Anlagenstandort.

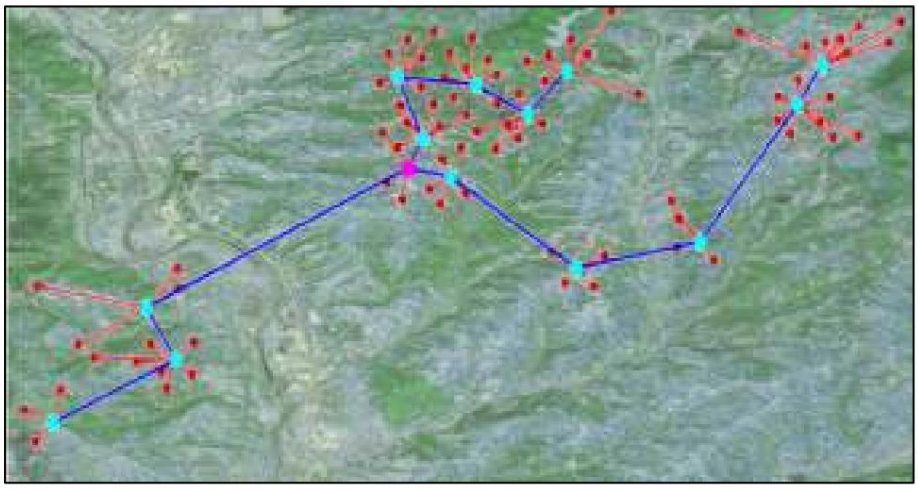


Abbildung 15. Entwurfsergebnis für die Stern-Baum-Anordnung des Rohrleitungsnetzes.

Machine Translated by Google

Mithilfe der Simulationssoftware wurde die Machbarkeit der Layoutstruktur des Rohrleitungsnetzes überprüft . Da die Anzahl der Brunnen im Rohrleitungsnetz zu groß war, wurde die Anzahl der Brunnen als Summe des Gasvolumens und des Knotens des Ventilverteilers für die Verbindungsüberprüfung bestimmt. Das Verbindungsdiagramm ist in Abbil Die entsprechende Druckabfallkalibrierung ist in Abbildung 17 dargestellt. Der Druckabfall lag innerhalb des zulässigen Druckabfallbereichs und erfüllte die Anforderungen der Gasübertrag

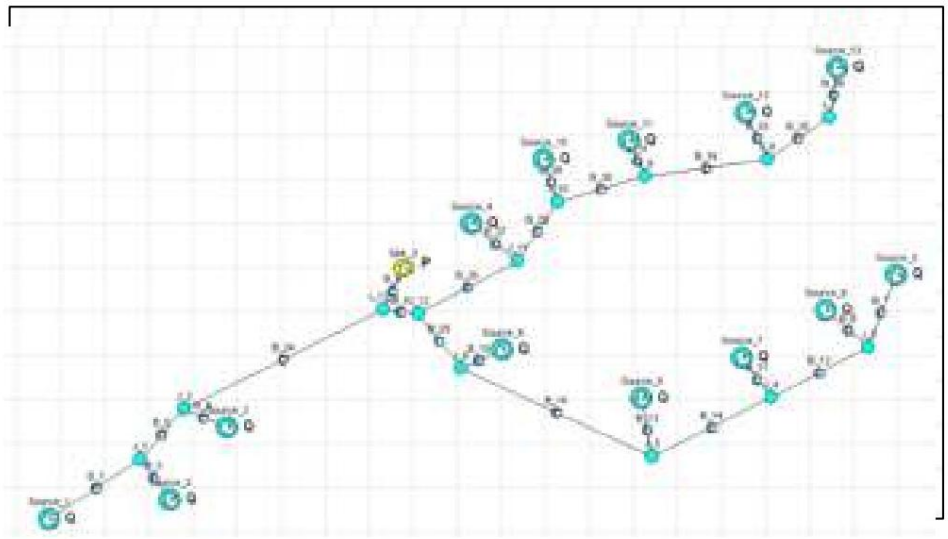


Abbildung 16. Darstellung der Simulation der Rohrleitungsnetzverbindung.

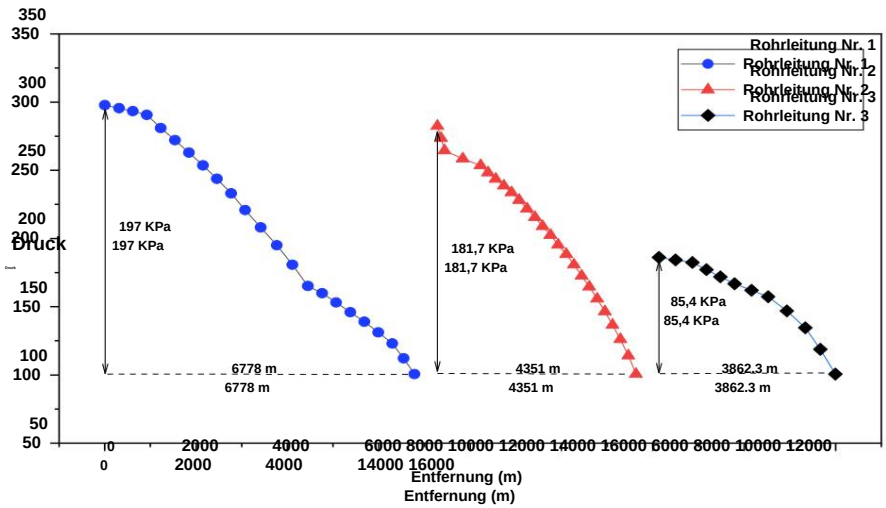
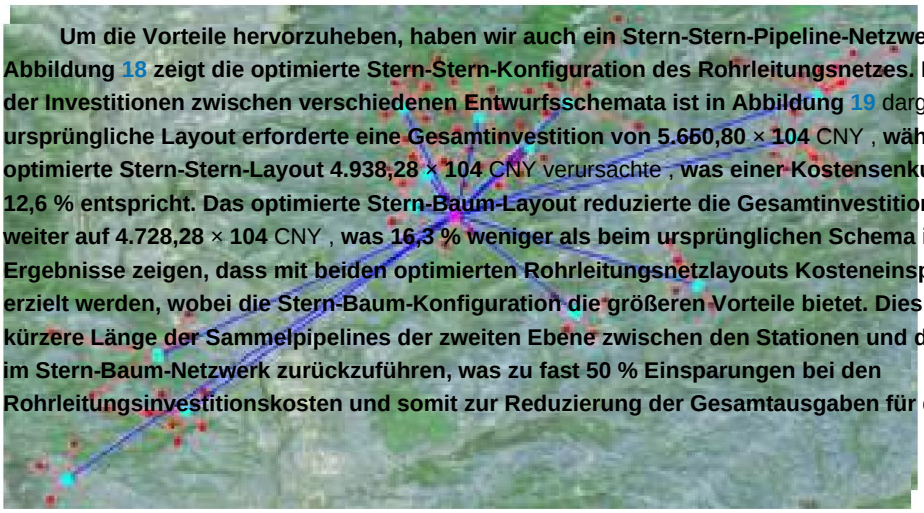


Abbildung 17. Ergebnisse der Druckabfallkalibrierung im Rohrleitungsnetzlayout.

Um die Vorteile hervorzuheben, haben wir auch ein Stern-Stern-Pipeline-Netzwerklayout evaluiert. Abbildung 18 zeigt die optimierte Stern-Stern-Konfiguration des Rohrleitungsnetzes. Ein Vergleich der Investitionen zwischen verschiedenen Entwurfsschemata ist in Abbildung 19 dargestellt. Das ursprüngliche Layout erforderte eine Gesamtinvestition von 5.660,80 × 104 CNY , während das optimierte Stern-Stern-Layout 4.938,28 × 104 CNY verursachte , was einer Kostensenkung von 12,6 % entspricht. Das optimierte Stern-Baum-Layout reduzierte die Gesamtinvestition , welche weiter auf 4.728,28 × 104 CNY , was 16,3 % weniger als beim ursprünglichen Schema ist. Diese Ergebnisse zeigen, dass mit beiden optimierten Rohrleitungsnetzlayouts Kosteneinsparungen erzielt werden, wobei die Stern-Baum-Konfiguration die größeren Vorteile bietet. Dies ist auf die kürzere Länge der Sammelpipelines der zweiten Ebene zwischen den Stationen und dem Werk im Stern-Baum-Netzwerk zurückzuführen, was zu fast 50 % Einsparungen bei den Rohrleitungsinvestitionskosten und somit zur Reduzierung der Gesamtausgaben für das Netzwerk führt .



Machine Translated by Google



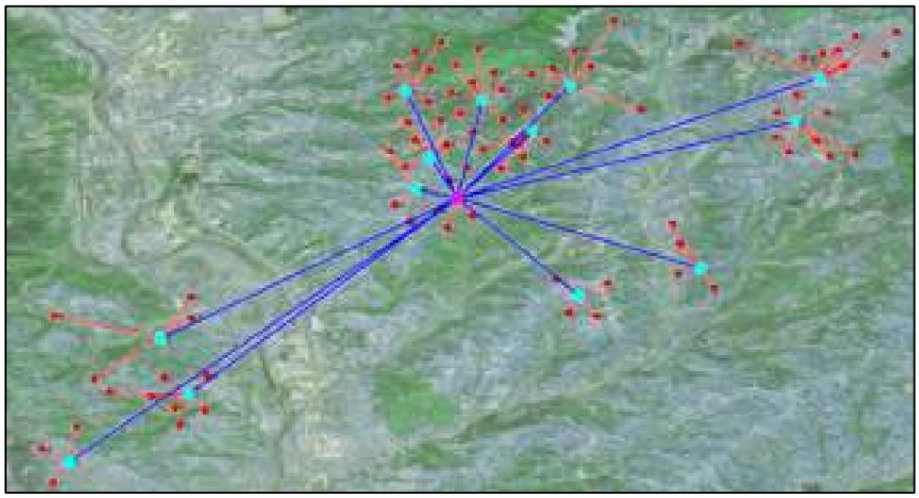


Abbildung 18. Entwurfsergebnis für die Stern-Stern-Anordnung des Rohrleitungsnetzes.

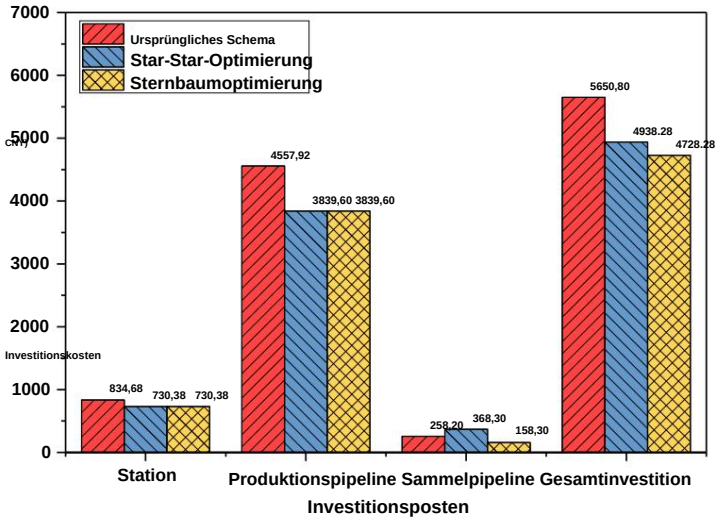


Abbildung 19. Vergleich der Investitionskosten.

5. Schlussfolgerungen

Die akademische Forschung zur Optimierung der Layouts von Gaspipeline-Netzwerken hat sich überwiegend auf einstufige, einstrukturierte Netzwerke konzentriert, während mehrstufigen und mehrstufigen kombinierten Gaspipeline-Netzwerken wenig Aufmerksamkeit geschenkt wurde. Um diese Lücke zu schließen, entwickelt dieses Dokument ein Designoptimierungsmodell für ein zweistufiges Stern-Baum-Gassammelpipeline- Netzwerk in diskreten Räumen. In Anbetracht der Komplexität des Problems und der Notwendigkeit rechnerischer Effizienz wird ein hierarchisches Optimierungsframework zusammen mit einem IGA entwickelt, um dieses Optimierungsproblem wirksam zu lösen. ffi Aufbauend auf diesem Modell haben wir das optimierte Design eines Stern-Baum-Pipeline -ff-

Netzwerks für ein Erdgasfeld in China ermittelt, was zu optimalen Designergebnissen führte. Die Erkenntnisse zeigen, dass mit dem optimierten Netzwerklayout die Investitionskosten im Vergleich zum ursprünglichen Schema um 16 % gesenkt werden, was einer Ersparnis von 922,52 × 104 CNY entspricht . Zusätzlich haben wir dieses optimierte Layout mit einer Stern-Stern-Topologie verglichen. Es wurde beobachtet, dass die Übernahme des Stern-Stern-Layouts die Investitionskosten im Vergleich zum Stern-Baum-Layout-Schema um 4 % erhöhen würde . Dieser Vergleich unterstreicht die überlegene Anwendbarkeit der Stern-Baum-Topologie für dieses spezielle Netzwerksystem.

Obwohl das in diesem Artikel entwickelte Modell das Problem der Optimierung des Stern-Baum-Pipeline-Netzwerkdesigns effektiv löst, weist es immer noch die folgenden Mängel auf: ff Das Modell berücksichtigt nicht die hydraulischen Strömungseigenschaften von Erdgas und kann dahe

ff

Machine Translated by Google

Beurteilen Sie, ob die Flüssigkeitsdruckänderungen im Rohrleitungsnetz den technologischen Anforderungen. Darüber hinaus wurde die hierarchische Optimierungsstrategie dieser Studie stellt nicht sicher, dass die erhaltene Lösung die globale Optimallösung ist. Daher werden zukünftige Die Forschung sollte die Gasbetriebseigenschaften umfassend in die Pipeline einbeziehen Netzwerkdesign-Optimierungsmodell. Darüber hinaus wird eine systematischere globale Um diese Herausforderung wirksam zu bewältigen, ist eine Optimierungsmethode von entscheidender Bedeutung.

Autorenbeiträge: Validierung, YQ; formale Analyse, HC; Untersuchung, JH; Software, PD; Daten Kuratierung, DL; Schreiben – Originalentwurf, YL und JZ; Schreiben – Überprüfung und Bearbeitung, YL und JZ Alle Die Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu.

Finanzierung: Diese Forschung wurde finanziert durch [die National Natural Science Foundation of China] Nummer [51704253].

Datenverfügbarkeitserklärung: Die Daten, die die Ergebnisse dieser Studie stützen, sind erhältlich bei der korrespondierende Autoren auf angemessene Anfrage.

Danksagung: Die Autoren danken für die finanzielle Unterstützung durch die National Natural Science Foundation of China (51704253) für dieses Projekt.

Interessenkonflikte: Die Autoren Yu Lin und Yanhua Qiu waren bei der Firma PetroChina beschäftigt Southwest Oil & Gasfield Company. Autor Hao Chen war bei der Firma Sichuan beschäftigt Shale Gas Exploration and Development Co., Ltd. Autor Dafan Liu war bei der Firma angestellt CNPC Chuanqing Drilling Engineering Co., Ltd. Die übrigen Autoren erklären, dass die Forschung wurde ohne jegliche kommerzielle oder finanzielle Beziehungen durchgeführt, die ausgelegt werden könnten als ein potenzieller Interessenkonflikt.

Nomenklatur

Sätze	
ICH	**Gut gesetzt,** i ÿ I = **{1, 2, . . . ,** NI}
J	**Satz von Kandidatenpositionen für Stationen,** j ÿ J = **1, 2, . . . ,** NJ
K	**Kandidatenpositionen für Pflanzen,** k ÿ K = **{1, 2, . . . ,** NK}
Parameter	
F	**Investitionen in das Pipeline-Netzwerk, CNY**
cij	**Preis der Pipeline zwischen Brunnen und Stationen, CNY/m**
Cjjÿ	**Preis der Pipeline zwischen den Stationen, CNY/m**
Dmax	**Maximal zulässige Rohrleitungslänge vom Brunnen zur Station, m**
fP	**Kosten der Anlage, CNY**
fS	**Stationskosten, CNY**
FC	**Anlageninvestitionen, CNY**
FS	**Bahnhofsinvestition, CNY**
FPP	**Investitionen in die Produktionspipeline, CNY**
FGP	**Pipeline-Investitionen sammeln, CNY**
Lij	**Rohrleitungslänge von Brunnen zu Stationen, m**
Ljjÿ	**Pipelinelänge zwischen Stationen, m**
qj	**Knotengasdurchfluss der Station j, m3/d**
qmax	**Kapazität der Station, m3/d**
qi	**Knotengasdurchfluss von Bohrloch i, m3/d**
qk	**Knotengasdurchfluss der Anlage k, m3/d**
ÿij	**Verbindungsentscheidungsvariable zwischen Brunnen und Stationen, 0 oder 1**
ÿj	**Stationsstandort-Entscheidungsvariable, 0 oder 1**
ÿjjÿ	**Verbindungsentscheidung variabel zwischen Stationen, 0 oder 1**
ÿk	**Werksstandort-Entscheidungsvariable, 0 oder 1**

Machine Translated by Google

Verweise

1. **Zhou, J.; Fu, T.; Chen, Y.; Xiao, Y.; Peng, J.; Liang, G. Ein gemischt-ganzzahliges nichtlineares Programmiermodell für die optimale Gestaltung natürlicher Gasspeicher-Oberflächen-Doppelrohrnetz.** J. Energy Storage **2021,** 44, **103379. [CrossRef]**
2. **Rios-Mercado, R.; Borraz-Sanchez, C. Optimierungsprobleme in Erdgastransportsystemen: Ein Überblick über den neuesten Stand der Technik.** Appl. Energy **2015,** 147, **536–555. [CrossRef]**
3. **Demissie, A.; Zhu, W.; Belachew, CT Ein multikriterielles Optimierungsmodell für den Betrieb von Gaspipelines.** Comput. Chem. Eng. **2017,** 100, **94–103. [CrossRef]**
4. **Hong, B.; Li, X.; Di, G.; Song, S.; Yu, W.; Chen, S.; Li, Y.; Gong, J. Ein integriertes MILP-Modell für die optimale Planung von Mehrperioden Onshore-Gasfeld-Sammelleitungssystem.** Comput. Ind. Eng. **2020,** 146, **106479. [CrossRef]**
5. **Sanaye, S.; Mahmoudimehr, J. Optimale Auslegung eines Erdgastransportnetzes.** Chem. Eng. Res. Des. **2013,** 91, **2465–2476. [CrossRef]**
6. **Rodriguez, DA; Oteiza, PP; Brignole, NB Simulated Annealing-Optimierung für Kohlenwasserstoff-Pipeline-Netzwerke.** Ind. Eng. Chem. Res. **2013,** 52, **8579–8588. [CrossRef]**
7. **Fu, YX Mehrzielparameter-Optimierungsdesign eines einrohrigen ringförmigen Mischwasser-Öl-Sammelrohrnetzes.** Schlüssel Eng. Mater. **2011,** 467, **1285–1290. [CrossRef]**
8. **Wang, B.; Yuan, M.; Zhang, H.; Zhao, W.; Liang, Y. Ein MILP-Modell für die optimale Auslegung eines mehrperiodischen Erdgasübertragungsnetzes .** Chem. Eng. Res. Des. **2018,** 129, **122–131. [CrossRef]**
9. **Zhou, J.; Zhou, X.; Liang, G.; Peng, J. Ein MINLP-Modell für die Netzwerkauslegung unterirdischer Erdgasspeicher.** J. Intell. Fuzzy Syst. **2020,** 38, **4619–4642. [CrossRef]**
10. **Daskin, M. Netzwerk- und diskrete Standortmodelle, Algorithmen und Anwendungen.** J. Oper. Res. Soc. **1997,** 48, **763–764. [CrossRef]**
11. **Posta, M.; Ferland, JA; Michelon, P. Eine exakte kooperative Methode für das Problem der Standortwahl ohne Kapazitäten. Mathe.** Programm. Comput **2014, 6, 199–231. [CrossRef]**
12. **Balinski, ML Integer programming: Methods, uses, calculations.** Manag. Sci. **1965,** 12, **253–313. [CrossRef]**
13. **Klincewicz, JG; Luss, H.; Rosenberg, E. Optimale und heuristische Algorithmen für die Standortbestimmung von Mehrproduktanlagen ohne Kapazitätsgrenzen.** Eur. J Oper. Res. **1986,** 26, **251–258. [CrossRef]**
14. **Efroymson, MA; Ray, TL Ein verzweigter Algorithmus zur Standortbestimmung von Anlagen.** Oper. Res. **1966,** 14, **361–368. [CrossRef]**
15. **Uester, H.; Dilaveroglu, S. Optimierung für Design und Betrieb von Erdgastransportnetzen.** Appl. Energy **2014,** 133, **56–69. [CrossRef]**
16. **Zhang, H.; Liang, Y.; Zhang, W.; Wang, B.; Yan, X.; Liao, Q. Ein einheitliches MILP-Modell für die topologische Struktur von Produktionsbohrungen Sammelleitungsnetz.** J. Pet. Sci. Eng. **2017,** 152, **284–293. [CrossRef]**
17. **Bosikov, II; Klyuev, RV; Silaev, IV; Stas, GV Umfassende Bewertung der formalisierten Belüftung Schwierigkeit und technologische Prozesse in Kohlebergwerken.** Sustain. Dev. Mt. Territ. **2023,** 15, **516–527. [CrossRef]**
18. **Nguyen, H.; Bui, XN; Choi, Y.; Topal, E. Anwendung künstlicher Intelligenz zur Vorhersage der Hangstabilität in Tagebauen: Eine Fallstudie mit einem neuartigen, auf imperialistischen Wettbewerbsalgorithmen basierenden neuronalen Netzwerk mit radialer Basisfunktion. In** Anwendungen künstlicher Intelligenz in Bergbau, Geotechnik und Geoengineering; **Elsevier: Amsterdam, Niederlande, 2024; S. 97–111.**
19. **Rosa, VR; Camponogara, E.; Ferreira Filho, VJM Designoptimierung von Unterwasserinfrastrukturen für Ölfelder mit Verteilerplatzierung und Rohrleitungslayout.** Comput. Chem. Eng. **2018,** 108, **163–178. [CrossRef]**
20. **Liu, Y.; Chen, S.; Guan, B.; Xu, P. Layoutoptimierung eines groß angelegten Öl- und Gassammelsystems basierend auf einer kombinierten Optimierungsstrategie .** Neurocomputing **2019,** 332, **159–183. [CrossRef]**
21. **Hong, B.; Li, X.; Di, G.; Li, Y.; Liu, X.; Chen, S.; Gong, J. Eine integrierte MILP-Methode zur Erfassung von Pipeline-Netzwerken unter Berücksichtigung hydraulische Eigenschaften.** Chem. Eng. Res. Des. **2019,** 152, **320–335. [CrossRef]**
22. **Montagna, AF; Cafaro, DC; Grossmann, IE; Burch, D.; Shao, Y.; Wu, XH; Furman, K. Pipeline-Netzwerkdesign zur Erfassung unkonventioneller Öl- und Gasproduktion mittels mathematischer Optimierung.** Optim. Eng. **2023,** 24, **539–589. [CrossRef]**
23. **Slowik, A.; Kwasnicka, H. Evolutionäre Algorithmen und ihre Anwendung auf technische Probleme.** Neural Comput. Appl. **2020,** 32, **12363–12379. [CrossRef]**
24. **Tohyama, H.; Ida, K.; Matsueda, J. Ein genetischer Algorithmus für das Problem der unbesetzten Standortwahl.** Electr. Commun. Jpn. **2011,** 94, **47–54. [CrossRef]**
25. **Beasley, JE OR-Bibliothek: Verteilen von Testaufgaben per E-Mail.** J. Oper. Res. Soc. **1990,** 41, **1069–1072. [CrossRef]**
26. **Al-Sultan, KS; Hussain, MF; Nizami, JS Ein genetischer Algorithmus für das Mengenüberdeckungsproblem.** J. Oper. Res. Soc. **1996,** 47, **702–709. [Querverweis]**
27. **Pettie, S.; Ramachandran, V. Ein optimaler minimaler Spannbaumalgorithmus.** J. ACM **2002,** 49, **16–34. [CrossRef]**
28. **Zhou, J.; Peng, J.; Liang, G.; Deng, T. Layoutoptimierung eines Tree-Tree-Gaspipeline-Netzwerks.** J. Pet. Sci. Eng. **2019,** 173, **666–680. [Querverweis]**

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultieren.

Machine Translated by Google

Algorithmen

MDPI

Artikel

Schnelle minimale Fehlerentropie für lineare Regression

Andreas Krüger [1], Xiao Liao [1], Wei Cui [1], Ying Wang [1], Hui Cao 2,* und Qingshu Guan [2]

[1] State Grid Information & Telecommunication Group Co., Ltd., Peking 102209, China; qiangli_sgcc@163.com (QL); xiaoliao_sgcc@163.com (XL); weicui_sgcc@163.com (WC); yingwang_sgcc@163.com (YW)
[2] Fakultät für Elektrotechnik, Xi'an Jiaotong-Universität, Xi'an 710049, China; qsguan@stu.xjtu.edu.cn
* Korrespondenz: huicao@mail.xjtu.edu.cn

Zusammenfassung: Das Kriterium der minimalen Fehlerentropie (MEE) findet in verschiedenen Anwendungen großen Nutzen , insbesondere in Kontexten, die durch nicht-Gaußsches Rauschen gekennzeichnet sind. Der Rechenaufwand ist jedoch beträchtlich und ist in erster Linie auf die doppelte Summationsoperation zurückzuführen, die bei der Berechnung der Wahrscheinlichkeitsdichtefunktion (PDF) des Fehlers erforderlich ist. Um dieses Problem anzugehen, führt unsere Studie einen neuartigen Ansatz ein, den sogenannten schnellen Algorithmus der minimalen Fehlerentropie (FMEE), der darauf abzielt, die Rechenkomplexität durch die Verwendung polynomischer Erweiterungen der Fehler-PDF zu verringern. Zunächst wird die PDF-Approximation einer Zufallsvariablen über die Gram-Charlier-Erweiterung abgeleitet. Anschließend ermitteln und optimieren wir die Entropie der Zufallsvariablen. Anschließend wird die dem linearen Regressionsmodell inhärente Fehlerentropie beschrieben und als Funktion des Regressionskoeffizientenvektors ausgedrückt . Schließlich berechnen wir mithilfe des Gradientenabstiegsalgorithmus den Regressionskoeffizientenvektor, der 2) Komplexität im Zusammenhang mit der minimalen Fehlerentropie entspricht. Eine theoretische Untersuchung zeigt, dass die Zeitkomplexität von FMEE bei O(n) liegt, was in starkem Kontrast zu O(n) bei MEE steht. Experimentell unterstreichen unsere Ergebnisse die bemerkenswerten Effizienzgewinne, die FMEE ermöglicht, wobei der Zeitaufwand weniger als 1‰ des bei MEE beobachteten Zeitaufwands beträgt. Erfreulicherweise wird dieser Effizienzsprung ohne Kompromisse bei der Genauigkeit erreicht, wie die vernachlässigbaren Unterschiede zwischen den Genauigkeiten von FMEE und MEE belegen. Darüber hinaus zeigen umfassende Regressionsexperim

Schlüsselwörter: minimale Fehlerentropie; lineare Regression; Gram–Charlier-Erweiterung; Rechenkomplexität

check for updates

Zitat: Li, Q.; Liao, X.; Cui, W.; Wang, Y.; Cao, H.; Guan, Q. Fast Minimum Error Entropy for Linear Regression. Algorithms **2024,** 17, **341. https://doi.org/10.3390/a17080341**

Wissenschaftlicher Redakteur: Frank Werner

Empfangen: 23. Juni 2024
Überarbeitet: 1. August 2024
Akzeptiert: 3. August 2024
Veröffentlicht: 6. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/ 4.0/).

1. Einleitung

Die Entropie dient als zentrale Messgröße zur Quantifizierung der inhärenten Unsicherheit innerhalb eines Systems und findet allgegenwärtige Anwendung in einer Vielzahl von Bereichen, darunter Dimensionsreduktion [1,2], Parameterschätzung [3,4], Identifikation [5,6], Rückkopplungskontrolle [7,8] und Anomalieerkennung [9–11]. Das Kriterium der minimalen Fehlerentropie (MEE) ist ein bemerkenswerter Rahmen, der darauf abzielt, die mit Schätzfehlern verbundene Entropie zu minimieren und so die Unsicherheit innerhalb des Schätzmodells zu verringern. MEE ist besonders relevant für Probleme bei der linearen Regressionsschätzung und weicht von herkömmlichen Methoden der kleinsten Quadrate ab, indem es nicht nur die Varianz der Vorhersagefehler berücksichtigt, sondern auch Kumulanten höherer Ordnung einbezieht, wodurch es für den Umgang mit nicht-Gaußschen Rauschverteilungen geeignet ist [12,13]. Angesichts der Häufigkeit von nicht-Gaußschem Rauschen in realen Szenarien wurde die Wirksamkeit von MEE in verschiedenen Anwendungen ordnungsgemäß validiert. Dazu gehören adaptive Filterung [14–16], Gesichtserkennung [17, 18], Identifizierung spärlicher Systeme [19–

Darüber hinaus wurden die Konvergenzeigenschaften von MEE in der bisherigen Literatur sorgfältig untersucht [25–27]. Detaillierte Analysen befassten sich mit der Konvergenzdynamik von Fixpunkt-MEE-Methoden [28], während Untersuchungen zum Zusamm

Machine Translated by Google

MEE und der minimale mittlere quadratische Fehler (MMSE) [29] sowie seine Beziehung zur maximalen Korrentropie [30] wurden ausführlich untersucht. Bemerkenswerte Erweiterungen des herkömmlichen MEE-Rahmens umfassen den Vorschlag von Kernel-MEE-Varianten [31], Erläuterungen zu Regularisierungsstrategien für MEE-Implementierungen [32] und die Entwicklung von halbüberwachten und verteilten Anpassungen [33]. Darüber hinaus wurden im Bereich der Modellierung von Fehlern in Variablen (EIV) Robustheitsanalysen von MEE-Methoden erläutert [34], neben der Vorstellung neuer Methoden wie dem Ansatz der minimalen Gesamtfehlerentropie [35].

Der Rechenaufwand, der mit dem MEE-Kriterium verbunden ist, ergibt sich aus seiner Komplexität, die quadratisch mit O(n) **skaliert.** ^{2}**Zeitkompliziertheit . Zentral für die Rechenlast von MEE ist Notwendigkeit einer doppelten Summationsoperation während der Berechnung des Gradienten bezüglich der Fehlerwahrscheinlichkeitsdichtefunktion (PDF), insbesondere bei Verwendung von Parzen- Fensterung [36] und bestimmten Arten von Kernelfunktionen. Darüber hinaus kann die Auswahl einer ungeeigneten Kernelfunktion oder ihrer Parameter die Genauigkeit der resultierenden PDF nachteilig beeinflussen. Um dieses Problem zu lösen, wurden Anstrengungen unternommen, den MEE-Ansatz des Gradientenabstiegs zu optimieren, wie die Normalisierung der Schrittweite in [37] zeigt, wobei die Leistung der Eingangsentropie ausgenutzt wird. Darüber hinaus haben innovative Methoden, wie die Verwendung eines Quantisierungsoperators, wie in [38] vorgeschlagen, dazu beigetragen, die Rechenkomplexität zu verringern. Durch die Abbildung des Fehlers auf eine Reihe realwertiger Codewörter transformiert dieser Ansatz die doppelte Summationsoperation effektiv in eine singuläre Form und reduziert so den Rechenaufwand de**

In diesem Dokument wird ein neuartiger Ansatz vorgestellt: der schnelle Algorithmus zur minimalen Fehlerentropieschätzung (FMEE), der speziell auf lineare Regressionsaufgaben zugeschnitten ist und die Technik der Polynomerweiterung nutzt, die auf die Fehlerwahrscheinlichkeitsdichtefunktion (PDF) angewendet wird. Im Gegensatz zur traditionellen Methode der minimalen Fehlerentropie (MEE), die Parzen- Fensterung zur Ableitung der Fehler-PDF verwendet, verwendet FMEE die Gram-Charlier-Erweiterungsmethode zur Approximation der ^{2}) **auf** O(n). **Insbesondere macht FMEE Kernelfunktionen oder die damit verbundenen Fehler-PDF, was zu einer deutlichen Reduzierung der Zeitkomplexität von** O(n **Parametern führt. Der vorgeschlagene Algorithmus umfasst mehrere wichtige Schritte: zunächst die Ableitung der PDF-Approximation für eine Zufallsvariable über die Gram-Charlier-Erweiterung, gefolgt von der Vereinfachung der Entropie der Zufallsvariable, die durch die Orthogonalität der Hermite-Polynome erleichtert wird. Anschließend wird der Fehler des linearen Regressionsmodells ermittelt, wobei die Fehlerentropie als Funktion des Regressionskoeffizientenvektors ausgedrückt wird. Schließlich wird die Gradientenabstiegstechnik verwendet, um den Regressionskoeffizientenvektor abzuleiten , der der minimalen Fehlerentropie entspricht. Die experimentelle Validierung unterstreicht die Wirksamkeit von FMEE und zeigt eine bemerkenswerte Reduzierung des Zeitaufwands mit einer Disparität von weniger als 1 ‰ im Vergleich zum MEE-Ansatz für identische Probleme. lem-Einstellungen. Wichtig ist, dass zwischen de**

Der Rest des Dokuments ist wie folgt gegliedert: Abschnitt 2 stellt die zugehörigen Algorithmen vor , Abschnitt 3 beschreibt FMEE im Detail , Abschnitt 4 enthält die Experimente und Abschnitt **5 enthält die Schlussfolgerungen .**

2. Die zugehörigen Algorithmen

2.1. MEE

MEE minimiert die im Fehler enthaltenen Informationen und maximiert die vom geschätzten Modell erfassten Informationen. Normalerweise wird die Renyi-Entropie zweiter Ordnung [39] verwendet, um die im Fehler des Modells enthaltene Entropie zu messen.

$$H(E) = \ddot{y} \log Ep(pE) = \ddot{y} \log p \quad {}^{2}_{E}(e)de, \qquad (1)$$

Machine Translated by Google

wobei E **die Fehlerzufallsvariable von MEE ist,** H(E) **die Entropie von** E ist **und** pE(e) **die PDF von E liefert. In diesem Artikel** liefert Ep(·) **den Erwartungswert, um Missverständnisse zu vermeiden. Die PDF von** E **kann dann durch Parzen-Fensterung geschät**

$$\hat{p}E(e) = \frac{1}{N} \sum_{i=1}^{N} Kh(e - ei), \quad (2)$$

wobei n **die Anzahl der beobachteten Daten,** K **die Kernelfunktion und** h **die Bandbreite ist. Normalerweise wird der Gauß-Kernel gewählt, und** $Kh(z) = \frac{1}{\sqrt{2\pi h}}$ Dann **ist die** empirische **Information über den Fehler** $\hat{H}(f)$ **kann wie folgt berechnet werden:**

$$\hat{H}(f) = -\log \frac{1}{2n} \sum_{i=1}^{N} \sum_{j=1}^{N} Kh(ei - ej). \quad (3)$$

Da die logarithmische Funktion monoton ansteigt, kann sie ohne Einfluss auf den Minimierungsprozess entfernt werden. Die transformierten empirischen Fehlerinformationen ergeben sich dann aus Gleichung (4):

$$R(e) = -\frac{1}{2n} \sum_{i=1}^{N} \sum_{j=1}^{N} Kh(ei - ej). \quad (4)$$

Bei der linearen Regression, $y = wTx + e$**, besteht das Ziel des Modells darin, w aus den begrenzten Stichproben zu schätzen. Da** $ei = yi - wTxi$ **ist, wird die entsprechende empirische Fehlerinformation mit Gleichung (5) berechnet:**

$$R(w) = \frac{1}{2n} \sum_{i=1}^{N} \sum_{j=1}^{N} Kh((yi - wTxi) - (yj - wTxj)). \quad (5)$$

Dann kann der MEE-Schätzer $\hat{w}$ durch Minimierung **von R(w) mit Gradientenabstieg erhalten werden . Die zeitliche Komplexität von MEE beträgt** $O(n^2)$ **aufgrund der doppelten Summationsoperation, wie in gezeigt Gleichung (5). Da der Zeitbedarf von MEE quadratisch zur Anzahl der beobachteten Daten ist, würde die Laufzeit von MEE mit der Zunahme der Menge der beobachteten Daten dramatisch ansteigen.**

2.2. Die Hermite-Polynome

Die Hermite-Polynome [40] sind eine klassische orthogonale Polynomfolge. Sie werden durch die Ableitungen der standardisierten Gaußschen PDF $\varphi(y)$ **als Gleichung (6) definiert:**

$$\frac{d^i \varphi(y)}{dy^i} = (-1)^i Hi(y)\varphi(y), \quad (6)$$

wobei H(i) das **Hermite-Polynom i-ter Ordnung liefert .**

Die Hermite-Polynome bilden ein orthogonales System wie in Gleichung (7):

$$\int \varphi(y) Hi(y) Hj(y) dy = \begin{cases} i! & \text{wenn } i = j \\ 0, & \text{wenn } i \neq j \end{cases}, \quad (7)$$

In diesem Dokument werden die ersten fünf Hermite-Polynome verwendet, die in Gleichung (8) dargestellt sind:

$$\begin{aligned} H0(x) &= 1, \\ H1(x) &= x, \\ H2(x) &= x^2 - 1, \\ H3(x) &= x^3 - 3x, \\ H4(x) &= x^4 - 6x^2 + 3. \end{aligned} \quad (8)$$

Machine Translated by Google

2.3. Die Gram–Charlier-Erweiterung

Die Gram–Charlier-Erweiterung liefert die Näherung der PDF einer Zufallsvariable, und es gibt keinen signifikanten Genauigkeitsverlust bei dieser Erweiterung, wie in [41] gezeigt wird. Unter der Annahme, dass die Zufallsvariable nahe der Normalverteilung liegt und denselben Mittelwert und dieselbe Varianz wie sie selbst hat, **kann die Gram–Charlier-Erweiterung der PDF von** x wie folgt ausgedrückt werden:

$$px(ÿ)\ ÿp\hat{}x(ÿ) = \exp ÿ\ ÿ\ \frac{1}{\sqrt{2ÿÿ\ 2ÿ}}\ \frac{(ÿ\ ÿ\ \mu)^2}{2}\ ÿ3(x)\ \frac{ÿ\ ÿ\ \mu}{ÿ}\ 1 + H3\ \frac{ÿ\ ÿ\ \mu}{ÿ}\ 3! + \frac{ÿ4(x)}{4!ÿ^4}\ H4\ \frac{ÿ\ ÿ\ \mu}{ÿ}\ , \quad (9)$$

wobei μ und $ÿ$ der Mittelwert und die Standardabweichung von x sind. $ÿ$ ist die gaußsche Zufallsvariable , deren Mittelwert und Variable dieselben sind wie die von x. ÿ3(x) **und** ÿ4(x) **sind die Kumulanten dritter und vierter Ordnung von x, die auch als Schiefe und Kurtosis bezeichnet werden können. H3 und H4 sind die Hermite-Polynome dritter und vierter Ordnung.**

3. Methodik In

diesem Abschnitt wird die Beziehung zwischen der differentiellen Entropie und der Gram- Charlier-Erweiterung der PDF des Fehlers für die lineare Regression hergeleitet. Darüber hinaus wird die Annahme umgesetzt, dass die PDF des Regressionsfehlers nahe der Gauß-Verteilung $ÿ(ÿ)$ liegt , **die denselben Mittelwert und dieselbe Varianz wie der Fehler hat.**

Die differentielle Entropie [42] einer Zufallsvariablen X wird in Gleichung (10) dargestellt:

$$H(x) = ÿ\ p(x)\log p(x)dx. \quad (10)$$

Setze Gleichung (9) in Gleichung (10) ein; das Ergebnis lässt sich aus Gleichung **(11) ableiten :**

$$H(x)\ ÿ\ ÿ\ p\hat{}x(ÿ)\log p\hat{}x(ÿ)dÿ$$

$$= ÿ \frac{1}{\sqrt{2ÿÿ\ 2ÿ}}\ \exp ÿ\ ÿ\ \frac{(ÿ\ ÿ\ \mu)^2}{2}\ ÿ3(x)\ 1 + H3\ 3!\ \frac{ÿ\ 3}{3!}\ \frac{ÿ\ ÿ\ \mu}{ÿ}\ + \frac{ÿ4(x)}{4!ÿ^4}\ H4\ \frac{ÿ\ ÿ\ \mu}{ÿ}\ ÿ$$

$$\text{Protokoll}\ \frac{1}{\sqrt{2ÿÿ\ 2ÿ}}\ \exp ÿ\ ÿ\ \frac{(ÿ\ ÿ\ \mu)^2}{2}\ ÿ3(x)\ 1 + H3\ 3!ÿ\ \frac{3}{3!}\ \frac{ÿ\ ÿ\ \mu}{ÿ}\ + \frac{ÿ4(x)}{4!ÿ^4}\ H4\ \frac{ÿ\ ÿ\ \mu}{ÿ}\ dÿ. \quad (11)$$

Dieses Integral ist ziemlich schwierig zu berechnen. Beachten Sie jedoch, dass, wenn die PDF von x **wie angenommen nahe der Normaldichte liegt ,** man davon ausgehen kann, dass ÿ3(x) **und** ÿ4(x) **sehr klein sind. Dann** kann **die Näherung in Gleichung (12) verwendet werden:**

$$\log(1 + ÿ)\ ÿ\ ÿ\ ÿ \quad \frac{2ÿ}{2}, \quad (12)$$

wobei $ÿ$ das Infinitesimal bezeichnet.

Dann kann Gleichung (11) in Gleichung **(13) umgeformt werden :**

Machine Translated by Google

$$H(x) = \ddot{y}\ \ddot{y}(\ddot{y})\ddot{y}\ \ddot{y}3(x)\ \mathbf{1} + \ddot{y}\ \mathbf{3}\ \frac{\mathbf{H3\ 3!}}{}\ \frac{\ddot{y}\ \ddot{y}\ \mu}{\ddot{y}} + \frac{\ddot{y}4(x)}{\mathbf{4!\ddot{y}\ 4}}\mathbf{H4}\ \frac{\ddot{y}\ \ddot{y}\ \mu}{\ddot{y}}\ \ddot{y} \log \ddot{y}(\ddot{y}) + \mathbf{H3}\ \frac{\ddot{y}3(x)}{\mathbf{3!\ddot{y}\ 3}}\ \frac{\ddot{y}\ \ddot{y}\ \mu}{\ddot{y}} + \frac{\ddot{y}4(x)}{\mathbf{4!\ddot{y}\ 4}}\mathbf{H4}\ \frac{\ddot{y}\ \ddot{y}\ \mu}{\ddot{y}}\ \frac{\mathbf{1}}{\mathbf{2}}\ \frac{\ddot{y}3(x)}{\mathbf{3!\ddot{y}\ 3}}\mathbf{H3}\ \frac{\ddot{y}\ \ddot{y}\ \mu}{\ddot{y}} + \frac{\ddot{y}4(x)}{\mathbf{4!\ddot{y}\ 4}}\mathbf{H4}\ \frac{\ddot{y}\ \ddot{y}\ \mu}{\ddot{y}}\ ^{\mathbf{2}}\ d\ddot{y}. \quad (13)$$

Da Hi **ein orthogonales System bildet, kann** Gleichung **(13)** als Gleichung **(14) vereinfacht werden,**

$$H(x) = \quad \mathbf{1}\ \log 2\ddot{y}e\ddot{y}\ ^{\mathbf{2}}\ \frac{^{\mathbf{2}}\ddot{y}3(x)}{\mathbf{2}\ \ddot{y}\ \mathbf{3!\ddot{y}\ 6}}\ \frac{^{\mathbf{2}}\ddot{y}4(x)}{\mathbf{2}\ \ddot{y}\ \mathbf{4!\ddot{y}\ 8}}. \quad (14)$$

Die Einzelheiten des Vereinfachungsprozesses der Gleichung (14) werden wie folgt veranschaulicht.

In diesem Teil geht es um die Herleitung von Gleichung (14). Zunächst muss gezeigt werden, dass

$$\mathbf{2H3}(\ \ddot{y}(\ddot{y})(\ )d\ddot{y} = \mathbf{0}\ \frac{\ddot{y}\ \ddot{y}\ \mu\ \ddot{y}\ \ddot{y}\ \mu}{\ddot{y}} \quad (15)$$

Und

$$\mathbf{2H4}(\ \ddot{y}(\ddot{y})(\ )d\ddot{y} = \mathbf{0}\ \frac{\ddot{y}\ \ddot{y}\ \mu\ \ddot{y}\ \ddot{y}\ \mu}{\ddot{y}} \quad (16)$$

Für Gleichung (15) nehmen wir an, dass t = *(ÿ* ÿ **µ)/ÿ,**

$$\mathbf{2H3}(\ \ddot{y}(\ddot{y})(\ )d\ddot{y}\ )\ \frac{\ddot{y}\ \ddot{y}\ \mu\ \ddot{y}\ \ddot{y}\ \mu}{\ddot{y}} = \frac{\mathbf{1}}{} \exp\ \ddot{y}\ 2\ddot{y}\ \frac{(\ddot{y}\ \ddot{y}\ \mu)}{\mathbf{2}\ \ddot{y}\ 2\ddot{y}\ddot{y}}\ ^{\mathbf{2}}\ (\ )\frac{\ddot{y}\ \ddot{y}\ \mu}{\ddot{y}}\ ^{\mathbf{3}}\ \ddot{y}\ 3(\ )\ d\ddot{y}\ \frac{\ddot{y}\ \ddot{y}\ \mu}{\ddot{y}} = \frac{\mathbf{1}}{\ddot{y}\ 2\ddot{y}}\ ^{\mathbf{2}\ \mathbf{t}}\ \mathbf{2\ t\ exp(\ddot{y}\ )(t}\ ^{\mathbf{2}}\ ^{\mathbf{3}}\ \ddot{y}\ 3t)dt \quad (17)$$

Da exp(ÿt $\frac{^{2}}{2}$) **ist eine gerade Funktion und** t $^{5\ 3\ \ddot{y}\ 3t}$ **ist eine ungerade Funktion, dann**

$$\frac{\mathbf{1}}{\mathbf{2\ t}\ \ddot{y}\ 2\ddot{y}}\ \mathbf{exp(-}\ \frac{^{\mathbf{2\ t}}}{\mathbf{2}})(t\ ^{\mathbf{3}}\ \ddot{y}\ 3t)dt = \mathbf{0} \quad (18)$$

und Gleichung (15) gilt.

Für Gleichung (16) nehmen wir an, dass t = *(ÿ* ÿ **µ)/ÿ,**

$$\begin{aligned}\mathbf{2H4}(\ \ddot{y}(\ddot{y})(\ )d\ddot{y}\ )\ \frac{\ddot{y}\ \ddot{y}\ \mu\ \ddot{y}\ \ddot{y}\ \mu}{\ddot{y}} &= \frac{\mathbf{1}}{2\ddot{y}\ddot{y}\ 2\ddot{y}}\ \mathbf{exp}\ \ddot{y}\ \ddot{y}\ \frac{(\ddot{y}\ \ddot{y}\ \mu)}{\mathbf{2}\ \ddot{y}\ \ddot{y}\ \mu\ (\ )}\ ^{\mathbf{2}}\ \ddot{y}\\ &\frac{}{\ddot{y}}\ ^{\mathbf{6}}\ \ddot{y}\ 6(\ )\frac{\ddot{y}\ \ddot{y}\ \mu}{\ddot{y}}\ ^{\mathbf{4}} + 3(\ )\frac{\ddot{y}\ \ddot{y}\ \mu}{\ddot{y}}\ ^{\mathbf{2}}\ d\ddot{y}\\ &= \frac{\mathbf{1}}{2\ddot{y}}\ \mathbf{exp(\ddot{y}\ \ddot{y}}\ \frac{^{\mathbf{2\ t}}}{\mathbf{2}})(t\ ^{\mathbf{6\ 4\ 2}\ \ddot{y}\ 6t + 3t}\ \mathbf{)dt}\\ &= 3\ \frac{\mathbf{1}}{2\ddot{y}}\ \mathbf{exp(\ddot{y}\ \ddot{y}}\ \frac{^{\mathbf{2\ t}}}{\mathbf{2}})t\ \mathbf{2dt}\ \ddot{y}\ \mathbf{6}\ \frac{\mathbf{1}}{2\ddot{y}}\ \mathbf{exp(\ddot{y}\ \ddot{y}}\ \frac{^{\mathbf{2\ t}}}{\mathbf{2}})t\ \mathbf{4dt}\\ &+ \frac{\mathbf{1}}{2\ddot{y}}\ \mathbf{exp(\ddot{y}\ \ddot{y}}\ \frac{^{\mathbf{2\ t}}}{\mathbf{2}})t\ \mathbf{6dt}\end{aligned} \quad (19)$$

Machine Translated by Google

Der erste Term in Gleichung (19) kann wie folgt berechnet werden :

$$\begin{aligned}
&\frac{1}{\ddot{y}\sqrt{2\ddot{y}}}\exp(\ddot{y}\ \frac{2t}{2})t\ 2dt \\
&= \frac{1}{\sqrt{2t\,(\ddot{y}t)d\ \exp(\ddot{y})\ \ddot{y}\ 2\ddot{y}\ 2}} — \\
&= \frac{1}{\sqrt{2\ddot{y}}}(\ddot{y}t)\ \exp(\ddot{y}\ \ddot{y}\ \frac{2t}{)\ 2}\ \Big|_{\ddot{y}\ddot{y}}^{+\ddot{y}} + \\
&\quad \frac{1}{\sqrt{2t\ \exp(\ddot{y}\ )dt\ \ddot{y}\ 2\ddot{y}\ 2}} \\
&= 1
\end{aligned} \tag{20}$$

Der zweite Term in Gleichung (19) kann wie folgt berechnet werden :

$$\begin{aligned}
&\frac{1}{\ddot{y}\sqrt{2\ddot{y}}}\exp(\ddot{y}\ \frac{2t}{2})t\ 4dt \\
&= \frac{1}{\ddot{y}\sqrt{2\ddot{y}}}(\ddot{y}t^{3})d\ \exp(\ddot{y}\ \frac{2t}{2}) \\
&= \frac{1}{\ddot{y}\sqrt{2\ddot{y}}}(\ddot{y}t^{3}\ _{2t\ )\ \exp(\ddot{y}\ )\ 2} — \Big|_{\ddot{y}\ddot{y}}^{+\ddot{y}} + \\
&\quad 3\ \frac{1}{\ddot{y}\sqrt{2\ddot{y}}}\exp(\ddot{y}\ \frac{2t}{2})t\ 2dt \\
&= 3
\end{aligned} \tag{21}$$

Der letzte Term in Gleichung (19) kann wie folgt berechnet werden :

$$\begin{aligned}
&\frac{1}{\sqrt{2\ddot{y}\ 2}}\exp(\ddot{y}\ \ddot{y}\ \frac{2t}{—})t\ 6dt \\
&= \frac{1}{\ddot{y}\sqrt{2\ddot{y}}}(\ddot{y}t^{5})d\ \exp(\ddot{y}\ \frac{2t}{2}) \\
&= \frac{1}{\ddot{y}\sqrt{2\ddot{y}}}(\ddot{y}t^{5})\ \exp(\ddot{y}\ \frac{2t}{2})\ \Big|_{\ddot{y}\ddot{y}}^{+\ddot{y}} + \\
&\quad 5\ \frac{1}{\ddot{y}\sqrt{2\ddot{y}}}\exp(\ddot{y}\ \frac{2t}{2})t\ 4dt \\
&= 15
\end{aligned} \tag{22}$$

Nach den Gleichungen (20)–(22)

$$2H4(\ \ddot{y}(\ddot{y})(\ )d\ddot{y} = 15\ \ddot{y}\ 6\ \ddot{y}\ 3 + 3 = 0 \tag{23}$$

und Gleichung (16) gilt.

Machine Translated by Google

Dann gilt für Gleichung (13):

$$H(x) = \ddot{y}\ \ddot{y}(\ddot{y})\log\ \ddot{y}(\ddot{y})d\ddot{y}\ddot{y}$$

$$\ddot{y}(\ddot{y}) \quad \frac{\ddot{y}3(x)\ \ddot{y}\ \ddot{y}\ \mu\ \ddot{y}4(x)\ \ddot{y}\ \ddot{y}\ \mu\ \mathbf{H3(\ )} + \mathbf{H4(\ )}}{\mathbf{4!\ddot{y}\ 4}}\ d\ddot{y} + \frac{\mathbf{3!\ddot{y}\ 3}}{\ddot{y}}$$

$$- \quad \frac{\ddot{y}3(x)\ \ddot{y}4(x)\ \mathbf{1}\ \ddot{y}\ \ddot{y}\ \mu\ \ddot{y}\ \ddot{y}\ \mu\ \ddot{y}(\ddot{y})}{\mathbf{H3(\ )} + \mathbf{H4(\ )}\ \mathbf{3!\ddot{y}\ 3\ 4!\ddot{y}\ 4\ 2}} \quad \frac{}{\ddot{y}}^{\mathbf{2}}\ d\ddot{y}\ddot{y}$$

$$\ddot{y}(\ddot{y}) \quad \frac{\ddot{y}3(x)\ \ddot{y}\ \ddot{y}\ \mu\ \ddot{y}(\ddot{y})\log}{\mathbf{H3(\ )} + \mathbf{3!\ddot{y}\ 3}}$$

$$\frac{\ddot{y}4(x)\ \ddot{y}\ \ddot{y}\ \mu\ \mathbf{H4(\ )}\ d\ddot{y}}{\mathbf{4!\ddot{y}\ 4}}\ \ddot{y} \qquad \textbf{(24)}$$

$$\frac{\ddot{y}3(x)\ \ddot{y}\ \ddot{y}\ \mu\ \ddot{y}\ \ddot{y}(\ddot{y})}{\mathbf{H3(\ )} + \mathbf{3!\ddot{y}\ 3}}$$

$$\frac{\ddot{y}4(x)\ \ddot{y}\ \ddot{y}\ \mu\ \mathbf{H4(\ )}}{\mathbf{4!\ddot{y}\ 4}}\ \ddot{y}^{\mathbf{2}}\ d\ddot{y} +$$

$$\frac{\mathbf{1}}{\mathbf{2}} \quad \frac{\ddot{y}3(x)\ \ddot{y}4(x)\ \ddot{y}\ \ddot{y}\ \mu\ \ddot{y}\ \ddot{y}\ \mu\ \ddot{y}(\ddot{y})}{\mathbf{H3(\ )} + \mathbf{H4(\ )}\ \mathbf{3!\ddot{y}\ 3\ 4!\ddot{y}\ 4}} \quad \frac{}{\ddot{y}}^{\mathbf{3}}\ d\ddot{y}$$

Beachten Sie, dass H0 = 1 ist, dann ist der zweite Term der Gleichung (24) 0. Da angenommen wird, dass x nahe an einer normalen Zufallsvariable liegt, sind ÿ3(x) **und** ÿ4(x) **sehr klein. Dann kann der letzte Term der Gleichung (24) als 0 weggelassen werden, da es sich um Monome dritter Ordnung handelt, die unendlich kleiner sind als die Terme, die nur Monome zweiter Ordnung enthalten. Berücksichtigen Sie dann für Gleichung (24) die Gleichungen (15) und (16) und beachten Sie erne**

$$H(x) = H(\ddot{y})\ \ddot{y} \quad \frac{\mathbf{1}}{\mathbf{2}}\ \ddot{y}\ \mathbf{3} \quad \frac{\ddot{y}3(x)\ \ddot{y}\ \ddot{y}\ \mu\ )]2+\ \ddot{y}(\ddot{y})}{\ddot{y}}[\ \mathrm{H3(\ 3!}$$

$$\ddot{y}\ \mathbf{4}\ddot{y}3(x) \quad \frac{\ddot{y}4(x)\ \ddot{y}\ \ddot{y}\ \mu\ )]2+}{\ddot{y}}[\ \mathrm{H4(\ 4!}$$

$$2\ \frac{\ddot{y}\ \ddot{y}\ \mu\ \ddot{y}\ \ddot{y}\ \mu\ \mathbf{H3(\ )}}{\ddot{y}\ \mathbf{4}} \quad \frac{}{\ddot{y}}\ \mathbf{H4(\ )}\ d\ddot{y}\ \mathbf{3!\ddot{y}\ 3\ 4!}$$

$$\ddot{y}\ \ddot{y}(\ddot{y})\ \mathbf{log} + \mathbf{log}\ \mathbf{exp}(\ddot{y}\ \ddot{y}\ 2\ddot{y}\ddot{y}\ \mathbf{2}\ \ddot{y}3(x)\ \ddot{y}\ \ddot{y}\ \mu \quad \frac{\ddot{y}^{\mathbf{2}}}{}\ )\ \ddot{y}$$

$$\frac{\mathbf{1}}{\ddot{y}4(x)\ \ddot{y}\ \ddot{y}\ \mu\ \mathbf{H3(\ )} + \mathbf{H4(\ )}\ d\ddot{y}}$$

$$\frac{\mathbf{3!\ddot{y}\ 3\ 4!\ddot{y}\ 4}}{\ddot{y}} \quad \frac{}{} \quad \frac{}{\ddot{y}} \qquad \textbf{(25)}$$

$$=\mathrm{H}(\ddot{y})\ \ddot{y} \quad \frac{{}^{\mathbf{2}}\ddot{y}\ \mathbf{3}(X)}{\mathbf{18\ddot{y}\ 6}} \quad \frac{{}^{\mathbf{2}}\ddot{y}\ \mathbf{3}(X)}{\mathbf{32\ddot{y}}}$$

$$\mathbf{log}\ \frac{\mathbf{1}}{\ddot{y}\ 2\ddot{y}\ddot{y}} \quad \ddot{y}(\ddot{y}) \quad \frac{\mathbf{8}\ \ddot{y}3(x)\ \ddot{y}\ \ddot{y}\ \mu\ \mathbf{H3(\ )}}{\ddot{y}} + \mathbf{3!\ddot{y}\ 3}$$

$$\frac{\ddot{y}4(x)\ \ddot{y}\ \ddot{y}\ \mu\ \mathbf{H4(\ )}\ d\ddot{y}}{\ddot{y}} + \mathbf{4!\ddot{y}\ 3} \quad \frac{\mathbf{1}}{\mathbf{2\ Zeile\ 2}} \quad \ddot{y}(\ddot{y})\ddot{y}^{\ \mathbf{2}\ \ddot{y}}$$

$$\frac{\ddot{y}3(x)\ \ddot{y}4(x)\ \ddot{y}}{\ddot{y}} \quad \frac{\mathbf{H4(\ )}\ \ddot{y}\ddot{y}\ \mu\ \mathbf{H3(\ )}\ \mathbf{4!\ddot{y}\ 3}}{\ddot{y}} \quad \frac{}{\ddot{y}}$$

$$= \quad \mathbf{1}\ \mathbf{log}\ 2\ddot{y}\mathbf{e}\ddot{y}\ \mathbf{2} \quad {}^{\mathbf{2}}\ddot{y}\ \mathbf{3}(X) \quad \frac{\mathbf{18\ddot{y}\ 6}}{} \quad {}^{\mathbf{2}}\ddot{y}\ \mathbf{4}(X) \quad \frac{\mathbf{32\ddot{y}\ 8}}{}$$

Daher gilt Gleichung (14) .

Für die lineare Regression, ei = yi **ÿ** w **Txi , ist die Standardabweichung des Fehlers** ew **in Gleichung (26):**

$$^{2\,\ddot{y}}\ (\mathrm{ew}) = \mathrm{Ep}(\mathrm{E} \quad \mathbf{2}\ \mathrm{w})\ \ddot{y}\ [\mathrm{Ep}(\mathrm{Ew})]2 = \mathrm{Ep}(\mathrm{E} \quad \mathbf{2}\ \mathrm{w})\ \ddot{y}\ (\bar{y}\ \ddot{y}\ \mathbf{wT}\ \bar{x}) \quad ^{\mathbf{2}}, \qquad \textbf{(26)}$$

Machine Translated by Google

wobei y¯der Mittelwert von y **und** ¯xden Mittelwert von x darstellt.

Um die Berechnung von ÿ4(ew) zu vereinfachen, **lassen Sie** ew **eine Zufallsvariable mit dem Mittelwert 0 sein:**

$$ewi = yi \ddot{y} \mathbf{wT} xi \ddot{y} \bar{y} + \mathbf{wT} \bar{x}. \quad (27)$$

Beachten Sie, dass die Transformation in Gleichung (27) die Standardabweichung nicht ändert und die Entropie von ew **als** $\ddot{y}^2(e + c) = \ddot{y}^2(e)$ **und** H(e + c) = H(e), **wenn** c **eine konstante Zahl ist.**
Um die Übersichtlichkeit zu wahren, wird im Folgenden $\ddot{y}^2$ **wird durch** var(·) ersetzt.

Dann kann für die lineare Regression die Fehlerentropie durch Gleichung (28) ausgedrückt werden:

$$H(ew) = 1 \frac{1}{2} \log 2\ddot{y}e + \log \text{var}(ew) \frac{2\ddot{y}_3 2\ddot{y}}{3!\text{var3}(ew)} \frac{2\ddot{y}_4(ew)}{2\ddot{y}4!\text{var4}(ew)}, \quad (28)$$

Wo

$$\ddot{y}3(ew) = Ep[(y \ddot{y} \mathbf{wT} x \ddot{y} \bar{y} + \mathbf{wT} \bar{x}) \quad ^3], \quad (29)$$

$$\ddot{y}4(ew) = Ep[(y \ddot{y} \mathbf{wT} x \ddot{y} \bar{y} + \mathbf{wT} \bar{x}) \quad ^4] \ddot{y} 3. \quad (30)$$

Um H(ew) **zu minimieren ,** wird **die Ableitung von** H(ew) **nach** w **berechnet in Gleichung (31):**

$$\frac{\ddot{y}H(ew)}{\ddot{y}w} = \frac{1}{2\text{var}(ew)\ln 2} \frac{\ddot{y}\text{var}(ew)}{\ddot{y}w} \frac{1}{2\ddot{y}3!\text{var6}(ew)} \ddot{y}$$
$$2\text{var3}(ew)\ddot{y}3(ew) \frac{\ddot{y}\ddot{y}3(ew)}{\ddot{y}w}$$
$$^2{}_{3\ddot{y}3}(ew)\text{var2}(ew) \frac{\ddot{y}\text{var}(ew)}{\ddot{y}w} \quad (31)$$
$$\frac{1}{2\ddot{y}4!\text{var8}(ew)\ddot{y}w} 2\text{var4}(ew)\ddot{y}4(ew) \frac{\ddot{y}\ddot{y}4(ew)}{}$$
$$^2{}_{4\ddot{y}4}(ew)\text{var3}(ew) \frac{\ddot{y}\text{var}(ew)}{\ddot{y}w},$$

Wo

$$\frac{\ddot{y}\text{var}(ew)}{} = \ddot{y}2Ep[(y \ddot{y} \mathbf{wT} x)x] + 2(\bar{y}\ddot{y} \mathbf{wT} \bar{x})\bar{x}\frac{\ddot{y}}{\ddot{y}w} \quad (32)$$

$$\frac{\ddot{y}\ddot{y}3(ew)}{} = 3Ep[(y \ddot{y} \mathbf{wT} x \ddot{y} \bar{y} + \mathbf{wT} \bar{x})\ddot{y}w \quad ^2(\ddot{y}x + \bar{x})], \quad (33)$$

$$\frac{\ddot{y}\ddot{y}4(ew)}{} = 4Ep[(y \ddot{y} \mathbf{wT} x \ddot{y} \bar{y} + \mathbf{wT} \bar{x})\ddot{y}w \quad ^3(\ddot{y}x + \bar{x})]. \quad (34)$$

Dann kann das optimale w für das Minimum von H(ew) **durch das Iterationsschema erhalten werden in Gleichung (35) mit Gradientenabstieg:**

$$\mathbf{w}\hat{}\, k+1 = \mathbf{w}\hat{}\, k \ddot{y} \ddot{y} \frac{\ddot{y}H(ewk)}{\ddot{y}w\hat{}\, k}, \quad (35)$$

Dabei ist $\ddot{y}$ **die Schrittweite, die durch die Armijo-Bedingung [43] bestimmt wird.**

Berechnung von $\frac{\ddot{y}H(ewk)}{\ddot{y}wk}$, $\frac{\ddot{y}\text{var}(ewk)}{\ddot{y}wk}$, $\frac{\ddot{y}\ddot{y}3(ewk)}{\ddot{y}wk}$ **und** $\frac{\ddot{y}\ddot{y}4(ewk)}{\ddot{y}wk}$ **werden**

benötigt. Aus den Gleichungen (28) und (29)–(34) ergibt sich die Zeitkomplexität für $\frac{\ddot{y}H(ewk)}{\ddot{y}wk}$ **, für** $\frac{\ddot{y}\text{var}(ewk)}{\ddot{y}wk}$ **,** $\frac{\ddot{y}\ddot{y}3(ewk)}{\ddot{y}wk}$ **und** $\frac{\ddot{y}\ddot{y}4(ewk)}{\ddot{y}wk}$ **,**

von var(ewk), ÿ3(ewk), ÿ4(ewk), **ist** O(n) . **Dann ist die Rechenkomplexität von FMEE ist** O(n). **Da die Zeitkomplexität von MEE** O(n) **für die Doppelsummationsoperation ist, läuft FMEE schneller als MEE.**

Machine Translated by Google

4. Experimente

In diesem Abschnitt wird eine umfassende Reihe von Experimenten durchgeführt, darunter sowohl numerische Simulationen als auch reale Szenarien. Die numerischen Simulationen dienen dazu, Validierung der Wirksamkeit des FMEE-Ansatzes und Überprüfung der Zeiteffizienz beider FMEE- und MEE-Methoden. Parallel dazu werden praktische Experimente durchgeführt, die darauf abzielen bei der Vorhersage von Stromausfällen in einer Stadt im Nordwesten Chinas.

In Bezug auf die numerischen Simulationen betrachten wir ein Szenario, in dem x ÿ R10 ist, mit das Modell definiert als $y = w^T x + e$, wobei wÿ = [1, ÿ1, 1, ÿ1, 1 , ÿ1, 1, ÿ1, 1, ÿ1] T und x folgt einer Normalverteilung N (0, I10). Zwei verschiedene Arten von Rauschen werden in unsere Experimente. Erstens wird Gaußsches Rauschen e ÿ N (0, 1) verwendet, während der zweite Typ besteht aus verallgemeinertem Gaußschen Rauschen, charakterisiert durch eine Wahrscheinlichkeitsdichtefunktion f(e) ÿ *exp(ÿ|e|)* 0,3. Diese hier verwendete Verteilung mit schweren Enden stimmt mit der vorherigen Werke, wie Referenzen [27,35]. Obwohl FMEE unter der Annahme abgeleitet wird, Bei nahezu gaußschem Rauschen ist zu erwarten, dass FMEE auch bei nicht- gaußschem Rauschen mit schweren Rändern wirksam bleibt. Dabei bezeichnet *ÿ* eine Konstante, die angepasst wird, um eine Varianz beizubeh von 1 für e. Für MEE wird die Schrittweite auf ÿ 0,005ÿ gesetzt und der Skalierungsparameter für die Gaußsche Kernel ist 10, was den Einstellungen in Referenz [27] entspricht . Über die Experimente hinweg Die Stichprobengrößen reichen von 100 bis 500, wobei jede Kombination aus Stichprobengröße und Rauschtyp 100 Mal wiederholt. Die daraus resultierenden Versuchsergebnisse sind unten aufgeführt.

Tabelle 1 bietet eine vergleichende Analyse des Zeitaufwands zwischen MEE und FMEE in Gegenwart von Gaußschem Rauschen. FMEE zeigt eine bemerkenswerte Beschleunigung in Rechenleistung im Vergleich zu MEE. Während die Rechenzeit von MEE weist ein quadratisches Wachstum relativ zur Stichprobengröße auf, der Rechenaufwand von FMEE skaliert linear mit der Anzahl der Samples. Insbesondere erreicht FMEE eine bemerkenswerte Geschwindigkeit, Verarbeitung von 500 Objekten in ca. 0,1 s, was es zu einem vielversprechenden Kandidat für zeitkritische Anwendungen.

Tabelle 1. Die Laufzeit von MEE und FMEE für Gaußsches Rauschen.

Stichprobengröße	Durchschnittliche Laufzeit		Optimale Laufzeit	
	ICH	FMEE	ICH	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

Tabelle 2 vergleicht den mittleren quadratischen Fehler zwischen MEE und FMEE im Kontext von Gaußsches Rauschen. Insgesamt zeigt MEE eine bessere Leistung als FMEE in Bezug auf von MSE. Besonders hervorzuheben ist der erhebliche Vorteil von MEE gegenüber FMEE, insbesondere offensichtlich, wenn die Stichprobengröße 100 beträgt. Wenn jedoch die Stichprobengröße zunimmt, wird die Diskrepanz im MSE zwischen den beiden Algorithmen verringert sich. Wenn die Stichprobengröße 500 erreicht, Der Unterschied schrumpft auf nur noch 2%. Diese Beobachtung legt nahe, dass FMEE zwar leicht hinter MEE, insbesondere bei größeren Datensätzen, bleibt eine solche Diskrepanz akzeptabel in praktischen Anwendungen, da FMEE die Rechenzeit im Vergleich zu zu MEE.

Machine Translated by Google

Tabelle 2. Der mittlere quadrierte Fehler von MEE und FMEE für Gaußsches Rauschen.

Stichprobengröße	Durchschnittlicher MSE		Optimaler MSE	
	ICH	FMEE	ICH	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

Tabelle 3 analysiert die für die Konvergenz zwischen MEE und FMEE erforderlichen Iterationszahlen im Kontext von Gaußschem Rauschen. Insbesondere erreicht FMEE Konvergenz mit deutlich weniger Iterationen im Vergleich zu MEE. Darüber hinaus wird beobachtet, dass die Iterationszahl von Der MEE nimmt mit zunehmender Stichprobengröße allmählich ab. Umgekehrt nimmt die Iterationszahl von Der FMEE bleibt unabhängig von Schwankungen der Stichprobengröße relativ konstant.

Tabelle 3. Die Anzahl der Iterationen von MEE und FMEE für Gaußsches Rauschen.

Stichprobengröße	Durchschnittliche Iterationszahl		Optimale Iterationszahl	
	ICH	FMEE	ICH	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

Tabelle 4 bietet eine vergleichende Bewertung des Zeitaufwands zwischen MEE und FMEE im Kontext von nicht-Gaußschem Rauschen. In allen fünf Versuchsgruppen FMEE zeigt durchweg eine bemerkenswerte Reduzierung des Zeitaufwands, durchgängig weniger als 1‰ des Zeitaufwands von MEE. Darüber hinaus theoretischen Vorhersagen zufolge weisen die Rechenkosten von MEE ein quadratisches Wachstum relativ zur Stichprobengröße, während der Zeitaufwand von FMEE linear mit der Stichprobengröße skaliert.

Tabelle 4. Die Laufzeit von MEE und FMEE für nicht-Gaußsches Rauschen.

Stichprobengröße	Durchschnittliche Laufzeit		Optimale Laufzeit	
	ICH	FMEE	ICH	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

Tabelle 5 bietet eine vergleichende Auswertung des mittleren quadratischen Fehlers (MSE) zwischen MEE und FMEE im Zusammenhang mit nicht-Gaußschem Rauschen. Im Allgemeinen ist die Ungleichheit zwischen Die MSE-Werte von FMEE und MEE sind relativ gering. Insbesondere übertrifft FMEE MEE wenn die Stichprobengröße 100 oder 400 beträgt, während MEE eine überlegene Leistung für Stichproben aufweist Größen von 200, 300 und 500. Darüber hinaus erreicht FMEE den optimalen MSE über alle Stichproben Größen, außer für $n = 100$. Darüber hinaus wird beobachtet, dass die MSE-Werte für nicht-Gaußsche Rauschen, wie in Tabelle 5 dargestellt, umfasst einen größeren Bereich als das Gaußsche Rauschen, Dies deutet auf eine größere Variabilität der Ergebnisse sowohl für MEE als auch für FMEE hin.

Machine Translated by Google

Tabelle 5. Der mittlere quadrierte Fehler von MEE und FMEE für nicht-Gaußsches Rauschen.

Stichprobengröße	Durchschnittlicher MSE		Optimaler MSE	
	ICH	FMEE	ICH	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

Tabelle 6 vergleicht die Anzahl der Iterationen, die für die Konvergenz zwischen MEE erforderlich sind und FMEE im Kontext von nicht-Gaußschem Rauschen. FMEE zeigt durchweg schnellere Konvergenz im Vergleich zu MEE. Darüber hinaus wird beobachtet, dass die Iterationszahl von MEE nimmt mit zunehmender Stichprobengröße ab. Im Gegensatz dazu bleibt die Iterationszahl von FMEE trotz Schwankungen in der Stichprobengröße relativ stabil.

Tabelle 6. Die Anzahl der Iterationen von MEE und FMEE für nicht-Gaußsches Rauschen.

Stichprobengröße	Durchschnittliche Iterationszahl		Optimale Iterationszahl	
	ICH	FMEE	ICH	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	35,35	571	12

Basierend auf den in den Tabellen **1–6 dargestellten Ergebnissen ist es offensichtlich, dass FMEE durchweg liefert vergleichbare Ergebnisse wie MEE, allerdings mit deutlich geringerem Zeitaufwand und Iterationszahlen. Insbesondere FMEE zeichnet sich durch seine bemerkenswerte Effizienz aus, es ist etwa 1000-mal schneller als MEE und erreicht Konvergenz innerhalb von etwa 0,1 s für 500 Instanzen. Angesichts seiner linearen Zeitkomplexität von** $O(n)$ **bietet FMEE erhebliche Versprechen für den Einsatz in Echtzeitszenarien, die durch nicht-Gaußsches Rauschen gekennzeichnet sind.**

Anschließend beginnen wir mit praktischen Experimenten, bei denen wir die Kenndaten von Transformatoren nutzen, um Ausfälle im Verteilnetz vorherzusagen. Die Transformatoren werden in folgende Kategor zwei Gruppen auf der Grundlage ihrer geografischen Lage innerhalb unterschiedlicher Ebenen städtischer Aktivität in einer bestimmten Stadt im Nordwesten Chinas. Zunächst richtet sich unsere Aufmerksamkeit auf eine Teilmenge von 1265 Transformatoren in einem kleineren, weniger dicht besiedelten Gebiet des oben genannte Stadt. Wir nehmen eine zufällige Aufteilung dieses Datensatzes in zwei Segmente vor: ein Trainingsset, das 80 % der Daten umfasst, insgesamt 1012 Instanzen, und ein Verifizierungsset umfasst die restlichen 20 %, bestehend aus 253 Instanzen. Genauer gesagt umfasst der Transformator-Datensatz verschiedene charakteristische Variablen, darunter die standardisierte schwer Überlastdauer, maximales Wirklastverhältnis, durchschnittliches Wirklastverhältnis, mittlere dreiphasige Unsymmetrie, standardisierte Dauer starker Dreiphasenunsymmetrie und so weiter.

Wir führen eine vergleichende Analyse durch und vergleichen dabei unser vorgeschlagenes FMEE mit drei etablierten Basismethoden: logistische Regression, neuronales Netzwerk und Support Vector Der Datensatz umfasst 1265 Transformatorkenngrößen im Zusammenhang mit erhebliche Belastungsniveaus, die 30 Runden zufälliger Partitionierung unterzogen werden. Anschließend wird jeder Der zufällig aufgeteilte Datensatz wird vier verschiedenen Algorithmen unterzogen. Die Vorhersagekraft dieser Algorithmen für den Verifizierungssatz wird anhand des F-Maßes und des Fehlers beurteilt.

Bewertungsmetriken.

Abbildung 1 zeigt die F-Maß-Bewertungen aus 30 Vorhersagefällen von vier verschiedenen Algorithmen durchgeführt. Bemerkenswert ist, dass jede Auswertung eine konsistente Aufteilung des Datensatzes auf die vier Algorithmen beinhaltet. Es zeichnet sich ein erkennbarer Trend a alle Algorithmen liefern F-Maß-Werte über 0,8. Besonders hervorzuheben ist, dass unsere vorgeschlagene FMEE-

Machine Translated by Google

weist die vielversprechendste Leistung auf und weist einen durchschnittlichen F-Wert von 0,904 auf, wodurch es die anderen drei Basismethoden bei weitem übertrifft.

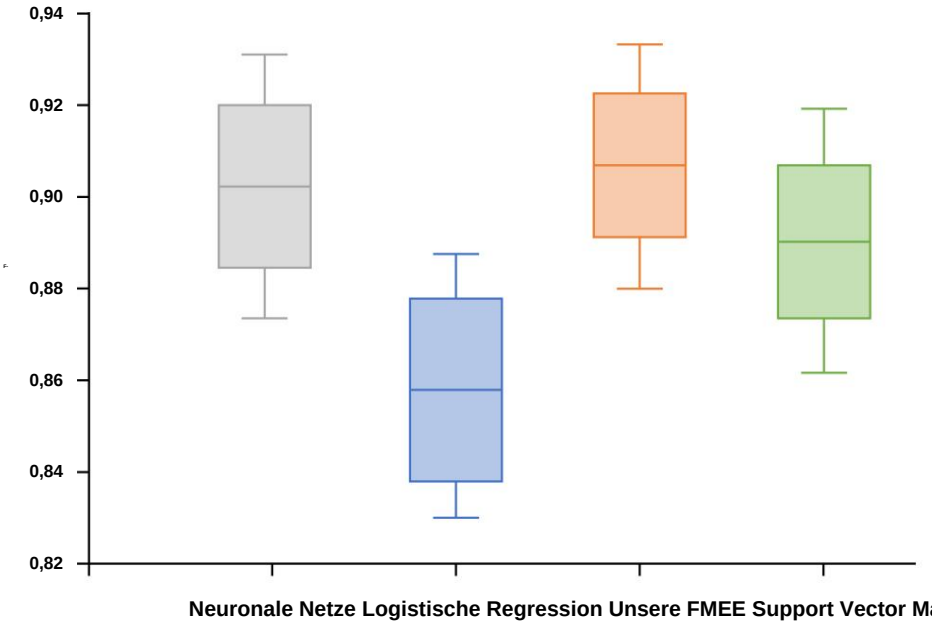


Abbildung 1. Das Boxplot des F-Maßes unter vier Algorithmen zur Vorhersage von Fehlerausfällen.

Um einen umfassenden Vergleich zwischen den vier Algorithmen zu ermöglichen, zeigt Abbildung **2** **die Fehlerrate und das F-Maß der Ergebnisse der Fehlerausfallvorhersage in Form von Linien- und Boxdiagrammen . Besonders hervorzuheben ist, dass FMEE als Spitzenreiter hervorgeht und in Bezug auf die Fehlerrate einen erheblichen Vorteil gegenüber der hochentwickelten Support Vector Machine aufweist. Konkret erreicht unser FMEE eine durchschnittliche Fehlerrate von 9,1 % und übertrifft das neuronale Netzwerk, die logistische Regression und die Support Vector Machine um jeweils 0,3 %, 4,2 % und 4,3 %. Diese Ergebnisse unterstrei**

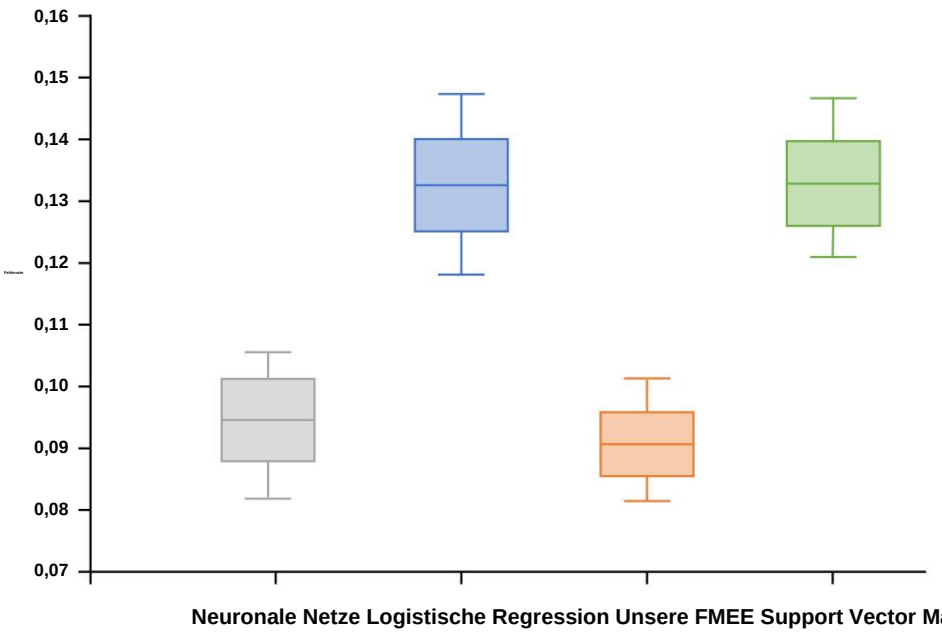


Abbildung 2. Das Boxplot der Fehlerrate unter vier Algorithmen zur Vorhersage von Fehlerausfällen.

Machine Translated by Google

Darüber hinaus weiten wir unsere Experimente auf eine dichter besiedelte Stadtregion innerhalb der nordwestchinesischen Stadt aus. Abbildung 3 zeigt die F-Maß-Bewertungen, die aus 30 unabhängigen Tests abgeleitet wurden, die mit den vier Vergleichsalgorithmen durchgeführt wurden. Auf einer umfassenden Skala erreicht unser vorgeschlagener FMEE ein F-Maß von 0,891 und übertrifft damit Basismethoden wie neuronale Netze, logistische Regression und Support Vector Machines um jeweils 0,013, 0,0

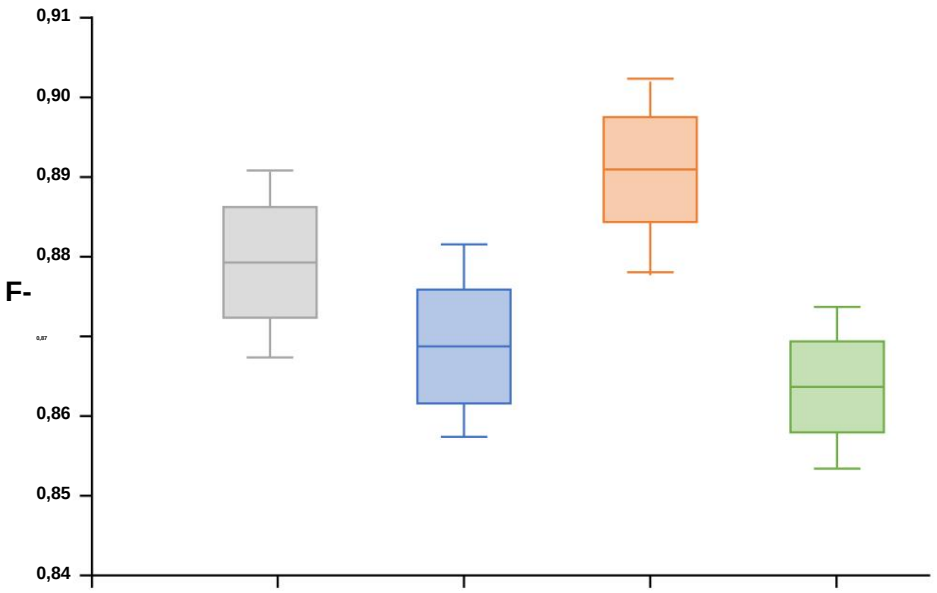


Abbildung 3. Das Boxplot des F-Maßes von vier Algorithmen zur Vorhersage von Fehlerausfällen.

Abschließend berichten wir über die Fehlerrate des F-Maßes in Bezug auf die Ergebnisse der Fehlerausfallvorhersage, die aus vier Vergleichsalgorithmen abgeleitet wurden, dargestellt durch Linien- und Boxdiagramme. Wie in Abbildung 4 dargestellt, erreicht unser vorgeschlagener FMEE eine durchschnittliche Fehlerrate von 11,2 % und zeigt damit eine bemerkenswerte Überlegenheit gegenüber drei Basismethoden. Diese Ergebnisse unterstreichen deutlich die Wirksamkeit unseres Algorithmus bei der Lösung von R

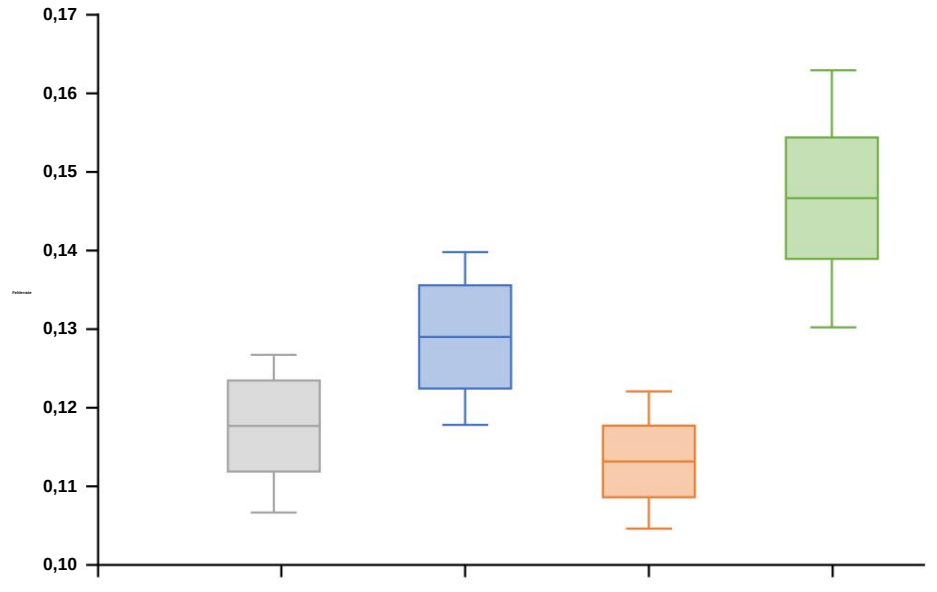


Abbildung 4. Das Boxplot der Fehlerrate von vier Algorithmen zur Vorhersage von Fehlerausfällen.

Machine Translated by Google

5. Schlussfolgerungen

Diese Studie stellt einen schnellen Algorithmus mit minimaler Fehlerentropie vor, der auf lineare Regression zugeschnitten ist. Aus unserer Untersuchung ergeben sich mehrere bemerkenswerte Schlussfolgerungen. Erstens weist das vorgeschlagene FMEE eine deutlich schnellere Konvergenz auf und erfordert im Vergleich zu MEE weniger Iterationen. Zweitens übertrifft FMEE zwar das traditionelle MEE in der Leistung bei gaußschem Rauschen nicht und liefert vergleichbare Ergebnisse bei nicht-gaußschem Rauschen, gemessen am mittleren quadratischen Fehler, aber seine Rechenleistung ist deutlich besser und benötigt weniger als 1 ‰ der von MEE benötigten Zeit. Drittens zeigt FMEE, obwohl es unter der Annahme einer Nähe des Rauschens zur gaußschen Verteilung abgeleitet wurde, Wirksamkeit bei der Handhabung nicht-gaußscher Rauschverteilungen mit schweren Rändern. Viertens ist FMEE aufgrund seiner außergewöhnlichen Zeiteffizienz vielversprechend für Echtzeitanwendungen in Szenarien mit nicht-gaußschem Rauschen. In Zukunft stellen weitere Analysen der Konvergenzbedingungen von FMEE und die Erforschung von Rauschnäherungsmethoden für beliebige Verteilungen Ansätze für zukünftige Forschung dar.

Beiträge der Autoren: Konzeptualisierung, QL und XL; Methodik, QL; Software, WC; Validierung , QL, XL und WC; formale Analyse, YW; Untersuchung, XL; Ressourcen, QL; Datenkuratierung, YW; Schreiben – Vorbereitung des Originalentwurfs, QG; Schreiben – Überprüfung und Bearbeitung, HC; Visualisierung, QG; Überwachung, HC; Projektverwaltung, WC; Mittelbeschaffung, QL und HC. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu.

Finanzierung: Diese Arbeit wird durch das wissenschaftliche und technologische Innovationsprojekt „Forschung zu digitalen Energieraumtechnologiesystemen und Schlüsseltechnologien“ (Zuschussnummer: SGIT0000XMJS2310456) der State Grid Information and Telecommunication Group unterstützt.

Erklärung des Institutional Review Board: Nicht zutreffend.

Einverständniserklärung: Nicht zutreffend.

Datenverfügbarkeitserklärung: Die in dieser Studie präsentierten Daten sind auf Anfrage beim entsprechenden Autor erhältlich. Die Daten sind nicht öffentlich verfügbar, da das im Versuchsprozess verwendete Stromsystem eine besondere Bedeutung hat und in einigen Forschungsbereichen die Privatsphäre gefährdet ist. Daher ist es nicht sinnvoll, die Daten vollständig offenzulegen.

Interessenkonflikte: Die Autoren (Qiang Li, Xiao Liao, Wei Cui und Ying Wang) waren bei der State Grid Information & Telecommunication Group Co., Ltd. angestellt. Die übrigen Autoren erklären, dass die Forschung ohne kommerzielle oder finanzielle Beziehungen durchgeführt wurde, die als potenzieller Interessenkonflikt ausgelegt werden könnten.

Verweise

1. **Baggenstoss, PM Maximum Entropy PDF Design mit Merkmalsdichtebeschränkungen: Anwendungen in der Signalverarbeitung.** IEEE Trans. Signal Process. **2015,** 11, **2815–2825.** [CrossRef]
2. **Li, Y.; Bai, X.; Yuan, D.; Yu, C.; San, X.; Guo, Y.; Zhang, L.; Ye, J. Cu-basiertes hochentropisches zweidimensionales Oxid als stabiler und aktiver photothermischer Katalysator.** Nat. Commun. **2023,** 14, **3171.** [CrossRef] [PubMed]
3. **Bisikalo, O.; Kharchenko, V.; Kovtun, V.; Krak, I.; Pavlov, S. Parametrisierung des stochastischen Modells zur Auswertung variabler kleiner Daten in der Shannon-Entropiebasis.** Entropy **2023,** 25, **184.** [CrossRef] [PubMed]
4. **Chen, Q.; Zhang, F.; Su, L.; Lin, B.; Chen, S.; Zhang, Y. Zustandsparameterfusionsschätzung für intelligente Fahrzeuge basierend auf IMM-MCCKF.** Appl. Sci. **2024,** 14, **4495.** [CrossRef]
5. **Carli, FP; Chen, T.; Ljung, L. Maximale Entropiekerne zur Systemidentifikation.** IEEE Trans. Autom. Control **2016,** 62, **1471–1477.** [Querverweis]
6. **Fan, X.; Chen, L.; Huang, D.; Tian, Y.; Zhang, X.; Jiao, M.; Zhou, Z. Von Einzelmetallen zu hochentropischen Legierungen: Wie maschinelles Lernen die Entwicklung von Metallelektrokatalysatoren beschleunigt.** Adv. Funct. Mater. **2024, 1, 2401887.** [CrossRef]
7. **Han, TT; Ge, SS; Lee, TH Persistent Dwell-Time Switched Nonlinear Systems: Variationsparadigma und Messgerätedesign.** IEEE Trans. Autom. Kontrolle **2010,** 55, **321–337.**
8. **Chesi, G. Stabilisierung und Entropiereduktion durch SDP-basiertes Design von Feedback-Reglern mit fester Ordnung und Abstimmungsparametern .** IEEE Trans. Autom. Control **2017,** 62, **1094–1108.** [CrossRef]
9. **Wei, P.; Li, HX Räumlich-zeitliche Entropie zur Anomalieerkennung und Lokalisierung von Lithium-Ionen-Akkupacks.** IEEE Trans. Ind. **2023,** 70 , **12851–12859.** [CrossRef]
10. **Wang, Z.; Li, G.; Yao, L.; Cai, Y.; Lin, T.; Zhang, J.; Dong, H. Intelligentes Fehlererkennungsschema für Windturbinen mit konstanter Geschwindigkeit basierend auf verbesserter multiskaliger Fuzzy-Entropie und adaptiver chaotischer Aquila-Optimierung basierend auf Support Vecto** **2023,** 138, **582–602.** [CrossRef] [PubMed]

Machine Translated by Google

11. **Daneshpazhouh, A.; Sami, A. Entropiebasierte Ausreißererkennung unter Verwendung eines halbüberwachten Ansatzes mit einigen positiven Beispielen.** Mustererkennung. Lett. **2014,** 49, **77–84.** [CrossRef]
12. **Qu, B.; Wang, Z.; Shen, B.; Dong, H. Dezentrale dynamische Zustandsschätzung für Mehrmaschinen-Stromversorgungssysteme mit nicht-Gaußschem Rauschen: Ausreißererkennung und -lokalisierung.** Automatica **2023,** 153, **111010.** [CrossRef]
13. **Feng, Z.; Wang, G.; Peng, B.; He, J.; Zhang, K. Neuartiger robuster Kalman-Filter mit minimaler Fehlerentropie und Wasserstein-Verteilung unter Modellunsicherheit und nicht-Gaußsches Rauschen.** Signal Process. **2023,** 203, **108806.** [CrossRef]
14. **Sun, M.; Davies, ME; Proudler, IK; Hopgood, JR Adaptiver Kernel-Kalman-Filter.** IEEE Trans. Signal Process. **2023,** 71, **713–726.** [Querverweis]
15. **Wu, L.; Lin, H.; Hu, B.; Tan, C.; Gao, Z.; Liu, Z.; Li, SZ Jenseits der Annahme von Homophilie und Homogenität: Relationenbasierte frequenzadaptive Graph-Neuralnetze.** IEEE Trans. Neural Netw. Learn. Syst. **2023,** 35, **8497–8509.** [CrossRef] [PubMed]
16. **Guo, L.; Wang, H. Minimale Entropiefilterung für multivariate stochastische Systeme mit nicht-Gaußschem Rauschen.** IEEE Trans. Autom. Control **2006,** 51, **695–700.**
17. **Wang, Y.; Tang, YY; Li, L. Auf minimaler Fehlerentropie basierende spärliche Darstellung für robustes Subraum-Clustering.** IEEE Trans. Signal Process. **2015,** 63, **4010–4021.** [CrossRef]
18. **Wang, M.; Deng, W. Tiefe Gesichtserkennung: Eine Untersuchung.** Neurocomputing **2021,** 429, **215–244.** [CrossRef]
19. **Lin, M.; Cheng, C.; Peng, Z.; Dong, X.; Qu, Y.; Meng, G. Nichtlineare dynamische Systemidentifikation mit den Methoden der spärlichen Regression und der separierbaren kleinsten Quadrate.** J. Sound Vib. **2021,** 505, **116141.** [CrossRef]
20. **Wu, Z.; Peng, S.; Ma, W.; Chen, B.; Principe, JC Minimale Fehlerentropiealgorithmen mit Sparsity-Strafbeschränkungen.** Entropie **2015,** 17, **3419–3437.** [CrossRef]
21. **Huang, S.; Tran, TD Spärliche Signalwiederherstellung durch Minimierung verallgemeinerter Entropiefunktionen.** IEEE Trans. Signal Process. **2018,** 67, **1322–1337.** [CrossRef]
22. **Yue, H.; Wang, H. Minimale Entropiekontrolle von Closed-Loop-Tracking-Fehlern für dynamische stochastische Systeme.** IEEE Trans. Autom. Control **2003,** 48, **118–122.**
23. **Wang, H. Minimale Entropiekontrolle nicht-Gaußscher dynamischer stochastischer Systeme.** IEEE Trans. Autom. Control **2002,** 47, **398–403.** [Querverweis]
24. **Mitra, R.; Bhatia, V. Auf dem Kriterium der minimalen Fehlerentropie basierende Kanalschätzung für Massive-MIMO in VLC.** IEEE Trans. Veh. Technol. **2018,** 68 , **1014–1018.**
25. **Erdogmus, D.; Principe, JC Konvergenzeigenschaften und Dateneffizienz des Kriteriums der minimalen Fehlerentropie beim Adaline- Training.** IEEE Trans. Signal Process. **2003,** 51, **1966–1978.**
26. **Chen, B.; Yuan, Z.; Zheng, N.; Principe, JC Kernel-Minimum-Error-Entropy-Algorithmus.** Neurocomputing **2013,** 121, **160–169.** [Querverweis]
27. **Hu, T.; Wu, Q.; Zhou, DX Konvergenz des Gradientenabstiegs für das Prinzip der minimalen Fehlerentropie bei der linearen Regression.** IEEE Trans. Signalprozess. **2016,** 64, **6571–6579.** [CrossRef]
28. **Zhang, Y.; Chen, B.; Liu, X.; Yuan, Z.; Principe, JC Konvergenz eines Fixpunkt-Minimalfehler-Entropiealgorithmus.** Entropie **2015,** 17, **5549–5560.** [CrossRef]
29. **Chen, B.; Zhu, Y.; Hu, J.; Zhang, M. Über optimale Schätzungen mit minimalem Fehlerentropiekriterium.** J. Frankl. Inst. **2010,** 347, **545–558.**
30. **Heravi, AR; Hodtani, GA Eine neue informationstheoretische Beziehung zwischen minimaler Fehlerentropie und maximaler Korrentropie.** IEEE Signal Process. Lett. **2018,** 25, **921–925.**
31. **Yang, S.; Tan, J.; Chen, B. Robustes, spike-basiertes kontinuierliches Meta-Lernen, verbessert durch eingeschränktes Kriterium der minimalen Fehlerentropie.** Entropie **2022,** 24, **455.** [CrossRef] [PubMed]
32. **Dang, L.; Chen, B.; Wang, S.; Ma, W.; Ren, P. Robuste Energiesystemzustandsschätzung mit minimaler Fehlerentropie ohne Kalman-Filter.** IEEE Trans. Instrument. Mess. **2020,** 69, **8797–8808.** [CrossRef]
33. **Wang, B.; Hu, T. Halbüberwachtes Minimalfehler-Entropieprinzip mit verteilter Methode.** Entropy **2018,** 20, **968.** [CrossRef] [PubMed]
34. **Chen, B.; Xing, L.; Xu, B.; Zhao, H.; Principe, JC Einblicke in die Robustheit der Entropieschätzung mit minimalem Fehler.** IEEE Trans. Neuronale Netze lernen. Syst. **2016,** 29, **731–737.** [CrossRef] [PubMed]
35. **Shen, P.; Li, C. Methode der minimalen Gesamtfehlerentropie zur Parameterschätzung.** IEEE Trans. Signal Process. **2015,** 63, **4079–4090.** [Querverweis]
36. **Kothari, P.; Kreiss, S.; Alahi, A. Vorhersage menschlicher Flugbahnen in Menschenmengen: Eine Deep-Learning-Perspektive.** IEEE Trans. Intell. Transp. Syst. **2021 ,** 23, **7386–7400.**
37. **Kim, N.; Kwon, K. Normalisierter Minimalfehler-Entropiealgorithmus mit rekursiver Leistungsschätzung.** Entropy **2016,** 18, **239.** [Querverweis]
38. **Chen, B.; Ma, R.; Yu, S.; Du, S.; Qin, J. Granger-Kausalitätsanalyse basierend auf dem quantisierten Mindestfehler-Entropiekriterium.** IEEE Signalprozess. Lett. **2019,** 26, **347–351.** [CrossRef]
39. **Niknam, M.; Santos, LF; Cory, DG Experimenteller Nachweis der Korrelations-Rényi-Entropie im zentralen Spinmodell.** Phys. Rev. Lett. **2021,** 127, **080401.** [CrossRef] [PubMed]
40. **Kim, T.; San Kim, D.; Jang, LC; Lee, H.; Kim, H. Darstellungen entarteter Hermite-Polynome.** Adv. Appl. Math. Mech. **2022,** 139, **102359.** [CrossRef]

Machine Translated by Google

41. Lin, W.; Zhang, JE Die gültigen Bereiche von Gram–Charlier-Dichten mit Kumulanten höherer Ordnung. J. Comput. Appl. Math. **2022,** 407, **113945.** **[Querverweis]**

42. Matsushita, R.; Brandão, H.; Nobre, I.; Da Silva, S. Differentialentropieschätzung mit einem Pareto-Kernel: Tail-Heaviness und Glättung. Phys. A Stat. Mech. Appl. **2024, 1, 129850. [CrossRef]**

43. Canales, C.; Galarce, C.; Rubio, F.; Pineda, F.; Anguita, J.; Barros, R.; Parragué, M.; Daille, LK; Aguirre, J.; Armijo, F.; et al. **Testen des Tests: Eine vergleichende Studie zur mikrobiellen Korrosion im Meer unter Labor- und Feldbedingungen.** ACS Omega **2021, 6, 13496–13507. [CrossRef] [PubMed]**

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Artikel

Optimierung der Genauswahl für die Krebsklassifizierung in Hochdimensionale Daten mit einem verbesserten afrikanischen Geier-Algorithmus

Mona G. Gafar 1,2,*, Amr A. Abohany 3,†, Ahmed E. Elkhouli 4,† und Amr A. Abd El-Mageed 5,†

1 Abteilung für Computertechnik und Information, Technische Hochschule in Wadi Alddawasir, Prinz-Sattam-bin-Abdulaziz-Universität, Kharj 16278, Saudi-Arabien.
2 Abteilung für maschinelles Lernen und Informationsabruf, Künstliche Intelligenz, Universität Kafrelsheikh, Kafrelsheikh 33511,
3 Ägypten. Fakultät für Computer und Information, Universität Kafrelsheikh, Kafrelsheikh 33511, Ägypten;
4 amrabohany8@gmail.com. Abteilung für Biomedizintechnik, Fakultät für Elektrotechnik, Universität Menofia, Menofia 32951, Ägypten;
5 eng.elkholy.ahmed@gmail.com. Abteilung für Informationssysteme, Universität Sohag, Sohag 82511, Ägypten; amr.atef@commerce.sohag.edu.eg. * Korrespondenz: m.gafar@psau.edu.sa
† oder dr.mona.gamal85@gmail.com. Diese Autoren haben zu gleichen Teilen zu dieser Arbeit beigetragen.

Zusammenfassung: Diese Studie stellt eine neuartige Methode namens RBAVO-DE (Relief Binary African Vultures Optimization based on Differential Evolution) vor, die sich mit der Herausforderung der Genselektion (GS) in hochdimensionalen RNA-Seq-Daten befasst, insbesondere mit dem normalisierten Level 3 RSEM-Gene-Datensatz rnaseqv2 lluminaHiSeq rnaseqv2 un edu, der mehr als 20.000 Gene enthält. RNA-Sequenzierung (RNA-Seq) ist ein bahnbrechender Ansatz, der eine umfassende Quantifizierung und Charakterisierung von Genexpressionen ermöglicht und die Fähigkeiten von Mikroarray-Technologien übertrifft, indem er eine detailliertere Ansicht von RNA-Seq-Genexpressionsdaten bietet. Die quantitative Genexpressionsanalyse kann von entscheidender Bedeutung sein, um Gene zu identifizieren, die normales von bösartigem Gewebe unterscheiden. Die Verwaltung dieser hochdimensionalen dichten Matrixdaten ist jedoch mit erheblichen Herausforderungen verbunden. Der RBAVO-DE-Algorithmus ist darauf ausgelegt, aus einem Datensatz mit mehr als 20.000 Genen sorgfältig die informativsten Gene auszuwählen und ihre Relevanz in 22 Krebsdatensätzen zu bewerten. Um die Wirksamkeit der ausgewählten Gene zu bestimmen, verwendet diese Studie die Klassifikatoren Support Vector Machine (SVM) und k-Nearest Neighbor (k-NN). Im Vergleich zu binären Versionen weithin anerkannter metaheuristischer Algorithmen zeigt RBAVO-DE eine bessere Leistung. Laut Wilcoxons Rangsummentest mit einem Signifikanzniveau von 5 % erreicht RBAVO-DE eine Klassifizierungsgenauigkeit von bis zu 100 % und reduziert die Merkmalsgröße in den meisten der 22 untersuchten Krebsdatensätze um bis zu 98 %. Dieser Fortschritt unterstreicht das Potenzial von RBAVO-DE, die Präzision der Genauswahl für d

Schlüsselwörter: RNA-Sequenzierung; Genselektion; hohe Dimensionalität; metaheuristische Algorithmen; Algorithmus der afrikanischen Geier; differenzielle Evolution; Relieftechnik

check for updates

Zitat: Gafar, MG; Abohany, AA; Elkhouli, AE; Abd El-Mageed, AA Optimierung der Genselektion für Krebsklassifikation in Hochdimensionale Daten mithilfe eines verbesserten Algorithmus für afrikanische Geier. Algorithmen **2024,** 17, **342. https://doi.org/10.3390/a17080342**

Akademische Herausgeber: Muhammad Adnan Khan und Frank Werner

Empfangen: 27. April 2024
Überarbeitet: 16. Juli 2024
Akzeptiert: 1. August 2024
Veröffentlicht: 6. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/ 4.0/).

1. Einleitung

Desoxyribonukleinsäure (DNA) bildet unseren genetischen Code und enthält das Rezept für unsere Existenz. Obwohl jede Zelle dieselbe DNA hat, ist jede Gewebestruktur einzigartig und dient einem bestimmten Zweck. Der RNA-Transkriptionsmechanismus bestimmt, welche Gene in einer Zelle aktiv sind, wodurch RNA in Proteine umgewandelt werden kann, die für die Struktur und Funktionalität der Zelle verantwortlich sind **[1]. Wir analysieren RNA-Seq-Genexpressionsdaten (GED), um genetische Veränderungen festzustellen und Krankheitsbiomarker zu bewerten. Die Analyse der differentiellen Expression identifiziert quantitative Unterschiede in der Genexpression und ermöglicht es uns, Gene zu kategorisieren, deren Expression sich unter verschiedenen Bedingungen**

Machine Translated by Google

Krankheiten und Wege zu ihrer Behandlung zu finden. Die Technologien zur Profilierung der Genexpression haben sich im Laufe der Jahre erheblich weiterentwickelt [2].

Zwei beliebte Techniken sind Microarray, das einen hybridisierungsbasierten Ansatz verwendet, und RNA-Seq, das auf Sequenzierung der nächsten Generation basiert. Diese Technologien haben ihren Wert für die Weiterentwicklung unseres Verständnisses der Genexpression unter Beweis gestellt [3]. Bemerkenswerte Studien wie die von Chen et al. [4] und Nunez et al. [5] haben das Potenzial dieser Techniken hervorgehoben, während Wang et al. [6] wertvolle Einblicke in den RNA-Seq-Ansatz geliefert haben .

Die beiden betrachteten Methoden dienen der Quantifizierung der Genexpression zur Klassifizierung und statistischen Untersuchung. In diesem Artikel wurden die Quantifizierungsdaten ausgewählt, die mit dem auf Next-Generation-Sequencing (NGS) basierenden RNA-Seq-Ansatz gewonnen wurden, da dieser im Vergleich zu Microarray-Daten eine höhere Genauigkeit bei der Ermittlung von RNA-Quantifizierungsniveaus bietet [7]. RNA-Seq hat mehrere Einschränkungen der Microarray-Analyse überwunden, wie etwa die Abhängigkeit von vorherigem Sequenzierungswissen, das den Im Gegensatz zu Mikroarrays erfordert RNA-Seq keine Vorkenntnisse und erweitert den dynamischen Erkennungsbereich erheblich. Diese Verbesserung verbessert die Genauigkeit der Ergebnisse und ermöglicht die Identifizierung eines umfassenden Satzes von Genen, was zu einem präzisen Verständnis der Krankheitsbiomarker führt. Infolgedessen hat sich RNA-Seq als wirksames Werkzeug für Forscher erwiesen, das wertvolle Einblicke in die zugrunde liegenden molekularen Mechanismen verschiedener Krankheiten bietet [10].

Der „Dimensionalitätsfluch“ [11] ist aufgrund der Fülle verfügbarer Daten in der heutigen Zeit zu einer weit verbreiteten Herausforderung geworden. Dies hat zu einem starken Anstieg der Entwicklung von Techniken und Algorithmen zur Merkmalsauswahl (FS) geführt. FS-Algorithmen können in vier verschiedene Methoden eingeteilt werden: Filter, Wrapper, eingebettet und Hybrid [12,13]. Diese Ansätze zielen darauf ab, die aussagekräftigsten Merkmale zu identifizieren, die zwischen versc In unserem Fall geht es uns darum, die mit Tumoren in Zusammenhang stehenden Gene zu identifizieren.

Der Filteransatz ist eine weit verbreitete Technik in der FS, bei der die Relevanz einzelner Gene anhand statistischer Werte bewertet wird. Diese Methode ist für ihre hohe Genauigkeit bei der Auswahl der zufriedenstellendsten Gengruppe bekannt. Dennoch weist der Filteransatz einige Einschränkungen auf, da bei der Arbeit an jedem einzelnen Gen die Wechselbeziehungen zwischen den Genen ignoriert werden, was zu einem Problem mit lokalen Optima führen kann [14]. Es ist erwähnenswert, dass der Filteransatz in zwei Untertypen unterteilt werden kann – univariat und multivariat – wobei letzterer die Korrelationen zwischen den Genen berücksichtigt. Einige Beispiele für den Filteransatz sind Relief [15], Fisher-Score [16], t-Test [17] und Informationsgewinn [18].

Die Wrapper-Methode untersucht alle möglichen Gen-Subsets, um eine Gen-Submenge zu testen und zu erstellen, um ihre Implikationen zu bestimmen. Ein spezifischer Klassifikator wird verwendet, um das Ergebnis jeder Submenge zu bestimmen, und die Kategorisierungstechnik wird für jede Bewertung mehrfach verwendet. Im Vergleich zum Filteransatz liefert der Wrapper-Ansatz eine bessere Leistung, indem er eine Kategorisierungstechnik verwendet, die den Lernvorgang steuert. Diese Methode erfordert jedoch viel Zeit und Rechenressourcen, insbesondere bei der Verarbeitung großer Datenmengen [19].

Metaheuristische Methoden (MHMs) sind hochrangige Heuristiken, die in der mathematischen Optimierung und in der Informatik verwendet werden, um zufriedenstellende Lösungen für Optimierungsprobleme zu finden, wenn Informationen unvollständig oder Ressourcen begrenzt sind [20]. MHMs untersuchen eine Teilmenge von Lösungen, was sie aufgrund ihrer wenigen Annahmen über das Problem für verschiedene Optimierungsprobleme nützlich macht [21]. Sie garantieren jedoch keine global optimalen Lösungen. Viele MHMs verwenden stochastische Optimierung, was bedeutet, dass die Lösung auf Zufallsvariablen basiert, die während des Prozesses generiert werden [22].

Metaheuristische Methoden (MHMs) sind praktischer als herkömmliche iterative Methoden und Optimierungsalgorithmen, da sie ein breiteres Spektrum möglicher Lösungen erkund Folglich sind MHMs zu einer bevorzugten Strategie zur Lösung von Optimierungsproblemen geworden [23]. Zahlreiche Forschungsarbeiten haben gezeigt, dass MHMs unter den verschiedenen Wrapper-Methoden am besten geeignet sind, um das Problem der Merkmalsauswahl (FS) zu lösen. Stochastische Methoden, einschließlich MHMs, können schnell optimale oder nahezu optimale Ergebnisse erz

Machine Translated by Google

bieten Vorteile wie Flexibilität, Selbstmanagement ohne detaillierte mathematische Eigenschaften und die Möglichkeit, mehrere Ergebnisse gleichzeitig zu bewerten. In letzter Zeit wurden verschiedene MHMs entwickelt, um das FS-Problem zu lösen. Sie bieten zuverlässige nahezu optimale Lösungen bei deutlich reduziertem Rechenaufwand [24].

Eingebettete FS-Methoden verwenden eine Lerntechnik, um geeignete Gene auszuwählen, die mit dem Klassifizierungsprozess interagieren. Diese Methode kombiniert die FS-Technik als Teil des Lernmodells. Der Lernalgorithmus wird mit einer anfänglichen Attribut-Teilmenge trainiert, um ein Maß für die Bewertung der Rangwerte von Attributen zu schätzen [25]. Das ultimative Ziel besteht darin, die Rechenzeit für die Neuklassifizierung verschiedener Teilmengen zu verringern, indem die FS- Phase in den Trainingsprozess integriert wird. Einige Techniken in diesem Ansatz führen eine Merkmalsgewichtung basierend auf Regularisierungsmodellen mit Zielfunktionen durch, die Anpassungsfehler minimieren und gleichzeitig erzwingen, dass Merkmalskoeffizienten klein oder genau null sind. Beispiele für eingebettete Methoden sind der regelbasierte Merkmalsteilmengenauswahlalgorithmus First-Order Inductive Learner (FOIL) und SVM basierend auf Recursive Feature Elimination (SVM-RFE) [26].

Die Hybridmethode ist eine gut durchdachte Technik, die Filter- und Wrapper -Methoden kombiniert, um die jeweiligen Stärken zu nutzen. Dieser Ansatz beginnt mit einem Filter, der die Dimensionalität des Merkmalsraums reduziert und mehrere Teilmengen mit mittlerer Komplexität generiert. Anschließend wird ein Wrapper als Lernmethode eingesetzt, um die am besten geeignete Kandidaten-Teilmenge auszuwählen **[27]. Der hybride Ansatz integriert die Genauigkeit von Wrappern und die Effizienz von Filtern, was zu einer optimalen Methodik führt [28,29].**

1.1. Motivation und Beiträge

In diesem Artikel wird ein effektiver globaler metaheuristischer Optimierungsalgorithmus vorgestellt, der als African Vultures Optimization (AVO)-Algorithmus [30] bezeichnet wird und die **Lebens- und Essgewohnheiten afrikanischer Geier nachahmt .** Der vorgeschlagene Algorithmus besteht aus vier grundlegenden **Phasen: Aufteilung der Population in drei Gruppen – die beste Lösung, die zweitbeste Lösung und die verbleibenden Lösungen; Messung des Hungerniveaus zur Formulierung eines mathematischen Modells für die Ausbeutung, Erkundung und Übertragung der Geier; Erkundung der Geier unter Einsatz von zwei Strategien, um es den Geiern zu ermöglichen, über längere Zeiträume große Entfernungen zurückzulegen, um an zufälligen Orten Nahrung zu finden; und Ausbeutung der Geier, die zwei Unterphasen umfasst. In der ersten Unterphase werden zwei unterschiedliche Strategien verwendet: Belagerungskampf für starke Geier und Rotationsflug, der eine spiralförmige Aktivität zwischen dem herausragenden Geier und dem Rest bildet. In der zweiten Unterphase werden zwei Strategien verwendet: das Versammeln der Geier um die Nahrungsquelle und ein feindlicher Belagerungskampf, bei dem die Geier aggressiver werden und versuchen, Nahrung zu stehlen, die von gesund**

Diese Studie schlägt eine verbesserte binäre Variante der AVO-Technik vor, nämlich die RBAVO-DE-Technik, ein produktiver Ansatz, der eine genaue Leistung bei der Handhabung des GS-Problems zeigt. Zunächst besteht eine hohe Wahrscheinlichkeit, dass die empfohlene Technik lokale Optima vermeidet und eine ausreichende Untersuchungspräzision, schnelle Konvergenz und verbesserte Stabilität erreicht. Im Vergleich zu aktuellen MHMs erreicht die vorgeschlagene RBAVO-DE-Technik eine verbesserte Wirksamkeit, indem sie optimale oder im Wesentlichen optimale Lösungen für viele der analysierten Situationen liefert. Der Relief-Algorithmus wird verwendet, um nur die verwandten Merkmale zu identifizieren und den endgültigen Klassifizierungsdatensatz zu erstellen. RBAVO-DE kombiniert den Relief-Algorithmus mit dem DE-Ansatz, um die Erkundungsfähigkeit zu erhöhen und durch Wiederholungen die besten Ergebnisse innerhalb des Lösungsraums zu erzielen. Es verwendet auch eine Übertragungsfunktion, um reale Positionswerte in binäre Werte umzuwandeln. Der RBAVO-DE- Ansatz ist bei GS sinnvoll, da er einfach zu verstehen und umzusetzen ist, verschiedene Optimierungsprobleme bewältigen kann, in akzeptabler Zeit wertvolle Ergebnisse

liefert , weniger Rechenleistung erfordert und eine kleine Anzahl von Steuerparametern verwendet. Die wichtigsten Beiträge **dieses Dokuments sind wie folgt: • Level-3-Daten basierend auf Next-Generation-Sequencing (RNA-Seq) wurden vorverarbeitet. • Die Fähigkeit des vorgeschlagenen AVO-Metaheuristik-Algorithmus, das GS-Problem zu lösen, wurd**

Machine Translated by Google

- **Um eine binäre Version namens RBAVO-DE-Algorithmus zu erstellen, wird der AVO-Algorithmus geändert und neu erstellt.**
- **Der vorgeschlagene RBAVO-DE-Algorithmus kombiniert die binäre Variante von AVO mit einem Relief-Ansatz und DE, um die Explorationsfähigkeit des Suchraums zu verbessern und die erzielten optimalen Ergebnisse zu steigern.**
- **Der vorgeschlagene RBAVO-DE-Algorithmus wird zum ersten Mal auf RNA-Seq GED angewendet.**
- Zur Bewertung der Ergebnisse werden **mehrere Leistungsindikatoren verwendet, darunter durchschnittliche Fitness, Klassifizierungsgenauigkeit, Anzahl ausgewählter Gene, Präzision, Rückruf und F1-Score.**
- Die Wirkung des vorgestellten RBAVO-DE-Algorithmus **unter Verwendung der beiden empfohlenen ML-Klassifikatoren (SVM und k-NN) wird mit anderen Algorithmen in der Literatur verglichen .**
- **Zur Bewertung der vorgeschlagenen RBAVO-DE-Methode werden 22 verschiedene Krebsdatensätze verwendet und die Ergebnisse werden angezeigt.**
- **Die ausgewählten Gene werden anhand von mit Krebs assoziierten Biomarkern untersucht.**

1.2. Struktur

Dieser Artikel ist in fünf Hauptabschnitte gegliedert. Abschnitt 2 gibt einen Überblick über frühere Forschungen zu FS unter Verwendung von RNA-Seq GED. Darauf folgt Abschnitt 3, der eine eingehende Erörterung des RBAVO-DE-Algorithmus, einer erweiterten Version von AVO, einschließlich seiner Parameter zur Behandlung von GS bietet. Abschnitt 4 zeigt die experimentellen Ergebnisse im Vergleich zu mehreren aktuellen MHMs. Schließlich schließt Abschnitt 6 diese Studie ab und schlägt Ansätze für zukünftige Untersuchungen vor.

2. Verwandte Werke

In diesem Abschnitt diskutieren wir die Literatur, die sich auf die Methoden konzentriert, die Forscher zur Klassifizierung von RNA-Seq GED verwenden, das typischerweise eine hohe Dimensionalität aufweist. Um eine optimale Leistung der Klassifizierungsmethoden zu erreichen, ist es entscheidend, irrelevante und nicht verwandte Gene zu ignorieren; daher ist die Auswahl geeigneter Gene ein entscheidender Schritt, bevor ML- und Deep- Learning-Techniken (DL) [31] oder andere Klassifizierungstechniken eingesetzt werden. In diesem Zusammenhang haben wir einige relevante Artikel in diesem Bereich untersucht, um das Ziel der RNA-Seq- Kategorisierung zur Krebsidentifizierung zu erreichen.

Yaqoob et al. [32] führten eine hochmoderne Technik für GS ein, die als Sine- Cosine–Cuckoo Search Algorithm (SCCSA) bekannt ist, eine hybride Methode, die maßgeschneidert ist, um neben etablierten ML-Modellen wie SVM zu funktionieren. Dieser innovative GS-Algorithmus wurde anhand eines Benchmark-Datensatzes zu Brustkrebs evaluiert, wobei seine Leistung sorgfältig analysiert und mit anderen GS-Methoden verglichen wurde. Um die Auswahl der Merkmale zu verfeinern, wurde zunächst die Methode der minimalen Redundanz und maximalen Relevanz als vorläufiger Filterschritt angewendet . Anschließend wurde der hybride SCCSA-Ansatz eingesetzt, um den GS-Prozess weiter zu verbessern und zu optimieren. Im letzten Schritt wurde der Datensatz mithilfe des SVM-Klassifikators anhand der ausgewählten Gene klassifiziert. In Anbetracht der entscheidenden Bedeutung von GS bei der Dekodierung komplexer biologischer Datensätze erweist sich SCCSA als entscheidendes Kapital für die Klassifizierung krebsbezogener Datensätze.

Joshi et al. [33] führten eine innovative Optimierungsstrategie namens PSO-CS ein, die in DL zur Klassifizierung von Hirntumoren integriert ist. Diese Methode verbessert die Effizienz der Particle Swarm Optimization (PSO)-Technik durch Einbeziehung des Cuckoo Search (CS)-Algorithmus und optimiert so den Klassifizierungsprozess. Nach dieser Optimierung verwendet PSO-CS DL zur Klassifizierung von GED im Zusammenhang mit Hirntumoren und identifiziert neben der PSO-CS-Optimierungsmethode verschiedene Klassen, die mit bestimmten Tumoren verbunden sind. Durch die Integration der PSO- CS-Technik mit DL übertraf sie andere DL- und ML-Modelle hinsichtlich der Klassifizierungsgenauigkeit deutlich, wie verschiedene Leistungskennzahlen belegen.

Mahto et al. [34] stellten einen bahnbrechenden Ansatz zur Krebsklassifizierung durch eine integrierte Methode basierend auf CS-SMO (CS and Spider Monkey Optimization) für

Zunächst wird die Fitnessfunktion des Spider Monkey Optimization (SMO)-Algorithmus mithilfe des CS-Algorithmus modifiziert. Diese Modifikation nutzt die Stärken beider MHMs

Machine Translated by Google

um eine Teilmenge von Genen zu identifizieren, die Krebs in einem frühen Stadium vorhersagen können. Um die Genauigkeit des CS-SMO-Algorithmus weiter zu verfeinern, wird ein Vorverarbeitungsschritt, MRMR, eingesetzt, um die Komplexität der Genexpressionsdatensätze für Krebs zu reduzieren. Anschließend werden diese Gen-Teilmengen mithilfe von DL verarbeitet, um verschiedene Krebsarten zu klassifizieren. Die Wirksamkeit des CS- SMO-Ansatzes in Verbindung mit DL wurde anhand von acht Benchmark-Mikroarray-Genexpressionsdatensätzen für Krebs bewertet, wobei seine Leistung anhand verschiedener Maßstäbe untersucht wurde. Die mit DL integrierte CS-SMO-Methode zeigte eine überlegene Klassifizierungsgenauigkeit für alle untersuchten groß angelegten Genexpressionsdatensätze für Krebs und übertraf bestehende DL- und ML-Modelle.

Neggaz et al. [35] führten eine verbesserte Version der Manta-Ray-Futtersuchoptimierung namens MRFO-SC ein, die trigonometrische Operatoren verwendete, die vom Sinus-Cosinus- Algorithmus (SC) inspiriert waren, um das GS-Problem zu lösen. Das k-NN-Modell wurde für die Gensatzauswahl verwendet. Darüber hinaus wurde die statistische Signifikanz des MRFO-SC mithilfe des Rangsummentests von Wilcoxon bei einem Signifikanzniveau von 5 % bewertet. Die Ergebnisse wurden ausgewertet und mit einigen aktuellen MHMs verglichen. Der Vergleich und die experimentellen Ergebnisse bestätigten die effektive Leistung des vorgeschlagenen MRFO-SC bei hoch- und niedrigdimensionalen Benchmark-Datensätzen, indem bei 85 % der GS-Benchmark-Datensätze die höchste Klassifizierungsgenauigkeit erzielt wurde.

Lyu et al. [36] untersuchten Krebsbiomarker, indem sie sich auf die Bedeutung der Gene in Bezug auf ihre Auswirkungen auf die Klassifizierung konzentrierten. Sie verfolgten einen zweiphasigen Ansatz – Datenvorverarbeitung und Nutzung eines Convolutional Neural Network – um den Tumortyp zu klassifizieren. In der zweiten Phase erstellten sie Heatmaps für jede Kategorie, um Gene zu identifizieren, die mit den Pixeln mit den höchsten Intensitäten in den Heatmaps in Zusammenhang stehen. Anschließend bewerteten sie die Pfade der ausgewählten Gene. Während der Vorverarbeitung entfernten sie die Ebenen der Genexpression, die während der GS-Phase nicht verändert worden waren, unter Verwendung einer Varianzschwelle von 1,19, wodurch die Anzahl der Gene von 19.531 auf 10.381 reduziert wurde. Die endgültige Klassifizierungsgenauigkeit betrug 95,59, was gut ist, aber durch eine effektivere FS-Methodik zur weiteren Reduzierung der Datendimensionalität noch verbessert werden könnte .

Khalifa et al. [37] bauten auf dem oben genannten Papier [36] auf. Der Schwerpunkt ihrer Studie lag auf fünf Arten von Krebsdaten: Uteruskorpus-Endometriumkarzinom (UCEC), Lungen-Plattenepithelkarzinom (LUSC), Lungenadenokarzinom (LUAD), Nieren-Nieren-Klarzellkarzinom (KIRC) und Brust-invasives Karzinom (BRCA). Der für die Studie verwendete Benchmark-Datensatz umfasste 2086 Datensätze und 972 Attribute. Jeder Datensatz lieferte detaillierte Probeninformationen, während jedes Attribut die RNA-Seq-Werte für ein bestimmtes Gen enthielt, dargestellt als RPKM (Reads Per Kilobase per Million) [38]. Die Forscher verwendeten einen gemischten Ansatz unter Verwendung von binärem PSO mit der Entscheidungsbaummethode (BPSO-DT), um die Daten vorzuverarbeiten. Aus 971 Attributen wurden 615 als die besten Attribute von RNA-SEquation ausgewählt. Die vorgeschlagene Methode erreichte eine Gesamttestklassifizierungsgenauigkeit von 96,90 %, wie die vorgeschlagenen Ergebnisse und die verwendeten Bewertungsmaßstäbe zeigen.

Xiao et al. [39] bewerteten ihre Methodik anhand von drei RNA-Seq-Datensätzen: Magenadenokarzinom (STAD), BRCA und LUAD. Ihr Ansatz stützte sich auf DL-Techniken, wobei sie fünf Klassifikatoren verwendeten und anschließend die DL-Technik nutzten, um die Ergebnisse der fünf Klassifikatoren zusammenzufassen. Dies führte zu einer Verbesserung der Klassifizierungsgenauigkeit aller Vorhersagen, wobei der BRCA-Datensatz eine Genauigkeit von 98,4 %, STAD eine Genauigkeit von 98,78 % und LUAD eine Genauigkeit von 99,20 % erreichte.

Liu et al. [40] verwendeten Mikroarray-Daten und einen hybriden Ansatz, um das Problem zu lösen. Im Gegensatz zu den zuvor erwähnten Arbeiten untersuchten sie jede Krebsklasse separat. Zur Leistungsbewertung verwendeten sie vier Gen-Benchmarks im Zusammenhang mit kleinen runden blauen Zelltumoren , Dickdarmkrebs, Lungenkrebs und Leukämie. Ihre vorgeschlagene Methode verwendete Relief als Vorverarbeitungstechnik, um Gene mit einer geringeren Korrelation mit der spezifischen Krebsklasse auszuschließen, gefolgt von PSO als Suchtechnik. Schließlich verwendeten sie das SVM- Modell, um die Klassifizierungsgenauigkeit der ausgewählten Gen-Untergruppe zu bewerten und die endgültige optimale Gen-Untergruppe für jeden Krebstyp zu erhalten.

Basierend auf der bestehenden Forschung scheinen sich die meisten Studien mit RNA-Seq GED noch in einem frühen Stadium zu befinden, wobei die Forscher versuchen, verschiedene Ideen umzusetzen und zu

Machine Translated by Google

in diesem vielversprechenden Bereich. Obwohl die Literatur eine Fülle von Experimenten enthält, bei denen mehrere Techniken verwendet werden, wie z. B. neuere FS- und DL-Methoden, ist aufgrund der hohen Dimensionalität von RNA-Seq GED keine einzelne Technik perfekt. FS von RNA-Seq GED spielt eine entscheidende Rolle bei der Bestimmung der Beziehung zwischen einem Gen und seiner Kategorie. Es ist eine kritische Vorverarbeitungsaufgabe zur Validierung von Gen-Biomarkern für Krebs und zur Ziel dieser Studie ist es daher, einen neuen Wrapper-Ansatz, den RBAVO-DE-Algorithmus, einzuführen und ihn erstmals auf RNA-Seq-Daten anzuwenden. Außerdem wird die Wirksamkeit des vorgeschlagenen Algorithmus mit der anderer FS-Techniken verglichen.

3. RBAVO-DE-Vorschlag für GS

In diesem Artikel wird eine verbesserte Variante von AVO namens RBAVO-DE vorgeschlagen, um die kleinsten relevanten Gen-Subsets aus dem Klassifizierungsprozess zu ermitteln und irrelevante Gene zu ignorieren. RBAVO-DE verwendet einen Relief-basierten binären AVO-Algorithmus in Kombination mit der DE-Technik. Das Hauptmerkmal von RBAVO-DE ist seine Fähigkeit, die Genauigkeit zu maximieren und gleichzeitig so wenig Merkmale wie möglich zu verwenden. Das vorgeschlagene RBAVO-DE besteht aus zwei Hauptschritten . Zunächst gibt es einen Vorverarbeitungsschritt, in dem der Relief-Algorithmus verwendet wird, um zu bestimmen, welche Merkmale signifikant sind, indem jedem Merkmal eine Gewichtung zugewiesen und dann die irrelevanten Merkmale mit der niedrigsten Gewichtung entfernt werden. Im zweiten Schritt werden der binäre AVO-Algorithmus und die DE-Technik angewendet, um die relevanteren und einzigartigeren Merkmale zu identifizieren. Der AVO-Algorithmus ist bei der Bearbeitung groß angelegter Probleme anfällig für

Der vorgeschlagene RBAVO-DE-Algorithmus zur Bewältigung der GS-Strategie erfordert die Anwendung des Relief-Algorithmus, Initialisierung, Positionssteigerung mithilfe des AVO-Algorithmus, binäre Konvertierung, Fitnessbewertung und Integration mit DE. Die folgenden Unterabschnitte erläutern diese Schrit

3.1. Anwendung des Relief-Algorithmus zur Merkmalsfiltration

Dieser Schritt zielt darauf ab, die Population mit dem Relief-Algorithmus [41] **vorzuverarbeiten , der als schnelle, einfache und effiziente Filtertechnik zum Auffinden miteinander verbundener Merkmale gilt . Das Hauptziel dieses Algorithmus besteht darin, Merkmale zu finden, die Stichprobenwerte unterscheiden und ähnliche Stichproben nahe beieinander gruppieren. Daher hängt die Methode von der gewichteten Rangfolge der Merkmale ab, wobei ein Merkmal mit einer höheren Gewichtung eine bessere Klassifizierungsleistung anzeigt.**

Nach der zufälligen Auswahl einer Stichprobe untersucht der Relief-Algorithmus zwei verschiedene Arten von am nächsten liegenden Stichproben: Near-Hit-Stichproben, die mit Stichproben derselben Klasse verknüpft sind, und Near-Miss-Stichproben, die mit Stichproben anderer Klassen verknüpft sind. Die Near-Hit- und Near-Miss-Werte können verwendet werden, um die Gewichte der Merkmale zu bestimmen. Um die Wichtigkeit im Klassifizierungsprozess zu bewerten, werden die Gewichte der Merkmale von der höchsten zur niedrigsten Gewichtung angeordnet. Schließlich werden die Merkmale mit den höchsten Gewichtungen ausgewählt. Die folgende Formel kann verwendet werden, um das Gewicht W **für Merkmal A zu bestimmen:**

$$WA = \sum_{j=1}^{N} {}^{x\,j}_{A} \ddot{y} NM(xj)_A \ddot{y}^2 \; {}^{\ddot{y}\,x\,j}_{A} \ddot{y} NH(xj)_A \ddot{y}^2 . \quad (1)$$

wobei WA **das Gewicht des Merkmals A bezeichnet,** x^{J}_{A} **j bezeichnet den Wert des Merkmals** A **für Daten** x **, Stichproben bezeichnet. Die den** x- **Klassen am nächsten liegenden** j **gleich oder unterschiedlich ist und N die Anzahl der Datenpunkte sind** NH(x j) **und** NM(x j).

Der Relief-Algorithmus beschränkt sich auf die notwendigen Merkmale und minimiert der Suchbereich, um dem AVO-Algorithmus zu helfen, bessere Funktionen schneller zu finden.

3.2. Initialisierung der Population

Der vorgeschlagene BAVO-Algorithmus beginnt mit der zufälligen Generierung einer Population von N **Positionen. Ein Vektor der Dimensionen** D, **der der Anzahl der Merkmale im ursprünglichen Datensatz entspricht, beschreibt jede Position als mögliche Lösung an ihrer eingeschränkten unteren und obe**

Machine Translated by Google

Grenzen. Dieser zufällig initialisierte Schritt, der innerhalb des Bereichs **[ÿ1,** 1] **für jede Variable des Positionsvektors liegt, verwendet die zufällig erzeugte Position.**

3.3. Stärkung der Positionen durch den AVO-Algorithmus

In diesem Artikel wird der AVO-Algorithmus [30] vorgestellt, ein metaheuristischer Optimierungsalgorithmus , der von afrikanischen Geiern und ihrem Lebens- und Fressverhalten inspiriert ist. Basierend auf grundlegenden Ideen über Geier ist der AVO-Algorithmus wie folgt konfiguriert: Der AVO-Algorithmus geht zunächst davon aus, dass die afrikanische Geierpopulation aus N Geiern **besteht . Die Population afrikanischer Geier wird dann in drei Gruppen unterteilt, um die primäre natürliche Funktion der Geier basierend auf der berechneten Fitnessfunktion widerzuspiegeln. Die erste Gruppe besteht aus dem stärksten Geier, was die beste Lösung darstellt; die zweite Gruppe enthält einen Geier, der schwächer ist als der erste, was die zweitbeste Lösung darstellt; und die letzte Gruppe umfasst die verbleibenden schwächeren Geier, was die schlechtesten Lösungen darstellt.**

Basierend auf dem oben Gesagten besteht der vorgeschlagene AVO-Algorithmus aus vier grundlegenden Phasen, um das Verhalten verschiedener Geierarten zu simulieren. Die folgenden Unterabschnitte erläutern diese Phasen.

3.3.1. Phase der Bevölkerungsaufteilung

Die anfängliche Population wird durch Bewertung der Fitnessfunktion der Lösung in Gruppen aufgeteilt. Die beste Lösung wird als bester Geier in der ersten Gruppe ausgewählt, während die zweite Gruppe die zweitbeste Lösung enthält. Die dritte Gruppe enthält die verbleibenden Lösungen. Da die Lösungen ständig danach streben, sich der besten und zweitbesten Lösung anzunähern, muss die Population für jede Iteration wie folgt neu bewertet werden:

$$R_g = \begin{cases} \text{BestVulture}_1, & \text{wenn } prg = \mathbf{L1}, \\ \text{SecondBestVulture}_2, & \text{wenn } prg = \mathbf{L2}, \end{cases} \quad (2)$$

$$prg = \ddot{y} \frac{\ddot{y}_g}{\ddot{y}_{g=1}^{N} \ddot{y}_g}. \quad (3)$$

Der beste Geier in der ersten Gruppe und der zweitbeste Geier in der zweiten Gruppe Iteration werden durch die Ausdrücke BestVulture_1 **und bei** g . th $\text{SecondBestVulture}_2$ **bezeichnet , Iteration**

Die Wahrscheinlichkeit, für jede Gruppe bei g **die beste Lösung auszuwählen, wird mit dem Rouletterad-Ansatz definiert, wie in Gleichung (3),** prg gezeigt . **Zwei Zufallsparameter im Bereich [0,** 1] **sind L1 und L2.**

3.3.2. Phase der Messung des Hungerniveaus

Diese Phase wird genutzt, um ein mathematisches Modell für die Ausbeutung, Erforschung und den Transfer unter den Geiern zu formulieren. Wenn sie nicht ausgehungert sind, können die Geier besser nach Nahrung suchen und weiter fliegen. Außerdem können Geier, wenn sie ausgehungert sind, auf der Suche nach Nahrung keine langen Strecken fliegen und können gewalttätig werden. Der Hungergrad des **Geiers** (F_i^{th}) **während der** g **-Iteration kann wie folgt geschrieben werden, was zur Entwicklung eines mathematischen Modells für die Ausbeutung, Erforschung und Übertragung durch die Geier verwendet wird:**

$$F_i^G = (\mathbf{2} \times \text{rand} + 1) \times z \times \mathbf{1}\, \ddot{y} \left(\frac{g_i}{G_{max}} \right) + \mathbf{t}, \quad (4)$$

$$t = h \times \mathbf{sin}\left(\frac{\ddot{y}}{\mathbf{2}} \times \frac{g_i}{G_{max}} \right) + \mathbf{cos}\left(\frac{\ddot{y}}{\mathbf{2}} \times \frac{g_i}{G_{max}} \right) \ddot{y}\, \mathbf{1} \cdot \quad (5)$$

Die Geier sind vermutlich voll, da die Variable F **von** g **zeigt den Wechsel der Geier von Exploration bis Exploitation reicht. Eine Zufallszahl zwischen 0 und 1 wird als** rand bezeichnet. z **bezeichnet einen Zufallswert innerhalb des Intervalls [ÿ1,** 1]. **Die aktuelle Iterationsnummer wird durch** gi , **bezeichnet, während die maximale Iterationsnummer durch** Gmax bezeichnet wird. **Gleichung (5) berechnet den** t- **Wert, um komplizierte Optimierungsprobleme effektiver zu lösen und das Erreichen eines lokalen Optimums zu verhindern. Ein beliebiger Wert innerhalb des Intervalls [ÿ2,** 2] **wird durch h dargestellt. Die Wahrscheinlichkeit, den Explorationsprozess durchzuführen, wird durch die vordefinierte gerege**

Machine Translated by Google

konstanter Parameter w; die Explorationswahrscheinlichkeit steigt mit steigendem Wert. Die Exploration ist weniger wahrscheinlich, wenn der Wert sinkt.

Gleichung (4) besagt, dass mit zunehmender Anzahl von Iterationen F **ein** $|F_{\vec{}}^{G}|$ **allmählich abnimmt. Ergebnis, der nächste Schritt im vorgeschlagenen AVO-Algorithmus kann wie folgt definiert werden:**

$$\begin{cases} \text{Explorationsphase (Nahrungssuche in unterschiedlichen Gebieten),} & \text{wenn } |F_{\vec{}}^{G}| \geq 1, \\ \text{Ausbeutungsphase (Nahrungssuche im umliegenden Gebiet),} & \text{wenn } |F_{\vec{}}^{G}| < 1. \end{cases} \quad (6)$$

3.3.3. Phase der Erkundung Die

Geier zeichnen sich durch ihre große optische Fähigkeit aus, in dieser Phase geeignete Nahrung zu finden. In dieser Phase werden zwei verschiedene Strategien angewendet, die es den Geiern ermöglichen, über längere Zeiträume große Entfernungen zurückzulegen, um an zufälligen Orten nach Nahrung zu suchen. Diese Orte werden mithilfe einer Zufallszahl randP1 **und eines voreingestellten Parameters P1 ausgewählt,** die **beide** Werte im Bereich **[0,** 1] **haben . In der Erkundungsphase ist der Hungergrad** |F **größer oder gleich 1. Die Erkundungstechniken** $|\vec{}^{G}|$ **ist werden wie folgt beschrieben:**

$$X_{ich}^{g+1} = \begin{cases} R_g \ddot{y} D_{gg} \times F_i, & \text{wenn randP1} \ddot{y} \\ R_g \ddot{y} F_i & \text{wenn } |F_{\vec{}}^{G}| \ddot{y} 1, \\ \text{P1 , + rand} \times ((UB \ddot{y} LB) \times \text{rand + LB}), & \text{wenn randP1 > P1 ,} \end{cases} \quad (7)$$

$$D_{\vec{}g} = |(2 \times \text{rand}) \times R_g \ddot{y} X_{\vec{}}^{G}|. \quad (8)$$

g+1 , **wobei die nächste aktualisierte Position bei der nächsten** (g + 1) th Iteration mit X **bezeichnet wird. Der** $_{\vec{}}$ **beste Geier, der für die aktuelle Iteration** g ausgewählt wurde, wird mit R g **bezeichnet. Dies wird mithilfe von Gleichung (2) bestimmt. Die Geier bewegen sich zufällig, um die Nahrung vor anderen Geiern zu schützen und ein hohes Maß an Zufälligkeit in ihrem Suchverhalten zu gewährleisten.** rand **ist ein zufälliger Wert zwischen null und eins. Die Obergrenze der Variablen wird mit** UB bezeichnet, **ihre Untergrenze mit** LB; **und die aktuelle Position bei** g Die g- **te Iteration ist** X i .

3.3.4. Phase der Nutzung

In dieser Phase $|F_{\vec{}}^{G}|$ **ist kleiner als 1. Die Wirksamkeit des vorgeschlagenen AVO-Algorithmus wird in zwei Unterphasen bewertet, die die Ausnutzungsphase bilden. Jede dieser Unterphasen verwendet zwei verschiedene Strategien. Für jede interne Unterphase wird die geeignete Strategie mithilfe von zwei vordefinierten Parametern ausgewählt: P2 für die erste Unterphase und P3 für die zweite, mit Werten zwischen 0 und 1. Diese beiden internen Unter**

- **Erste Unterphase der Nutzung: Diese Unterphase wendet zwei verschiedene Strategien wenn** |F $_{\vec{}}^{G}$ **an | ist kleiner als 1 und größer oder gleich 0,5. Die Auswahl einer dieser beiden Strategien erfolgt mithilfe eines Zufallswerts** randP2 **innerhalb des Bereichs [0,** 1] **und eines angegebenen Parameters P2.**

  Die anfängliche Strategie dieser Unterphase wird Belagerungskampf genannt und umfasst ausreichend starke und einigermaßen gesättigte Geier. Robustere und gesündere Geier versuchen, Nahrung nicht mit anderen zu teilen, da sie sich um eine einzige Nahrungsquelle herum versammeln. Indem sie in der Nähe der gesunden Geier schwärmen und sich in kleine Kämpfe verwickeln, versuchen die schwächeren Geier, ihnen Nahrung wegzunehmen. Die zweite Strategie wird umgekehrt Rotationsflug genannt; sie erzeugt eine Spiralbewegung zwischen einem überlegenen Geier und den anderen. Im Folgenden werden die Strategien für die erste Ausbeutungsunterphase veranschaulicht:

$$X_{ich}^{g+1} = \begin{cases} D_{gi} \times (F \quad + \text{rand}) \ddot{y} d_g \quad _T, & \text{wenn randP2} \ddot{y} \text{ P2,} \\ R_g \ddot{y} (S_{\text{Auftritt 1}} + S_2^{G}), & \text{wenn randP2 > P2,} \end{cases} \quad \text{wenn } 1 > |F_{\vec{}}^{G}| \ddot{y} 0{,}5, \quad (9)$$

$$d_g = R_g \ddot{y} X_{ich,}^{G} \quad (10)$$

$$S^{G} = R_g \times 1 \quad \frac{\text{rand} \times X_i^{g}}{2\ddot{y}} \times \cos(X_{\vec{}}^{G}), \quad (11)$$

Machine Translated by Google

$$S_2^G = R\,g \times \frac{\text{rand} \times X^g}{2\ddot{y}} \times \text{Sünde}(X_1^G). \quad (12)$$

wobei die nächste aktualisierte Position des Geiers beim nächsten (g + 1) Die **Iteration wird bezeichnet durch** X^{g+1} **mit Gleichung (8)** definiert wird **und** rand **ein beliebiger Wert zwischen 0 und 1 ist. Die Distanz** d_T^g **zwischen dem Geier und einem der beiden besten Geier wird mit Gleichung (10) geschätzt;** R g **bezeichnet den jeweils besten Geier in der berechneten Position bezeichnet die aktuelle Position Strom** g th Iteration, die mit Gleichung **(2) berechnet wird ;** S_1^G **und** S_2^g **unter Verwendung der Gleichungen (11) bzw. (12) ; und** X^G **am** g Iteration .

- **Zweite Unterphase der Nutzung: Diese Unterphase wird durchgeführt, wenn** $|F^{gi}|$ **ist kleiner als 0,5. Verschiedene Geierarten versammeln sich um die Nahrungsquelle und liefern sich während dieser Unterphase zahlreiche Belagerungen und Kämpfe. Diese Unterphase verwendet zwei verschiedene Strategien. Ein vordefinierter Parameter P3 und ein beliebiger Wert** randP3 **mit Werten zwischen 0 und 1 werden verwendet, um zu entscheiden, welche dieser beiden Strategien verwendet werden soll Die anfängliche Strategie dieser Unterphase besteht darin, Geier um die Nahrungsquelle herum zu versammeln.**

Bei dieser Strategie suchen verschiedene Geierarten nach Nahrung und können in der Nähe einer einzigen Quelle miteinander konkurrieren. Die zweite Strategie ist als feindlicher Belagerungskampf bekannt. Bei dieser Strategie werden die Geier aggressiver und versuchen, den gesunden Geiern ihre Nahrungsreste zu stehlen, indem sie sie auf unterschiedliche Weise umkreisen. Die gesunden Geier hingegen verfallen und sind nicht in der Lage, die anderen Geier abzuwehren. Im Folgenden werden die Strategien für die zweite Ausbeutungsunterphas

$$X_{ich}^{g+1} = \begin{cases} \frac{\text{Ein}_1^g + A_2^G}{2}, & \text{wenn randP3} \ddot{y} \textbf{ P3,} \\ R\,g \ddot{y} |d_T^G| \times F^G \times \text{Levyd}, & \text{randP3} > \textbf{P3,} \end{cases} \quad \text{wenn } |F^G| < \textbf{0,5,} \quad (13)$$

$$\text{Ein}_1^g = \text{BestVultureg}_1 \frac{\text{1 von 1}}{\text{BestVultureg}_1 \ddot{y} (X^g)^2} \times X\,ig^g \times F_{ich}^g, \quad (14)$$

$$\text{Ein}_2^g = \text{ZweitbesterGeier}_2 \frac{\text{ZweitbesterVultureg } \mathbf{2}\times X^G}{\text{ZweitbesterGeierg } 2\ddot{y}(X^G)^2} \times F_{ich}^g, \quad (15)$$

$$\text{Levyd} = \mathbf{0{,}01} \times \frac{\mu \times \ddot{y}}{|\ddot{y}|^{\frac{1}{\ddot{y}}}}, \quad (16)$$

$$\text{Sigma} = \left(\frac{\ddot{y}(1+\ddot{y}) \times \sin\left(\frac{\ddot{y}\ddot{y}\,2}{2}\right)}{\ddot{y}(1+\ddot{y}\mathbf{2}) \times \ddot{y} \times \mathbf{2}^{\left(\frac{\ddot{y}\ddot{y}1)\,2}{2}\right)}} \right)^{\frac{1}{\ddot{y}}}. \quad (17)$$

wobei die nächste aktualisierte Position des Geiers am nächsten (g + 1) g+1 A g **und** A g , **der Versammlung** th **Iteration, Reflexion** wird, werden unter Verwendung **der Gleichungen (14) bzw.** (15) **ausgewertet . Um die Effektivität des AVO-Algorithmus zu erhöhen,** ist Levyd **die Flugverteilungsfunktion, die unter Verwendung von Gleichung (16) erhalten wird.** d **ist der Dimensionsraum;** $\ddot{y}$ **wird durch Gleichung (17) definiert, wobei** $\ddot{y}$ = **1,5 ein konstanter Wert ist; und** μ **und** $\ddot{y}$ **Zufallszahlen sind, die gleichmäßig innerhalb des Bereichs [0,** 1] verteilt sind.

3.4. Konvertierung in die binäre Natur

Im vorgestellten AVO-Algorithmus werden die Positionen als reale Werte angezeigt. Daher sind sie nicht unmittelbar auf das binäre GS-Problem anwendbar. Daher müssen diese realen Positionswerte in binäre Werte umgewandelt werden, um der binären Natur von GS zu entsprechen und die Struktur des ursprünglichen Algorithmus beizubehalten. Bei dieser Umwandlung stellen 1en die realen Werte der relevanten ausgewählten Gene im Binärisierungsvektor dar, während 0en die realen Werte der nicht ausgewählten Gene darstellen, die nicht relevant sind. Bei jeder Iteration g kann der folgende mathematische Ausdruck verwendet werden, um die reale Position X^G **in eine binäre Position** X_{bin}^G umzuwandeln (X

Machine Translated by Google

$$(X^{G}_{..})bin = \begin{cases} \textbf{1 wenn } X^{G}_{..} > \ddot{y}, \\ \textbf{andernfalls } i\ 0. \end{cases} \quad \textbf{(18)}$$

wobei ein beliebiger Schwellenwert innerhalb von [0, 1] **durch ÿ dargestellt wird . Gemäß diesem grundlegenden Ansatz zur binären Konvertierung ersetzt die binäre „1“ (ausgewähltes Merkmal) ihren realen Wert, wenn $X^{G}_{..}$ größer als ÿ ist. Andererseits, wenn $X^{G}_{..}$ kleiner ist, weist sie den Wert auf die binäre „0“ (ein Merkmal, das nicht ausgewählt wurde) gesetzt.**

3.5. Beurteilung des Fitnessfunktionswertes

Das Finden der geringsten Anzahl ausgewählter Merkmale und das Optimieren der Klassifizierungsgenauigkeit der verfügbaren Klassifikatoren (k-NN- und SVM-Modelle) sind zwei widersprüchliche Ziele , die ausgeglichen werden müssen, um die beste Lösung zu erreichen und ihre Qualität zu bestimmen. Da die Genauigkeit der k-NN- und SVM-Klassifikatoren beeinträchtigt werden kann, wenn die Anzahl ausgewählter Merkmale geringer als optimal ist, gleicht die Fitnessfunktion die Größe und Genauigkeit der ausgewählten Merkmale aus . Die Fitnessfunktion konzentriert sich auf die Senkung der Klassifizierungsfehlerrate und nicht auf die Genauigkeit, und zwar wie fo

$$\text{f it} = \textbf{w1} \times \text{Errrate} + \textbf{w2} \times |D| \frac{|\text{f eatpicked}|}{\textbf{w1 ÿ [0,}}, \quad \textbf{(19)}$$

1], **w2** = **1 ÿ w1.**

Dabei bezeichnet f eatpicked **die Anzahl der ausgewählten Merkmale,** D **die Gesamtzahl der Merkmale im Datensatz und** Errrate **die Klassifizierungsfehlerrate der k-NN- und SVM-Klassifikatoren. Die Wichtigkeit der Klassifizierungsgenauigkeit und die Anzahl der ausgewählten Merkmale werden durch die Gewichtungsparameter w1 bzw. w2** angegeben . **Die Werte von w1 (0,99) und w2 (0,01) wurden durch umfangreiche Versuche ermittelt, die in früheren Studien durchgeführt wurden [21,42,43].**

3.6. Einbeziehung der DE-Technik **DE [44]**

zeichnet sich durch seine hohe Effizienz und Einfachheit bei der Suche nach einer geeigneten Lösung für komplexe Optimierungsprobleme aus. Es können schnell wertschöpfende Ergebnisse erzielt werden. Drei Hauptprozesse – Mutation, Crossover und Selektion – sind für DE notwendig. Der Prozess der differentiellen Mutation zielt darauf ab, für **den Lösungsvektor jeder Iteration** einen modifizierten Vektor ÿi zu erzeugen. Dies kann mathematisch wie folgt berechnet werden:

$$ÿÿÿ_{..} = \textbf{ÿÿX}_{\textbf{r1}} + \text{WM}(ÿÿX_{\textbf{r2}} \textbf{ ÿ ÿÿX}_{\textbf{r3}}) \quad \textbf{(20)}$$

Innerhalb des Bereichs [1, Populationsgröße] werden drei zufällig ausgewählte asymmetrische Vektoren durch die Symbole Xr1 , Xr2 **und** Xr3 gekennzeichnet . WM **bezeichnet den Mutationsgewichtungsfaktor innerhalb des Intervalls [0,** 1].

Nach der Mutationsoperation führt DE eine Crossover-Operation durch, um die Vielfalt der Population zu erhöhen. Ein Nachkommenvektor ui **wird durch die Kombination von Werten aus dem Zielvektor** Xi **und dem geänderten Vektor** ÿi erzeugt . **Der am häufigsten verwendete und grundlegende Faktor der Crossover-Suche ist das binäre Crossover, das den folgenden mathematischen Ausdruck hat:**

$$\textbf{ui,d} = \begin{cases} \textbf{ÿi,d ,} & \textbf{wenn } \text{rand } \textbf{ÿ } \text{CR oder d} = \text{jrand}, \\ \textbf{Xi,d ,} & \textbf{andernfalls.} \end{cases} \quad \textbf{(21)}$$

wobei eine gleichmäßig verteilte Zufallszahl jrand **ÿ [1, 2, ...,** DX] **verwendet wird, um sicherzustellen, dass der mutierte Vektor mindestens eine Dimension hat. Die Wahrscheinlichkeit der Kreuzung jedes Elements wird mithilfe der Kreuzungsrate** CR berechnet, **die häufig als großer Wert bestimmt wird** (CR **= 0,9).**

Anschließend wird die Auswahloperation wie in Gleichung (22) gezeigt durchgeführt. Dabei wird ein Vergleich zwischen der Fitnessfunktion f(Xi) **des Zielvektors und dem entsprechenden Nachkommenvektor** f(ui) **durchgeführt und die Fitnessfunktion mit dem niedrigsten Wert wird für die nächste Iteration beibehalten.**

Machine Translated by Google

$$Xi = \begin{cases} \overline{Xi}, & \textbf{wenn } f(ui) < f(Xi) \\ Xi, & \textbf{ansonsten.} \end{cases} \quad \textbf{(22)}$$

3.7. Der vollständige RBAVO-DE-Algorithmus

Zur Handhabung der GS-Strategie werden in den folgenden Unterabschnitten die Schritte des empfohlenen RBAVO-DE-Algorithmus beschrieben. Der Pseudocode für den vorgeschlagenen RBAVO-DE- Algorithmus ist in Algorithmus 1 enthalten . In Abbildung 1 ist **ein Flussdiagramm des vorgeschlagenen RBAVO-DE-Algorithmus dargestellt , das die wichtigsten Schritte veranschaulicht.**

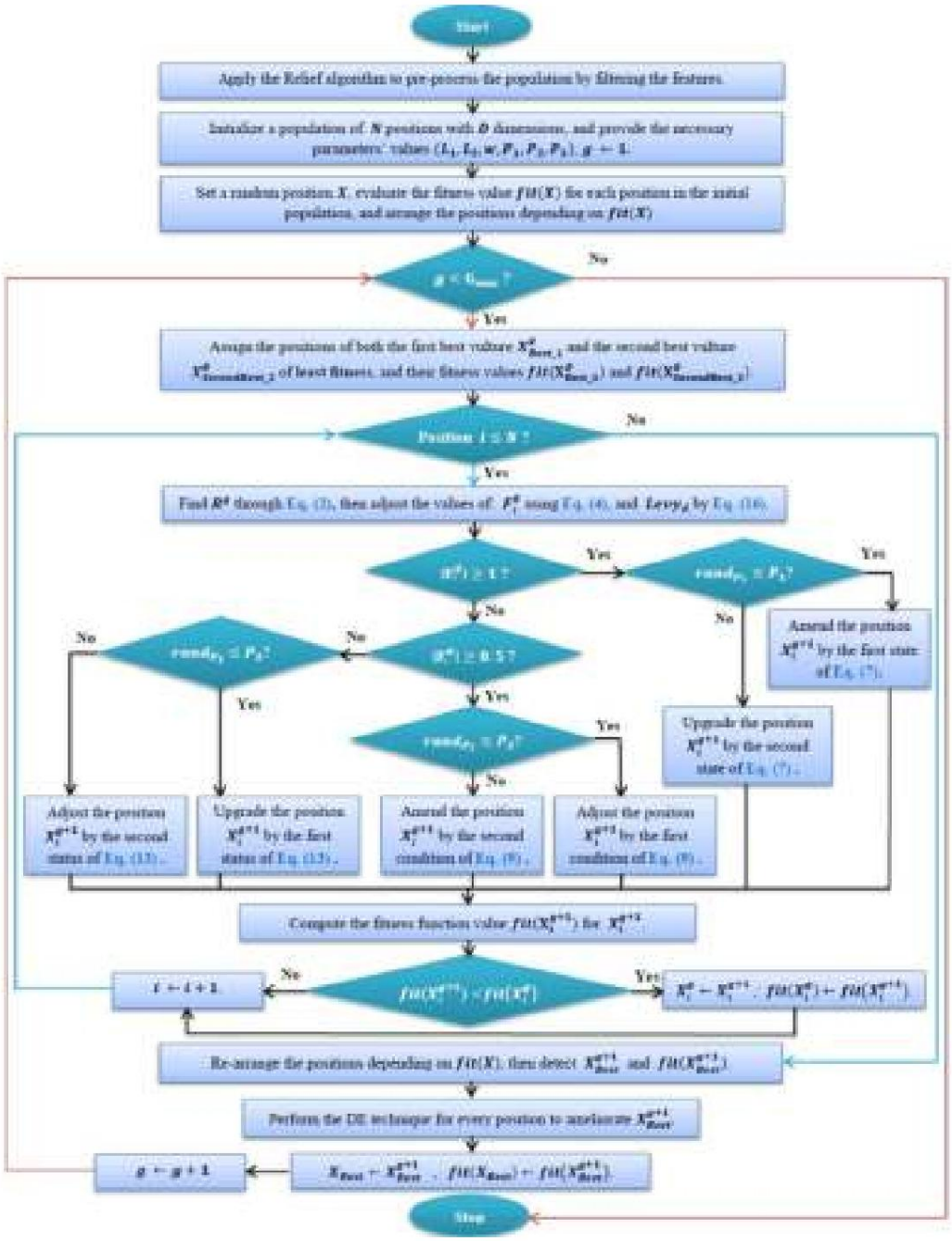


Abbildung 1. Flussdiagramm des RBAVO-DE-Algorithmus.

Machine Translated by Google

Algorithmus 1 Der vorgeschlagene RBAVO-DE-Algorithmus.

Eingang:

N – Gesamtzahl der Stellen (Größe der Bevölkerung)
Gmax – Maximalzahl der erlaubten Iterationen
D – dimensionaler Raum des Problems
LB – Untergrenzen von Variablen
UB – Obergrenzen von Variablen
CR – Übergangsrate
WM – Mutationsgewichtungsfaktor

Ausgabe:

XBest – die **weltweit beste Position des Geiers, die bei der Suche gefunden wurde**
f it(XBest) – der **global beste gefundene Fitnessfunktionswert, der verringert werden sollte**

1: Starten
2: Wenden Sie den Relief-Ansatz zum Auswählen und Filtern der zugehörigen Features an, wie in Abschnitt 3.1 gezeigt.
3: Initialisieren Sie eine Population von N **Positionen und geben Sie die Werte der erforderlichen Parameter an (L1, L2, w, P1, P2 und P3).**
4: Legen Sie eine zufällige Position X **in der Anfangspopulation fest;**
5: Bewerten Sie die Fitnesswerte fit (X) **für jede Position in der Anfangspopulation.**
6: Ordne die Positionen in aufsteigender Reihenfolge entsprechend ihrer Fitnessfunktion fit (X);
7: g ÿ **1; 8:** ÿ **Aktuelle Anzahl Iterationen**
solange g < Gmax **do**
9: **Weisen Sie die Positionen der beiden erstbesten Geier** X zu X^{G}_{Beste1} **und der zweitbeste Geier** X $X^{G}_{Zweitbeste2}$, **sowie ihre Fitness Werte** für (X^{G}_{Beste1}) **und** fit ($X^{G}_{Zweitbeste2}$), **unter allen Positionen in der Bevölkerung;**
10: **für die Position des Geiers** i = **1 :** N **do**
11: **Ermitteln Sie den besten Geier** R g **mithilfe von Gleichung (2).**
12: **Passen Sie die Hungersnot des Geiers** an F^{G}_{-} **unter Verwendung von Gleichung (4);**
13: **Modifizieren Sie die Levy-Flug-Verteilungsfunktion** Levyd **mithilfe von Gleichung (16).**
14: **wenn** $|F^{G}_{-}|$ **ÿ 1 dann**
15: **wenn** randP1 **ÿ P1 dann**
16: **Ändere die Position des Geiers** X g+1 **basierend auf der ersten Stufe der Explorationsphase unter Verwendung von Gleichung (7);**
17: **sonst wenn** randP1 > **P1 dann**
18: **Erhöhe die Position des Geiers** um x g+1 **basierend auf der zweiten Stufe der Explorationsphase unter Verwendung von Gleichung (7);**
19: **Ende, wenn**
20: **sonst wenn** $|F^{G}_{-}|$ **< 1 dann**
21: **wenn** $|F^{G}_{-}|$ **ÿ 0,5 dann**
22: **wenn** randP2 **ÿ P2 dann**
23: **Passe die Position des Geiers** an X g+1 **basierend auf der ersten Bedingung der ersten Unterphase der Nutzung unter Verwendung Gleichung (9);**
24: **sonst wenn** randP2 > **P2 dann**
25: **Ändere die Position des Geiers** X $^{g+1}_{-}$ **basierend auf der zweiten Bedingung der ersten Unterphase der Nutzung unter Verwendung von Gleichung (9);**
26: Ende, wenn
27: **sonst wenn** $|F^{G}_{-}|$ **< 0,5 dann**
28: **wenn** randP3 **ÿ P3 dann**
29: **Erhöhe die Position des Geiers** um x g+1 **basierend auf dem ersten Status der zweiten Unterphase der Nutzung unter Verwendung Gleichung (13);**
30: **sonst wenn** randP3 > **P3 dann**
31: **Passe die Position des Geiers** an X g+1 **basierend auf dem zweiten Status der zweiten Unterphase der Nutzung unter Verwendung Gleichung (13);**
32: **Ende, wenn**
33: **Ende, wenn**
34: **Ende, wenn**
35: f es(X^{g+1}_{-}) **ÿ Berechnen Sie den Fitnessfunktionswert für** X $^{g+1}_{-}$;
36: **wenn** f es(X^{g+1}_{-}) < passen (X^{G}_{-}) **Dann**
37: X^{G}_{-} **ÿ** X $^{g+1}_{-}$;
38: f es(X^{G}_{-}) **ÿ passen Sie es an** (X^{g+1}_{-});
39: **Ende, wenn**
40: **Ende für**
41: **Ordnen Sie die Positionen in aufsteigender Reihenfolge neu an, abhängig von ihrer Fitnessfunktion** fit (X).
42: **Ermitteln der global besten Position** X $^{g+1}_{Am\ besten}$ **und sein weltweit bester Fitnesswert** für (X $^{g+1}_{Am\ besten}$), **wenn die aktuelle Iteration** g + **1 beendet ist;**
43: **Führen Sie die DE-Technik für jede Position durch, um** X zu verbessern $^{g+1}_{Am\ besten}$, **wie in Abschnitt 3.6 gezeigt;**
44: XBeste **ÿ** X $g+1_{Am\ besten}$;
45: passen (XBest) **ÿ** passen (X $^{g+1}_{Am\ besten}$);
46: g **ÿ** g + **1;**
47: Ende während
48: Ende

4. Experimentelle Ergebnisse und Diskussion

In diesem Abschnitt werden die empirischen Ergebnisse der vorgeschlagenen RBAVO-DE und ihrer Gegenstücke: Binary Artificial Bee Colony (BABC) [45], Binary Salp Swarm Algorithm (BSSA) [46], Binäres PSO (BPSO) [47], Binärer Fledermausalgorithmus (BBA) [48], Binärer Grauwolf Optimierung (BGWO) [49], Binärer Grasshopper-Optimierungsalgorithmus (BGOA) [50], Binärer Whale Optimization Algorithm (BWOA) [51], Binary ASO (BASO) [52], Binary Bird Schwarmalgorithmus (BBSA) [53], Binärer HGSO (BHGSO) [54] und Binärer Harris Hawks Op-

Machine Translated by Google

timization (BHHO) [55]. Die Optimierer werden durch Training und Tests evaluiert.
Benchmarks, wobei schlüssige Ergebnisse aus den Durchschnittswerten der Bewertungsmetriken
abgeleitet werden . Die Benchmarks, die zur Beurteilung der Leistung des vorgeschlagenen Modells verwendet w
ausführlich in Abschnitt 4.1. Die in den Betriebsumgebungen verwendeten Parameter sind
in Abschnitt 4.2. Die zur Bewertung verwendeten Metriken werden in Abschnitt 4.3 beschrieben. Die Analyse von
Die experimentellen Ergebnisse werden in Abschnitt 4.4 erörtert.

4.1. Beschreibung des Datensatzes

Es wurden tiefgreifende experimentelle Ansätze und verschiedene Wrapper-Algorithmen angewendet, um
22 Datensätze mit Genbeschreibungen; die Daten umfassen normalisierte Level 3 RNA-Seq
Genexpressionsdaten für 22 Tumorarten vom Broad Institute. Diese Daten
sind öffentlich zugänglich und können in [56] gefunden werden. Wir haben uns an die beschriebene Methodik gehalten
in [36] und beobachtete Diskrepanzen zwischen den im Papier referenzierten Daten von GitHub
und die im Dokument angegebenen Zahlen, die von der Website stammen. Die
Die Daten der Website enthielten eine Mischung aus Tumor- und Standardproben, während das Papier behandelte
die Daten einheitlich als Tumorproben. Daher führten wir eine detaillierte Untersuchung durch
der Daten. Zunächst bot die Website verschiedene Formate des gleichen Datensatzes an, den wir
zu analysieren. Beim Eintauchen in die Daten stießen wir auf die folgenden Herausforderungen:

- Bestimmte Gene werden durch eine ID identifiziert, es fehlt aber ein zugehöriges Symbol.
- Einige Gene fehlen in der Annotation-Datei.
- Es kommt zu einer Vermischung der Proben, darunter sowohl normale Proben als auch Tumorproben.

Folglich war eine Vorverarbeitung notwendig, um die Proben zu trennen und zu
identifizieren, zwischen normalen und Tumorproben für die binäre Klassifizierung zu unterscheide
den GS-Prozess. Wir haben die oben genannten Herausforderungen folgendermaßen angegangen:

- Wir haben für jede in der Annotation-Datei gefundene ID das entsprechende Gensymbol nachgeschlagen.
- Nach dem Querverweis mit der Anmerkungsdatei wurden über einhundert Gene eliminiert.
- Basierend auf dem Beispielbericht wurde jede Zeile in der Excel-Tabelle nach den
  Art der Probe für die binäre Klassifizierung.

Darüber hinaus wurde der Relief-Ansatz, wie in Abschnitt 3.1 beschrieben, in der
Vorverarbeitungsphase verwendet , um das Gewicht jedes Gens innerhalb des Benchmarks zu berechnen.
wurden anschließend von der höchsten zur niedrigsten geordnet. Gene mit niedrigerem Gewicht wurden dann
entfernt. Der Relief-Ansatz war nützlich, um Gene zu verwerfen, die keinen Beitrag leisten
zur Klassifizierung.

Nach der Vorverarbeitungsphase wurden die Daten verfeinert und waren bereit für den GS-Prozess.
Im Gegensatz zu dem in [36] verwendeten Ansatz , der die Mehrfachklassifizierung aller Arten von
Da wir die Krebsarten nicht alle gemeinsam betrachten, haben wir uns dazu entschieden, jede Krebsart einzeln zu analysieren, um eine höhere Spezifität zu erreichen.
Tabelle 1 [57] bietet einen umfassenden Überblick über alle 22 Tumortypen und ihre
jeweilige Stichprobenanzahl. Die Anzahl der in den Benchmark-Datensätzen verwendeten Merkmale ist
32.000 Merkmale (Gene).

Tabelle 1. In diesem Papier verwendete Datensätze [57].

# Krebs	Normal Aufzeichnungen	Tumor Aufzeichnungen
1 Urothelkarzinom der Harnblase (BLCA)	19	408
2 Schilddrüsenkrebs (THCA)	59	501
3 Gebärmutterhalskrebs und endozervikaler Krebs (CESCs)	3	304
4 Cholangiokarzinom (CHOL)	9	36
5 Kolon-Adenokarzinom (COAD)	41	458
6 Speiseröhrenkrebs (ESCA)	13	184
7 Glioblastoma multiforme (GBM)	5	153

Machine Translated by Google

Tabelle 1. Fortsetzung

# Krebs	Normal Aufzeichnungen	Tumor Aufzeichnungen
8 Thymom (THYM)	2	120
9 Plattenepithelkarzinom im Kopf-Hals-Bereich (HNSC)	44	520
10 Nierenchromophobiker (KICH)	25	66
11 KIRC	72	533
12 Nierenpapilläre (KIRP)	32	290
13 Leber-Hepatozelluläres Karzinom (LIHC)	50	371
14 LUAD	59	515
15 LUSC	51	501
16 Pankreasadenokarzinom (PAAV)	4	178
17 UCEC	35	176
18 Phäochromozytom und Paragangliom (PCPG)	3	179
19 Rektumadenokarzinom (READ)	10	94
20 Sarkom (SARC)	2	259
21 Haut-kutanes Melanom (SKCM)	1	103
22 STAD	37	415

4.2. Parametereinstellung

Der vorgeschlagene RBAVO-DE-Algorithmus wurde mit binären Varianten verschiedener metaheuristische Optimierer wie BABC, BSSA, BPSO, BBA, BGWO, BWOA, BBSA, BGOA, BHHO, BASO und BHGSO. Die kritischen Parameter für die in diesem Die Ergebnisse der Studie sind in Tabelle 2 aufgeführt.

Tabelle 2. Die wichtigsten Parameter der verwendeten ML-Modelle.

Modell	Parameter
SVM	Polynomkernel = 2
k-NN	Euklidische Distanzmetrik k = 5 [21,58,59]

Innerhalb des vorgeschlagenen Rahmens wurde die Gültigkeit der Ergebnisse mithilfe einer 10-fachen Kreuzvalidierungsansatz, um die Zuverlässigkeit der Ergebnisse sicherzustellen. Dabei wurden zufällig Aufteilung jedes Datensatzes in zwei unterschiedliche Teilmengen: 80 % für das Training und die restlichen 20 % zum Testen. Der Trainingsteil wurde verwendet, um die vorgestellten Klassifikatoren zu trainieren mit Optimierungstechniken, während der Testteil dazu diente, die ausgewählten Gene zu bewerten. Wirksamkeit. Die am häufigsten bei allen verglichenen Algorithmen verwendeten Parameter sind zusammengefasst in Tabelle 3. Zur Durchführung aller Experimente in diesem Artikel wurde Python in einem Rechenumgebung mit einer Dual Intel® Xeon® Gold 5115 2,4 GHz CPU und 128 GB RAM auf dem Betriebssystem Microsoft Windows Server 2019.

Tabelle 3. Parameterkonfigurationen aller verwendeten Optimierer.

Optimierer	Parameter
Alle Optimierer	Ausführungsanzahl = 30
	Maximale Anzahl Iterationen Gmax = 100
	Bevölkerungsgröße N = 10

Machine Translated by Google

Tabelle 3. Fortsetzung

Optimierer	Parameter
Vorschlag für die RBAVO-DE	**Dimensionalität** D **= Anzahl der Gene in den verwendeten Datensätzen** **Obere Grenzen** UB **Untere Grenzen** LB **Parameter L1** = **0,7 Parameter L2** = **0,2 Parameter** w = **2**
BSSA	**Parameter P1** = **0,6** **Parameter P2** = **0,6** **Parameter P3** = **0,5** **Anzahl der Schmarotzer** SD = **0,1 ×** N **Anzahl der Produzenten** PD = **0,2 ×** N **Sicherheitsschwelle** ST = **0,8**
BABC	**Anzahl der angestellten Bienen** = **16 Anzahl der Kundschafterbienen** = **3 Anzahl der Zuschauerbienen** = **4 Trägheitsgewicht** (ÿmax = *0,9ÿmin* = **0,4)**
BPSO	**Beschleunigungskoeffizienten** (c2 = **c1** = **1,2)**
BBA	**Lautstärke** A = **0,8** **Impulsemissionsrate** r = **0,95**
BWOA	**Untere und obere Impulsfrequenz** = **0, 10** a **wird linear von 2 auf 0 verringert** p = **0,5** b = **1,0**
BHHO	**Kaninchenenergie** E **ÿ [ÿ1,** 1]
BGWO	a **wird linear von 2 auf 0 verringert Cmin**
BGOA	= **0,00004 und Cmax** = **1**
BBSA	**Flugfrequenz** ff = **10 Auswirkung auf die Wachsamkeit der Vögel Folgekoeffizient** fl **= 0,5** **Wahrscheinlichkeit der Nahrungssuche** p = **0,8** **Beschleunigungskoeffizienten** (c1 = **c2** = **1,5)**
BASO	**Verhalten** (a1 = **a2** = **1,0)** **Tiefengewicht** *ÿ* = **50**
BHGSO	**Multiplikatorgewicht** *ÿ* = **0,2** **Anzahl der Cluster** = **2** *ÿ* = *ÿ* = **0,1** K = **1,0, und l3** = 1E **ÿ 02 l1** = 5E **ÿ 03, l2** = 1E + **02**

4.3. Bewertungskriterien

Um die Wirksamkeit der vorgeschlagenen RBAVO-DE im Vergleich zu anderen Methoden zu bewerten, wurde jede Strategie unabhängig voneinander dreißig Mal anhand jedes Benchmarks getestet, um eine statistische Validierung der Ergebnisse sicherzustellen. Zu diesem Zweck wurden die folgenden etablierten Leistungsmetriken für die GS-Probleme verwendet:

- **Mittlere Klassifizierungsgenauigkeit** (MeanAC): **Dieses Maß stellt die Genauigkeit der korrekten Klassifizierung von Daten dar, berechnet durch dreißig unabhängige Ausführungen des Algorithmus . Es wird folgendermaßen ermittelt:**

$$\text{MittlererAC} = \frac{1}{30}\sum_{k=1}^{30}\frac{1}{M}\sum_{r=1}^{M} \text{Übereinstimmung}(PLr, ALr) \quad \textbf{(23)}$$

In dieser Formel bezeichnet m **die Anzahl der Proben im Testdatensatz, während** PLr **und** ALr **die vom Klassifikator vorhergesagten Beschriftungen bzw. die tatsächlichen Klassenbeschriftungen für Probe r** bezeichnen . **Die Funktion** match(PLr , ALr) **dient als Vergleichsmechanismus. Wenn** PLr **gleich** ALr ist , wird match(PLr , ALr) **der Wert 1 zugewiesen. Wenn sie nicht übereinstimmen, wird der Wert 0 zugewiesen.**

Machine Translated by Google

• **Mittlere Fitness** (MeanFi): **Dieses Maß berechnet den durchschnittlichen** MeanFi, **der erreicht wird, indem der vorgeschlagene Ansatz dreißig Mal unabhängig voneinander ausgeführt wird. Es misst das Gleichgewicht zwischen der Senkung der Klassifizierungsfehlerrate und der Minimierung der Anzahl ausgewählter Gene. Ein niedrigerer Wert weist auf eine bessere Lösung hin, die anhand der Fitness**

$$\text{MittelwertFi} = \frac{1}{30}\sum_{k=1}^{30} f_{\hat{y}}^{k} \quad (24)$$

wobei $f_{\hat{y}}^{k}$ **zeigt den optimalen Fitnesswert, der bei der** k **-** ten **Ausführung erreicht wurde.**

• **Mittlere Anzahl ausgewählter Gene** (MeanGe): **Diese Metrik berechnet die durchschnittliche Anzahl ausgewählter Gene (oder das GS-Verhältnis), indem der vorgeschlagene Ansatz dreißig Mal unabhängig voneinander ausgeführt wird, und ist wie folgt definiert:**

$$\text{MittlererFe} = \frac{1}{30}\sum_{k=1}^{30} \frac{|d_{\hat{y}}^{k}|}{|D|} \quad (25)$$

Gene . $\hat{y}^{k}$ **| bezeichnet die Gesamtzahl der in der optimalen Lösung für das** k -te ausgewählte **ex-Dabei** stellen |d **ecution und** |D| **die Gesamtzahl der im anfänglichen Benchmark vorhandenen Gene dar.**

• **Standardabweichung** (STDE): **Basierend auf den obigen Ergebnissen wurden die Gesamtdurchschnittsergebnisse aus dreißig separaten Ausführungen jedes Optimierers bei jedem Benchmark folgendermaßen auf Stabilität geprüft:**

$$\text{STDE} = \sqrt{\frac{1}{29}\sum_{k=1}^{30} (Y_{\hat{y}}^{k} - \text{MittelwertY})^2} \quad (26)$$

Dabei zeigt Y **das zu verwendende Maß,** $Y_{\hat{y}}^{k}$ **bezeichnet den Maßwert** Y **im** k-ten **Durchlauf und** MeanY **ist der Mittelwert des Maßes aus dreißig unabhängigen Ausführungen.**

Die Daten in den folgenden Tabellen stellen die Mittelwerte aus dreißig unabhängigen Ausführungen dar, wobei der Schwerpunkt auf Klassifizierungsgenauigkeit, Anzahl der ausgewählten Gene, durchschnittlicher Fitness, Präzision, Rückruf und F1-Score liegt. Die folgenden Unterabschnitte untersuchen und diskutieren diese experimentellen Ergebnisse gründlich, wobei fettgedruckte Abbildungen die

4.4. Ergebnisse des Vergleichs der vorgeschlagenen RBAVO-DE mit verschiedenen metaheuristischen Algorithmen

Der vorgeschlagene RBAVO-DE-Ansatz mit den SVM- und k-NN-Modellen wurde mit anderen aktuellen metaheuristischen Methoden verglichen, die unter denselben Bedingungen ausgeführt wurden, um seine Überlegenheit gegenüber seinen Gegenstücken zu zeigen . Der vorgeschlagene RBAVO-DE-Algorithmus wurde mit **binären Varianten mehrerer Optimierer verglichen ,** darunter BABC, BSSA, BPSO, BBA, BGWO, BWOA, **BBSA, BGOA, BHHO, BASO und BHGSO.**

4.4.1. Ergebnisse unter Verwendung des k-NN-Modells

Tabelle 4 zeigt die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und anderer Optimierungstechniken, die das k-NN-Modell verwenden, basierend auf Genauigkeitsmetriken, die unter ähnlichen Bedingungen verglichen wurden. Die experimentellen Ergebnisse zeigen, dass der vorgeschlagene RBAVO-DE in vier Benchmarks die vielversprechendsten Ergebnisse erzielte. Es ist bemerkenswert, dass alle konkurrierenden Methoden, einschließlich des vorgeschlagenen RBAVO-DE, das das k-NN-Modell verwendet , in achtzehn Benchmarks vergleichbare Ergebnisse lieferten .

Die restlichen Kennzahlen für das k-NN-Modell sind in Anhang A.1 aufgeführt. In Bezug auf die durchschnittlichen Fitnesswerte zeigte der vorgeschlagene RBAVO-DE-Algorithmus eine höhere Effizienz im Vergleich zu seinen Gegenstücken, die das k-NN-Modell unter ähnlichen Bedingungen verwenden. Der RBAVO-DE-Algorithmus lieferte die niedrigsten Fitnessergebnisse und den wettbewerbsfähigsten STDE in allen Benchmarks. Außerdem sind alle verwendeten Benchmarks hochdimensional, was zeigt, dass der vorgeschlagene RBAVO-DE auf allen Benchmarks unabhängig von der Größe effektiv ausgeführt werden kann. Der RBAVO-DE-Algorithmus ist vielversprechend, da er die Nutzung und Erkundung des Suchraums effektiv ausbalanciert und es vermeidet, während der Iterationen in lokalen Optima gefangen zu sein. Im Gegensatz zu

Machine Translated by Google

Wie viele andere Algorithmen, die dazu neigen, gefangen zu werden, zeigte es die Fähigkeit, zu entkommen solche Fallen.

In Bezug auf die mittleren Ergebnisse für die Gene, die durch den vorgeschlagenen RBAVO-DE-Algorithmus und seine Gegenstücke, die k-NN verwenden, ausgewählt wurden, übertraf der vorgeschlagene RBAVO-DE- die anderen Methoden in allen Benchmarks hinsichtlich der Anzahl der ausgewählten Gene. Darüber hinaus beruht die Fähigkeit der RBAVO-DE, signifikante Gene zu identifizieren, auf ihrer effektiven Erkundung möglicher Bereiche bei gleichzeitiger Verbesserung der Genauigkeit.

In Bezug auf die mittlere Präzision des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke mit k-NN übertraf der vorgeschlagene RBAVO-DE alternative Ansätze für drei der zweiundzwanzig Datensätze. Für neunzehn Datensätze erzielte BWOA ähnliche Ergebnisse, BABC, BPSO, BGWO, BGOA und BHHO erzielten ähnliche Ergebnisse wie die vorgeschlagene RBAVO-DE. BASO und BHGSO, die zu den gleichen Ergebnissen führten wie die vorgeschlagene RBAVO-DE auf siebzehn Datensätzen, belegte den vierten Platz. Letztlich schnitt BBA unter allen Ansätzen am schlechtesten ab und führte zu ähnlichen Ergebnissen wie die vorgeschlagene RBAVO-DE für fünfzehn Datensätze.

In Bezug auf den mittleren Rückruf des vorgeschlagenen RBAVO-DE und seiner Gegenstücke mit k -NN übertraf RBAVO-DE die alternativen Ansätze für drei der 22 Datensätze. Andererseits waren BSSA, BABC, BPSO, BGWO, BWOA, BGOA und BBSA führte zu den gleichen Ergebnissen wie RBAVO-DE für 19 Datensätze, während BHHO ähnliche Ergebnisse für 18 Datensätze erzielte. BHGSO, das die gleichen Ergebnisse erzielte wie die vorgeschlagene RBAVO-DE für siebzehn Datensätze, Platz vier. Letztlich belegten BASO und BBA den Der niedrigste Wert aller Ansätze und die Ergebnisse waren für sechzehn Datensätze mit RBAVO-DE vergleichbar.

In Bezug auf den mittleren F1-Score des vorgeschlagenen RBAVO-DE und seiner Gegenstücke Unter Verwendung von k-NN übertraf der vorgeschlagene RBAVO-DE die alternativen Ansätze für drei der zweiundzwanzig Datensätze. Im Gegensatz dazu für achtzehn Datensätze, BSSA, BABC, BPSO, BGWO, BGOA und BBSA erzielten die gleichen Ergebnisse wie RBAVO-DE, während BWOA und BHHO lieferte vergleichbare Ergebnisse für siebzehn Datensätze. BASO und BHGSO, die erzielte für 16 Datensätze die gleichen Ergebnisse wie die vorgeschlagene RBAVO-DE und belegte den vierten Platz. Letztlich schnitt BBA unter allen Ansätzen am schlechtesten ab und erzielte ähnliche Ergebnisse wie die RBAVO-DE für vierzehn Datensätze vorgeschlagen.

Tabelle 4. Die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke mit k-NN bezüglich Ergebnisse der durchschnittlichen Klassifizierungsgenauigkeit.

Datensatzmetrik	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA Durchschnitt 0,998 STDE	0,993	0,998	0,993	0,989	0,995	0,994	0,993	0,991	0,994	0,994	0,989	
0,004 Durchschnitt	0,006	0,005	0,006	0,003	0,006	0,006	0,006	0,005	0,006	0,006	0,002	
CESC	0,984 STDE 0,000	0,984	0,984	0,984	0,984	0,984	0,984	0,984	0,984	0,984	0,984	0,984
CHOL Durchschnitt	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
1,000 STDE 0,000 COAD	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
Durchschnitt 1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
STDE 0,000 ESCA Durchschnitt	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
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 Durchschnitt 1,000 STDE	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974	
0,000 HNSC	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
Durchschnitt 0,993 STDE 0,004	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
KICH Durchschnitt	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
1,000 STDE 0,000 Durchschnitt	0,991	0,991	0,991	0,991	0,992	0,991	0,991	0,991	0,991	0,991	0,991	
0,998 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	
Durchschnitt 1,000 STDE 0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
Durchschnitt 1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
KIRC	STDE 0,000 LUAD	0,993	0,994	0,992	0,992	0,994	0,992	0,994	0,993	0,993	0,993	0,992
Durchschnitt 0,991	0,003	0,004	0,002	0,000	0,004	0,002	0,004	0,003	0,003	0,003	0,000	
KIRP	STDE LUSC	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
Durchschnitt 1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
LIHC	STDE PAAD	1 .000	1 .000	1 .000	1 .000	1 .000	1 .000	1 .000	1 .000	1 .000	1 .000	1.000
Durchschnitt 1,000	0,000	0,000	0,000	0,000	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	0,991	0,991	
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	
	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	

Machine Translated by Google

Tabelle 4. Fortsetzung

Datensatzmetrik	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
PCPG Durchschnitt 1,000 STDE	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0,000 READ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0,000
Durchschnitt 1,000 STDE SARC	1.000	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	0,000
Durchschnitt 0,981 STDE	0,981	0,981	0,981	0,981	0,981	0,981	0,981	0,981	0,981	0,981	0,981	0,981
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SKCM Durchschnitt 1,000 STDE	1.000	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	0,000
STAD Durchschnitt 0,999	0,996	0,999	0,996	0,989	0,996	0,995	0,997	0,992	0,995	0,996	0,989	
STDE 0,003 THCA	0,005	0,004	0,005	0,000	0,005	0,006	0,005	0,005	0,006	0,005	0,000	
Durchschnitt 1,000 STDE 0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
THYM Durchschnitt	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
1,000 STDE 0,000 UCEC	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
Durchschnitt 0,998	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
STDE 0,008	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	
Rangliste S|T|N	4|18|0	0|18|4	0|18|4	0|18|4	Fettgedruckte	0|18|4	0|18|4	0|18|4	0|18|4	0|18|4	0|18|4	0|18|4

Werte kennzeichnen die besten Ergebnisse.

4.4.2. Ergebnisse unter Verwendung des SVM-Modells

Tabelle 5 zeigt die Ergebnisse der vorgeschlagenen RBAVO-DE und verschiedene Optimierungsmaßnahmen. Techniken, die das SVM-Modell hinsichtlich der Klassifizierungsgenauigkeit verwenden, bewerteten Ergebnisse unter identischen Betriebsbedingungen. Die experimentellen Ergebnisse zeigen, dass die Der vorgeschlagene RBAVO-DE-Algorithmus übertraf die anderen Ansätze, indem er die meisten vielversprechende Werte für vier Datensätze. Alle konkurrierenden Methoden, einschließlich der vorgeschlagenen RBAVO-DE erzielte unter Einsatz von SVM in 18 Benchmarks gleichwertige Werte.

Die restlichen Metriken, die SVM verwenden, sind in Anhang A.2 aufgeführt. In Bezug auf die Mittlere Fitnesswerte des vorgeschlagenen RBAVO-DE und seiner Gegenstücke unter Verwendung des SVM Das vorgeschlagene RBAVO-DE-Modell erwies sich unter gleichen Betriebsbedingungen als effizienter als die anderen Methoden. Das vorgeschlagene RBAVO-DE unter Verwendung des SVM-Modells ergab die niedrigsten Fitnesswerte und den wettbewerbsfähigsten STDE über alle Benchmarks hinweg. Darüber hinaus sind alle verwendeten Benchmarks hochdimensional, was zeigt, dass die vorgeschlagenen RBAVO-DE kann auf allen Benchmarks unabhängig von der Größe effektiv ausgeführt werden. Der RBAVO-DE Algorithmus ist vielversprechend, indem er die Nutzung und Erkundung der Suche effektiv in Einklang bringt Raum, wodurch vermieden wird, während der Iterationen in lokalen Optima gefangen zu sein. Im Gegensatz zu vielen anderen Algorithmen, die dazu neigen, in eine Falle zu tappen, wurde die Fähigkeit demonstriert, solchen Fallen zu entkommen.

Bezüglich der mittleren Ergebnisse für die von der vorgeschlagenen RBAVO-DE ausgewählten Gene und ihrer Gegenüber den SVM-Anwendern erzielte die vorgeschlagene RBAVO-DE vielversprechendere Ergebnisse als andere Techniken in allen in dieser Studie verwendeten Benchmarks. Auch die Überlegenheit der vorgeschlagenen RBAVO-DE mit SVM in diesem Zusammenhang zeigt ihre Fähigkeit um wertvolle Bereiche des Suchraums effektiv zu erkunden und dabei Bereiche zu vermeiden, mit nicht umsetzbare Lösungen.

Im Hinblick auf die durchschnittlichen Präzisionswerte der vorgeschlagenen RBAVO-DE und ihrer Gegenüber den Konkurrenten, die SVM einsetzen, war die vorgeschlagene RBAVO-DE der Alternative überlegen. Ansätze für drei der zweiundzwanzig Datensätze. Andererseits für achtzehn Datensätze, BABC, BGWO, BWOA, BBSA und BASO lieferten vergleichbare Ergebnisse. Für siebzehn Datensätze BSSA, BPSO, BGOA und BHHO lieferten ähnliche Ergebnisse wie RBAVO-DE. Platz vier. Letztlich schnitt BBA unter allen Ansätzen am schlechtesten ab und erzielte ähnliche Ergebnisse denen der vorgeschlagenen RBAVO-DE für zwölf Datensätze.

In Bezug auf den durchschnittlichen Rückruf der vorgeschlagenen RBAVO-DE-Methode und ihrer Gegenstücke Mit SVM übertraf der vorgeschlagene RBAVO-DE die alternativen Ansätze bei drei von die zweiundzwanzig Datensätze. Im Gegensatz dazu zeigten BPSO und BHHO eine ähnliche Leistung für achtzehn Datensätze, während BSSA, BABC, BGWO, BWOA, BGOA und BBSA die gleichen Ergebnisse lieferten als die vorgeschlagene RBAVO-DE für neunzehn Datensätze. BHGSO, das die gleichen Ergebnisse wie RBAVO-DE für siebzehn Datensätze, Platz vier. Am Ende BASO und BBA rangiert am niedrigsten unter allen Ansätzen und führt zu ähnlichen Ergebnissen wie RBAVO-DE für sechzehn Datensätze.

Machine Translated by Google

In Bezug auf den mittleren F1-Score des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke, die SVM verwenden, übertraf der vorgeschlagene RBAVO-DE die alternativen Ans für fünf der 22 Datensätze. Andererseits lieferte BWOA ähnliche Ergebnisse denen der vorgeschlagenen RBAVO-DE für siebzehn Datensätze, während BSSA, BABC, BPSO, BGWO, BGOA und BBSA lieferten vergleichbare Ergebnisse für sechzehn Datensätze. BHHO, das erzielte für fünfzehn Datensätze die gleichen Ergebnisse wie die vorgeschlagene RBAVO-DE und belegte den vierten Platz. Letztlich schnitt BBA unter allen Ansätzen am schlechtesten ab und führte zu ähnlichen Ergebnissen wie der RBAVO-DE für 12 Datensätze.

Tabelle 5. Die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke unter Verwendung von SVM bezüglich der durchschnittlichen Klassifikationsgenauigkeitswerte.

Datensatzmetrik	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA Durchschnitt 1.000 STDE	1.000	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	0,000
CESC Durchschnitt 1.000	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	0,000
CHOL Durchschnitt 1,000 STDE	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
0,000 COAD	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Durchschnitt 1,000 STDE 0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000
Durchschnitt 0,986	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 STDE 0,013 GBM	0,975	0,977	0,974	0,974	0,977	0,974	0,975	0,975	0,975	0,974	0,974	
Durchschnitt 1,000	0,004	0,008	0,000	0,000	0,008	0,000	0,005	0,005	0,005	0,000	0,000	
STDE 0,000 HNSC Durchschnitt	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	
Durchschnitt 1,000 STDE 0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
Durchschnitt 1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
KICH STDE 0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
Durchschnitt 1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
KIRC STDE 0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
Durchschnitt 1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
KIRP STDE 0,000 LUAD	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
Durchschnitt 0,993	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
LIHC STDE 0,003	1,000	1,000	1,000	0,9992	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
Durchschnitt 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 0,000 PAAD Durchschnitt	0,991.	0,991.	0,991.	0,000	0,991.	0,991.	0,991.	0,991.	0,991.	0,992.	0,991	
1,000 STDE 0,000	0 .000	0 .000	0 .000	1,000	0 .000	0 .000	0 .000	0 .000	0 .000	0 .002	0,000	
LUSC PCPG Durchschnitt	1.000	1.000	1.000	0,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
1,000 STDE READ	0.000	0.000	0.000	1,000	0.000	0.000	0.000	0.000	0.000	0.000	0,000	
Durchschnitt 1,000 STDE	1.000	1.000	1.000	0,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	0.000	0.000	0.000	1,000	0.000	0.000	0.000	0.000	0.000	0.000	0,000	
	1.000	1.000	1.000	0,000	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	
SARC Durchschnitt 0,994 STDE	0,986	0,989	0,986	0,981	0,989	0,987	0,986	0,984	0,986	0,984	0,981	
0,008 SKCM	0,008	0,009	0,008	0,000	0,009	0,009	0,008	0,006	0,008	0,007	0,000	
Durchschnitt 1,000 STDE	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	
STAD Durchschnitt 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 THCA	0,000	0,000	0,000	0,002	0,000	0,000	0,000	0,002	0,000	0,000	0,000	
Durchschnitt 1,000 STDE 0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
THYM Durchschnitt	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
1,000 STDE 0,000 UCEC	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1.000	
Durchschnitt 1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
STDE 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	
Rangliste S|T|N	4|18|0	0|18|4	0|18|4	0|18|4	0|18|4	0|18|4	Fettgedruckte Zahlen zeigen die		0|18|4	0|18|4	0|18|4	0|18|4	0|18|4	0|18|4
		besten Ergebnisse an.												

4.5. Konvergenzanalyse

Aus Anhang B [57] **geht hervor , dass die vorgeschlagene RBAVO-DE unter Verwendung des SVM und k-NN-Modelle erreichten in allen Benchmarks optimales Konvergenzverhalten. Daher die Konvergenzleistung des vorgeschlagenen RBAVO-DE unter Verwendung von SVM und k-NN Modelle demonstrieren ihre Fähigkeit, optimale Ergebnisse zeitnah zu erreichen und dabei ein effizientes Gleichgewicht zwischen Suche und Nutzung.**

Machine Translated by Google

4.6. Wilcoxons Rangsummentest

In diesem Artikel werden die Fitnesswerte des vorgeschlagenen RBAVO-DE und seiner Gegenstücke paarweise mithilfe des Wilcoxon-Rangsummentests [60] verglichen, um festzustellen, ob es einen statistisch signifikanten Unterschied zwischen den verschiedenen Ansätzen gibt. Dieser Test ist ein wichtiges Instrument zur Bewertung des Erfolgs des vorgeschlagenen Algorithmus. Der Wilcoxon-Test wird bei Hypothesentests zum Vergleich übereinstimmender Daten eingesetzt. Dabei werden die absoluten Unterschiede in den Ergebnissen gepaarter Verfahren für das j **von** N Problemen sortiert. Anschließend wird der niedrigste Wert du th **Die Nullhypothese wird abgelehnt, wenn das resultierende Signifikanzniveau weniger als 5 % beträgt, andernfalls wird sie nicht abgelehnt.**

Bei der Untersuchung der Ergebnisse des paarweisen Vergleichs zwischen dem RBAVO-DE und seinen Gegenstücken mithilfe des Wilcoxon-Tests haben wir festgestellt, dass alle R + **-Werte 253, alle** R **ÿ -Werte 0 und alle p-Werte 4,768 × 10ÿ7** betragen , **was weniger als 0,05 ist (Signifikanzniveau 5 %). Daher kann der Schluss gezogen werden, dass der vorgeschlagene RBAVO-DE unter Verwendung von SVM und k-NN in allen Fällen besser abschneidet als jeder andere Algorithmus. Daher liefern alle p-Werte unter 0,05 (Signifikanzniveau 5 %) überzeugende Beweise dafür, dass die Ergebnisse der vorgeschlagenen Strategie statistisch signifikant und nicht nur zufällig sind.**

4.7. Rechenkomplexität des RBAVO-DE-Algorithmus und verschiedener metaheuristischer Algorithmen

4.7.1. Rechenkomplexität der Ausführungszeit des RBAVO-DE-Algorithmus

Jeder der fünf Kernschritte von RBAVO-DE kann separat analysiert werden, um seine Rechenkomplexität zu bestimmen . Diese Schritte umfassen das Filtern von Funktionen, das Initialisieren der Population, das Boosten und Ändern der Position, das Bewerten der Fitnessfunktion und die Verwendung von DE. Anschließend kann Oexecute_time(RBAVO **ÿ** DE) **verwendet werden, um die gesamte Rechenkomplexität des vorgeschlagenen RBAVO-DE-Algorithmus darzustellen .** Dies kann mithilfe der folgenden **Formeln in Big-O-Notation berechnet werden :**

$$\begin{aligned} \text{Oexecute_time(RBAVO } \ddot{y} \text{ DE)} = \text{Oexecute_time(Features } \textbf{filtern)+} \\ \text{Oexecute_time(Population initialisieren)+} \\ \text{Oexecute_time(Position steigern } \textbf{und ändern)+} \\ \text{Oexecute_time(Fitnessfunktion bewerten)+} \\ \text{Oexecute_time(DE-Technik integrieren).} \end{aligned} \quad (27)$$

Da N **die Populationsgröße bestimmt,** ist Gmax **die maximale Anzahl von Iterationen zulässig, und** D **ist der dimensionale Raum des Problems. Daher gilt:**

Oexecute_time(Filterfunktionen) = **Oexecute_time(D).**
Oexecute_time(Population **wird initialisiert)** = **Oexecute_time(N).**
Oexecute_time(Erhöhung **und Änderung der Position)** = Oexecute_time(Gmax **×** N **×** D).
Oexecute_time(Bewertung **der Fitnessfunktion)** = Oexecute_time(Gmax **×** N).
Oexecute_time (unter Einbeziehung **der DE-Technik)** = Oexecute_time (N **×** D).

Dafür,

$$\begin{aligned} \text{OAusführungszeit(RBAVO } \ddot{y} \text{ DE)} = \textbf{OAusführungszeit(D)+} \\ \textbf{OAusführungszeit(N)} + \text{OAusführungszeit(Gmax } \times \text{ N } \times \text{ D)+} \\ \text{OAusführungszeit(Gmax } \times \text{ N) + OAusführungszeit(N } \times \text{ D) =} \\ \text{OAusführungszeit(Gmax } \times \text{ N } \times \text{ D).} \end{aligned}$$

Machine Translated by Google

4.7.2. Rechenaufwand für den Speicherbedarf des RBAVO-DE-Algorithmus

Hierbei wird die Speichermenge gemessen, die ein Algorithmus zur Bearbeitung eines Problems benötigt, wenn die Menge der Eingaben zunimmt. Dies wird häufig als der zusätzliche Speicher ausgedrückt, den der Algorithmus über die Eingabe hinaus benötigt. Dabei werden die folgenden zwei Hauptelemente zusammengeführt: 1. Speichernutzung der Eingabevariablen: Dies

ist die Speichermenge, die der Algorithmus zur Speicherung der Eingabedaten benötigt. Es gibt 13 Eingabevariablen im Zusammenhang mit dem vorgeschlagenen RBAVO-DE-Algorithmus, und zwar: N, Gmax, **D,** LB, UB, CR, WM, **L1, L2, w, P1, P2 und P3. Da jede Variable numerische Werte speichert, werden von jeder 4 Byte Speicher genutzt . Folglich beträgt die gesamte Speichernutzungskomplexität dieser 13 Eingabevariablen 52 Byte (13 × 4 Byte** = **52 Byte). Die Speichernutzungskomplexität der Eingabewerte ist konstant.**

2. Zusätzlicher Speicherbedarf: Hier wird angegeben, wie viel Speicher der Algorithmus zusätzlich zum Input benötigt. Dazu gehört der Speicherbedarf für Datenstrukturen, interne Variablen und andere Komponenten des Algorithmus. Unabhängig von der Größe des Inputs benötigt der RBAVO-DE-Algorithmus eine bestimmte Menge an zusätzlichem Variablen sind beteiligt:

- Die von Xinitial **verbrauchte Speicherkomplexität beträgt (4 ×** N **× D) Bytes. Dies liegt daran, dass jede Position im Positionsvektor** Xinitial **4 Bytes Speicher benötigt und ihre Größe** (N **× D)** beträgt, **proportional zur anfänglichen Population von** N **Positionen mit Dimensionsgröße D. Da die erforderliche Speichermenge linear mit dem Wert** (N **× D)** steigt, **ist die Speicherkomplexität linear. • Die Variablen** pr, **g,** Fi , **t, R,** Di , dt , **S1, S2, A1, A2,** sigma, fit (Xi), fit (XBest1), fit (XSecondBest2) **und** fit (XBest) **benötigen jeweils 4 Bytes Speicherplatz, da sie nur numerische Werte darstellen. Daraus ergibt sich, dass die gesamte Speicherkomplexität dieser 16 Variablen 64 Bytes beträgt (16 × 4 Bytes** = **64 Bytes). Diese Speicherkomplexität ist konstant.**

- **Die Ortsvektoren** BestVulture1, SecondBestVulture2, Xi , Levyd , **μ, ÿ,** X bin , ÿi , Xr1 , Xr2 , Xr3 , ui **und** XBest **benötigen jeweils 4 Byte Speicherplatz und die Größe jedes Vektors ist D, proportional zur Dimensionsgröße der erfassten Positionen.**
  Als Ergebnis haben die Vektoren an diesen 13 Punkten eine Gesamtspeichernutzungskomplexität von (52 × D) **Bytes (13 × 4 ×** D **Bytes). Da der benötigte Speicher linear mit dem Wert D wächst, ist die Speichernutzungskomplex**

Daraus ergibt sich für alle oben aufgeführten zusätzlichen Variablen eine Gesamtkomplexität des Speicherbedarfs von (4 × N **× D) + 64** + **(52 ×** D) **Bytes.**

Schließlich kann die folgende Formel verwendet werden, um die Rechenkomplexität der gesamten Speichernutzung des vorgeschlagenen RBAVO-DE-Algorithmus zu bestimmen:

Speichernutzungskomplexität (RBAVO-DE) = **Eingabewerte Speichernutzung +**
Zusätzlicher Speicherverbrauch = **52** + **(4 ×** N **× D) + 64** + **(52 ×** D) **Bytes.**

Beachten Sie, dass es konstante Bytes gibt, die nicht berücksichtigt werden. In Big-O- Notation kann die Rechenkomplexität des gesamten Speicherverbrauchs von RBAVO-DE als Omemory_usage(RBAVO ÿ DE) **dargestellt werden . Dies kann dann in Big-O-Notation berechnet werden, nachdem alle Konstanten wie folgt eliminiert wurden:**

Omemory_usage(RBAVO **ÿ** DE) = Omemory_usage(Speichernutzung **der Eingabewerte)**
+ Omemory_usage(Zusätzliche **Speichernutzung)**
= Omemory_usage(1) + **Omemory_usage(N ×** D) + Omemory_usage(1) + Omemory_usage(D)
= **Omemory_usage(N ×** D).

Machine Translated by Google

Es kann schwierig sein, einen umfassenden Vergleich der Speichernutzung und der Komplexität der Ausführungszeit verschiedener metaheuristischer Algorithmen zusammenzustellen, da diese Komplexitäten je nach Implementierung, Größe des Problems und anderen Operatoren unterschiedlich sein können. Darüber hinaus sind nicht für alle aufgeführten Algorithmen umfassende Bewertungen der Komplexität der Ausführungszeit und der Speichernutzung verfügbar, und ihre Eigenschaften können je nach vorliegendem Problem variieren.

5. Vor- und Nachteile des RBAVO-DE-Algorithmus

Dieser Teil präsentiert eine ausgewogene Diskussion über die Vor- und Nachteile der RBAVO-DE-Algorithmus. Die Vorteile des RBAVO-DE-Algorithmus können wie folgt aufgelistet werden: • Hohe Genauigkeit bei der Klassifizierung: Der RBAVO-DE-Algorithmus hat eine hohe Klassifizierungsgenauigkeit bewiesen, die in einigen Fällen bis zu 100 % erreichte. Dies ist eine bemerkenswerte Leistung, insbesondere in Bezug auf Krebsdatensätze, da eine genaue Genauswahl die diagnostischen und therapeutischen Ergebnisse direkt beeinflussen kann. • Effektive Reduzierung der Merkmalsgröße: Der Algorithmus hat eine bemerkenswerte Fähigkeit gezeigt, die Merkmalsgröße um bis zu 98 % zu reduzieren und gleichzeitig die Klassifizierungsgenauigkeit beizubehalten oder zu verbessern . Diese Dimensionsreduzierung ist entscheidend für die effiziente Verarbeitung hochdimensionaler Datensätze. • Robustheit über verschiedene Datensätze hinweg: Die Wirksamkeit des RBAVO-DE-Algorithmus über 22 Krebsdatensätze hinweg weist auf seine Robustheit und Anpassungsfähigkeit an verschiedene genetische Dateneigenschaften hin. Diese Vielseitigkeit ist für breitere Anwendungen in der Genomforschung von Vorteil. • Überlegene Leistung gegenüber Mitbewerbern: Im Vergleich mit binären Varianten weithin anerkannter metaheuristischer Algorithmen sticht RBAVO-DE durch seine herausragende Leistung in Bezug auf Genauigkeit und Merkmalsreduzierung hervor und unterstreicht seinen innovativen Ansatz bei der Genauswahl.

Auf der anderen Seite können die Nachteile der vorgeschlagenen Methodik wie folgt aufgelistet werden: • Rechenkomplexität: Die hohe Dimensionalität der Datensätze und die iterative Natur metaheuristischer Algorithmen legen nahe, dass RBAVO-DE einen erheblichen Rechenaufwand verursachen kann. Dieser Aspekt könnte seine Anwendbarkeit in Umgebungen mit begrenzten Rechenressourcen einschränken. • Bedenken hinsichtlich der Generalisierbarkeit: Trotz seines Erfolgs bei verschiedenen Krebsdatensätzen muss die Generalisierbarkeit des Algorithmus auf andere Arten biologischer Daten oder Krankheiten noch gründlich untersucht werden. Es ist entscheidend, RBAVO-DE in breiteren Kontexten zu testen, um seine Anwendbarkeit über die untersuchten Datensätze hinaus zu bestätigen . • Parametersensitivität: Wie viele metaheuristische Algorithmen könnte RBAVO-DE empfindlich auf seine Parametereinstellungen reagieren, was sich auf Leistung und Effizienz auswirkt. Detaillierte Studien über die Auswirkungen verschiedener Parameterkonfigurationen und Strategien zu ihrer Optimierung könnten den Nutzen des Algorithmus steigern.

Zusammenfassend stellt der RBAVO-DE-Algorithmus einen bedeutenden Fortschritt in der Genauswahl für die Krebsklassifizierung dar und weist bemerkenswerte Stärken in Bezug auf Genauigkeit und Effizienz auf. Die Beseitigung seiner potenziellen Nachteile durch weitere Forschung könnte jedoch seine Anwendbarkeit erweitern und seine Leistung verbessern, was ihn zu einem noch wertvolleren Werkzeug in der Genomik und personalisierten Medizin machen würde.

6. Schlussfolgerungen und zukünftige Richtungen

Der in diesem Artikel vorgeschlagene RBAVO-DE-Ansatz war der erste, der zur Behandlung von GS-Problemen in RNA-Seq-Genexpressionsdaten und zur Bestimmung potenzieller Biomarker für verschiedene Tumorklassen implementiert wurde, um die beste gefundene Lösung zu verbessern. Die Ergebnisse waren vielversprechend und zeigten, dass die Wirksamkeit und Fähigkeiten des vorgeschlagenen RBAVO-DE-Algorithmus erheblich verbessert wurden. SVM und k-NN, zwei weit verbreitete Klassifizierungsmodelle , wurden verwendet, um die Wirksamkeit jedes Satzes ausgewählter Gene zu bewerten. Die Leistungsfähigkeit des vorgeschlagenen RBAVO-DE-Algorithmus wurde mit binären Varianten von elf weit verbreiteten

Machine Translated by Google

metaheuristische Techniken, um es anhand verschiedener Instanzen anhand unterschiedlicher Tumorklassen zu beurteilen. Die Bewertung wurde mithilfe einer Kombination von Bewertungsmaßstäben durchgeführt, wie z. B. durchschnittliche Fitness, Klassifizierungsgenauigkeit, Anzahl der ausgewählten Gene, Präzision, Rückruf und F1-Score. Der vorgeschlagene RBAVO-DE-Algorithmus unter Verwendung der SVM- und k-NN-Klassifikatoren erzielte bei der Behandlung von GS-Problemen vielversprechendere Ergebnisse als andere Optimierer. Trotz der vielversprechenden Ergebnisse eröffnet diese Forschung mehrere Möglichkeiten für zukünftige Untersuchungen, um das Feld weiter voranzutreiben :

- Algorithmische Verbesserungen: Verbesserung der Effizienz des Algorithmus und Reduzierung der Rechenkomplexität .
- Krankheitsübergreifende Anwendbarkeit: Testen des vorgeschlagenen Algorithmus anhand genetischer Daten aus verschiedene Krankheiten außer Krebs.
- Vergleichende Analysen mit Deep Learning Modellen: Evaluierung von RBAVO-DE gegenüber Fortgeschrittene Deep-Learning-Modelle für die genetische Datenanalyse.
- Klinische Validierung in der Praxis: Zusammenarbeit mit klinischen Experten, um den praktischen Nutzen ausgewählter Gene in der Krebsbehandlung zu validieren.
- Skalierbarkeit und Parallelisierung: Verbessern der Skalierbarkeit des Algorithmus durch parallele Berechnungen, um größere genetische Datensätze effizient zu verarbeiten.
- Interdisziplinäre Anwendungen: Erkunden des Potenzials des Algorithmus in anderen Bereichen, in denen hochdimensionale Daten verarbeitet werden, wie etwa in der Finanz- und Umweltmodellier

Zusammenfassend lässt sich sagen, dass der RBAVO-DE-Algorithmus zwar einen bedeutenden Fortschritt auf dem Gebiet der Genauswahl für die Krebsklassifizierung darstellt, die für die künftige Forschung skizzierten Wege jedoch das Potenzial für weitere Fortschritte und breitere Anwendungen dieser Arbeit unterstreichen. Um das volle Potenzial der Methoden der Genselektion auszuschöpfen und so die Ergebnisse im Gesundheitswesen zu verbessern und unser Verständnis komplexer Erkrankungen zu erweitern, sind kontinuierliche interdisziplinäre Zusammenarbeit und Innovation von entscheidender Bedeutung .

Autorenbeiträge: MGG: Ressourcen, Validierung und Schreiben – Überprüfung und Bearbeitung. AAA: Ressourcen, formale Analyse, Datenkuratierung, Validierung, Schreiben – Originalentwurf und Schreiben – Überprüfung und Bearbeitung. AEE: Untersuchung, Visualisierung, Datenkuratierung, Validierung, Schreiben – Originalentwurf und Schreiben – Überprüfung und Bearbeitung. AAAE-M.: Konzeptualisierung, Methodik, Software, formale Analyse, Untersuchung, Datenkuratierung, Validierung, Schreiben – Originalentwurf und Schreiben – Überprüfung und Bearbeitung. Alle Autoren haben die veröffentlichte Version des M

Finanzierung: Die Prince Sattam bin Abdulaziz University hat diese Forschungsarbeit unter der Projektnummer PSAU/2023/01/25612 finanziert.

Datenverfügbarkeitserklärung: Um die Anwendbarkeit und Reproduzierbarkeit dieser Forschung zu verbessern, sind die entwickelte Software und der relevante Python-Code dieser Forschung öffentlich verfügbar und erhältlich unter [61].

Danksagungen: Die Autoren danken der Prinz-Sattam-bin-Abdulaziz-Universität für die finanzielle Unterstützung dieses Forschungsprojekts.

Interessenkonflikte: Die Autoren erklären, dass ihnen keine konkurrierenden finanziellen Interessen oder persönlichen Beziehungen bekannt sind, die die in diesem Artikel berichtete Arbeit beeinflusst haben könnten.

Abkürzungen

In dieser Untersuchung werden folgende Abkürzungen verwendet:

AVO	Optimierung für Afrikanische Geier
HNSC	Plattenepithelkarzinom im Kopf-Hals-Bereich
DE	Differenzielle Evolution
KICH	Nierenchromophobiker
RBAVO-DE	Relief Binäre Afrikanische Geier Optimierung basierend auf Differential Evolution
KIRP	Niere, Nierenpapillär
GS	Genselektion
LIHC	Leber-Hepatozelluläres Karzinom
RNA-Seq	RNA-Sequenzierung

Machine Translated by Google

PAAD	**Pankreasadenokarzinom**
DNA	**Desoxyribonukleinsäure**
PCPG	**Phäochromozytom und Paraganglion**
SVM	**Support-Vektor-Maschine**
LESEN	**Rektumadenokarzinom**
k-NN	**k-Nächster Nachbar**
SARC	**Sarkom**
FS	**Funktionsauswahl**
SKCM	**Haut Kutanes Melanom**
MHMs	**Metaheuristische Methoden**
MeanFi	**Mittlere Fitness**
ML	**Maschinelles Lernen**
MeanGe	**Durchschnittliche Größe ausgewählter Gene**
DL	**Tiefes Lernen**
STDE	**Standardabweichung**
SA	**Simuliertes Abkühlen**
RF	**Zufälliger Wald**
FOLIE	**Induktiver Lerner erster Ordnung**
GSO	**Gravitationsoptimierer**
SCCSA	**Sinus-Cosinus-Kuckuckssuchalgorithmus**
SMO	**Klammeraffen-Optimierung**
BHGSO	**Optimierung der binären Henry-Gas-Löslichkeit**
BPSO	**Binäre Partikelschwarmoptimierung**
BABC	**Binäres künstliches Bienenvolk**
BSSA	**Binärer Salpenschwarm-Algorithmus**
BBA	**Binärer Bat-Algorithmus**
BGWO	**Binäre Grey-Wolf-Optimierung**
BGOA	**Binärer Grasshopper-Optimierungsalgorithmus**
BWOA	**Binärer Wal-Optimierungsalgorithmus**
BASO	**Binäre Atomsuchoptimierung**
BBSA	**Binärer Vogelschwarm-Algorithmus**
BHHO	**Binäre Harris Hawks Optimierung**
BBBO	**Binäre Braunbär-Optimierung**
BAO	**Binäre Aquila-Optimierung**
BMOA	**Binärer Meerkat-Optimierungsalgorithmus**
LUAD	**Lungenadenokarzinom**
LUSC	**Plattenepithelkarzinom der Lunge**
BRCA	**Invasives Brustkarzinom**
KIRC	**Nierenklarzellkarzinom**
UCEC	**Endometriumkarzinom des Uteruscorpus**
RPKM	**Lesevorgänge pro Kilobase pro Million zugeordneter Lesevorgänge**
STAD	**Magenadenokarzinom**
SDAE	**Gestapelter Denoising-Autoencoder**
CGA	**Krebsgenomatlas**
BLCA	**Urothelkarzinom der Blase**
THCA	**Schilddrüsenkrebs**
CESC	**Gebärmutterhals- und Endozervikalkrebs**
CHOL	**Cholangiokarzinom**
COAD	**Kolon-Adenokarzinom**
ESCA	**Speiseröhrenkrebs**
GBM	**Glioblastoma multiforme**
THYM	**Thymom**

Machine Translated by Google

Anhang A. Vergleichsergebnisse der vorgeschlagenen RBAVO-DE und ihrer Gegenstücke

Anhang A.1. Vergleichsergebnisse basierend auf dem k-NN-Modell

Tabelle A1. Die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke mit k-NN bezüglich der durchschnittlichen Fitnesswerte.

Datensatzmetrik	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA Durchschnitt 0,00190 STDE	0,01110	0,00710	0,01120	0,01470	0,00940	0,01070	0,01140	0,01310	0,01030	0,01080	0,01590
0,00430 Durchschnitt	0,00530	0,00440	0,00540	0,00240	0,00550	0,00540	0,00540	0,00440	0,00550	0,00540	0,00180
CESC 0,01610 STDE 0,00000	0,01990	0,02020	0,02050	0,01980	0,02000	0,02020	0,02020	0,01970	0,02020	0,02030	0,02060
CHOL Durchschnitt	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	0,00390	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 Durchschnitt 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
ESCA Durchschnitt 0,02560	0,02930	0,02960	0,02990	0,02930	0,02940	0,02970	0,02960	0,02910	0,02960	0,02980	0,03000
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
GBM -Durchschnitt 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 Durchschnitt 0,00710 STDE	0,01250	0,01290	0,01330	0,01280	0,01240	0,01310	0,01320	0,01270	0,01290	0,01320	0,01360
0,00350 Durchschnitt	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 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
Durchschnitt 0,00250	0,00020	.00010	0 .00010	0,00020	0,00020	0,00020	.00010	.00010	.00010	.00010	0,00010
KIRC STDE 0,00380	0,01110	0,01040	0,01220	0,01200	0,01020	0,01200	0,01070	0,01040	0,01120	0,01170	0,01290
Durchschnitt 0,00000	0,00270	0,00340	0,00190	0,00020	0,00340	0,00190	0,00320	0,00300	0,00280	0,00220	0,00010
KIRP 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
Durchschnitt 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 STDE 0,00000 LUAD	.00390	0,00430	0,00450	0 .00390	0 .00400	.00420	0,00430	0,00380	0,00430	0,00420	0,00460
Durchschnitt 0,00870	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 Durchschnitt 0,00000	0,01260	0,01290	0,01310	0,01240	0,01270	0,01290	0,01290	0,01240	0,01290	0,01300	0,01340
STDE 0,00000 PAAD	0,00020	0,00010	0,00010	0,00020	0,00020	0,00010	0,00010	0,00020	0,00010	0,00010	0,00010
LUSC Durchschnitt 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 PCPG	0,00020	0,00010	0,00010	0,00020	0,00010	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
Durchschnitt 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 Durchschnitt	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 SARC	0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
Durchschnitt 0,01890	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 SKCM Durchschnitt	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
Durchschnitt 0,00070 STDE 0,00280	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
	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
STAD	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 -Durchschnitt 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 -Durchschnitt 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 -Durchschnitt 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
Rangliste S[T]N 22|0|0	0|0|22	0|0|22	0|0|22	Fettgedruckte	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22

Werte kennzeichnen die besten Ergebnisse.

Tabelle A2. Die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke mit k-NN bezüglich der Durchschnittswerte ausgewählter Gene.

Datensatzmetrik	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA Durchschnitt 142,03 STDE	210,17	237,70	235,00	196,03	218,17	228,50	226,30	195,23	228,27	232,03	237,13
017,18 Durchschnitt	020,50	019,07	017,12	023,83	018,27	019,43	014,81	013,17	014,60	018,90	014.14
CESC 122,50 STDE 005,45	195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	213,33	215,90	233,43
CHOL Durchschnitt	011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	005,88	006,02	005.64
122,13 STDE 006,18 COAD	195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63
Durchschnitt 120,30	011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36
STDE 005,82 Durchschnitt 120,17	195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63
STDE 006,81 GBM	011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	006,44	005,08	004.36
ESCA Durchschnitt 120,73	195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,57	218,87	230,07
STDE	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 Durchschnitt 133,43 STDE	202,87	222,67	227,43	203,00	211,37	217,97	220,87	195,93	219,83	223,63	241,27
017,63 Durchschnitt	009,26	011,96	003.91	013,24	007,65	007,41	005,32	008,26	007,71	005,68	006.21
KICH 121,43 STDE 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

Tabelle A2. Forts.

Datensatzmetrik RBAVO-DE BSSA BABC BPSO BBA BGWO BWOA BGOA BHHO BBSA BASO BHGSO												
KIRC	Durchschnitt 137,57	200,83	220,17	227,70	192,17	209,87	216,13	222,30	193,27	218,73	218,60	234,10
	STDE 017,10	009,63	011,74	008,27	008,73	014,08	010,11	015,18	017,36	013,82	012,78	004.21
KIRP	Durchschnitt 122,33	195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211,57	232,63
	STDE 009.08	011,17	005,31	003.11	009,66	009.30	008,35	005,66	006,49	006,44	005.08	004.36
LIHC	Durchschnitt 123,40	197,00	214,03	225,67	193,33	200,63	212,33	213,63	189,30	213,10	212.03	232,43
	STDE 006,59 LUAD	007,78	005,22	003,36	009,17	010,02	009,35	004,89	006,89	005,69	004.81	004.97
Durchschnitt 127,03 STDE 008,62		197,47	216,13	222,93	190,83	205,33	212,57	215,67	191,80	215,27	217.83	237,90
	Durchschnitt 125,73	008,98	004,74	006,16	008,61	010,57	006,15	004,43	008,50	006,13	004.71	004.94
LUSC	STDE 006,35 PAAD	198,87	215,63	223,00	189,13	203,60	213,23	216,37	194,67	215,53	215.23	241,23
	Durchschnitt 121,10	008,77	003,54	004,89	010,69	007,20	009,70	003,44	006,40	005,54	006.16	004.66
STDE 007,71 PCPG Durchschnitt		195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211.57	232,63
	122,17 STDE 006,42	011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	006,44	005.08	004.36
READ Durchschnitt 120,13 STDE		195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211.57	232,63
	007,72 SARC	011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	006,44	005.08	004.36
Durchschnitt 122,47 STDE 005,79		195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211.57	232,63
	SKCM Durchschnitt	011,17	005,31	003,11	009,66	009,30	008,35	005,66	006,49	006,44	005.08	004.36
123,03 STDE 008,83 Durchschnitt		195,13.	213,03.	225,30	194,10	199,50	213,63	212,67	185,60.	212,90.	216.43	231,80
	142,60 STDE 013,31	01 1,17	00 5,31	003,11	009,66	009,30	008,35	005,66	00 6,49	00 6,30	004.98	003.40
THCA Durchschnitt 117,97 STDE		195,13	213,03	225,30	194,10	199,50	213,63	212,67	185,60	211,00	211.57	232,63
	006,67 THYM	011,17	005,31	003,11	009,66.	009,30	008,35	005,66	006,49	006,44	005.08	004.36
STAD	Durchschnitt 120,37	215,00	236,17	243,33	18 7,50	221,17.	4,87	232,77	201,73	231,13	7,10	236,97
	STDE 004,89 UCEC	018,43	013,94	016,29.	010,27	01 9,24	022,63	014,25	014,00	014,04	015,35	005.86
Durchschnitt 143,93 STDE 013.18		195,43	213,40	22 5,13	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	21 2,67	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
Rangliste S|T|N	**22|0|0**	**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**	**0|0|22**	**0|0|22**

Fettgedruckte Werte kennzeichnen die besten Ergebnisse.

Tabelle A3. Die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke mit k-NN bezüglich der durchschnittlichen Präzisionswerte.

Datensatz 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 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 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 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 0,98110	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	0,99770	0,99740	0,99840	0,99870	0,99770	0,99800	0,99540	0,99710	0,99740	0,99670	0,99800
0,99920 THCA 1,00000	1,00000	1,00000	1,00000	0,99790	1,00000	0,99970	1,00000	1,00000	1,00000	0,97900	0,99930
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	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.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
Rang (S|U|N)	**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**

Fettgedruckte Werte kennzeichnen die besten Ergebnisse.

Tabelle A4. Die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke mit k-NN bezüglich der durchschnittlichen Recall-Werte.

Datensatz 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

Tabelle A4. Forts.

Datensatz 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	0,99240	0,99270	0,99170	0,99140	0,99240	0,99210	0,99470	0,99310	0,99270	0,99340	0,99210
PAAD 1,00000 PCPG	1,00000	1,00000	1,00000	0,99580	1,00000	1,00000	1,00000	0,99930	1,00000	0,99310	0,99270
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	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
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	1,00000
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 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
Rang (S|U|N) 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

Fettgedruckte Werte kennzeichnen die besten Ergebnisse.

Tabelle A5. Die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke mit k-NN bezüglich der durchschnittlichen F1-Scorewerte.

Datensatz 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	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 HNSC	1,00000	1,00000	1,00000	0,99940	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
0,99590 0,98670	0,98670	0,98670	0,98670	0,98670	0,98670	0,98670	0,98670	0,98670	0,98670	0,98670	0,98670
KICH KIRC 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	0,97970	0,97970	0,97970	0,97940	0,97970	0,97970	0,97970	0,97970	0,97970	0,97980	0,97970
0,99500 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 0,99860	0,99520	0,99520	0,99520	0,99480	0,99520	0,99520	0,99520	0,99520	0,99520	0,99500	0,99520
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 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 0,99050	0,99500	0,99500	0,99500	0,99500	0,99500	0,99500	0,99500	0,99500	0,99500	0,99500	0,99500
SKCM 1,00000 STAD	1,00000	1,00000	1,00000	0,99690	1,00000	0,99980	1,00000	0,99970	1,00000	0,99630	0,99600
0,99960 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 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
Rang (S|U|N) 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

Fettgedruckte Werte kennzeichnen die besten Ergebnisse.

Anhang A.2. Vergleichsergebnisse basierend auf dem SVM-Modell

Tabelle A6. Die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke mit SVM bezüglich der durchschnittlichen Fitnesswerte.

Datensatzmetrik	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA Durchschnitt 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 Durchschnitt		0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
CESC		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 CHOL Durchschnitt		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 COAD		0,00390	0,00430	0,00450	0,00390	0,00400	0,00430	0,00430	0,00370	0,00420	0,00420	0,00470
Durchschnitt 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 Durchschnitt 0,01370 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
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 -Durchschnitt 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
HNSC -Durchschnitt 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

Tabelle A6. Forts.

Datensatzmetrik	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
KICH Durchschnitt 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
	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
KIRC Durchschnitt 0,00000 STDE 0,00000	0,00390	0,00430	0,00450	0,00380	0,00400	0,00430	0,00430	0,00380	0,00430	0,00430	0,00460
	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00020
KIRP Durchschnitt 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
	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
LIHC Durchschnitt 0,00000 STDE 0,00000	0,00410	0,00440	0,00450	0,00500	0,00430	0,00440	0,00450	0,00400	0,00450	0,00450	0,00500
	0,00010	0,00010	0,00010	0,00280	0,00010	0,00020	0,00010	0,00020	0,00010	0,00010	0,00020
LUAD Durchschnitt 0,00840 STDE 0,00160 Durchschnitt	0,01250	0,01290	0,01310	0,01250	0,01260	0,01290	0,01290	0,01230	0,01290	0,01260	0,01330
	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 STDE 0,00000 PAAD Durchschnitt	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
0,00000 STDE 0,00000 PCPG Durchschnitt 0,00000	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
STDE 0,00000 READ Durchschnitt 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
	0,00020	0 .00010	0 .00010	0,00020	0,00020	0,00020	0 .00010	0,00010	0 .00010	0 .00010	0,00010
SARC Durchschnitt 0,00570 STDE 0,00860 SKCM	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
Durchschnitt 0,00000 STDE 0,00000 Durchschnitt 0,01040	0,01780.	0,01570	0,01840	0,02260	0 .01480	0 .01760	0,01870	0,02000.	0,01870	0,02000	0,02330
	0 .00800	0,00880	0,00800	.00020	0,00900	0,00820	0,00770	0 .00610	0,00770	0,00670	0,00010
STDE 0,00280 THCA Durchschnitt 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
	0,00020	0,00010	0,00010	0,00020	0,00020	0,00020	0,00010	0,00010	0,00010	0,00010	0,00010
STAD THYM Durchschnitt 0,00000 STDE 0,00000	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
UCEC Durchschnitt 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
	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
Rangliste S|T|N **22|0|0**	**0|0|22**	**0|0|22**	**0|0|22**	**Fettgedruckte**	**0|0|22**	**0|0|22**	**0|0|22**	**0|0|22**	**0|0|22**	**0|0|22**	**0|0|22**

Werte kennzeichnen die besten Ergebnisse.

Tabelle A7. Die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke mit SVM

bezüglich der durchschnittlichen ausgewählten Genwerte.

Datensatzmetrik	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA Durchschnitt 121,37 STDE 007,49 Durchschnitt	194.70	213.03	225,30	194,10	199,50	213,40	212,67	186.27	211.00	211,83	232,37
	011.44	005.18	003,11	009,66	009,30	008,44	005,66	005.77	006.44	005,25	004.42
CESC 119,83 STDE 007,02 CHOL Durchschnitt	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
125,57 STDE 007,14 COAD Durchschnitt 122,63	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 007,73 Durchschnitt 125,50 STDE 011,18 GBM	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 Durchschnitt 119,73 STDE 008,07 HNSC	195.13	215.40	225,30	194,10	201,77	213,63	213,13	185.70	212.40	218,87	230,07
	011.17	009.58	003,11	009,66	012,33	008,35	006,44	006.47	008.98	004,68	007.64
Durchschnitt 122,17 STDE 008,29 Durchschnitt 121,73	194.93	213.03	225,30	193,90	198,77	213,70	212,83	185.50	211.00	211,00	232,33
	011.27	005.31	003,11	009,47	010,28	008,36	005,69	006.51	006.44	004,92	004.15
STDE 008,49 Durchschnitt 123,87 STDE 007,57	196.40	213.90	225,13	194,03	202,10	213,83	213,50	186.93	211.47	210,87	232,17
	008.49	005.22	003,19	009,56	010,76	009,34	005,22	005.74	005.66	004,79	005.02
KICH Durchschnitt 120,07 STDE 007,87	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
KIRC Durchschnitt 122,63 STDE 008,01 LUAD	197.10	214.70	225,00	191,93.	200,30.	212,97.	213,13.	187.57	213.83	214,63.	231,27
	009.42	005.09	4,23	00 8,70	01 1,59	00 7,64	00 4,94	005.31	005.81	00 4,20	008.02
KIRP Durchschnitt 121,17 STDE 007,56	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
LIHC Durchschnitt 122,67 STDE	204.27	220.37	226,40	212,83	214,07	219,07	223,70	200.60	223.03	225,17	249,90
	006.99	006.29	006,80	017,60	007,06	007,88	004,13	007.66	006.48	004,81	007.97
	195.13	213.03	225,30	194,10	199,43	213,63	212,67	185.87	212.40	214,50	233,00
	011.17	005.31	003,11	009,66	009,37	008,35	005,66	006.40	006.10	005,98	004.34
LUSC	195.13	213.03	225,30	194,10	199,50	213,63	212,67	185.60	211.00	211,57	232,63
006.28	011.17	005.31	003.11	009,66	009,30	008,35	005,66	006.49	006.44	005,08	004.36
PAAD Durchschnitt 119,23 STDE 006,90 PCPG	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
Durchschnitt 122,43 STDE 006,68 READ Durchschnitt	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
123,93 STDE 007,45 SARC Durchschnitt 135,27	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 018,57 SKCM Durchschnitt 122,47 STDE 007,66	202.90	226.90	233,77	194,10	210,53	223,87	217,80	189.93	220.57	220,87	231,53
	020.86	020.25	014,44	009,66	017,74	017,39	013,90	013.41	016.40	011,76	003.70
	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

Tabelle A7. Forts.

Datensatzmetrik RBAVO-DE BSSA BABC BPSO BBA BGWO BWOA BGOA BHHO BBSA BASO BHGSO											
STAD Durchschnitt 123,53 STDE 010,65 THCA	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
Durchschnitt 119,47 STDE 005.58	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
THYM -Durchschnitt 121,80 STDE 007,21 UCEC	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
-Durchschnitt 122,23 STDE 005,93	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
Rangliste S|T|N 22|0|0	0|0|22 0|0|22 0|0|22 Fettgedruckte			0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22	0|0|22

Werte kennzeichnen die besten Ergebnisse.

Tabelle A8. Die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke unter Verwendung von SVM bezüglich der durchschnittlichen Präzisionswerte.

Datensatz 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	0,99890	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	0,99890	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	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
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
Rang (S|U|N) 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

Fettgedruckte Werte kennzeichnen die besten Ergebnisse.

Tabelle A9. Die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke mit SVM bezüglich der durchschnittlichen Recall-Werte.

Datensatz 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	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
KIRC KIRP 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 1,00000 PAAD	1,00000	1,00000	0,99970	0,99930	1,00000	1,00000	1,00000	1,00000	1,00000	0,99930	0,99930
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
PCPG. 1,00000	0,99040	0,99040	0,99040	0,99010	0,99040	0,99040	0,99040	0,99040	0,99040	0,98780	0,99040
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
	0,99550	0,99640	0,98380	0,98200	0,99460	0,99100	0,99010	0,98470	0,99370	0,97840	0,97930
Rang (S|U|N) 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

Fettgedruckte Werte kennzeichnen die besten Ergebnisse.

Machine Translated by Google

Tabelle A10. Die Ergebnisse des vorgeschlagenen RBAVO-DE-Algorithmus und seiner Gegenstücke mit SVM bezüglich der durchschnittlichen F1-Scorewerte.

Datensatz	RBAVO-DE	BSSA	BABC	BPSO	BBA	BGWO	BWOA	BGOA	BHHO	BBSA	BASO	BHGSO
BLCA	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
CESC	1,00000	1,00000	1,00000	1,00000	0,99950	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	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
ESCA	1,00000	1,00000	1,00000	1,00000	0,99910	1,00000	1,00000	1,00000	1,00000	1,00000	0,99970	1,00000
GBM	1,00000	1,00000	1,00000	1,00000	0,99910	1,00000	1,00000	1,00000	1,00000	1,00000	0,99820	0,99600
HNSC	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
KICH	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
KIRC	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	0,95730	0,96420
KIRP	0,96980	0,97000	0,96760	0,96250	0,96980	0,96900	0,96970	0,96970	0,96940	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	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	0,99520	0,99520	0,99520	0,99470	0,99520	0,99520	0,99520	0,99520	0,99520	0,99320	0,99520
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	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
SARC	0,99900	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	0,99690	1,00000	1,00000	1,00000	0,99970	1,00000	0,99880	0,99850
STAD	0,99410	0,99480	0,99480	0,99480	0,99480	0,99480	0,99480	0,99480	0,99480	0,99480	0,99480	0,99480
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	1,00000	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
Rang (S|U|N)	**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**

Fettgedruckte Werte kennzeichnen die besten Ergebnisse.

Anhang B. Konvergenzgraphen des vorgeschlagenen RBAVO-DE und Wettbewerbsmethoden

Anhang B.1. Konvergenzgraphen unter Verwendung des SVM-Modells

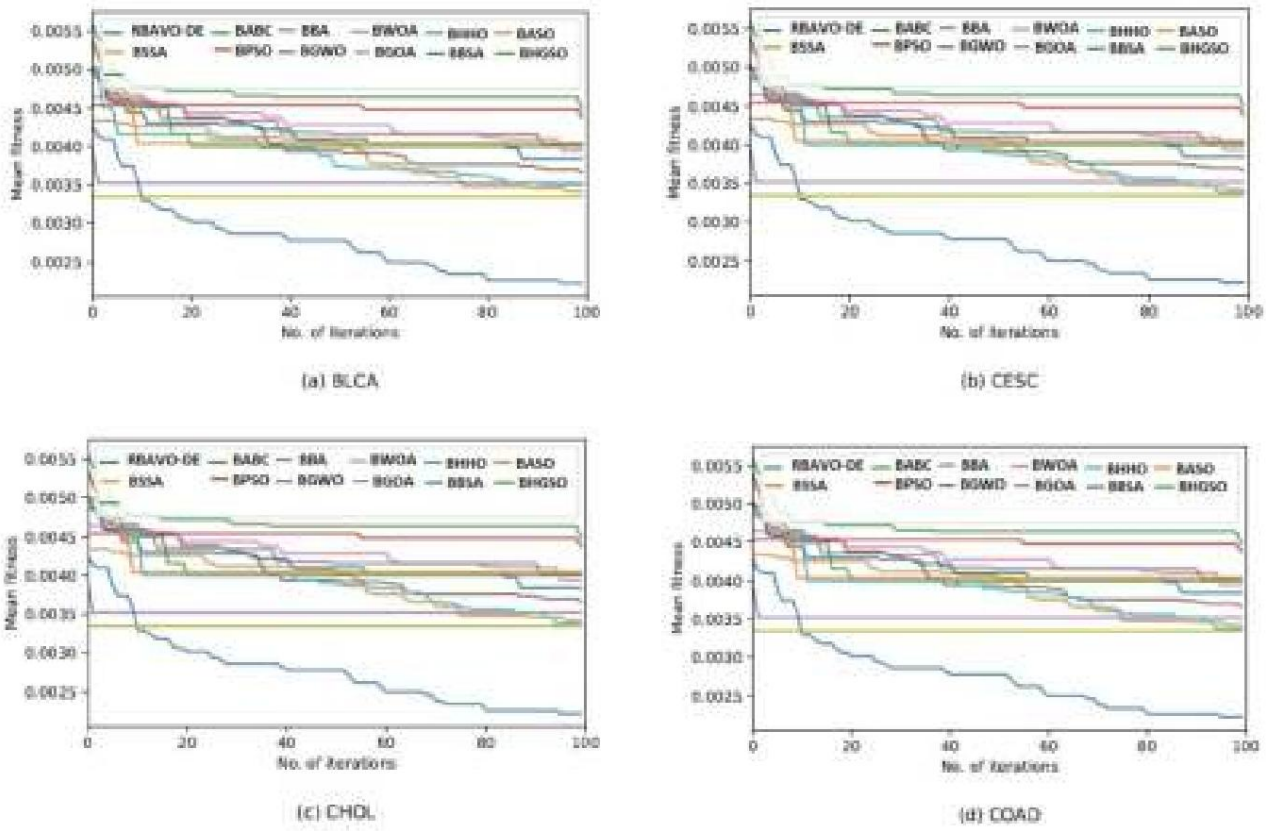
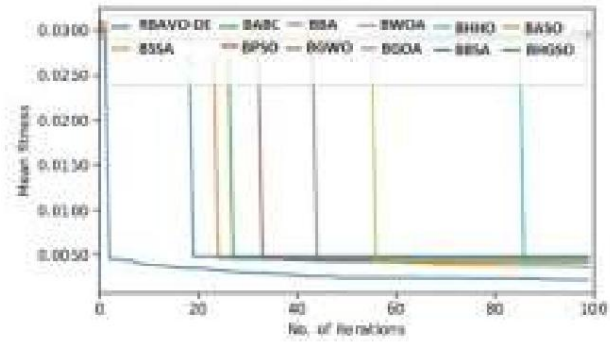
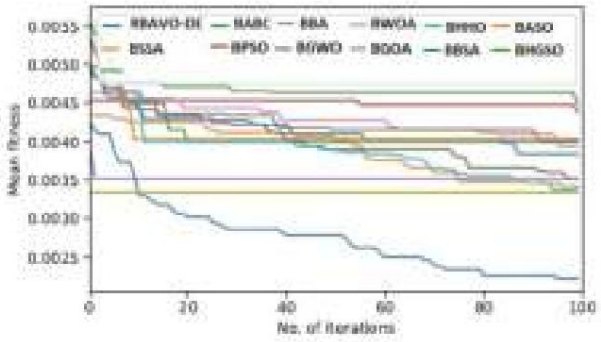


Abbildung A1. Fortsetzung

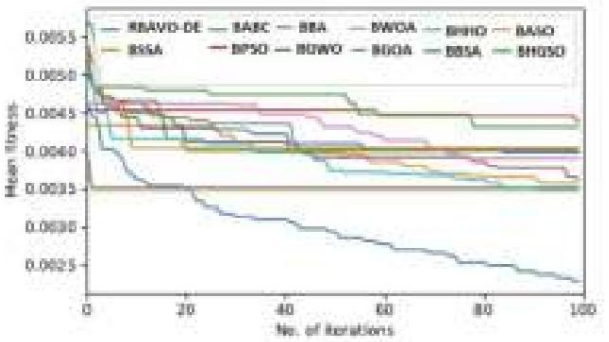
Machine Translated by Google



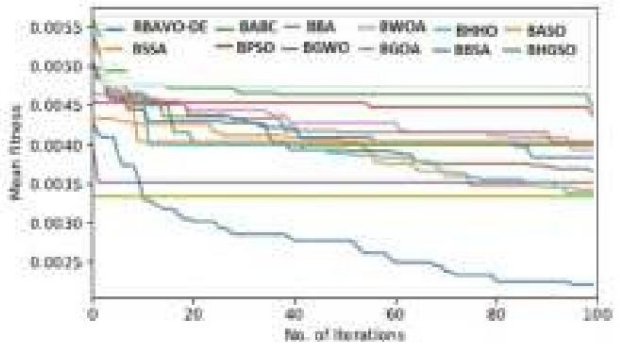
(e) ESCA



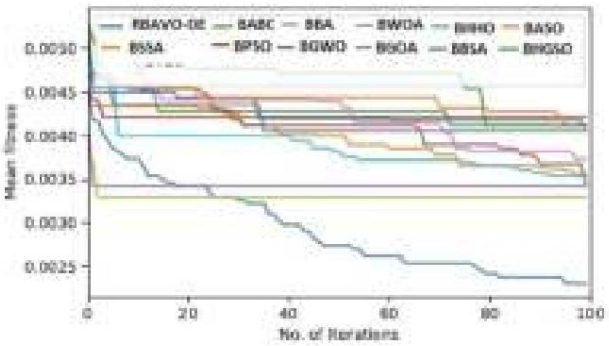
(f) GBM



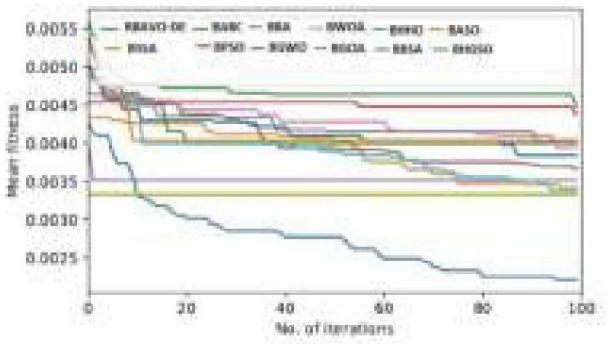
(g) HNSC



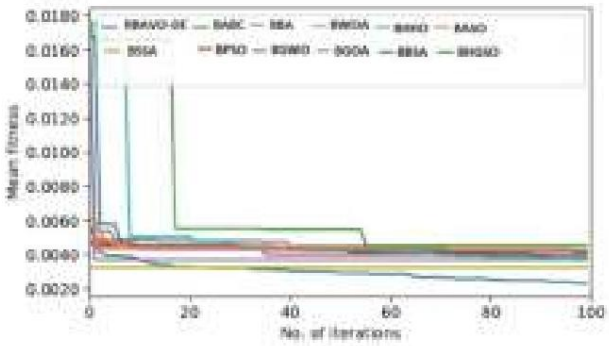
(h) KICH



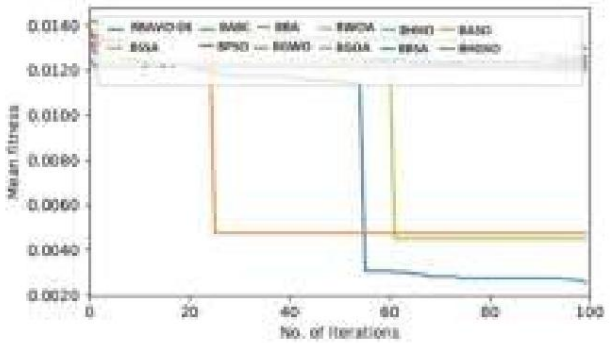
(i) KIRC



(j) KIRP



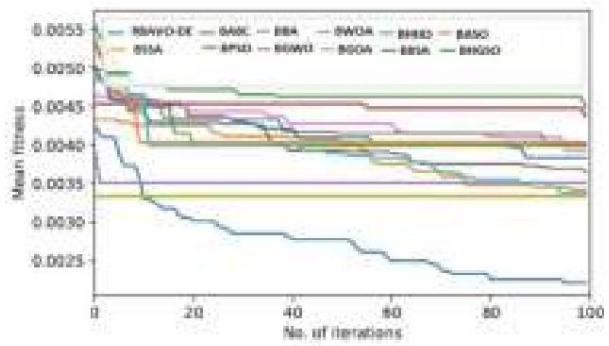
(k) LIHC



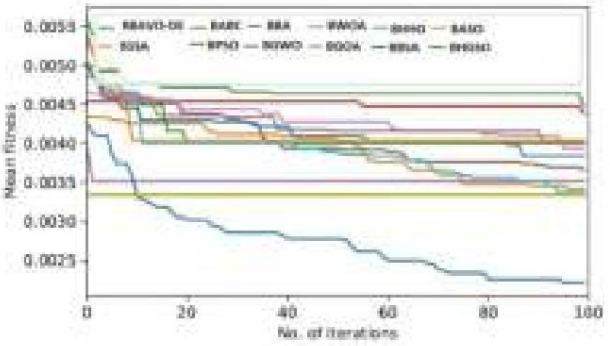
(l) LUAD

Abbildung A1. Fortsetzung

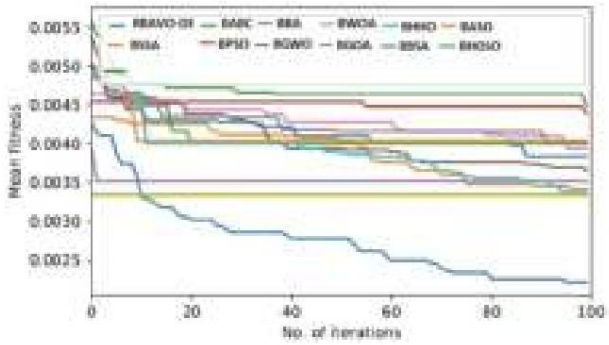
Machine Translated by Google



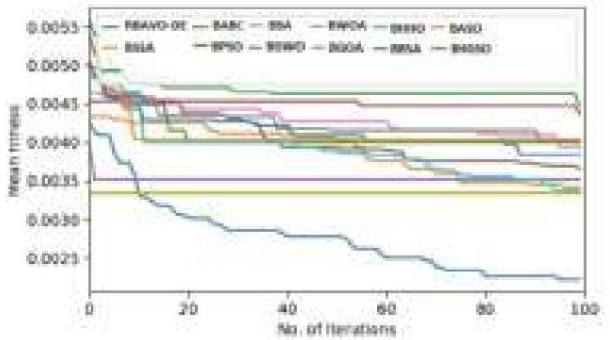
(m) LUSC



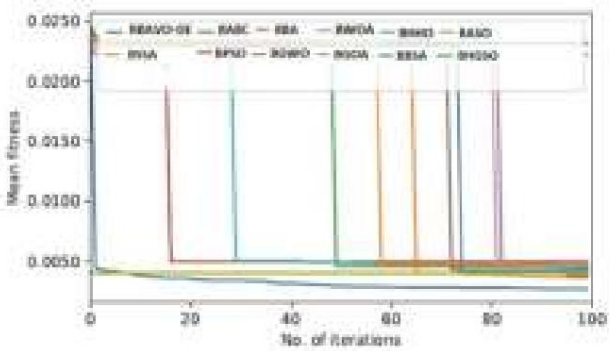
(n) PAAD



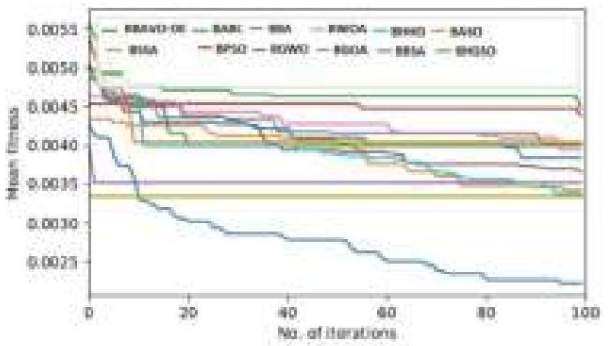
(o) PCPG



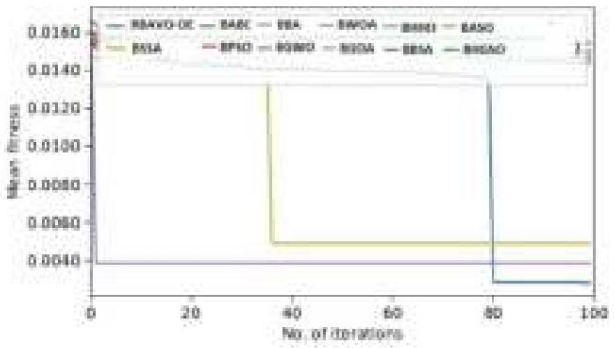
(p) READ



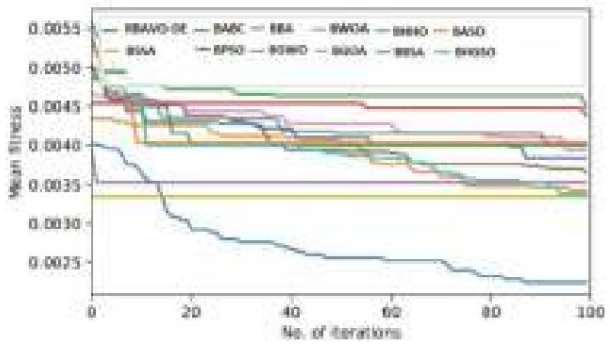
(q) SARC



(r) SKCM



(s) STAD



(t) THCA

Abbildung A1. Fortsetzung

Machine Translated by Google

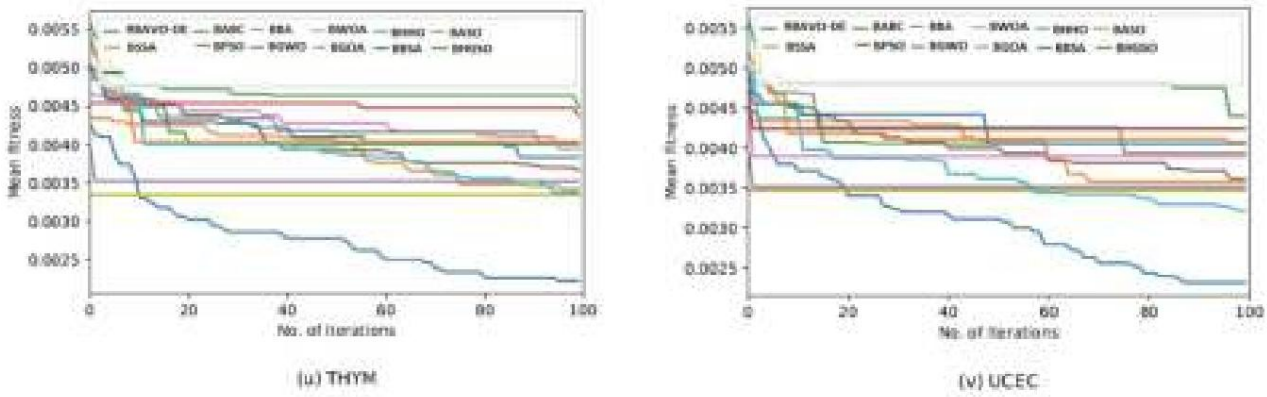


Abbildung A1. Konvergenzgraphen des vorgeschlagenen RBAVO-DE-Algorithmus und konkurrierender Methoden, die das SVM-Modell verwenden, bei den allgemeinen Benchmarks [57].

Anhang B.2. Konvergenzgraphen unter Verwendung des k-NN-Modells

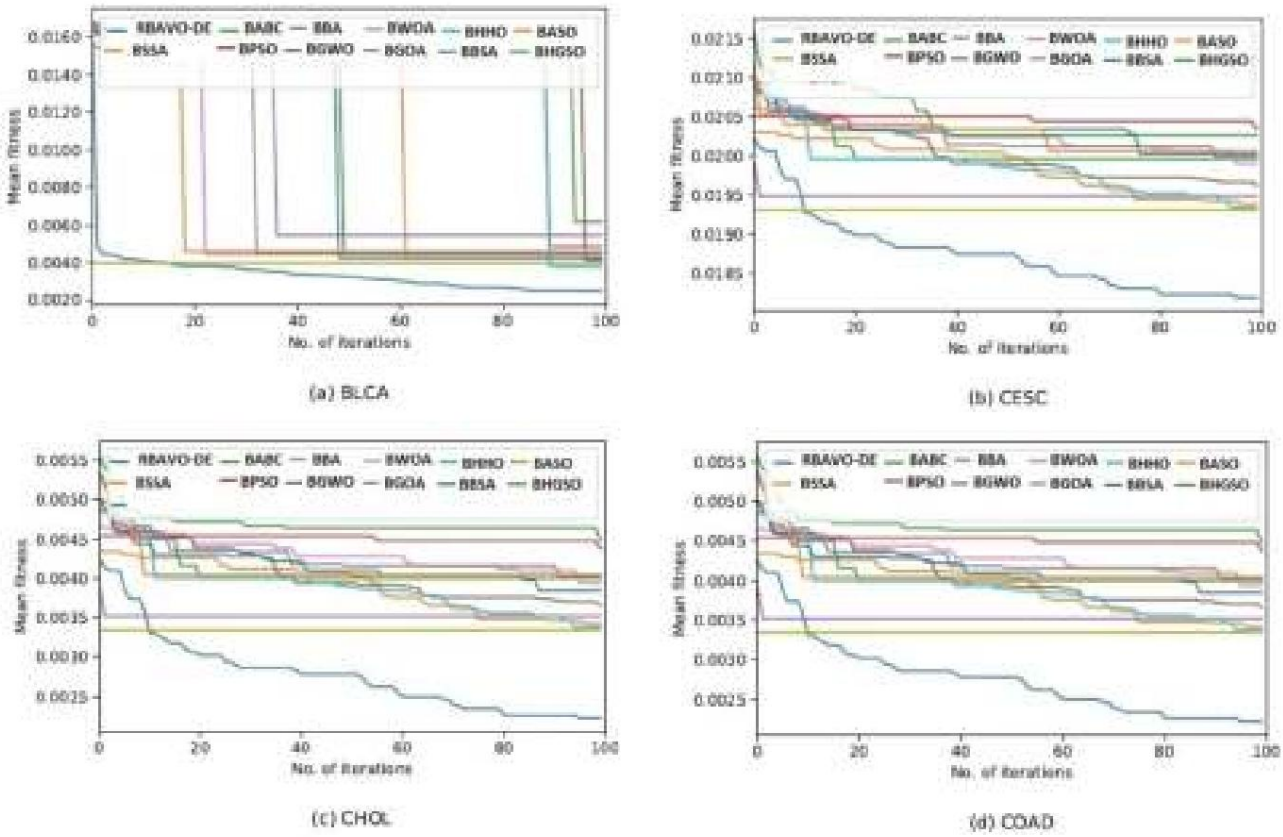
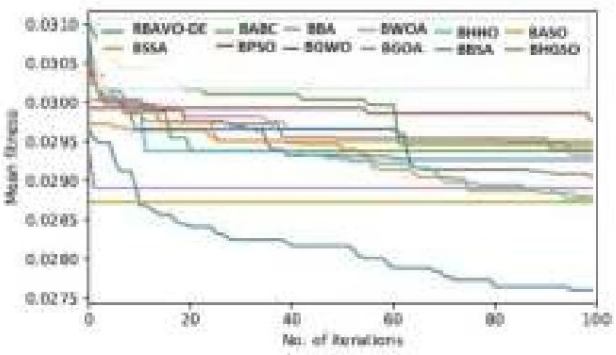
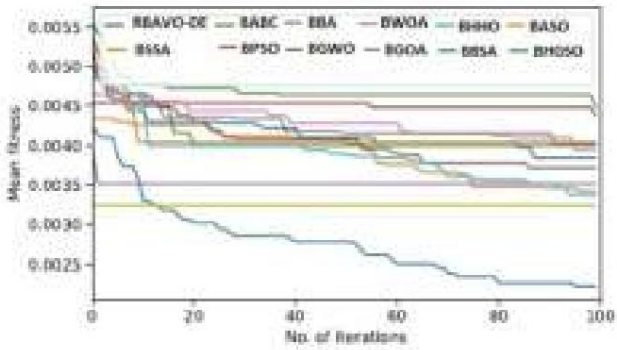


Abbildung A2. Fortsetzung

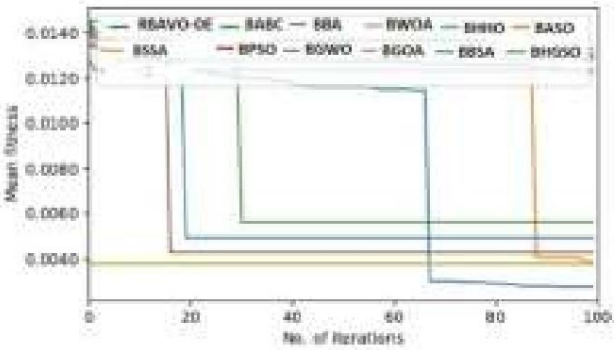
Machine Translated by Google



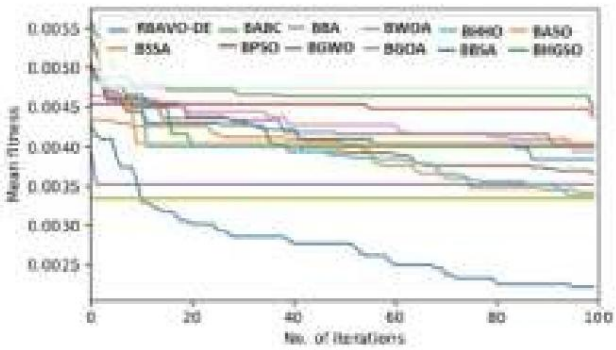
(e) ESCA



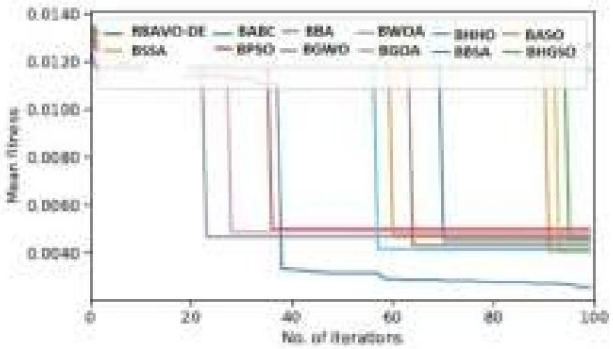
(f) GBM



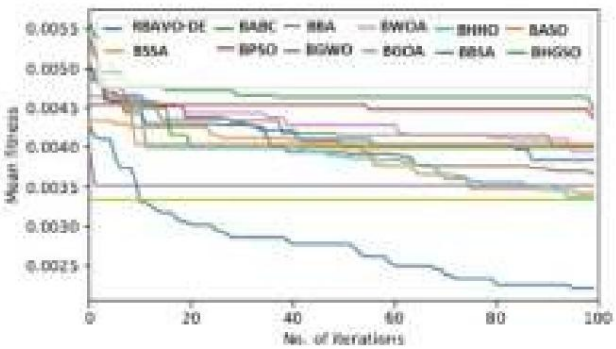
(g) HNSC



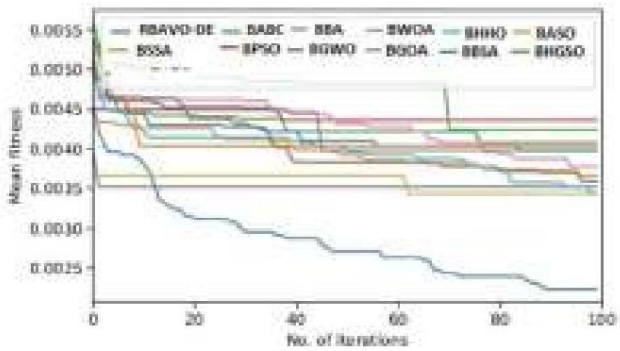
(h) KICH



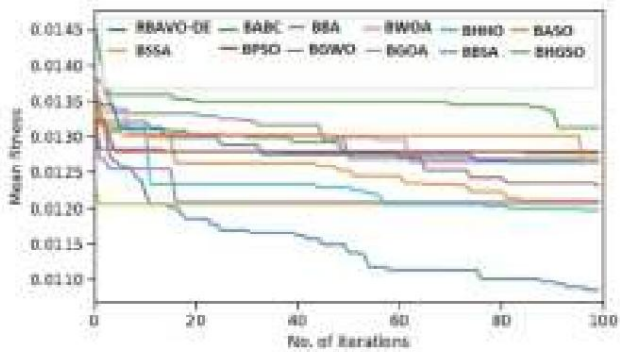
(i) KIRC



(j) KIRP



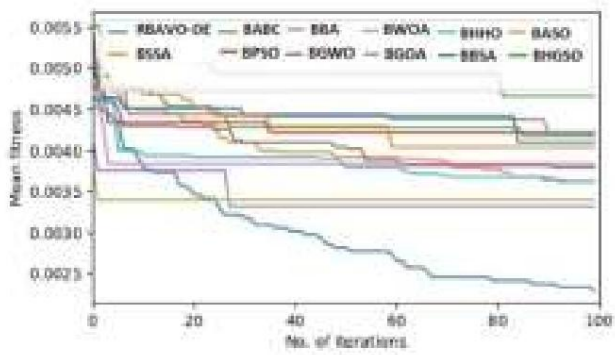
(k) LIHC



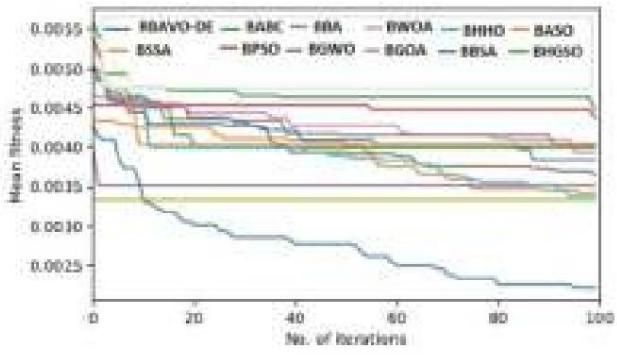
(l) LUAD

Abbildung A2. Fortsetzung

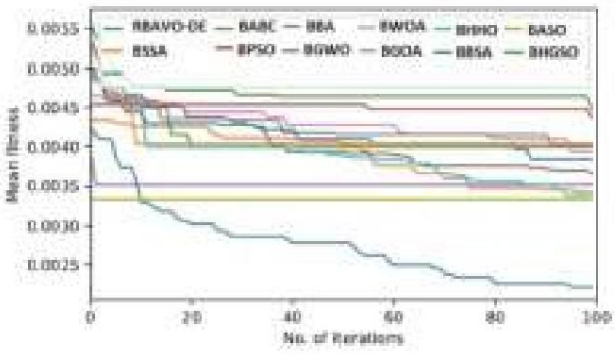
Machine Translated by Google



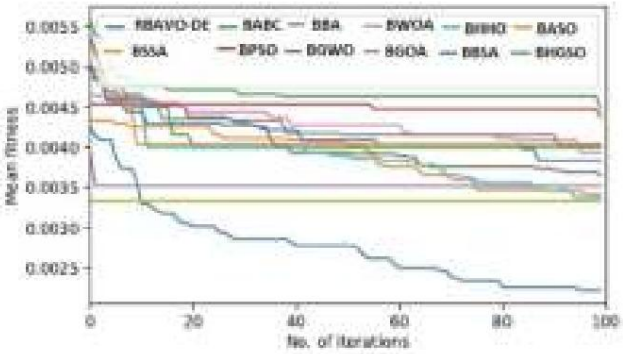
(m) LUSC



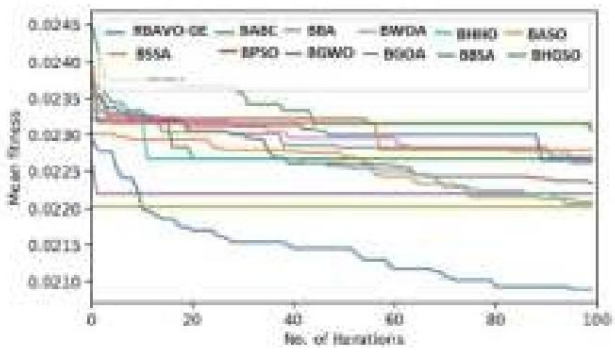
(n) PAAD



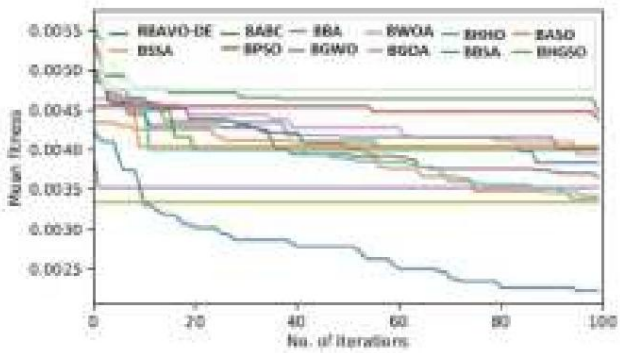
(o) PCPG



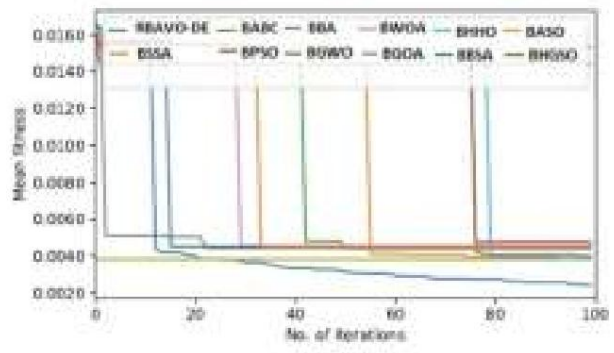
(p) READ



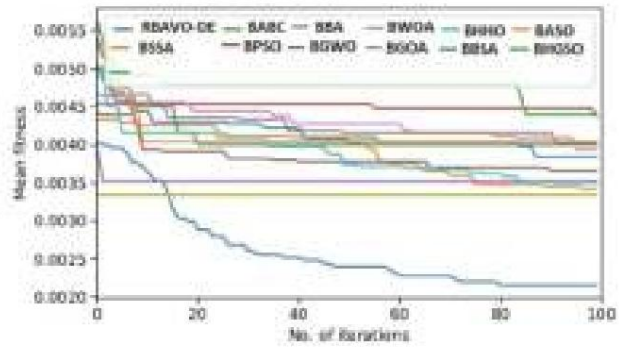
(q) SARC



(r) SKCM



(s) STAD



(t) THCA

Abbildung A2. Fortsetzung

Machine Translated by Google

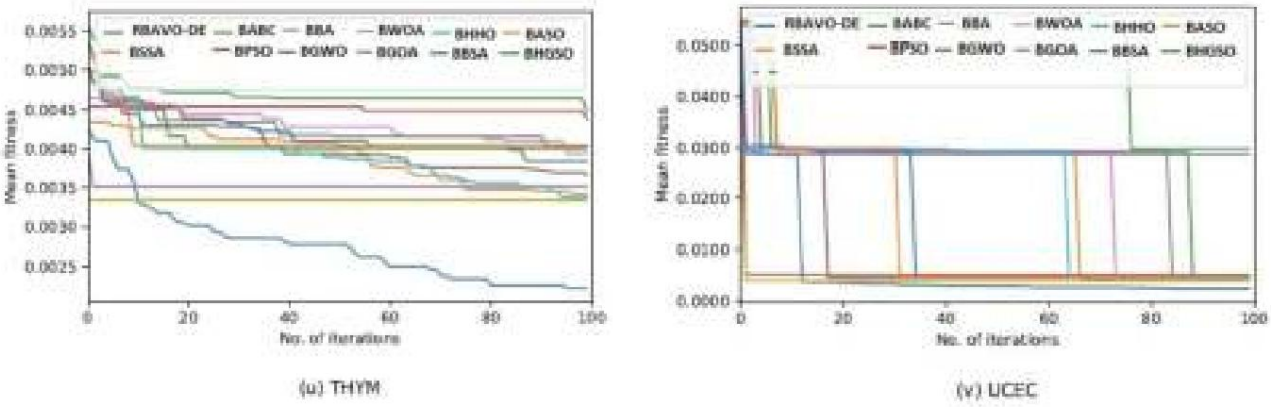


Abbildung A2. Konvergenzgraphen des vorgeschlagenen RBAVO-DE-Algorithmus und konkurrierender Methoden, die das k-NN-Modell verwenden, bei den allgemeinen Benchmarks (Forts.) [57].

Verweise

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. Neue Erkenntnisse zur durch Mikro-RNA und lange nicht-kodierende RNA vermittelten Regulierung von Stammzellpopulationen.** 3 Biotech **2022,** 12, **270.** [CrossRef] [PubMed]
2. **Kakati, T.; Bhattacharyya, DK; Kalita, JK; Norden-Krichmar, TM DEGnext: Klassifizierung von differentiell exprimierten Genen aus RNA-Sequenzdaten unter Verwendung eines Convolutional Neural Network mit Transfer Learning.** BMC Bioinform. **2022,** 23, **17.** [CrossRef]
3. **Zhao, S.; Fung-Leung, WP; Bittner, A.; Ngo, K.; Liu, X. Vergleich von RNA-Seq und Microarray im Transkriptom-Profiling aktivierter T-Zellen.** PLoS ONE **2014, 9, e78644.** [CrossRef]
4. **Chen, Z.; Luo, Z.; Zhang, D.; Li, H.; Liu, X.; Zhu, K.; Zhang, H.; Wang, Z.; Zhou, P.; Ren, J.; et al. TIGER: Ein Webportal mit Ressourcen zur Genexpression bei Tumorimmuntherapien.** Genom. Proteom. Bioinform. **2023,** 21, **337–348.** [CrossRef] [PubMed]
5. **Nunez-Garcia, J.; AbuOun, M.; Storey, N.; Brouwer, M.; Delgado-Blas, J.; Mo, SS; Ellaby, N.; Veldman, K.; Haenni, M.; Châtre, P.; et al. Harmonisierung von In-silico-Sequenzierungsmethoden der nächsten Generation für Diagnostik und Überwachung.** Sci. Rep. **2022,** 12, **14372.** [CrossRef]
6. **Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: Ein revolutionäres Werkzeug für die Transkriptomik.** Nat. Rev. Genet. **2009,** 10, **57–63.** [CrossRef]
7. **Kim, WJ; Choi, BR; Noh, JJ; Lee, YY; Kim, TJ; Lee, JW; Kim, BG; Choi, CH Vergleich von RNA-Seq und Microarray im Vorhersage der Proteinexpression und Überlebensvorhersage.** Front. Genet. **2024,** 15, **1342021.** [CrossRef]
8. **Wang, M.; Chen, X.; Dai, Y.; Wu, D.; Liu, F.; Yang, Z.; Song, B.; Xie, L.; Yang, L.; Zhao, W.; et al. Konkordanzstudie eines auf 520-Gen-Sequenzierung der nächsten Generation basierenden genomischen Profilierungstests von Gewebe- und Plasmaproben.** Mol. Diagn. Ther. **2022,** 26 [Querverweis]
9. **Metzker, ML Sequenzierungstechnologien – Die nächste Generation.** Nat. Rev. Genet. **2010,** 11, **31–46.** [CrossRef] [PubMed]
10. **Pandey, D.; Onkara Perumal, P. Ein Überblick über Deep Learning für die RNA-Seq-Datenanalyse der nächsten Generation.** Funkt. Integr. Genom. **2023,** 23, **134.** [CrossRef]
11. **Liu, S.; Yao, W. Vorhersage von Lungenkrebs durch Genexpression und Deep Learning mit KL-Divergenzgenauswahl.** BMC Bioinform. **2022,** 23, **175.** [CrossRef] [PubMed]
12. **Houssein, EH; Oliva, D.; Celik, E.; Emam, MM; Ghoniem, RM Verstärkter Rußseeschwalben-Optimierungsalgorithmus für globale Optimierung und Merkmalsauswahl.** Expert Syst. Appl. **2023,** 213, **119015.** [CrossRef]
13. **Joshi, AA; Aziz, RM Ein zweiphasiger, auf Cuckoo-Suche basierender Ansatz zur Genauswahl und Deep-Learning-Klassifizierung von Krebserkrankungen unter Verwendung von Genexpressionsdaten mit einer neuartigen Fitnessfunktion.** Multimed. Tools Appl. **2024,** 83, **71721–71752.** [CrossRef]
14. **Ramaswamy, R.; Kandhasamy, P.; Palaniswamy, S. Merkmalsauswahl für Alzheimer-Genexpressionsdaten unter Verwendung modifizierter binärer Partikelschwarmoptimierung.** IETE J. Res. **2023,** 69, **9–20.** [CrossRef]
15. **Cui, X.; Li, Y.; Fan, J.; Wang, T. Ein neuartiger Filtermerkmalauswahlalgorithmus basierend auf Relief.** Appl. Intell. **2022,** 52, **5063–5081.** [Querverweis]
16. **Alhenawi, E.; Al-Sayyed, R.; Hudaib, A.; Mirjalili, S. Methoden zur Merkmalsauswahl auf Basis von Genexpression-Microarray-Daten zur Krebsklassifizierung : Eine systematische Übersicht.** Comput. Biol. Med. **2022,** 140, **105051.** [CrossRef] [PubMed]
17. **Parlak, B.; Uysal, AK Eine neuartige Filtermerkmalsauswahlmethode für die Textklassifizierung: Extensive Feature Selector.** J. Inf. Sci. **2023,** 49, **59–78.** [CrossRef]
18. **Albulayhi, K.; Abu Al-Haija, Q.; Alsuhibany, SA; Jillepalli, AA; Ashrafuzzaman, M.; Sheldon, FT IoT-Einbruchserkennung mittels maschinellem Lernen mit einer neuartigen leistungsstarken Merkmalsauswahlmethode.** Appl. Sci. **2022,** 12, **5015.** [CrossRef]
19. **Fatima, A.; Nazir, T.; Nazir, AK; Din, AMU Eine effiziente inkrementelle Wrapper-basierte Methode zur Auswahl von Gensubgruppen mit Informationsgewinn (IG basierend auf IWSSr) zur Tumorunterscheidung.** Multimed. Tools Appl. **2024,** 83, **64741–64766.** [CrossRef]
20. **Kaur, S.; Kumar, Y.; Koul, A.; Kumar Kamboj, S. Eine systematische Übersicht über metaheuristische Optimierungstechniken für Merkmalsauswahlen in der Krankheitsdiagnose: Offene Fragen und Herausforderungen.** Arch. Comput. Methods Eng. **2023,** 30, **1863–1895.** [CrossRef]

Machine Translated by Google

21. **Abd El-Mageed, AA; Gad, AG; Sallam, KM; Munasinghe, K.; Abohany, AA Verbesserte binäre adaptive windgesteuerte Optimierungsalgorithmus-basierte Dimensionsreduktion für überwachte Klassifizierung.** Comput. Ind. Eng. **2022,** 167, **107904.** [Querverweis]
22. **Gad, AG; Sallam, KM; Chakrabortty, RK; Ryan, MJ; Abohany, AA Ein verbesserter binärer Sparrow-Suchalgorithmus für Merkmals Selektion bei der Datenklassifizierung.** Neural Comput. Appl. **2022,** 34, **15705–15752.** [CrossRef]
23. **Hussien, RM; Abohany, AA; Abd El-Mageed, AA; Hosny, KM Verbesserter binärer Meerkat-Optimierungsalgorithmus für eine effiziente Merkmalsauswahl bei der Klassifizierung durch überwachtes Lernen.** Knowl.-Based Syst. **2024,** 292, **111616.** [CrossRef]
24. **Abd El-Mageed, AA; Abohany, AA; Elashry, A. Effektive Merkmalsauswahlstrategie für überwachte Klassifizierung basierend auf einer Verbesserter binärer Aquila-Optimierungsalgorithmus.** Comput. Ind. Eng. **2023,** 181, **109300.** [CrossRef]
25. **Yin, Y.; Jang-Jaccard, J.; Xu, W.; Singh, A.; Zhu, J.; Sabrina, F.; Kwak, J. IGRF-RFE: Eine hybride Merkmalsauswahlmethode für MLP-basierte Netzwerk-Intrusion-Erkennung auf dem UNSW-NB15-Datensatz.** J. Big Data **2023,** 10, **15.** [CrossRef]
26. **Nakao, H.; Imaoka, M.; Hida, M.; Imai, R.; Nakamura, M.; Matsumoto, K.; Kita, K. Bestimmung der einzelnen Faktoren im Zusammenhang mit Hallux valgus mittels SVM-RFE.** BMC Musculoskelet. Disord. **2023,** 24, **534.** [CrossRef] [PubMed]
27. **Sarafrazi, S.; Nezamabadi-Pour, H. Die Klassifizierung binärer Probleme mit einem GSA-SVM-Hybridsystem bewältigen.** Mathe. Comput. Modell. **2013,** 57, **270–278.** [CrossRef]
28. **Cadenas, JM; Garrido, MC; MartíNez, R. Feature-Subset-Auswahlfilter–Wrapper basierend auf Daten niedriger Qualität.** Expert Syst. Appl. **2013,** 40, **6241–6252.** [CrossRef]
29. **Oh, IS; Lee, JS; Moon, BR Hybride genetische Algorithmen zur Merkmalsauswahl.** IEEE Trans. Pattern Anal. Mach. Intell. **2004,** 26, **1424–1437.**
30. **Abdollahzadeh, B.; Gharehchopogh, FS; Mirjalili, S. African vultures optimization algorithm: Ein neuer, von der Natur inspirierter metaheuristischer Algorithmus für globale Optimierungsprobleme.** Comput. Ind. Eng. **2021,** 158, **107408.** [CrossRef]
31. **El-Shafeiy, E.; Hassanien, AE; Sallam, KM; Abohany, A. Ansatz zum Training eines Quantenneuronalen Netzwerks zur Vorhersage der Schwere von COVID-19 bei Patienten.** Comput. Mater. Contin. **2020,** 66, **1745–1755.** [CrossRef]
32. **Yaqoob, A.; Verma, NK; Aziz, RM Optimierung der Genauswahl und Krebsklassifizierung mit hybriden Sinus-Cosinus- und Kuckuckssuchalgorithmen .** J. Med. Syst. **2024,** 48, **10.** [CrossRef]
33. **Joshi, AA; Aziz, RM Deep-Learning-Ansatz zur Klassifizierung von Hirntumoren mittels metaheuristischer Optimierung mit Gen Expressionsdaten.** Int. J. Imaging Syst. Technol. **2023,** 34, **e23007.** [CrossRef]
34. **Mahto, R.; Ahmed, SU; Rahman, Ru; Aziz, RM; Roy, P.; Mallik, S.; Li, A.; Shah, MA Ein neuartiges und innovatives Rahmenwerk zur Klassifizierung von Krebs durch eine aufeinanderfolgende Nutzung der hybriden Merkmalsauswahl.** BMC Bioinform. **2023,** 24, **479.** [Cr [PubMed]
35. **Neggaz, N.; Neggaz, I.; Abd Elaziz, M.; Hussien, AG; Abulaigh, L.; Damaševiÿcius, R.; Hu, G. Verstärkung des Futtersuchoptimierers für Mantas durch Trigonometrieoperatoren: Eine Fallstudie zu medizinischen Datensätzen.** Neural Comput. Appl. **2024,** 36, **9405–9436.** [CrossRef]
36. **Lyu, B.; Haque, A. Deep Learning-basierte Tumortypklassifizierung unter Verwendung von Genexpressionsdaten. In Proceedings der 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, Washington, DC, USA, 29. August–1. September 2018; S. 89–96.**
37. **Khalifa, NEM; Taha, MHN; Ali, DE; Slowik, A.; Hassanien, AE Künstliche Intelligenztechnik für die Genexpression durch Tumor-RNA-Seq-Daten: Ein neuartiger optimierter Deep-Learning-Ansatz.** IEEE Access **2020, 8, 22874–22883.** [CrossRef]
38. **Dillies, MA; Rau, A.; Aubert, J.; Hennequet-Antier, C.; Jeanmougin, M.; Servant, N.; Keime, C.; Marot, G.; Castel, D.; Estelle, J.; et al. Eine umfassende Bewertung von Normalisierungsmethoden für die Hochdurchsatz-RNA-Sequenzierungsdatenanalyse von Illumina.** Briefings Bioinform. **2013,** 14, **671–683.** [CrossRef]
39. **Xiao, Y.; Wu, J.; Lin, Z.; Zhao, X. Eine auf Deep Learning basierende Multi-Modell-Ensemble-Methode zur Krebsvorhersage.** Computermethoden Programs Biomed. **2018,** 153, **1–9.** [CrossRef]
40. **Liu, M.; Xu, L.; Yi, J.; Huang, J. Eine Methode zur Merkmalsgenauswahl basierend auf ReliefF und PSO. In Proceedings der 10. Internationalen Konferenz für Messtechnik und Mechatronik-Automatisierung (ICMTMA) 2018, Changsha, China, 7.–8. Januar 2018; IEEE: New York, NY, USA, 2018; S. 298–301.**
41. **Kononenko, I. Schätzung von Attributen: Analyse und Erweiterungen von RELIEF. In Proceedings der European Conference on Machine Learning, Catania, Italien, 6.–8. April 1994; Springer: Berlin/Heidelberg, Deutschland, 1994; S. 171–182.**
42. **Faris, H.; Mafarja, MM; Heidari, AA; Aljarah, I.; Ala'M, AZ; Mirjalili, S.; Fujita, H. Ein effizienter binärer Salpenschwarmalgorithmus mit Crossover-Schema für Merkmalsauswahlprobleme.** Knowl.-Based Syst. **2018,** 154, **43–67.** [CrossRef]
43. **Abdel-Basset, M.; Ding, W.; El-Shahat, D. Ein hybrider Harris Hawks-Optimierungsalgorithmus mit simuliertem Annealing zur Merkmalsauswahl.** Artif. Intell. Rev. **2020,** 54, **593–637.** [CrossRef]
44. **Storn, R.; Price, K. Differentialevolution – Eine einfache und effiziente Heuristik für die globale Optimierung über kontinuierliche Räume.** J. Glob Optim. **1997,** 11, **341–359.** [CrossRef]
45. **Karaboga, D.; Basturk, B. Ein leistungsstarker und effizienter Algorithmus zur numerischen Funktionsoptimierung: Artificial Bee Colony (ABC) Algorithmus.** J. Glob. Optim. **2007,** 39, **459–471.** [CrossRef]
46. **Mirjalili, S.; Gandomi, AH; Mirjalili, SZ; Saremi, S.; Faris, H.; Mirjalili, SM Salp Swarm Algorithm: Ein bioinspirierter Optimierer für technische Konstruktionsprobleme.** Adv. Eng. Softw. **2017,** 114, **163–191.** [CrossRef]
47. **Khanesar, MA; Teshnehlab, M.; Shoorehdeli, MA Eine neuartige binäre Partikelschwarmoptimierung. In Proceedings der 2007 Mediterranean Conference on Control & Automation, Athen, Griechenland, 27.–29. Juni 2007; IEEE: New York, NY, USA, 2007; S. 1–6.**

Machine Translated by Google

48. Mirjalili, S.; Mirjalili, SM; Yang, XS Binärer Bat-Algorithmus. Neural Comput. Appl. **2014,** 25, **663–681. [CrossRef]**

49. Emary, E.; Zawbaa, HM; Hassanien, AE Binäre Grauwolf-Optimierungsansätze zur Merkmalsauswahl. Neurocomputing **2016,** 172, **371–381. [CrossRef]**

50. Hichem, H.; Elkamel, M.; Rafik, M.; Mesaaoud, MT; Ouahiba, C. Ein neuer binärer Grashüpfer-Optimierungsalgorithmus für das Merkmalsauswahlproblem . J. King Saud Univ. Comput. Inf. Sci. **2022,** 34, **316–328. [CrossRef]**

51. Hussien, AG; Hassanien, AE; Houssein, EH; Bhattacharyya, S.; Amin, M. S-förmiger binärer Wal-Optimierungsalgorithmus zur Merkmalsauswahl. In Recent Trends in Signal and Image Processing: ISSIP 2017; **Springer: Berlin/Heidelberg, Deutschland, 2019; S. 79–87.**

52. Zhao, W.; Wang, L.; Zhang, Z. Atomsuchoptimierung und ihre Anwendung zur Lösung eines hydrogeologischen Parameterschätzungsproblems . Knowl.-Based Syst. **2019,** 163, **283–304. [CrossRef]**

53. Meng, XB; Gao, XZ; Lu, L.; Liu, Y.; Zhang, H. Ein neuer bioinspirierter Optimierungsalgorithmus: Bird Swarm Algorithm. J. Exp. Theor. Artif. Intell. **2016,** 28, **673–687. [CrossRef]**

54. Hashim, FA; Houssein, EH; Mabrouk, MS; Al-Atabany, W.; Mirjalili, S. Henry Optimierung der Gaslöslichkeit: Eine neuartige Physik-basierender Algorithmus. Future Gener. Comput. Syst. **2019,** 101, **646–667. [CrossRef]**

55. Heidari, AA; Mirjalili, S.; Faris, H.; Aljarah, I.; Mafarja, M.; Chen, H. Harris Hawks-Optimierung: Algorithmus und Anwendungen. Future Gen. Comput. Syst. **2019,** 97, **849–872. [CrossRef]**

56. Normalisierter Level3-RNA-Seq-Genexpressionsdatensatz. Online verfügbar: https://gdac.broadinstitute.org/ (aufgerufen am 20. Dezember 2023).

57. El-Mageed, AAA; Elkhouli, AE; Abohany, AA; Gafar, M. Genauswahl über einen verbesserten Algorithmus zur Optimierung der Kernreaktion zur Krebsklassifizierung in hochdimensionalen Daten. J. Big Data **2024,** 11, **46. [CrossRef]**

58. Mafarja, M.; Mirjalili, S. Whale-Optimierungsansätze für die Auswahl von Wrapper-Funktionen. Appl. Soft Comput. **2018,** 62, **441–453. [Querverweis]**

59. Thaher, T.; Heidari, AA; Mafarja, M.; Dong, JS; Mirjalili, S. Binärer Harris Hawks-Optimierer für hochdimensionale Merkmalsauswahl mit geringer Stichprobengröße . In Evolutionary Machine Learning Techniques; **Springer: Berlin/Heidelberg, Deutschland, 2020; S. 251–272.**

60. Derrac, J.; García, S.; Molina, D.; Herrera, F. Ein praktisches Tutorial zur Verwendung nichtparametrischer statistischer Tests als Methode zum Vergleich von evolutionären und Schwarmintelligenzalgorithmen. Swarm Evol. Comput. **2011, 1, 3–18. [CrossRef]**

61. Python-Code für die Genauswahl durch Relief Binary African Vultures Optimization, integriert mit Differential Evolution. Verfügbar online: https://github.com/D-Amr-Atef/Gene_Selection_RBAVO-DE.git (abgerufen am 18. Mai 2024).

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Rezension

Ein Überblick über bestärkendes Lernen in der Produktionsplanung: Eine inferenzielle Perspektive

Vladimir Modrak [1,*], Ranjitharamasamy Sudhakarapandian [2], Arunmozhi Balamurugan [2] und Zuzana Soltysova [1]

[1] Fakultät für Fertigungstechnologien, Technische Universität Kosice, 080 01 Prešov, Slowakei;
[2] zuzana.soltysova@tuke.sk Schule für Maschinenbau, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, Indien; sudhakarapandian.r@vit.ac.in (RS); arunmozhi.b@vit.ac.in (AB)
* Korrespondenz: vladimir.modrak@tuke.sk

Zusammenfassung: In dieser Studie wurde eine systematische Überprüfung der Produktionsplanung auf der Grundlage von Techniken des bestärkenden Lernens (RL) unter Verwendung insbesondere bibliometrischer Analysen durchgeführt. Ziel dieser Arbeit ist es unter anderem, auf das wachsende Interesse an diesem Bereich hinzuweisen und den Einfluss von RL als Form des maschinellen Lernens auf die Produktionsplanung darzulegen. Um dies zu erreichen, untersucht das Papier die Produktionsplanung mithilfe von RL, indem es die beschreibenden Metadaten relevanter Veröffentlichungen in den Datenbanken Scopus, ScienceDirect und Google Scholar untersucht. Die Studie konzentriert sich auf ein breites Spektrum von Veröffentlichungen aus den Jahren zwischen 1996 und 2024. Die Ergebnisse dieser Studie können als neue Erkenntnisse für zukünftige Forschungsvorhaben im Bereich der Produktionsplanung mithilfe von RL-Techniken dienen.

Schlüsselwörter: Bibliometrische Analyse; Produktionsplanung; bestärkendes Lernen

check for updates

Zitat: Modrak, V.; Sudhakarapandian, R.; Balamurugan, A.; Soltysova, Z. Ein Überblick über Reinforcement Learning in der Produktionsplanung : Eine inferentielle Perspektive. Algorithms 2024, 17, 343. https://doi.org/10.3390/a17080343

Wissenschaftliche Herausgeber: Alexandre Dolgui, David Lemoine, María I. Restrepo und Frank Werner

Erhalten: 6. Juni 2024
Überarbeitet: 28. Juli 2024
Akzeptiert: 2. August 2024
Veröffentlicht: 7. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.
Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/ 4.0/).

1. Einleitung

Die Produktionsplanung gilt als eines der wichtigsten Elemente des Fertigungsmanagements, wenn es darum geht, die Produktionsaktivitäten an die Unternehmensziele anzupassen, einen reibungslosen Warenfluss zu gewährleisten und die Wettbewerbsfähigkeit des Unternehmens auf dem Markt zu sichern. Planungsalgorithmen spielen eine wichtige Rolle bei der Verbesserung der Effizienz und Effektivität der Produktion und sind daher seit langem Gegenstand umfangreicher Forschung in verschiedenen interdisziplinären Bereichen wie Wirtschaftsingenieurwesen, Automatisierung und Managementwissenschaften [1]. Die Aufgaben der Produktionsplanung können mithilfe von drei Haupttypen schrittweiser Verfahren gelöst werden, nämlich exakten Algorithmen, heuristischen Algorithmen und metaheuristischen Algorithmen [2,3]. Obwohl ein exakter Algorithmus theoretisch die optimale Lösung garantieren kann, macht die NP-Härte größerer Probleme ihre effektive und effiziente Lösung unmöglich [4]. Heuristiken verwenden eine Reihe von Regeln, um Planungslösungen schnell und effektiv zu erstellen, ohne die globale Optimierung zu b
Während Metaheuristiken gute Planungslösungen in einer angemessenen Rechenzeit erzeugen können , hängt die Art und Weise, wie Suchoperatoren diese erstellen, stark von der jeweiligen Situation ab [9–13]. Darüber hinaus stellt der iterative Suchprozess Herausforderungen in Bezug auf Zeitaufwand und Anwendbarkeit in Echtzeitszenarien bei der Behandlung groß angelegter Probleme dar. Planungsansätze, die auf Reinforcement Learning (RL) basieren, haben sich in dieser Hinsicht als nützliches Werkzeug erwiesen. Reinforcement Learning ist ein Teilgebiet des breiteren Bereichs des maschinellen Lernens. RL gilt als einer der perspektivischsten Ansätze für eine robuste kooperative Planung, die es Produktionsmanagern ermöglicht, mit einer komplexen Fertigungsumgebung zu interagieren, aus früheren Erfahrungen zu lernen und optimale Entscheidungen zu treffen. Dabei handelt es sich um den Prozess, bei dem ein Agent autonom Aktionen auswählt und ausführt,

Machine Translated by Google

Belohnungen, die es in bestimmten Szenarien erhält. Das Hauptziel von RL besteht darin, die kumulative Belohnung zu optimieren, die ein Agent durch die Bewertung und Auswahl von Aktionen in einer dynamischen Umgebung erhält **[14–20]. Die jüngste Entwicklung in der künstlichen Intelligenztechnologie hat eine erfolgreiche Anwendung von RL bei sequentiellen Entscheidungsfindungsproblemen mit mehreren Zielen ermöglicht, die für die Roboterplanung und -steuerung verwendet werden können [21,22]. Die Forschung zur Produktionsplanung mit RL hat sich seit 1998 im Vergleich zu Metaheuristiken zu fortschrittlichen Optimierungstechniken entwickelt. RL verbessert die Rechenleistung bei der Lösung von Planungsproblemen erheblich. Seit seiner Einführung im Jahr 1996 wurden zahlreiche Forschungsartikel zu RL zur Produktionsplanung verfasst, die eine solide und wertvolle Grundlage bilden (siehe einige davon, z. B. [23–27]), aber nur zwei aktuelle Übersichtsartikel [28,29] widmeten sich diesem Thema. Die beiden im Jahr 2021 veröffentlichten Artikel konzentrierten sich jedoch auf andere Perspektiven von RL-basierten Planungsalgorithmen als in diesem Artikel dargestellt. Mit anderen Worten steht der Inhalt des vorge**

Die Forschungsfrage dieses Artikels ist vielschichtig und umfasst mehrere Aspekte dieses Bereichs, die separate Antworten erfordern. Die wichtigsten Aspekte, die in diesem Artikel behandelt werden, sind: Was sind die wichtigsten neuen Forschungsbereiche in diesem Bereich? Welche verwandten Themen werden in der Produktionsplanung auf der Grundlage von RL behandelt? Was sind typische Implementierungsbereiche innerhalb von RL, die auf die Produktionsplanung angewendet werden?

Seine Neuheit besteht in der Bereitstellung zusätzlicher Analysen und Bewertungen zu RL in der Produktionsplanung. Es umfasst beispielsweise den Zitierungstrend für RL zur Produktionsplanung, die einflussreichsten Autoren in diesem Bereich, die relevantesten Quellen auf diesem Gebiet sowie einen Vergleich deterministischer Planungsmethoden und unsicherer Planungsmethoden .

2. Materialien und Methoden

Der hier verwendete bibliometrische Ansatz stellt als Hauptforschungsmethode ein quantitatives Instrument zur Überwachung und Darstellung des wissenschaftlichen Fortschritts durch die Untersuchung und Visualisierung wissenschaftlicher Informationen dar. Die wachsende Akzeptanz bibliometrischer Methoden in mehreren akademischen Disziplinen deutet darauf hin, dass ihre Verwendung die erwarteten Effekte bri Die vorliegende Untersuchung der RL-basierten Produktionsplanung wird mithilfe des aus fünf zusammenhängenden Phasen bestehenden Verfahrens durchgeführt, wie in Abbildung 1 dargestellt.

Darüber hinaus folgt diese Forschung einem inferentiellen Ansatz, bei dem die Stichprobe der Bevölkerung untersucht wird, um ihre Merkmale zu bestimmen [34,35]. Darüber hinaus wurde das von **Suárez [36]** vorgeschlagene **inferentielle Konzept der wissenschaftlichen Darstellung hier angewendet, um Forschungsergebnisse zu formulieren .** Sein Wesen besteht darin, alternatives Denken anzuwenden, um Ergebnisse zu erzielen, die sich von der **isomorphen Sichtweise der wissenschaftlichen Darstellung in dem Sinne unterscheiden,** dass empirisches Wissen eine **wichtige Rolle beim induktiven Denken** spielt **[37,38].**

Machine Translated by Google

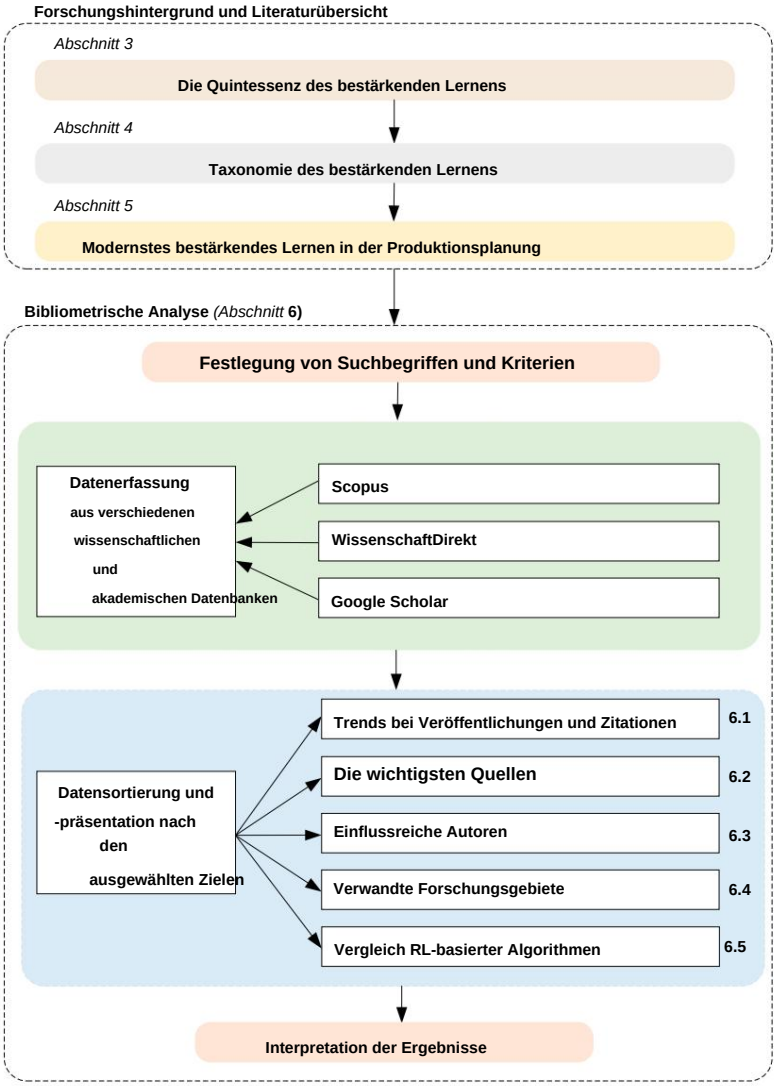


Abbildung 1. Rahmen der Forschungsmethodik.

3. Die Quintessenz des bestärkenden Lernens

Künstliche Intelligenz (KI) ist am Ende dieses Jahrzehnts zu einem integralen und bedeutenden Aspekt unseres täglichen Lebens geworden. Künstliche Intelligenz gibt es bereits seit sieben Jahrzehnten, hat aber in den letzten drei Jahrzehnten deutlich an Dynamik gewonnen, was zu umfangreicher Forschung auf diesem Gebiet geführt hat [39]. KI bezieht sich auf verschiedene Techniken , die es Computern ermöglichen, durch die Analyse von Daten Wissen zu erwerben und Entscheidungen zu treffen. Von diesen Techniken hat das maschinelle Lernen (ML) in den letzten zwei Jahrzehnten bedeutende Fortschritte gemacht und sich von einer bloßen Kuriosität in Laboren zu einer weit verbreiteten praktischen Technologie in kommerziellen Anwendungen entwickelt. In der KI ist ML zum bevorzugten Ansatz für die Erstellung funktionaler Software in Bereichen wie Computer Vision, Verarbeitung natürlicher Sprache, Spracherkennung, Robotersteuerung und verschiedener

Viele KI-Entwickler erkennen mittlerweile an, dass es in vielen Fällen bequemer ist, ein System zu trainieren, indem man ein ideales Input-Output-Verhalten demonstriert, als es manuell so zu programmieren, dass es für jeden möglichen Input die gewünschte Reaktion vorwegnimmt. ML hat verschiedene Bereiche der Informatik und Sektoren, die sich mit datengesteuerten Problemen befassen, erheblich beeinflusst. Dazu gehören Verbraucherdienstleistungen, die Diagnose von Defekten in komplexen Systemen und die Verwaltung von Logistikketten. ML-Ansätze hatten unterschiedliche Auswirkungen auf mehrere empirische Wissenschaften wie Biologie, Kosmologie und Sozialwissenschaften. Diese Ansätze wurden kreativ eingesetzt, um große experimentelle Datenmengen

Machine Translated by Google

Drei wichtige ML-Paradigmen sind überwachtes Lernen, unüberwachtes Lernen und bestärkendes Lernen [15]. Überwachtes Lernen umfasst den Prozess des Trainierens eines Modells anhand eines Datensatzes, der gekennzeichnet wurde, was bedeutet, dass jeder Eingabe die entsprechende genaue Ausgabe zugeordnet ist. Ziel ist es, Wissen über eine Transformation von gegebenen Eingaben in entsprechende Ausgaben zu erlangen, das dann verwendet werden kann, um genaue Prognosen für neue, unbeobachtete Daten zu erstellen. Unüberwachtes Lernen umfasst die Identifizierung verborgener Muster oder inhärenter Strukturen in unmarkierten Eingabedaten. Das Modell zielt darauf ab, Wissen über die grundlegende Verteilung oder Gruppierung der Daten zu erlangen, ohne explizite Anweisungen zum gewünschten Ergebnis zu erhalten [28]. Beim RL geht es darum, einem Agenten beizubringen, zahlreiche Entscheidungen zu treffen, indem er sich mit einer Umgebung beschäftigt und Belohnungen oder Bestrafungen als Feedback erhält . Der Agent erlangt die Fähigkeit, seine Aktivitäten zu optimieren, um die Gesamtbelohnungen, die er über einen bestimmten Zeitr

Aufgrund seiner dynamischen und adaptiven Methodik ist RL äußerst effektiv bei der Lösung von Problemen in den Bereichen Fertigungssysteme und Robotikforschung [22,28]. RL unterscheidet sich vom überwachten Lernen, bei dem ein Modell anhand eines vorgegebenen Datensatzes trainiert wird, und vom unüberwachten Lernen, bei dem das Modell verborgene Muster in den Daten entdeckt. Beim bestärkenden Lernen wird ein Agent dazu erzogen, eine Reihe von Entscheidungen zu treffen, indem er sich mit einer Umgebung auseinandersetzt. Das Ziel des Agenten besteht darin, eine Strategie zu entwickeln, die manchmal auch als Richtlinie bezeichnet wird und die im Laufe der Zeit die insgesamt erzielten Belo

RL ist ein Entscheidungsprozess, der von der Verhaltenspsychologie beeinflusst wird. Dabei lernt ein Agent, ein Ziel zu erreichen, indem er handelt und Feedback aus der Umgebung in Form von Belohnungen oder Strafen erhält. Diese Versuch-und-Irrtum-Methode ähnelt der Art und Weise, wie Menschen und Tiere von ihrer Umgebung lernen. Die kritischen Komponenten von RL sind der Agent, die Umgebung, Aktionen, Zustände und Belohnungen. Der Agent beobachtet die Umgebung, ergreift Aktionen und erhält Belohnungen, die als Leitfaden für zukünftige Aktionen dienen [14]. RL ist besonders effektiv bei Aufgaben mit ungewissem Ausgang, wie z. B. Spielen, Steuern von Robotern und Verwalten von Ressourcen. In diesen Situationen muss der Agent Erkundung (Ausprobieren neuer Aktionen) und Ausbeutung (Verwendung etablierter Aktionen, die zu erheblichen Belohnungen führen) ausbalancieren, um effizient eine optimale Strategie zu entwickeln [42].

Der mathematische Rahmen des RL wird üblicherweise mithilfe von Markov-Entscheidungsprozessen (MDPs) dargestellt, die einen systematischen Ansatz zur Modellierung von Entscheidungsproblemen bieten , die sowohl zufällige als auch agentengesteuerte Ergebnisse umfassen. Das Lösen eines MDPs beinhaltet die Festlegung einer Richtlinie, die die optimale Vorgehensweise in jeder Phase vorschreibt, mit dem Ziel, die prognostizierten kumulativen Belohnungen in der Zukunft zu maximieren. Die jüngsten Fortschritte im Bereich des Deep Learning haben die Kapazitäten des RL dramatisch verbessert [15,16]. Die Kombination aus tiefen neuronalen Netzwerken und RL-Algorithmen hat zur Entstehung des Deep Reinforcement Learning (DRL) geführt, das in komplizierten Situationen außergewöhnliche Erfolge gezeigt hat. Prominente Beispiele sind AlphaGo, entwickelt von Google, das im Spiel Go über menschliche Champions triumphierte, sowie verschiedene Implementierungen im autonomen Fahren, bei denen DRL-Algorithmen die Fähigkeit erlangen, in komplizierten und sich ständig ändernden Umgebungen erfolgreich zu navigieren. Reinforcement Learning bleibt ein dynamischer und sich schnell entwickelnder Bereich des maschinellen Lernens, in dem die laufende Forschung die Grenzen dessen, was künstliche Agenten leisten können, ständig erweitert. Wenn wir uns mit den Komplexitäten des RL befassen, entdecken wir neue Möglichkeiten für intelligente Systeme, die in der Lage sind, Wissen zu erwerben und ihr Verhalten in Echtzeit anzupassen. RL ebnet den Weg für eine Zukunft, in der Computer nahtlos mit Menschen zusammenarbeiten und auf fortgeschrittenere Weise autonom funktionieren können [42].

4. Taxonomie von Reinforcement-Learning-Algorithmen

Sutton und Barto [42] betonen die entscheidende Rolle der drei wesentlichen Komponenten im RL-Prozess: •

Eine Richtlinie bestimmt die in jedem Umgebungszustand auszuführenden Aktionen.
• Ein Belohnungssignal kategorisiert diese Aktionen als vorteilhaft oder nachteilig, basierend auf dem unmittelbaren Ergebnis des Übergangs zwischen Zuständen.

Machine Translated by Google

ff
• **Eine Wertfunktion beurteilt die langfristige Wirksamkeit von Maßnahmen, indem sie nicht nur die unmittelbare Belohnung eines Zustands, sondern auch die erwarteten zukünftigen Belohnungen berücksichtigt.**

Das Hauptziel des Agenten ist die Optimierung der kumulativen Belohnung. Folglich ist das Belohnungssignal die Grundlage für die Änderung der Richtlinie und der Wertfunktion. In bestimmten RL-Systemen gibt es eine zusätzliche Komponente, die als Modell der Umgebung bezeichnet wird und optional ist. Es repliziert das Verhalten des Klimas oder ermöglicht umfassendere Annahmen über das Verhalten der Umgebung. Modelle werden für die strategische Planung verwendet, bei der Entscheidungen auf der Grundlage der Erwartung zukünftiger Szenarien vor ihrem tatsächlichen Eintreten getroffen werden **[24]. Modellbasierte Methoden werden verwendet, um Probleme des bestärkenden Lernens durch die Verwendung von Modellen und Planung zu bewältigen, im Gegensatz zu einfacheren modellfreien Methoden. Sie weisen ein offensichtliches Versuch-und-Irrtum-Lernverhalten auf und werden als das Gegenteil von Planung betrachtet [26,27]. RL-Prozesse können mithilfe eines mathematischen Rahmens namens MDP dargestellt und analysiert werden. Das stochastische mathematische Modell kann mithilfe eines 5-Tupels (s, a, p, r, ÿ) formalisiert** werden, in dem das erste Symbol „s“ die endliche **Sammlung aller möglichen Umgebungszustände darstellt ,** während **st—der 'a' stellt die Menge aller möglichen Aktionen dar, während die zum Zeitpunkt 't' ausgeführte Aktion als bei** markiert ist . **Symbol 'p' bezieht sich auf die Übergangswahrscheinlichkeitsmatrix, die die bedingte Wahrscheinlichkeit des Übergangs in einen neuen Zustand 's' mit einer Belohnung 'r' bei gegebenem aktuellen Zustand 's' und Aktion 'a' (für alle Zustände und Aktionen)** definiert **[44]. Abbildung 2 zeigt das RL des Produktionsplanungszyklus .**

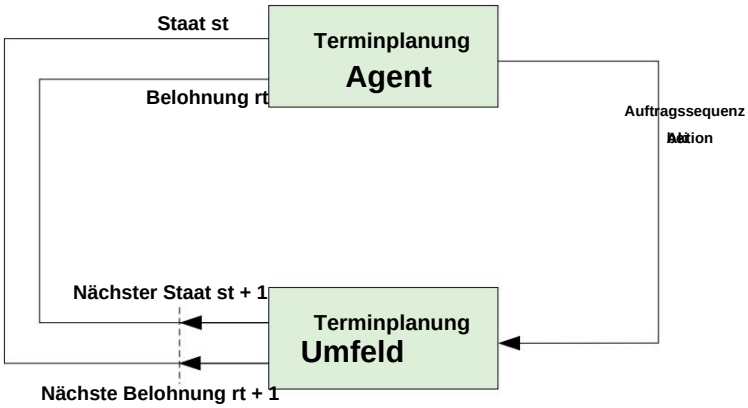


Abbildung 2. RL-Produktionsplanungszyklus.

Die Lernmethoden von RL-Algorithmen oder -Agenten können in zwei unterschiedliche Kategorien eingeteilt werden: modellbasiert und modellfrei. Modellbasiertes RL wird oft als indirektes Lernen bezeichnet. Der Agent verwendet ein prädiktives Modell, um die Kontrollrichtlinie aus der Umgebung durch eine begrenzte Anzahl von Interaktionen zu lernen. Der Agent wendet dieses Modell dann auf nachfolgende Episoden an, um Belohnungen zu erhalten [15]. Modellfreies RL oder direktes Lernen bezieht sich auf Lernen, bei dem ein Agent lernt, Entscheidungen zu treffen, ohne explizit ein Modell der Umgebung zu erstellen. Der Agent erwirbt Wissen über die Kontrollrichtlinie durch Erfahrungslernen aus der Umgebung und verwendet Versuch und Irrtum, um Belohnungen zu optimieren, ohne sich auf ein bereits vorhandenes Modell zu verlassen. Dieser Ansatz zeigt die Anpassungsfähigkeit von RL-Algorithmen, die es ihnen ermöglicht, in dynamischen Umgebungen zu lernen und sich weiterzuentwickeln [45].

Da es viele verschiedene RL-Algorithmen gibt, ist es sinnvoll, die Unterschiede zwischen ihnen zu verstehen . Zu diesem Zweck wurden Klassifizierungssysteme entwickelt, um sie nach verschiedenen Kriterien zu kategorisieren. Beispielsweise schlugen ALMahamid und Grolinger [46] vor, RL-Algorithmen basierend auf dem Umgebungstyp zu kategorisieren. RL-Algorithmen können auch aus der Perspektive der Richtlinie klassifiziert werden: On-Policy- vs. Off-Policy-Lernen [47,48]. RL werden hauptsächlich nach ihren Lernansätzen klassifiziert [15,45]. In dieser Hinsicht ist **in Abbildung 3 eine aktualisierte Klassifizierung von RL-Algorithmen aus der Perspektive der verwendeten Lernansätze dargestellt .**

Machine Translated by Google

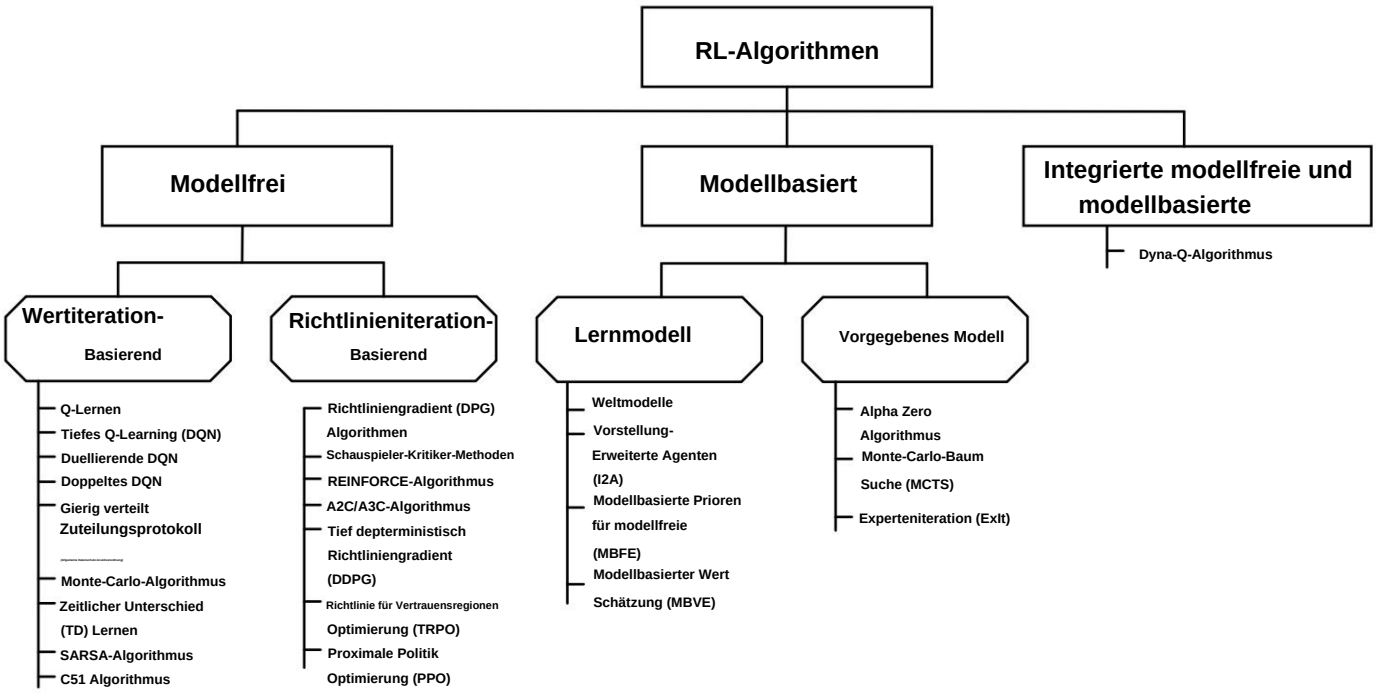


Abbildung 3. Klassifizierung von RL-Algorithmen.

5. Stand des Reinforcement Learning in der Produktionsplanung

Die deterministische Planung umfasst die Organisation und Planung von Aufgaben unter festen Parametern und bekannten Einschränkungen. Ihr Ziel ist die Optimierung der Produktionseffizienz und die Minim Eine umfassende Arbeit [49] untersucht grundlegende Prinzipien und fortgeschrittene Techniken der deterministischen Planung und gibt Einblicke in verschiedene Algorithmen und ihre Anwendungen. Pinedos Arbeit [50] befasst sich eingehend mit deterministischer Planung Theorie, Algorithmen und praktische Systeme, wobei ihre Bedeutung in der Fertigung hervorgehoben wird und Service-Operationen. Die Flowshop-Planung umfasst die Sequenzierung von Operationen auf mehreren Maschinen in einer festgelegten Reihenfolge, um die Durchlaufzeit oder die Gesamtfertigstellungszeit zu minimieren. Allahverdi [51] diskutiert verschiedene Flowshop-Planungsprobleme, einschließlich Setups, und überprüft Algorithmen und Ansätze zu ihrer Bewältigung. Panwalkar und Smith [52] bieten eine grundlegende Übersicht klassischer und zeitgenössischer Forschung zur Flowshop-Planung, die sowohl die exakte als auch die heuristische Methoden.

Techniken des Reinforcement Learning (RL) haben an Popularität gewonnen und bestätigen das Interesse an agentenbasierten Modellen. Frühere Studien konzentrierten sich hauptsächlich auf die Verwendung von lösen Herausforderungen bei der Auftragsplanung. Der Fertigungssektor steht vor Herausforderungen wie wie Kundenzufriedenheit, Systemverschlechterung, Nachhaltigkeit, Lagerbestand und Effizienz, was sich auf die Nachhaltigkeit und Rentabilität der Anlage auswirkt. Industrie 4.0 und Smart Manufacturing Lösungen für optimierte Abläufe und qualitativ hochwertige Produkte. Paraschos et al. [53] integrieren RL mit Lean Green Manufacturing, um eine nachhaltige Produktionsumgebung zu schaffen. Reduzierung der Umweltbelastung durch minimierten Materialverbrauch und Lebenszyklus Erweiterung durch Pull-Produktion, vorausschauende Wartung und Praktiken der Kreislaufwirtschaft. Eine strenge experimentelle Analyse bestätigt die Wirksamkeit bei der Verbesserung der Nachhaltigkeit und Materialwiederverwendung.

In jüngster Zeit wurden bedeutende Fortschritte bei der Verwendung von RL zur Lösung verschiedener kombinatorischer Optimierungsprobleme erzielt, darunter Produktionsplanung, Fahrzeug-Routing Problem des Handlungsreisenden [54–59]. Im RL kann eine Produktionsplanungsaufgabe wie folgt betrachtet werden: eine Umgebung, in der ein Agent arbeitet, Entwicklung einer Richtlinie durch Offline-Schulung durch Interaktion mit dieser Umgebung. Dieser Ansatz bietet eine neuartige Möglichkeit, die Planung Herausforderungen, die strenge Echtzeitbeschränkungen erfordern, wie z. B. eine dynamische Job-Shop-Planung Probleme [22,60–64].

Machine Translated by Google

Bei Produktionsplanungsproblemen werden häufig wertbasierte RL-Algorithmen verwendet, darunter Q-Learning, Temporal Difference TD(ÿ) -Algorithmus, SARSA, ARL, informiertes Q-Learning, duales Q-Learning, approximatives Q-Learning, Gradient Descent TD(ÿ) -Algorithmus, Revenue Sharing, Q-III-Lernen, relationales RL, Relaxed SMART und TD(ÿ)-Lernen. Beim Deep Reinforcement Learning (DRL) werden viele wertbasierte Ansätze verwendet, wie DQN (Deep Q-Learning Networks), lose gekoppeltes DRL, Multiclass DQN und der Q-Network- Algorithmus [43,44,65–73].

Qu et al. [25] verwendeten die Multi-Agenten-Approximation Q-Learning und demonstrierten deren Wirksamkeit durch numerische Experimente in statischen und dynamischen Umgebungen und verschiedenen Flow-Shop-Szenarien. Ziel war die Erstellung und Implementierung eines effizienten Fertigungsplans unter Berücksichtigung realistischer Wechselwirkungen zwischen Arbeitsfähigkeiten und adap Luo [74] verwendet die DRL-Methode, um die dynamische flexible Werkstattplanung zu behandeln, wobei er sich auf Szenarien zur Reduzierung von Verspätungen konzentriert. Luo et al. [75] verwendeten erstmals die „hierarchische Multi-Agent-Proximal-Policy-Optimierung“ (HMAPPO) für das sich ständig ändernde partielle, wartefreie, multiobjektive flexible Werkstattproblem (MOFJSP) mit neuen Stelleneinfügungen und Maschinenausfällen.

Maschinenverschleiß während der Produktion erhöht die Betriebskosten und verursacht Arbeitsunterbrechungen, was kostspielige Korrekturwartungen erforderlich macht. Vorbeugende Wartung verlängert die Lebensdauer der Maschine, bringt aber Ausfallzeiten und Kosten mit sich. Die Studie [76] befasst sich mit diesen Herausforderungen, indem sie Produktions- und Wartungspläne in Mehrmaschinensystemen optimiert und wissensbasiertes Reinforcement Learning einführt, um die Effektivität von RL bei der Steuerung von Produktionsentscheidungen zu verbessern und die Maschinenzusammenarbeit zu fördern. Vergleichende Bewertungen in deterministischen und stochastischen Szenarien unterstreichen die Fähigkeit des Algorithmus, Geschäftserträge zu maximieren und R

Laut Wang et al. [54] umfasst der Produktionsplanungsprozess die Herstellung verschiedener Artikel unter Verwendung eines hybriden Produktionsmusters, das das Multi-Agenten- Modell Deep Reinforcement Learning (MADRL) beinhaltet. Popper et al. [60] schlugen MADRL vor, um flexible Produktionsanlagen unter Berücksichtigung von Faktoren wie Effizienz und Umweltzielen zu optimieren . Du et al. [77] verwendeten die Deep Q-Network (DQN)-Methode, um das Problem der flexiblen Task-Shop -Planung (FJSP) bei sich ändernden Verarbeitungsraten, Rüstzeiten, Leerlaufzeiten und Aufgabentransporten zu lösen. Dieser Ansatz integriert Zustandsindikatoren und Aktionen, um die Effizienz der DQN-Komponente zu verbessern. Darüber hinaus enthält er eine problemorientierte explorative Datenanalyse (EDA)-Komponente zur Verbesserung der Datenexploration.

Li et al. [78] verwendeten eine Methode des Deep Reinforcement Learning (DRL), um das diskrete flexible Job-Shop-Problem mit Wiederverwendbarkeit zwischen Werkzeugen (DFJSP-ITR) zu lösen, und befassten sich mit dem Problem der mehrzieligen Optimierung, nämlich der Minimierung der kombinierten Fertigungsdauer und des Gesamtenergieverbrauchs . Die vorgeschlagene Lösung umfasst einige generische Zustandsmerkmale, einen auf genetischer Programmierung basierenden Aktionsraum und eine Belohnungsfunktion. Zhou et al. [79] schlugen vor , auf RL basierende Online-Planungsstrategien mit zusammengesetzten Belohnungsfunktionen zu verwenden , um die Effektivität und Robustheit industrieller Systeme zu verbessern. Die Arbeit [18] verwendete einen fortschrittlichen DRL- Algorithmus, um die Produktionsplanung in komplexen Job-Shops zu optimieren, und hob Vorteile wie verbesserte Anpassungsfähigkeit, globale Sichtbarkeit und Optimierung hervor. Einige Autoren haben sich auf die Entwicklung verschiedener DRL-Algorithmen konzentriert, mit denen komplexe Strategien für die Produktionsplanung formuliert werden können . So erstellten Luo et al. [26] ein Online-Umplanungsframework für das dynamische mehrzielige flexible Job-Shop-Planungsproblem, mit dem die Gesamtverspätung minimiert oder die Maschinenauslastung maximiert werden kann. Zhou et al. [27] verwendete die Deep-Q-Learning-Technik, um die dynamische Planung in der intelligenten Fertigung zu adressieren. Eine weitere dynamische Planungsmethode unter Verwendung von Deep RL wird in [80] vorgeschlagen, wobei proximale Richtlinienoptimierung verwendet wird, um die ideale Planungsrichtlinie zu bestimmen. Wang und Usher [81] untersuchten die Verwendung des Verstärkungs-Q-Learning-Algorithmus für die agentenbasierte Produktionsplanung. Die Integration von RL in die Produktionsplanung bedeutet einen großen Fortschritt bei der Optimierung von Fe

Machine Translated by Google

bieten wertvolle Einblicke in Forschungstrends, einflussreiche Arbeiten und wichtige Mitwirkende in diesem sich schnell entwickelnden Bereich.

Industrie 4.0 und Smart Industry sind die Vorreiter in der modernen Fertigung und Produktion. Industrie 4.0 integriert fortschrittliche Technologien wie das Internet der Dinge (IoT), künstliche Intelligenz (KI), Big Data und cyber-physische Systeme, um hochautomatisierte und vernetzte Produktionsumgebungen zu schaffen. Das Konzept der intelligenten Industrie betont Echtzeit-Datenanalyse, vorausschauende Wartung und anpassbare Fertigungsprozesse, um Effizienz, Flexibilität und Produktion zu verbessern [27]. Die Implementierung von RL-Algorithmen ist in diesen Paradigmen von entscheidender Bedeutung, da sie maschinelles Lernen und Zukunftsvorhersagen durch Interaktionen in der Produktionsumgebung erleichtert. Dies verbessert die Planung, Ressourcenzuweisung und Qualitätskontrolle und ermöglicht Unternehmen eine hohe Automatisierung und Präzision bei der Entscheidungsfindung , was zu einer widerstandsfähigeren und reaktionsfähigeren modernen Fertigung führt [54]. Tabelle 1 veranso

Tabelle 1. Beiträge des bestärkenden Lernens zur Produktionsplanung.

Planungstyp	Belohnen	Beitrag
Kategorie: Umwelt		
Komplexe Werkstatt [18]	**Strafen**	**Erweiterter DRL-Algorithmus optimiert die Produktionsplanung und hebt Anpassungsfähigkeit, Sichtbarkeit und Optimierung hervor.**
Flow Shop [25]	**Abgeschlossene Bestellung vs. Verzugsstrafe**	**Effizienter Fertigungsplan unter Berücksichtigung der Fähigkeiten der Arbeitskräfte und anpassungsfähiger Maschinen.**
Flexibler Job Shop [74]	**Verspätungen/Maschinenauslastungsgrad; Energiekosten; verspätete Produkte**	**Online-Umplanung zur Minimierung von Verspätungen oder Maximierung der Maschinenauslastung. MARL-Methode zur Optimierung hinsichtlich Effizienz und Umweltzielen.**
Lohnfertigung [54]	**Strafen**	**DQN-Technik zur Minimierung der Aufgabenerledigungszeit bei dynamischer Planung. MADRL optimiert die Leistung auf Systemebene unter Berücksichtigung von Werkstückinteraktionen.**
Kategorie: Methode		
Komplexe Werkstatt [18]	**Strafen**	**DRL-Algorithmus zur Optimierung der Planung, Verbesserung der Anpassungsfähigkeit, Sichtbarkeit und Optimierung.**
Flow Shop [25]	**Abgeschlossene Bestellung vs. Verzugsstrafe**	**Effizienter Zeitplan unter Berücksichtigung der Fähigkeiten der Arbeitskräfte und der Interaktion mit den Maschinen.**
Flexibler Job Shop [24]	**Verspätungen/Maschinenauslastungsgrad; Energiekosten; verspätete Produkte**	**Bestärkendes Lernen minimiert die Fertigstellungszeiten bei der dynamischen Jobplanung.**
Kategorie: Belohnung		
Komplexe Lohnfertigung [62]	**Strafen**	**Jobplanung mit Multiagenten im Resource Preemption Learning.**
Flow Shop [60]	**Fertiger Auftrag vs. Verzugsstrafe; Energiekosten; verspätete Produkte**	**Optimierung von Produktionsanlagen für Effizienz und Nachhaltigkeit.**
Flexibler Job Shop [19]	**Verspätung/Maschinenauslastung; Energiekosten; verspätete Produkte**	**Befasst sich mit dem Verstärkungsalgorithmus bei der Planung, mit dem Ziel, Entscheidungen in der flexiblen Werkstatt zu treffen.**

6. Bibliometrische Analyse von Studien zum RL in der Produktionsplanung

Die bibliometrische Analyse von RL im Kontext der Produktionsplanung konzentriert sich hier auf seine theoretischen Grundlagen und praktischen Auswirkungen zwischen den Jahren 1996 und 2024. Die ersten beiden Phasen umfassen die Sammlung bibliografischer Daten, die über die Scopus-Datenbank gesammelt wurden. Die Suche wurde auf diese Datenbank beschränkt, da sie als umfassendes Archiv wissenschaftlicher Literatur bekannt ist und häufig in akademischen Bewertungen verwendet wird . Die in den Abschnitten **6.1–6.4 vorgestellten Einschlusskriterien für die Zwecke der Analyse umfassten Veröffentlichungen, die die Begriffe „Reinforcement Learning“ und „Produktion“ und „Planung“ enthielten . Darüber hinaus wurden die Begriffe „Reinforcement Learning“ und „deterministische Planung“ oder „Reinforcement Learning“ und „unsichere Planung“ für die** in Abschnitt **6.5 durchgeführte** Analyse verwe

Machine Translated by Google

Zur Extraktion der erforderlichen **Informationen, wie etwa der jährlichen wissenschaftlichen Ergebnisse, der wichtigsten Quellen, der am häufigsten zitierten Autoren und des gemeinsamen Auftretens von Schlüsselwörtern, wurden Microsoft Excel und die VOSviewer-Software 1.6.20 eingesetzt.**

Der Prozess der ausgewählten Studien für die genannte Forschungsmethodik wird zusammengefasst in Abbildung 4, gemäß PRISMA-Richtlinien.

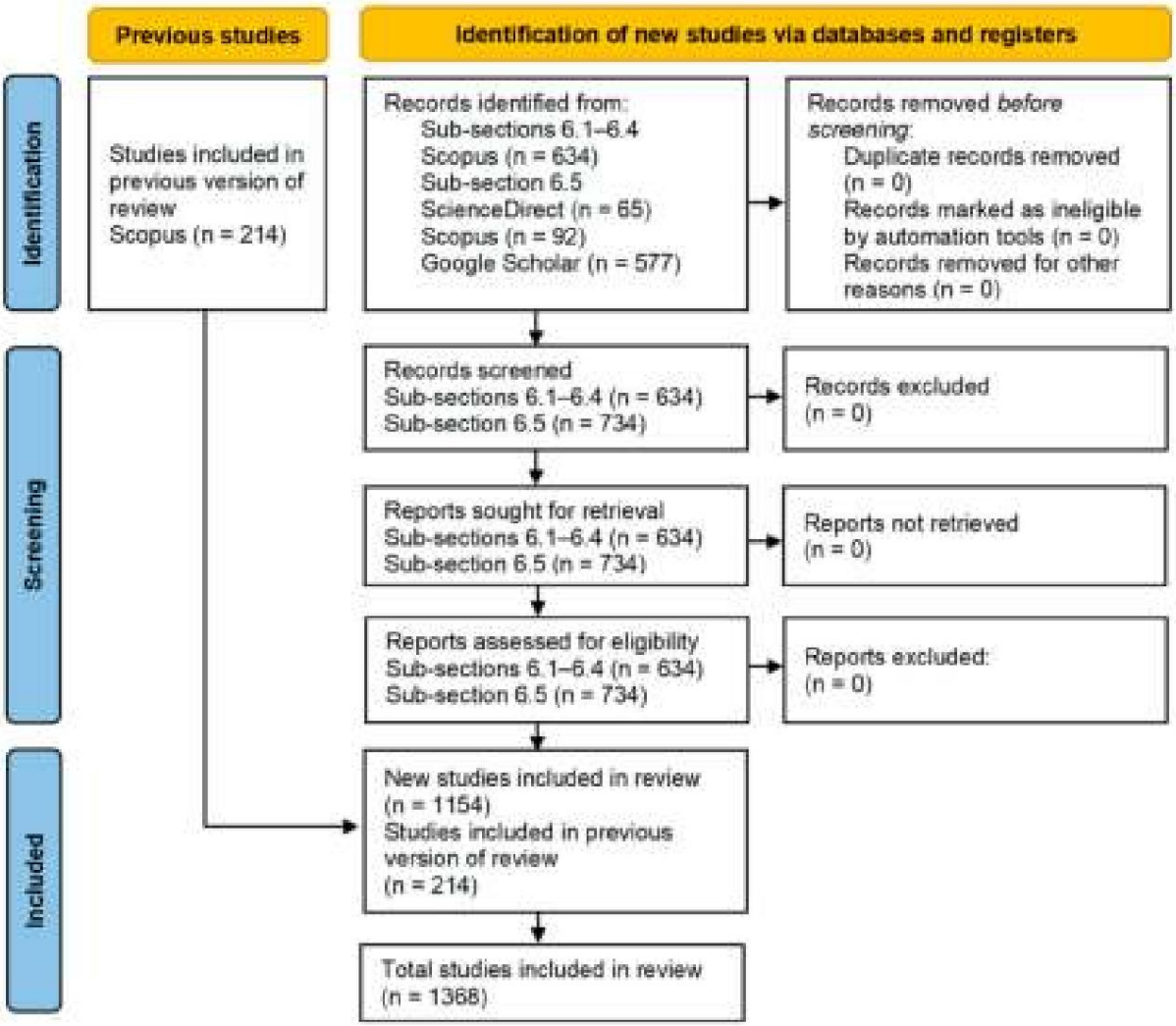


Abbildung 4. PRISMA 2020-Flussdiagramm für aktualisierte systematische Überprüfungen, die Datenbanksuchen beinhalteten.

Mithilfe der oben angegebenen Schlüsselwörter und Suchkriterien (für die Abschnitte 6.1–6.4 n = 634 Veröffentlichungen und für Abschnitt 6.5 n = 734 Veröffentlichungen) **wurden bis zum 27. Mai 2024, als alle Dokumenttypen in die Analyse des Reviews einbezogen wurden, insgesamt 1368 Artikel mit gemeinsamer Autorschaft** gefunden **(neue Studien n = 1154 und in der vorherigen Version des Reviews enthaltene**

Die Überprüfungsanalyse der Abschnitte 6.1–6.4 umfasste 634 gemeinsam verfasste Artikel, die unter „Artikeltitel, Zusammenfassung und Schlüsselwörter“ durchsucht wurden. Die wichtigsten Forschungsbereiche, die in den Artikeln besonders behandelt werden, sind Informatik, Ingenieurwissenschaften, Mathematik, Entscheidungswissenschaften , Wirtschaft, Management und Zählung, Energie, Chemieingenieurwesen, Materialwissenschaften und andere Themen. Die 634 Veröffentlichungen in unserer Stichprobe wurden in 19 verschiedene Forschungsthemen kategorisiert. Die acht wichtigsten Forschungsbereiche sowie ihre Artikelverteilung sind in Tabelle 2 aufgeführt.

Machine Translated by Google

Tabelle 2. Zugehörige Forschungsdisziplinen.

Forschungsgebiete	Anzahl der Publikationen	Prozentsatz
Informatik	429	32,7 %
Maschinenbau	402	30,6 %
Mathematik	159	12 %
Entscheidungswissenschaften	85	6,5 %
Wirtschaft, Management und Zählen	49	3,7 %
Energie	44	3,4 %
Chemieingenieurwesen	34	2,6 %
Materialwissenschaft	25	2 %
Sonstiges	85	6,5 %

6.1. Trends bei Veröffentlichungen und Zitationen

Die Anzahl der veröffentlichten Artikel und ihre Zitierungen liefern in der Regel ausreichend zuverlässige Informationen, um die weitere Entwicklung des untersuchten Forschungsbereichs vorherzusehen. die Anzahl der Veröffentlichungen und Zitate im Bereich Produktionsplanung mit Lernmethoden Algorithmen, die Daten aus rund 29 Jahren speichern, haben die Trendanalysegrafik abgeleitet. Für Zu diesem Zweck wurden die gleichen Suchbegriffe und Schlüsselwörter verwendet wie bei der Identifizierung der wichtigsten Forschungskategorien, aber die Dokumenttypen wurden auf alle Typen erweitert. Der Grund für die Änderung der Dokumenttypen war, die ersten Forschungsinitiativen in diesem Bereich herauszufinden. Domäne. Die gleichen Suchbedingungen wurden im Rest des Papiers angewendet, um die größere Stichprobe von Veröffentlichungen für den Zweck der Untersuchung. Die jährliche Verteilung von Publikationen (von insgesamt 634 Einträgen) und deren Zitierungen aus derselben Datenbank während der Zeitraum von 1996 bis 2024 ist in Abbildung 5 dargestellt.

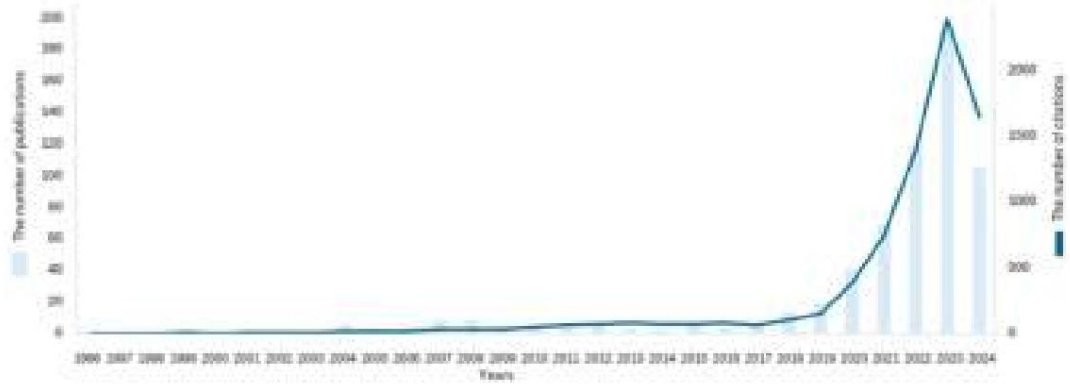


Abbildung 5. Publikations- und Zitationstrends für RL im Bereich Produktionsplanung von 1996 bis 2024.

Eine Untersuchung der jährlichen wissenschaftlichen Veröffentlichungen zwischen 1996 und 2018 zeigte deutlich, dass die Zahl der jährlich veröffentlichten Artikel relativ stabil niedrig war. zwei Jahrzehnte, die Notwendigkeit von RL in der Produktionsplanung unter realen Bedingungen offenbar nicht erschienen. Im Zeitraum 2019–2023 gab es einen deutlichen Anstieg der Zahl der Veröffentlichungen, die in der anerkanntesten Datenbank für Peer-Review-Inhalte registriert wurden. Dieses Phänomen ist vor allem auf die Fortschritte im Bereich der künstlichen Intelligenz zurückzuführen. Es ist bemerkenswert, dass, wenn dieser exponentielle Trend der Zunahme der Zahl der Veröffentlichungen wie im Jahr 2024 zu sehen ist, dann kann man davon ausgehen, dass während der Im nächsten Jahrzehnt wird die Bedeutung von RL in der Fertigungsplanung deutlich zunehmen.

6.2. Wichtigste Quellen

Eine Identifizierung der relevantesten Publikationen aus einem ersten Datensatz stellt gängiger Ansatz in der bibliometrischen Forschung, da solche Quellen in der Regel einflussreiche Forschung veröffentlichen, die breites Interesse weckt. In der Regel sind die produktivs haben den größten Einfluss auf die Entwicklung der Wissenschaft in einem bestimmten Bereich, da sie

Machine Translated by Google

mehr Artikel veröffentlichen und mehr Zitate generieren [77–79]. Was die Relevanz der Literatur im untersuchten Bereich betrifft, werden hier die zehn wichtigsten Zeitschriften mit den meisten veröffentlichten Artikeln aufgeführt. In diesem Unterabschnitt werden auch die zehn am häufigsten zitierten wissenschaftlichen Zeitschriften erwähnt. Die Abbildungen 6 und 7 kategorisieren Zeitschriften nach diesen beiden Kriterien, um zu zeigen, dass sie Dokumente darstellen, die für RL in der Proc

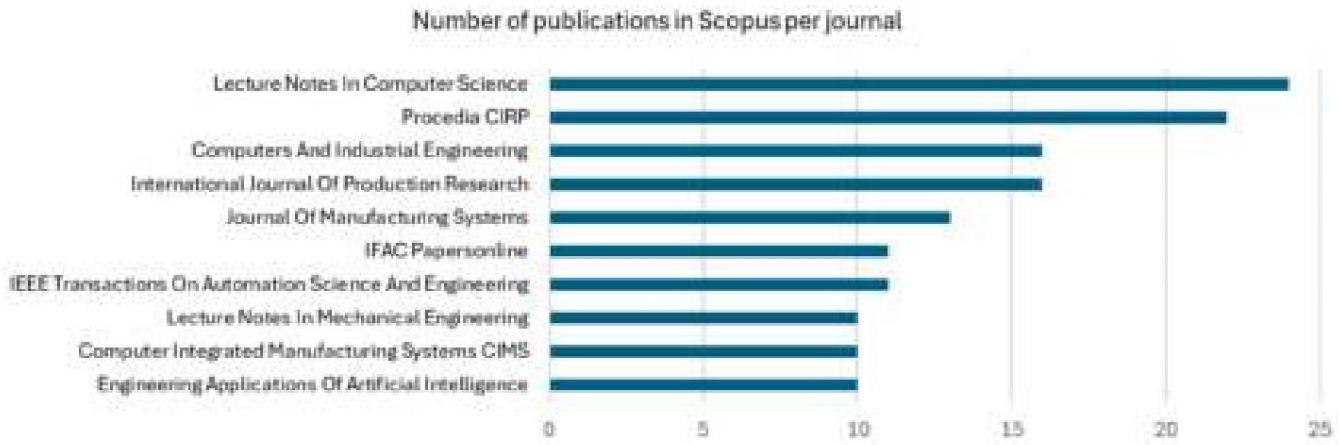


Abbildung 6. Die zehn meistveröffentlichten Zeitschriften für RL im Bereich Produktionsplanung von 1996 bis 2024.

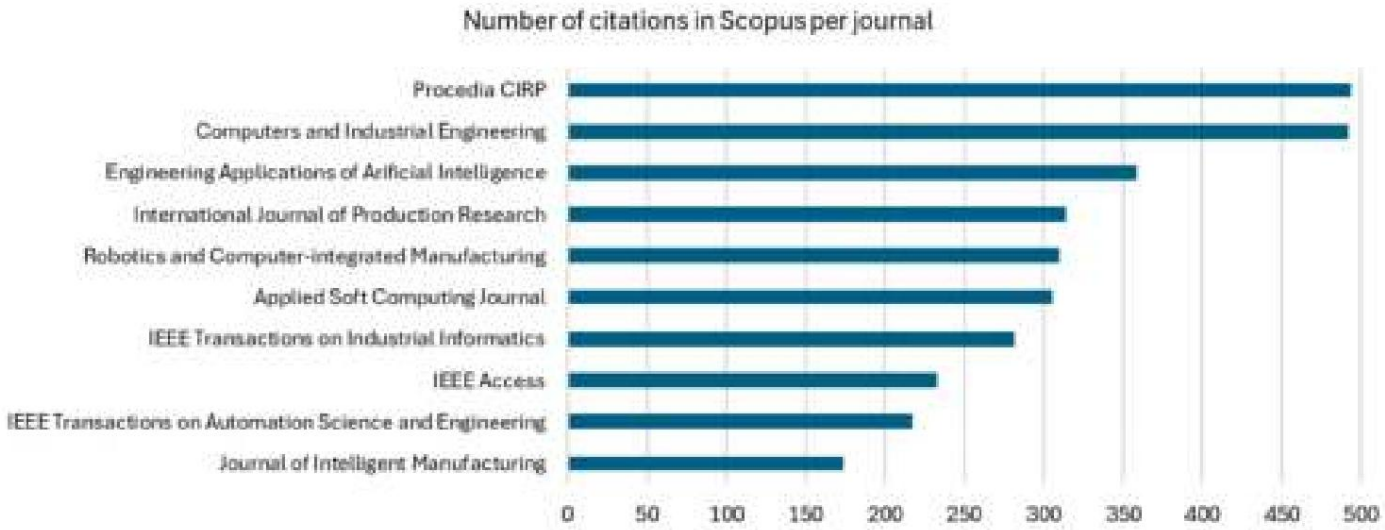


Abbildung 7. Die zehn am häufigsten zitierten Zeitschriften für RL im Bereich Produktionsplanung von 1996 bis 2024.

Die Lecture Notes in Computer Science, Procedia CIRP und Computers and Industrial Engineering gelten zusammen mit den zwölf anderen Zeitschriften als die relevantesten wissenschaftlichen Veröffentlichungen in diesem Bereich. Es ist auch festzustellen, dass alle fünfzehn in den Abbildungen 6 und 7 aufgeführten Zeitschriften, unabhängig von den Ergebnissen der verwendeten Metriken, empirisch als weithin anerkannt für die Verbreitung fortgeschrittener Forschung zum Thema Reinforcement Learning in der Produktionsplanung eingestuft werden können. Darüber hinaus wird ihre wissenschaftliche Genauigkeit auch dadurch belegt, dass von den fünfzehn identifizierten Zeitschriften acht hohe Qualitätsstandards erfüllten, da sie für Current Content Connect-Zeitschriften mit nachweisbarem Einfluss auf die Steuerung von Forschungspraktiken und -v

6.3. Meistzitierte Autoren

Seit den 1996er Jahren haben viele Autoren bedeutende Beiträge zur Entwicklung dieses Bereichs geleistet. In diesem Unterabschnitt sollen einige der Autoren vorgestellt werden, die bedeutende intellektuelle Beiträge zur Forschung geleistet haben. Die Analyse der am häufigsten zitier

Machine Translated by Google

wurde mit Daten aus der Scopus-Datenbank durchgeführt. Zusätzlich wurde für jede Publikationsquelle (von insgesamt 634 Elementen) eine Co-Zitationsanalyse ausgeführt, um das Netzwerk zwischen den Studien aufzudecken. Zu diesem Zweck wurde die Software VOSviewer [82] verwendet. Um relevante Informationen und eine klare grafische Darstellung komplexer Beziehungen zu erhalten, wurden die folgenden Filter eingesetzt. Filter 1: Maximale Anzahl von Autoren pro Dokument – 15; Filter 2: Minimale Anzahl von Dokumenten eines Autors – 3; Filter 3: Minimale Anzahl von Zitaten eines Autors – 10. Darüber hinaus wurde die Methode der Vollzählung angewendet, d. h. Veröffentlichungen mit Co-Autoren aus mehreren Ländern werden für jedes dieser Länder als vollständige Veröffentlichung gezählt . Das Co-Zitationsnetzwerk der ausgewählten Stichprobe von Wissenschaftlern unter Verwendung dieser Einstellungen ist in Abbildung 8 visualisiert. **tt**

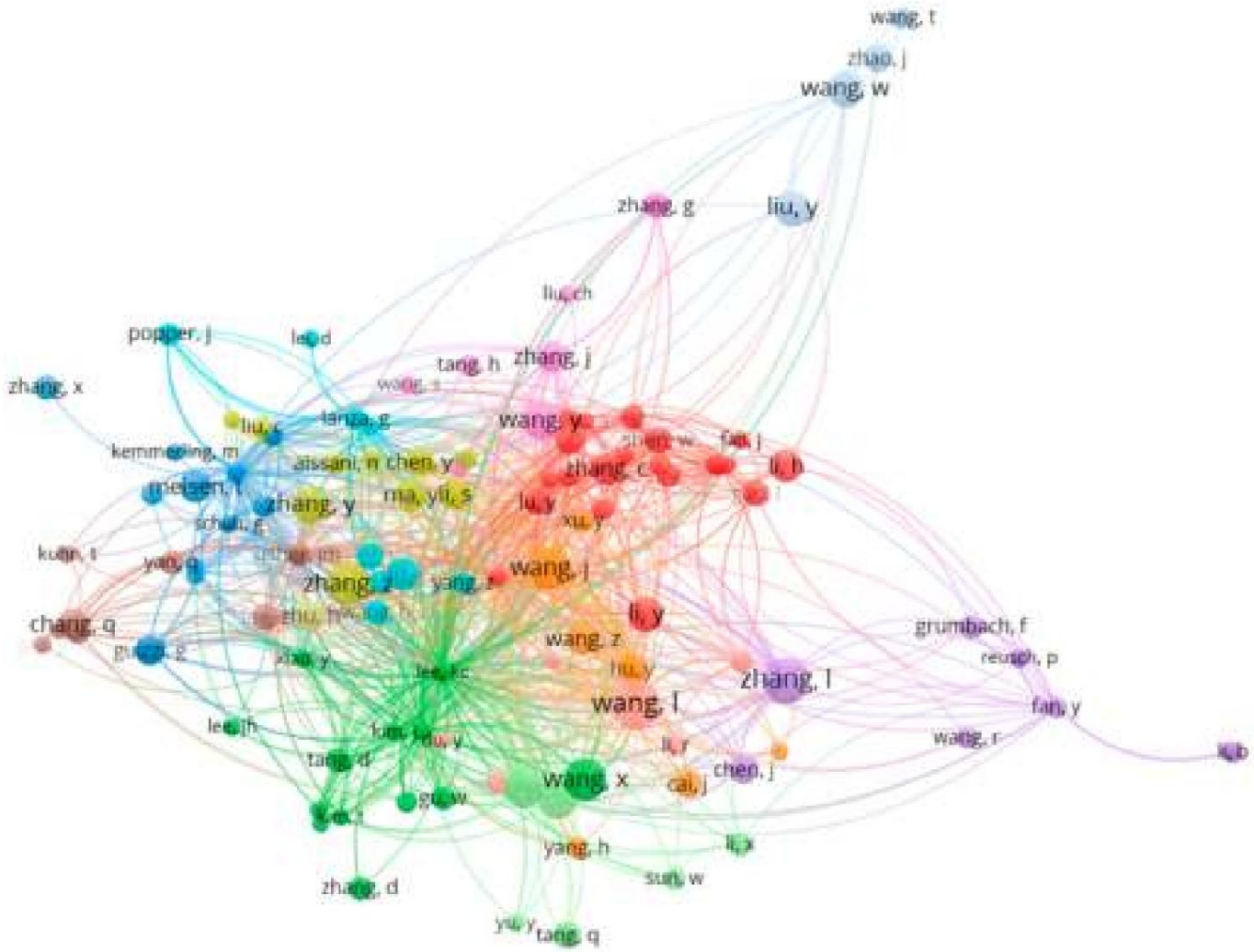


Abbildung 8. Co-Zitationsnetzwerk der Autoren, die von 1996 bis 2024 bedeutende Beiträge zur Entwicklung dieses Fachgebiets geleistet haben.

Diese Co-Citation-Netzwerkkarte zeigt unter anderem Wissenschaftler, die in den letzten 29 Jahren die meisten Zitate erhalten haben. Von den 1376 zitierten Autoren erfüllen 138 die oben genannten Kriterien. Jeder Wissenschaftler aus jeder enthaltenen Publikation wird in diesem Netzwerk durch einen Knoten dargestellt. Die Größe jedes Knotens gibt die Häufigkeit der Zitierung der wissenschaftlichen Arbeiten des betreffenden Themas an. Zwischen zwei Knoten wird eine Kante gezeichnet, wenn die beiden Wissenschaftler in einem gemeinsamen Dokument zitiert wurden. Um einflussreiche Wissenschaftler in dem jeweiligen Bereich anhand ihrer Zitierungsraten zu ordnen , wurden die zehn am häufigsten zitierten Autoren au

Machine Translated by Google

Tabelle 3. Einflussreichste Autoren.

Name	Anzahl der Zitate	Land	Name	Anzahl der Zitate	Land
Waschenck, B.	386	Deutschland	Wang, Y.C.	265	Taiwan
Wang, L.	375	China	Wang, H.	224	China
Wang, J.	354	China	Lin, CC	216	Taiwan
Bauernhansl, T.	342	Deutschland	Schröder, M.	209	China
Usher, JM	265	USA	Beldjilali, B.	207	Algerien

6.4. Identifizierung der verwandten Forschungsbereiche

Das Ziel dieses Teils des Artikels ist es, Forschungsbereiche zu identifizieren, in denen die Frage der Produktionsplanung auf Basis von RL ist von Interesse. Zu diesem Zweck ist die Schlüsselwortanalyse Anwendung wurde eingesetzt, um wichtige Forschungsthemen zu trennen, was erntete großes Interesse von Forschern aus weniger bedeutenden Kreisen. Um relevante und repräsentativen Kategorien, ausgenommen weniger bedeutende, wurde folgende Einstellung verwendet: Mindestanzahl des Vorkommens von Schlüsselwörtern – 14. Basierend auf dieser Einschränkung, 88 Schlüsselwörter erfüllen den Schwellenwert von insgesamt 3767 Schlüsselwörtern. Die wichtigsten Schlüsselwörter, die wurden automatisch aus den Titeln der Zeitungen über die Produktionsplanung erstellt, basierend auf RL sowie deren Vorkommen sind in Abbildung 9 dargestellt.

Wie man sieht, können die Themen auf der Grundlage eines Computeralgorithmus in fünf Cluster unterteilt werden. wobei jeder Cluster eine andere Farbe hat, wie in Abbildung 9 dargestellt. Das gemeinsame Auftreten der Schlüsselwörter Die Karte hebt diese Cluster hervor. Dabei gilt: Je dunkler die Farbe, desto höher der Dichtewert. Es ermöglicht die Identifizierung relevanter Forschungsthemen und ihrer gegenseitigen Beziehungen. Basierend auf dem erhaltene bibliometrische Ergebnisse aus VOSviewer, zehn verwandte Themen, die sehr In der Nähe des untersuchten Forschungsbereichs wurden folgende Bereiche ermittelt, wie in Tabelle 4 dargestellt.

Die Ergebnisse aus Tabelle 4 weisen auf ihre Übereinstimmung mit der praktischen Realität und dem Betrieb hin. Forschungsziele. So gab es in den letzten Jahren offensichtlich ein gesteigertes Interesse bei der Verwendung von bestärkendem Lernen zur Optimierung von Echtzeit-Job-Scheduling-Aufgaben [76–86]. Diese Tatsache kann mit dem anhaltenden Trend zur Massenanfertigung in der Produktion von Konsumgütern in Zusammenhang gebracht werden [87,88]. Wie bekannt ist, ist Massenanfertigung charakterist sich dynamisch ändernde Benutzeranforderungen rechtzeitig erfüllen, während maßgeschneiderte Produkte zu unterschiedlichen Terminen fertigzustellen. Dementsprechend ist eine effiziente Jobplanung in Echtzeit Algorithmen auf Basis von DRL werden unverzichtbar. Die nächste wichtige Methode, die unter den Top 10 der Schlüsselwörter mit gemeinsamen Vorkommen ist die Produktionssteuerung. Sie verwendet verschiedene Steuerungs Techniken zur Erreichung von Produktionszielen hinsichtlich Produktionsplänen und Qualität (siehe zB [89–92]). Das nächste wichtige gemeinsam vorkommende Schlüsselwort in Tabelle 3 – Multi-Agenten-Systeme. Generell gilt: Multi-Agenten-Systeme in das bestärkende Lernen für die Produktion integrieren Planung bietet zahlreiche Vorteile in Bezug auf Flexibilität, Skalierbarkeit und Anpassungsfähigkeit [93–95]. Indem sie dezentrale Entscheidungsfindung und kontinuierliches Lernen ermöglichen, Systeme können die Komplexität und Dynamik moderner Produktionsumgebungen effektiv bewältigen, was zu effizienteren und belastbareren Planungslösungen führt [96]. Zu den Mit anderen Schlüsselwörtern ist es möglich, auch „Smart Manufacturing“ hervorzuheben, das den Implementierungsbereich der Produktionsplanung auf der Grundlage von RL darstellt. Obwohl Smart Manufacturing ist auch zu einem Schlagwort geworden, das auch seine Nachteile hat. Konzeption etabliert sich allmählich als neues Fertigungsparadigma. Andererseits wird die Komplexität der Netzwerkinfrastrukturen für intelligente Fertigung höher und höher, und die Unsicherheit dieser Fertigungsumgebung wird zu einem ernsten Problem [25]. Diese Tatsachen führen zur Notwendigkeit, fortschrittliche dynamische Planungslösungen anzuwenden, die auch die Produktionsplanung mit RL umfassen. Dieser Absatz Antworten auf die im Abschnitt **1** formulierte zentrale Forschungsfrage .

Darüber hinaus wird die Anwendung von RL in der Produktionsplanung hier analysiert durch bibliometrische Mittel aus der Sicht verschiedener Planungsprobleme.

Machine Translated by Google

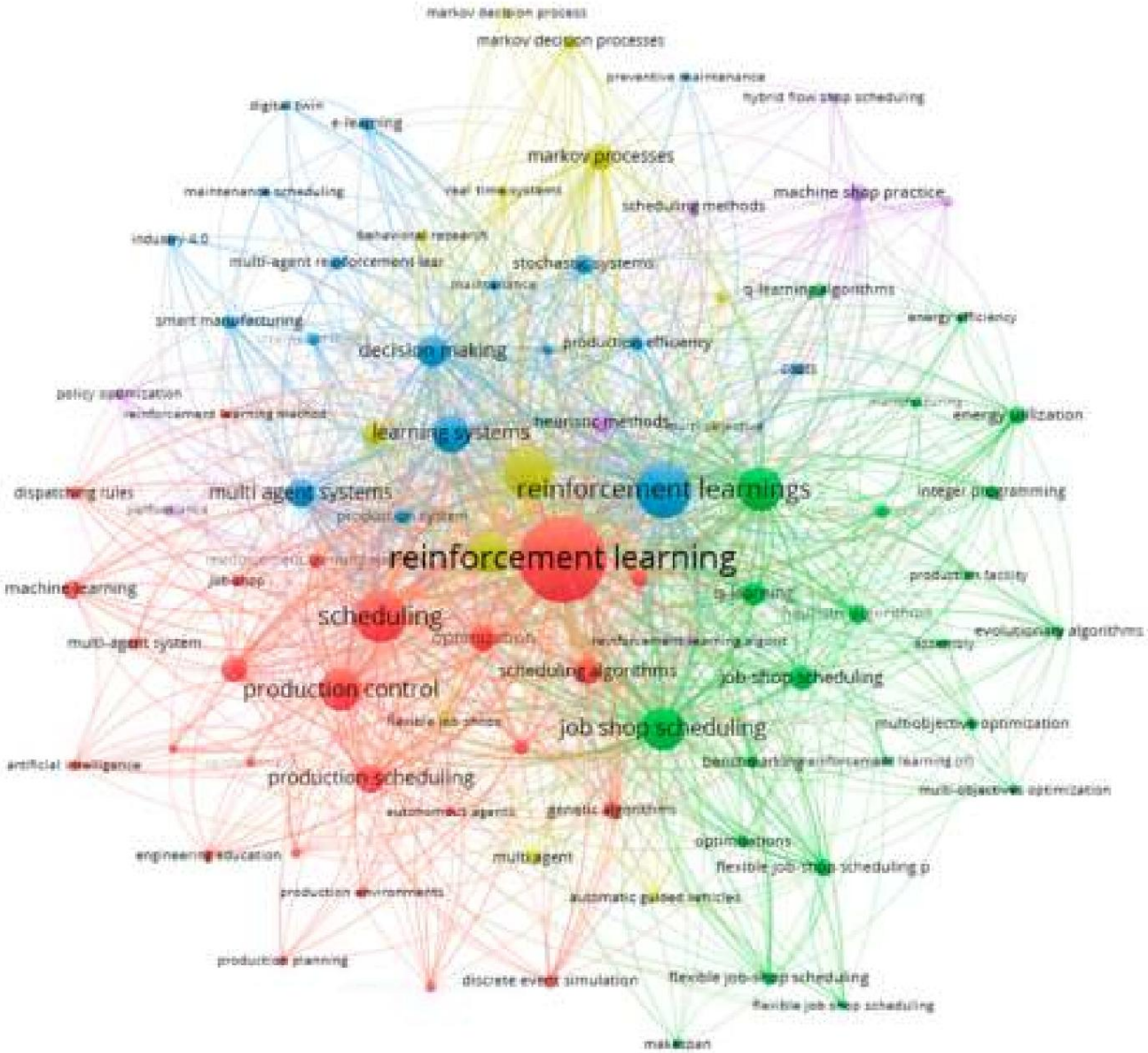


Abbildung 9. Ko-Auftrittsschlüsselwörter für RL im Feld Produktionsplanung von 1996 bis 2024.

Tabelle 4. Die meisten verwandten Methoden und Implementierungsbereiche der Produktionsplanung basieren auf RL.

Verwandte Forschungsbereiche	Anzahl der gemeinsamen Vorkommen	Verwandte Begriffe/Themen	Anzahl der gemeinsamen Vorkommen
Bestärkendes Lernen	526	Produktionskontrolle	161
Terminplanung	220	Tiefes bestärkendes Lernen	148
Tiefes Lernen	215	Lernsysteme ffi	134
Lernalgorithmen	178	Entscheidungsfindung	99
Werkstattplanung	168	Multiagentensysteme	86

6.5. Vergleich von RL-Anwendungen bei verschiedenen Planungsproblemen

Im Allgemeinen werden Planungsmethoden nach dem Zeitpunkt kategorisiert, zu dem die Entscheidung getroffen wird. in dynamische und statische unterteilt. Dynamische Planung bezieht sich auf Echtzeitsysteme, die erfordern die Reaktion auf sich ändernde Nachfrageanforderungen, während die statische Planung offline erfolgt und konzentriert sich auf den kurzfristigen Zeithorizont, indem ein fester Zeitrahmen für den Abschluss des Prozesses festgelegt wird. Wang et al. [28] **haben festgestellt, dass die Produktionsplanung auf Basis von RL-Ansätzen**

Machine Translated by Google

werden weitgehend zur Lösung dynamischer Planungsprobleme eingesetzt. Planungsprobleme können auch durch die Art der Planungsumgebung als deterministisch eingestuft werden, bei der Verarbeitung Parameter sind bekannt und unveränderlich und nicht deterministisch, wenn Eingabeparameter sind unsicher [97]. Probleme bei der Produktionsplanung nach diesem zweiten Kriterium können wie in Abbildung 10 dargestellt klassifiziert .

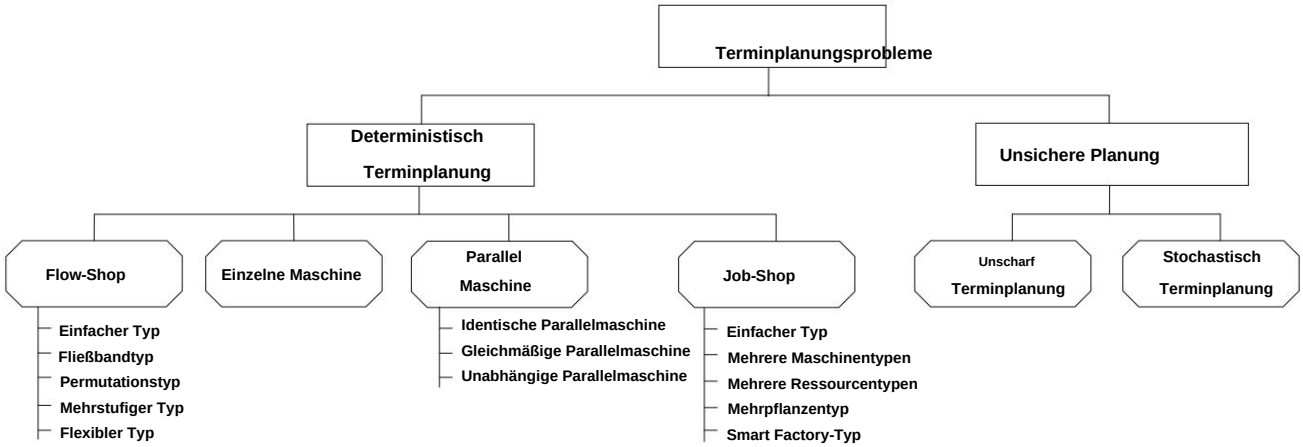


Abbildung 10. Grundlegende Klassifizierung von Planungsproblemen.

Das Verständnis der Unterschiede zwischen bestimmten Planungsproblemen ist entscheidend für die Auswahl Entwicklung geeigneter Optimierungs- und Planungstechniken mit dem Ziel, Effizienz und Produktivität in verschiedenen Branchen. Der Hauptunterschied zwischen der Planung ffi- Probleme zeigen sich beispielsweise in der Flexibilität und Komplexität der Routenplanung. So Probleme haben eine hohe Routing-Flexibilität und Komplexität, während für Flow-Shop und Single Maschinenprobleme sind typischerweise geringe Flexibilität, und parallele Maschinenprobleme haben moderate Routing-Flexibilität und Komplexität [98,99]. Ihre Unterschiede liegen auch in der Art der Anwendung Domänen, für die sie bestimmt sind. Zum Beispiel für eine kundenspezifische Fertigung ist typische Job-Shop-Planung [100,101], für Montagelinien wird meist als Flow-Shop betrachtet Planungsprobleme [102,103], und die offene Werkstattplanung ist häufig anwendbar in Gesundheitswesen [104,105].

Die Klassifizierung wird weiterhin verwendet, um eine Anwendungshäufigkeit von RL in den identifizierten Kategorien der Planungsprobleme. Erstens: deterministische versus unsichere Die Terminplanung wird nach dieser Sichtweise verglichen. Die Suchbegriffe wurden dabei definiert durch die Kombination der folgenden Schlüsselwörter: „Reinforcement Learning“ zusammen mit „Deterministische Planung“ bzw. „Unsichere Planung“. Daten zusammen mit Die Einschlusskriterien wurden durch folgende Suche ermittelt: (i) ScienceDirect-Datenbank – Alle Felder, Alle Jahre und Alle Dokumenttypen; (ii) Scopus-Datenbank – Alle Felder, Alle Jahre, Alle und (iii) Google Scholar-Datenbank – Alle Jahre und Alle Dokumenttypen. Zusammen 734 Publikationen wurden analysiert. Web of Science wurde bei diesem Verfahren nicht verwendet, da die geringe Anzahl wissenschaftlicher Arbeiten zu diesen Themen. Die so gewonnenen Daten sind Die Angaben sind in Tabelle 5 aufgeführt.

Tabelle 5. Vergleich deterministischer und unsicherer Planungsmethoden basierend auf RL.

Verwendete Suchbegriffe	Wissenschaft direkt Datenbank	Scopus Datenbank	Google Scholar Datenbank
„Deterministisches Scheduling“ und „Verstärkendes Lernen“	53	68	474
„Unsichere Terminplanung“ und „Verstärkendes Lernen“	12	24	103

Aus Tabelle 5 ist ersichtlich, dass deterministisches Scheduling auf Basis von RL in einer dominanten Position gegen unsichere Planung mit RL. Um daher die vielversprechendsten

Machine Translated by Google

Im Folgenden werden **die Bereiche der Produktionsplanung im Kontext von RL, deterministischer Planung und unsicheren Planungsansätzen untersucht. Die dafür verwendeten Schlüsselwörter waren:**

(a) Terminplanung UND „verstärkendes Lernen“ UND spezifische deterministische Terminplanungsprobleme („Job Shop“, „Flow Shop“, „Open Shop“, „Einzelmaschine“ und „Parallelmaschinen “).

(b) „Verstärkendes Lernen“ UND spezielle unsichere Planungsprobleme („Fuzzy-Planung “ und „stochastische Planung“).

Die Daten wurden aus der ScienceDirect-Datenbank mit den Filtern „Forschungsartikel“ und „Alle Jahre“, aus der Scopus-Datenbank mit den Filtern „Artikeltitel“, „Zusammenfassungen“, „Schlüsselwörter“, „Alle Jahre“ und „Alle Dokumenttypen“ und aus der Google Scholar-Datenbank – „Alle Jahre“ und „Alle Dokument
Die erzielten Ergebnisse werden in **Abbildung** 11 **grafisch dargestellt und in Prozent miteinander verglichen .**

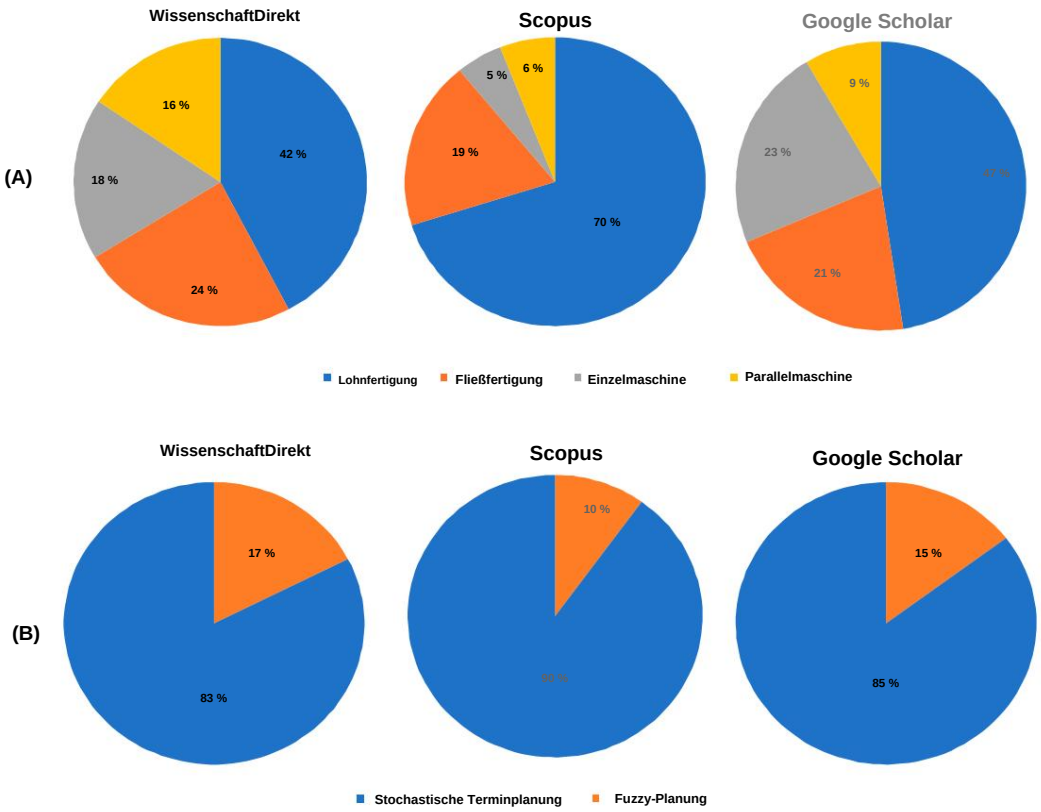


Abbildung 11. Vergleich von (a) deterministischen Planungsmethoden und (b) unsicheren Planungsmethoden anhand ihrer Vorkommen in der Literatur.

Aus Abbildung **11a ist ersichtlich , dass RL-basierte Planungsalgorithmen hauptsächlich zur Lösung von Job-Shop-Planungsproblemen** verwendet werden **[106], während die Flow-Shop-Produktionsumgebung der zweitwichtigste Typ von Planungsproblemen ist, bei dem RL angewendet wird. Natürlich finden RL-basierte Algorithmen auch Anwendung bei anderen Arten von Planungsproblemen. Als die auffälligsten davon sind die Energieeffizienzplanung, die Mehrzielplanung und die distributive Planung zu nennen tt Das letztgenannte Planungsproblem wird insbesondere bei intelligenten Fertigungssystemen ausgenutzt [107]. Bei der Analyse der Ergebnisse aus Abbildung 11b ist festzustellen, dass RL überwiegend bei stochastischen Planungsmethoden verwendet wird, verglichen mit seiner Anwendung bei Fuzzy-Planungsmethoden. Darüber hinaus wurde die Wirksamkeit der stochastischen Planung mit RL auch durch Benchmarking-Studien nachgewiesen [29,108,109]. Obwohl die Anwendung von RL bei der Fuzzy-Planung vielversprechend ist, ist sie auf**

ÿ

Machine Translated by Google

mit RL-Methoden [110,111]. Die Ergebnisse dieses quantitativen Ansatzes zum Vergleich der Anwendung von RL in verschiedenen Planungsproblemen liefern jedoch einige nützliche Manag Einblicke in die Planungspraxis. Dennoch sind weitere Untersuchungen und Forschungen zu qualitativen Aspekten der RL-basierten Planung, insbesondere im Kontext von Smart Manufacturing, ist erforderlich.

7. Vergleich der vorgestellten Forschung und früherer Übersichtsartikel

Wie bereits in der Einleitung erwähnt, setzt dieser Übersichtsartikel die Arbeit des zwei frühere Artikel [28,29] durch die Analyse einiger neuer relevanter Aspekte der Produktionsplanung Methoden mit RL und bietet einige aktuelle Informationen zu den jüngsten Entwicklungen in diesem Bereich. Der Zweck dieses Abschnitts besteht darin, einen klaren Überblick über die Unterschiede und Gemeinsamkeiten der drei Werke in Bezug auf ihre Thematik. Die folgende Tabelle 6 enthält eine zusammenfassende Gegenüberstellung der in den genannten Arbeiten behandelten Untersuchungsgebiete.

Tabelle 6. Der Vergleich der Übersichtsartikel zu RL in der Produktionsplanung.

Der Schwerpunkt der Untersuchung	Die verglichenen Übersichtsartikel		
	Wang et al., 2021 [28]	Kayhan und Yildiz, 2021 [29]	Angebotene Forschung
Der Zitierungstrend für RL in der Produktion Planung von 1996 bis 2024			ÿ
Die wichtigsten Veröffentlichungen auf diesem Gebiet	ÿ		ÿ
Die einflussreichsten Autoren auf diesem Gebiet			ÿ
Der Vergleich von deterministischen und unsicheren Scheduling-Methoden basierend auf RL			ÿ
Der Vergleich von stochastischen und Fuzzy Scheduling-Methoden basierend auf RL			ÿ
Der Publikationstrend für RL in der Produktion Terminplanung	ÿ	ÿ	ÿ
Die wichtigsten Quellen veröffentlichten Forschungsarbeiten zu RL in der Produktionsplanung			ÿ
Der Vergleich der häufigsten deterministische Scheduling-Methoden basierend auf RL	ÿ	ÿ	ÿ
Der Vergleich der wertorientierten und richtlinienbasierte Planungsmethoden basierend auf RL	ÿ		
Der Vergleich von statischen und dynamischen Scheduling-Methoden basierend auf RL	ÿ		
Die Kategorisierung der Literatur zu RL in Produktionsplanung nach mehreren Aspekte		ÿ	
Der Publikationstrend für RL in der Produktion Planung nach Fertigungs Umfeld		ÿ	
Der Vergleich verschiedener RL-Typen Algorithmen zur Produktionsplanung		ÿ	
Der Vergleich von stochastischen und deterministischen Scheduling-Methoden basierend auf RL		ÿ	

Aus dieser Tabelle ist ersichtlich, dass die verglichenen Arbeiten sich größtenteils ergänzen gegenseitig.

Machine Translated by Google

8. Schlussfolgerungen

Der bestehende Forschungsbestand zur Produktionsplanung besteht hauptsächlich aus Studien aus den Bereichen Informatik, Ingenieurwissenschaften, Mathematik, Entscheidungswissenschaften , Betriebswirtschaft, Management und Berechnung, Energie, Chemieingenieurwesen und Materialwissenschaften. Diese wissenschaftlichen Disziplinen, insbesondere „Ingenieurwissenschaften“ und „Informatik“, haben einen erheblichen Einfluss auf die Entwicklung der Produktionsplanung auf der Grundlage

Die Untersuchung bibliometrischer Ergebnisse zeigt häufig, dass eine Zunahme der Anzahl veröffentlichter Artikel mit der Erkennung fortschrittlicher Trends im Fachgebiet einhergeht . Aus diesem Grund wird eine bibliometrische Analyse **in vielen akademischen Bereichen immer nützlicher,** da sie die Abbildung wissenschaftlicher Informationen und die Analyse **der Forschungsentwicklung objektiv und wiederholbar macht. Die Verwendung dieser Methode ermöglicht es uns, die Netzwerke wissenschaftlicher Zusammenarbeit zu identifizieren, Verbindungen zwischen neuen Studienthemen und Forschungsströmen herzustellen sowie die Verbindungen zwischen Zitaten, Co-Zitaten und der Veröffentlichungsprodukti**

Der Hauptbeitrag dieses Artikels kann in zwei Schritten zusammengefasst werden:

(i) Dieser Bericht liefert zusätzliche Einblicke in RL in der Produktionsplanung, indem er die folgenden neuen Features: •

Der Zitierungstrend für RL im Bereich Produktionsplanung von 1996 bis 2024 (siehe Abbildung 5) wurde ermittelt und zeigt, dass sich der Aufwärtstrend von 2019 bis heute mit einem deutlichen Anstieg der Zahl fortsetzt. • Die Analyse der relevantesten Quellen im Hinblick auf die Anzahl ihrer Zitierungen (siehe Abbildung 7) wurde durchgeführt, um die einflussreichsten Quellen zu ermitteln, denen die neuesten Erkenntnisse auf diesem Gebiet zur Verfügung stehen. • Die einflussreichsten Autoren auf diesem Gebiet wurden ermittelt (siehe Tabelle 3), um den neuesten Stand der RL-basierten Produktionsplanung zu bestimmen. • Der quantitative Vergleich deterministischer und unsicherer Planungsmethoden auf RL-Basis wurde durchgeführt (siehe Tabelle 5). Er zeigte, dass deterministische Planungsmethoden mit RL gegenüber RL-basierten unsicheren Planungstechniken eine beherrschende Stellung einnehmen. • Der quantitative Vergleich stochastischer und Fuzzy-Planungsmethoden ist in Abbildung **11b dargestellt . Aus dem Vergleich ging hervor, dass RL überwiegend in stochastischen Planungsmethoden verwendet wird, im Vergleich zu seiner Anwendung in Fuzzy-Planungsmethoden.**

(ii) Der Beitrag dieses Papiers zeigt sich auch in den folgenden Aktivitäten und den daraus resultierenden Aussagen:

• Der Publikationstrend für RL im Bereich Produktionsplanung wurde aktualisiert, indem der Zeitraum von 1996 bis 2024 abgebildet wurde (siehe Abbildung 5). Es zeigte sich, dass das exponentielle Wachstum der Veröffentlichungen ab 2019 [28] schnell anhält und ihre Anzahl deutlich zunimmt.

• Die aktualisierte Liste der wichtigsten veröffentlichten Quellen für Forschung zu RL in der Produktionsplanung wurde in Abbildung 6 zusammengestellt. Es wurde festgestellt, dass fünf von zehn bis 2021 identifizierten Quellen zu den zehn wichtigsten bis 2024 identifizierten Dies bedeutet, dass aufgrund des wachsenden Interesses an diesem Forschungsfeld gleichzeitig die Anzahl der Zeitschriften zunimmt, die sich mit diesem Thema befassen. • Eine Aktualisierung des quantitativen Vergleichs der am häufigsten auf RL basierenden deterministischen Planungsmethoden ist in Abbildung 11a dargestellt. In diesem Zusammenhang wurde der zuvor festgestellte Trend bestätigt, dass RL-basierte Planungsalgorithmen hauptsächlich zur Lösung von Job-Shop-Planungsproblemen verwendet werden, während die Flow-Shop- Produktion den zweitwichtigsten Planungsproblemtyp darstellt. • Die aktualisierte Kategorisierung neu entwickelter RL-Algorithmen wurde ausgearbeitet und in Abbildung 3 dargestellt . **Sie kann eine bessere Implementierungsunterstützung für die Entscheidungsfindung bei realen Problemen bieten.**

Machine Translated by Google

- **Die aktualisierte Klassifizierung der Planungsmethoden ist in Abbildung 10** dargestellt. Sie kann **grundsätzlich zur einfacheren Auswahl der Planungstechniken zur Lösung bestimmter Problemtypen verwendet werden.**

Zusätzlich zu den oben genannten Erkenntnissen ist es erforderlich, sich auf andere Herausforderungen zu konzentrieren , die in Zukunft berücksichtigt werden müssen, wie z. B. die Fertigungsprozessplanung mit integrierter Unterstützung für den Wissensaustausch, die zunehmende Nachfrage nach Verbesserungen der allgegenwärtigen „Intelligenz“ in Fertigungsprozessen, einschließlich der Entwicklung und Implementierung intelligenter Algorithmen, und der Bedarf an robusten Plar

Beiträge der Autoren: Konzeptualisierung, VM; Methodik, RS und AB; Validierung, ZS; formale Analyse, VM und RS; Untersuchung, VM und AB; Schreiben – Vorbereitung des Originalentwurfs, AB und VM; Schreiben – Überprüfung und Bearbeitung, RS und VM; Visualisierung, ZS; Projektverwaltung, VM. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu.

Finanzierung: Diese Forschung wurde durch das Projekt SME 5.0 mit Mitteln aus dem Forschungs- und Innovationsprogramm Horizon der Europäischen Union im Rahmen der Marie Skÿodowska-Curie- Fördervereinbarung Nr. 101086487 und dem KEGA-Projekt Nr. (Cultural and Education Grant Agency 044TUKE-4/2023, erteilt vom Bildungsministerium der Slowakischen Republik.

Erklärung des Institutional Review Board: Nicht zutreffend.

Einverständniserklärung: Nicht zutreffend.

Datenverfügbarkeitserklärung: Die Daten sind im Artikel enthalten.

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen. Die Geldgeber hatten keinen Einfluss auf die Gestaltung der Studie, auf die Erhebung, Analyse oder Interpretation der Daten, auf das Verfassen des Manuskripts oder auf die Entscheidung, die Ergebnisse zu veröffentlichen.

Verweise

1. **Pinedo, M.** Planung und Terminierung in Fertigung und Dienstleistung; **Springer: New York, NY, USA, 2005. 2.**
Beheshti, Z.; Shamsuddin, SMH Eine Übersicht über populationsbasierte metaheuristische Algorithmen. Int. J. Adv. Soft Comput. Appl. **2013, 5, 1–35.**
3. **Xhafa, F.; Abraham, A. (Hrsg.)** Metaheuristiken für die Planung in Industrie- und Fertigungsanwendungen; **Springer: Berlin/ Heidelberg, Deutschland, 2008; Band 128.**
4. **Abdel-Kader, RF Partikelschwarmoptimierung für eingeschränkte Befehlsplanung.** VLSI Des. **2008,** 2008, **930610. [CrossRef]**
5. **Balamurugan, A.; Ranjitharamasamy, SP Eine modifizierte Heuristik für die Batchgrößenoptimierung mit kombinierter Zeit in einem System zur kundenspezifischen Massenfertigung.** Int. J. Ind. Eng. Theory Appl. Pract. **2023,** 30, **1090–1115.**
6. **Ghassemi Tari, F.; Olfat, L. Heuristische Regeln für das Verspätungsproblem im Flow Shop mit Zwischenterminen.** Int. J. Adv. Manuf. Technol. **2014, 71 , 381–393.**
7. **Modrak, V.; Pandian, RS Flow-Shop-Planungsalgorithmus zur Minimierung der Fertigstellungszeit für das n-Jobs-m-Maschinen-Problem.** Teh. Vjesn. **2010,** 17, **273–278.**
8. **Thenarasu, M.; Rameshkumar, K.; Rousseau, J.; Anbuudayasankar, SP Entwicklung und Analyse von Prioritätsentscheidungsregeln unter Verwendung des MCDM-Ansatzes für eine flexible Werkstattplanung: Eine Simulationsstudie.** Simul. Model. Pract. Theory **2022,** 114, **102416. [CrossRef**
9. **Pandian, S.; Modrak, V. Möglichkeiten, Hindernisse und Herausforderungen des genetischen Algorithmus bei der Herstellung von Zellbildung.** Adv. Logist Syst. **2009, 3, 63–70.**
10. **Abdulredha, MN; Bara'a, AA; Jabir, AJ Heuristische und metaheuristische Optimierungsmodelle für die Aufgabenplanung im Wolken-Nebel Systeme: Eine Übersicht.** Iraqi J. Electr. Electron. Eng. **2020,** 16, **103–112. [CrossRef]**
11. **Modrak, V.; Pandian, RS; Semanco, P. Kalibrierung von GA-Parametern für Layout-Design-Optimierungsprobleme unter Verwendung von Experimente.** Appl. Sci. **2021,** 11, **6940. [CrossRef]**
12. **Keshanchi, B.; Souri, A.; Navimipour, NJ Ein verbesserter genetischer Algorithmus für die Aufgabenplanung in Cloud-Umgebungen unter Verwendung von Prioritätswarteschlangen: Formale Verifizierung, Simulation und statistische Tests.** J. Syst. Softw. **2017,** 124, **1–21. [CrossRef]**
13. **Jans, R.; Degraeve, Z. Metaheuristiken für dynamische Losgrößenbestimmung: Eine Überprüfung und ein Vergleich von Lösungsansätzen.** Eur. J. Oper. Res. **2007,** 177, **1855–1875. [CrossRef]**
14. **Han, BA; Yang, JJ Eine auf tiefem Verstärkungslernen basierende Lösung für das Problem der flexiblen Werkstattplanung.** Int. J. Simul. Model. **2021,** 20, **375–386. [CrossRef]**
15. Shyalika , **C .; Silva, T.; Karunananda, A. Reinforcement Learning in Dynamic Task Scheduling: Eine Übersicht . 306. [CrossRef]**
16. **Wang, X.; Zhang, L.; Ren, L.; Xie, K.; Wang, K.; Ye, F.; Chen, Z. Kurzer Überblick über die Anwendung von Reinforcement Learning auf Job Shop Scheduling-Probleme.** J. Syst. Simul. **2022,** 33, **2782–2791.**

Machine Translated by Google

17. **Dima, IC; Gabrara, J.; Modrak, V.; Piotr, P.; Popescu, C. Einsatz von Expertensystemen im operativen Produktionsmanagement. In Proceedings der 11. WSEAS International Conference on Mathematics and Computers in Business and Economics (MCBE'10), Iasi, Rumänien, 13.–15. Juni 2010; WSEAS Press: Stevens Point, WI, USA, 2010.**
18. **Waschneck, B.; Reichstaller, A.; Belzner, L.; Altenmüller, T.; Bauernhansl, T.; Knapp, A.; Kyek, A. Optimierung globaler Produktionsplanung mit Deep Reinforcement Learning.** Procedia CIRP **2018,** 72, **1264–1269.** [CrossRef]
19. **Yan, J.; Liu, Z.; Zhang, T.; Zhang, Y. Autonome Entscheidungsfindungsmethode des Transportprozesses für flexible Werkstattplanungsprobleme basierend auf bestärkendem Lernen. In Proceedings der 2021 International Conference on Machine Learning and Intelligent Systems Engineering, MLISE, Chongqing, China, 9.–11. Juli 2021; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA; S. 234–238.**
20. **Modrak, V.; Pandian, RS** Operations Management Research und zellulare Fertigungssysteme; **IGI Global: Hershey, PA, USA, 2010.**
21. **Huang, Z.; Liu, Q.; Zhu, F. Hierarchisches Verstärkungslernen mit adaptiver Planung für die Robotersteuerung.** Eng. Appl. Artif. Intell. **2023,** 126, **107130.** [CrossRef]
22. **Arviv, K.; Stern, H.; Edan, Y. Kollaboratives Verstärkungslernen für ein Zwei-Roboter-Jobtransfer-Flow-Shop-Planungsproblem.** Int. J. Prod. Res. **2016,** 54, **1196–1209.** [CrossRef]
23. **Wen, X.; Zhang, X.; Xing, H.; Ye, G.; Li, H.; Zhang, Y.; Wang, H. Ein verbesserter genetischer Algorithmus basierend auf bestärkendem Lernen für das Planungsproblem der Flugzeugmontage.** Comput. Ind. Eng. **2024,** 193, **110263.** [CrossRef]
24. **Aydin, ME; Öztemel, E. Dynamische Werkstattplanung mit Hilfe von Reinforcement-Learning-Agenten.** Robot. Auton. Syst. **2000,** 33, **169–178.** [Querverweis]
25. **Qu, S.; Wang, J.; Govil, S.; Leckie, JO Optimierte adaptive Planung eines Fertigungsprozesssystems mit vielseitig qualifizierter Belegschaft und mehreren Maschinentypen: Ein ontologiebasierter, Multi-Agent-Reinforcement-Learning-Ansatz.** Procedia CIRP **2016,** 57, **55–60.** [CrossRef]
26. **Luo, S.; Zhang, L.; Fan, Y. Dynamische Mehrzielplanung für flexible Werkstätten durch Deep Reinforcement Learning.** Comput. Ind. Ing. **2021,** 159, **107489.** [CrossRef]
27. **Zhou, L.; Zhang, L.; Horn, BKP Dynamische Planung auf Basis von Deep Reinforcement Learning in der intelligenten Fertigung.** Procedia CIRP **2020,** 93, **383–388.** [CrossRef]
28. **Wang, L.; Pan, Z.; Wang, J. Eine Überprüfung der intelligenten Optimierung auf Basis von bestärkendem Lernen für die Fertigungsplanung.** Komplexes Systemmodell. Simul. **2021, 1, 257–270.** [CrossRef]
29. **Kayhan, BM; Yildiz, G. Anwendungen des Reinforcement Learning bei Maschinenplanungsproblemen: Eine umfassende Literaturübersicht .** J. Intell. Manuf. **2021,** 34, **905–929.** [CrossRef]
30. **Broadus, RN Auf dem Weg zu einer Definition der „Bibliometrie“.** Scientometrics **1987,** 12, **373–379.** [CrossRef]
31. **Arunmozhi, B.; Sudhakarapandian, R.; Sultan Batcha, Y.; Rajay Vedaraj, IS Eine Inferenzanalyse von Edelstahl in additiver Fertigung unter Verwendung bibliometrischer Indikatoren.** Mater Today Proc. **2023,** im Druck. [CrossRef]
32. **Randhawa, K.; Wilden, R.; Hohberger, J. Eine bibliometrische Überprüfung offener Innovation: Festlegung einer Forschungsagenda.** J. Prod. Innov. Manag. **2016,** 33, **750–772.** [CrossRef]
33. **van Raan, A. Fortgeschrittene bibliometrische Methoden als quantitativer Kern von Peer-Review-basierten Evaluierungs- und Foresight-Übungen.** Scientometrics **1996,** 36, **397–420.** [CrossRef]
34. **Brandom, RB** Articulating Reasons: Eine Einführung in den Inferentialismus; **Harvard University Press: Cambridge, MA, USA, 2001.**
35. **Kothari, CR** Forschungsmethodik: Methoden und Techniken; **New Age International: Neu-Delhi, Indien, 2004.**
36. **Suárez, M. Eine inferentielle Konzeption wissenschaftlicher Repräsentation.** Philos. Sci. **2004,** 71, **767–779.** [CrossRef]
37. **Contessa, G. Wissenschaftliche Darstellung, Interpretation und Surrogatdenken.** Philos. Sci. **2007,** 74, **48–68.** [CrossRef]
38. **Govier, T.** Probleme bei der Argumentanalyse und -bewertung; **University of Windsor: Windsor, ON, Kanada, 2018; Band 6.**
39. **Munusamy, R.; Mukherjee, A.; Vasudevan, K.; Venkateswaran, B. Design und Simulation eines künstlichen Intelligenz (KI) Gehirns für ein 2D-Fahrzeugnavigationssystem.** INCAS Bull. **2022,** 14, **53–64.** [CrossRef]
40. **Dunjko, V.; Briegel, HJ Maschinelles Lernen und künstliche Intelligenz im Quantenbereich: Ein Überblick über die jüngsten Fortschritte.** Rep. Prog 2018 , 81, **074001.**
41. **Horvitz, E.; Mulligan, D. Daten, Privatsphäre und das Gemeinwohl.** Science **2015,** 349, **253–255.** [CrossRef]
42. **Sutton, RS; Barto, AG** Reinforcement Learning: Eine Einführung; **MIT: New York, NY, USA, 2018.**
43. **Kuhnle, A.; May, MC; Schäfer, L.; Lanza, G. Erklärbares Verstärkungslernen in der Produktionssteuerung von Lohnfertigungssystemen .** Int. J. Prod. Res. **2022,** 60, **5812–5834.** [CrossRef]
44. **Esteso, A.; Peidro, D.; Mula, J.; Díaz-Madroñero, M. Verstärkungslernen angewendet auf die Produktionsplanung und -steuerung.** Int. J. Prod. Res. **2023,** 61, **5772–5789.** [CrossRef]
45. **Khan, MAM; Khan, MRJ; Tooshil, A.; Sikder, N.; Mahmud, MP; Kouzani, AZ; Nahid, AA Eine systematische Übersicht über auf bestärkendem Lernen basierende Robotik im letzten Jahrzehnt.** IEEE Access **2020, 8, 176598–176623.** [CrossRef]
46. **AlMahamid, F.; Grolinger, K. Reinforcement-Learning-Algorithmen: Ein Überblick und eine Klassifizierung. In Proceedings der 2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Kingston, ON, Kanada, 12.–17. September 2021; IEEE: Piscataway, NJ, USA; S. 1–7.**
47. **Akalin, N.; Loutfi, A. Reinforcement-Learning-Ansätze in der sozialen Robotik.** Sensors **2021,** 21, **1292.** [CrossRef]
48. **Zhang, H.; Yu, T. Taxonomie von Reinforcement-Learning-Algorithmen. In** Deep Reinforcement Learning: Grundlagen, Forschung und Anwendungen; **Springer: Singapur, 2020; S. 125–133.**

Machine Translated by Google

49. Baker, KR; Trietsch, D. Prinzipien der Sequenzierung und Planung; **John Wiley & Sons: Hoboken, NJ, USA, 2009.**

50. Pinedo, ML Scheduling: Theorie, Algorithmen und Systeme; **Springer: Berlin/Heidelberg, Deutschland, 2016.**

51. Allahverdi, A.; Ng, CT; Cheng, TCE; Kovalyov, MY Eine Untersuchung von Planungsproblemen mit Rüstzeiten oder -kosten. Eur. J. Oper. Res. **2008,** 187, **985–1032. [CrossRef]**

52. Panwalkar, SS; Smith, ML Überblick über die Planungsforschung in Fließgeschäften. Oper. Res. **1977,** 25, **45–84. [CrossRef]**

53. Paraschos, PD; Koulinas, GK; Koulouriotis, DE Reinforcement Learning-basierte Optimierung für nachhaltige und schlanke Produktion im Kontext von Industrie 4.0. Algorithms **2024,** 17, **98. [CrossRef]**

54. Wang, S.; Li, J.; Luo, Y. Intelligente Planung für flexible und hybride Produktion mit Multi-Agent Deep Reinforcement Learning. In Proceedings der 2. internationalen IEEE-Konferenz 2021 zu Informationstechnologie, Big Data und künstlicher Intelligenz, ICIBA, Chongqing, China, 17.–19. Dezember 2021; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA; S. 288–294.

55. Tang, J.; Haddad, Y.; Salonitis, K. Rekonfigurierbare Fertigungssystemplanung: Ein Deep-Reinforcement-Learning-Ansatz. Procedia CIRP **2022,** 107, **1198–1203. [CrossRef]**

56. Shahrabi, J.; Adibi, MA; Mahootchi, M. Ein Reinforcement-Learning-Ansatz zur Parameterschätzung in der dynamischen Werkstattplanung . Comput Ind Eng. **2017,** 110, **75–82. [CrossRef]**

57. Yang, J.; You, X.; Wu, G.; Hassan, MM; Almogren, A.; Guna, J. Anwendung von Reinforcement Learning bei UAV-Clusteraufgaben Planung. Future Gen. Comput. Syst. **2019,** 95, **140–148. [CrossRef]**

58. Yuan, X.; Pan, Y.; Yang, J.; Wang, W.; Huang, Z. Studie zur Anwendung von Reinforcement Learning bei der Betriebsoptimierung von HLK-Systemen. In Building Simulation; **Tsinghua University Press: Peking, China, 2021; Band 14, S. 75–87.**

59. Kurinov, I.; Orzechowski, G.; Hämäläinen, P.; Mikkola, A. Automatisierter Bagger basierend auf Reinforcement Learning und Mehrkörper Systemdynamik. IEEE Access **2020, 8, 213998–214006. [CrossRef]**

60. Popper, J.; Motsch, W.; David, A.; Petzsche, T.; Ruskowski, M. Nutzung von Multi-Agent Deep Reinforcement Learning für flexible Werkstattplanung unter nachhaltigen Gesichtspunkten. In Proceedings der International Conference on Electrical, Computer, Communications and Mechatronics Engineering 2021, ICECCME, Mauritius, Mauritius, 7.–8. Oktober 2021; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA.

61. Xiong, H.; Fan, H.; Jiang, G.; Li, G. Eine simulationsbasierte Studie von Dispatching-Regeln in einem dynamischen Job-Shop-Planungsproblem mit Batch-Release und erweiterten technischen Vorrangbeschränkungen. Eur. J. Oper. Res. **2017,** 257, **13–24. [CrossRef]**

62. Palacio, JC; Jiménez, YM; Schietgat, L.; Van Doninck, B.; Nowé, A. Ein Q-Learning-Algorithmus für flexible Werkstattplanung in einem realen Fertigungsszenario. Procedia CIRP **2022,** 106, **227–232. [CrossRef]**

63. Chang, J.; Yu, D.; Hu, Y.; He, W.; Yu, H. Deep Reinforcement Learning für dynamische flexible Job Shop Scheduling mit Random Prozesse 2022, 10 , **760. [CrossRef]**

64. Liu, R.; Piplani , R.; Toro, C. Deep Reinforcement Learning für die **dynamische Planung einer flexiblen Werkstatt.** **4049–4069. [CrossRef]**

65. Samsonov, V.; Kemmerling, M.; Paegert, M.; Lütticke, D.; Sauermann, F.; Gützlaff, A.; Schuh, G.; Meisen, T. Fertigungssteuerung in Job-Shop-Umgebungen mit bestärkendem Lernen. In Proceedings der 13. Internationalen Konferenz über Agenten und künstliche Intelligenz (ICAART 2021), Online, 4.–6. Februar 2021; SciTePress: Setúbal, Portugal; S. 589–597.

66. Cunha, B.; Madureira, AM; Fonseca, B.; Coelho, D. Deep Reinforcement Learning als Job-Shop-Scheduling-Solver: Eine Literaturübersicht . In Proceedings der 18. Internationalen Konferenz über Hybrid Intelligent Systems (HIS 2018), Porto, Portugal, 13.–15. Dezember 2018; Madureira, AM, Abraham, A., Gandhi, N., Varela, ML, Hrsg.; Springer International Publishing: Cham, Schweiz, 2018.

67. Wang, X.; Zhang, L.; Lin, T.; Zhao, C.; Wang, K.; Chen, Z. Lösen von Job-Scheduling-Problemen in einer Umgebung mit Ressourcenpräemption mit Multi-Agent-Reinforcement-Learning. Robot. Comput. Integr. Manuf. **2022,** 77, **102324. [CrossRef]**

68. Oh, SH; Cho, YI; Woo, JH Distributionales Verstärkungslernen mit unabhängigen Lernern für flexible Werkstattplanungsprobleme mit hoher Variabilität. J. Comput. Des. Eng. **2022, 9, 1157–1174. [CrossRef]**

69. Zhang, Y.; Zhu, H.; Tang, D.; Zhou, T.; Gui, Y. Dynamische Werkstattplanung basierend auf Deep Reinforcement Learning für Multi-Agenten-Fertigungssysteme. Robot. Comput. Integr. Manuf. **2022,** 78, **102412. [CrossRef]**

70. Liang, Y.; Sun, Z.; Song, T.; Chou, Q.; Fan, W.; Fan, J.; Rui, Y.; Zhou, Q.; Bai, J.; Yang, C.; et al. Lenovo-Terminpläne für Laptops Fertigung mittels Deep Reinforcement Learning. Interfaces **2022,** 52, **56–68. [CrossRef]**

71. Chen, Y.; Guo, W.; Liu, J.; Shen, S.; Lin, J.; Cui, D. Ein Multi-Sollwert-Kühlregelungsansatz für luftgekühlte Rechenzentren unter Verwendung der tiefer Q-Netzwerk-Algorithmus. Meas. Control **2024,** 57, **782–793. [CrossRef]**

72. Théate, T.; Ernst, D. Eine Anwendung von Deep Reinforcement Learning im algorithmischen Handel. Expert Syst. Appl. **2021,** 173, **114632. [Querverweis]**

73. Sanaye, S.; Sarrafi, A. Eine neuartige Energiemanagementmethode basierend auf dem Deep Q Network-Algorithmus für niedrige Betriebskosten eines integriertes Hybridsystem. Energy Rep. **2021, 7, 2647–2663. [CrossRef]**

74. Luo, S. Dynamische Planung für flexible Werkstätten mit neuen Stelleneinfügungen durch Deep Reinforcement Learning. Appl. Soft Comput. J. **2020,** 91, **106208. [CrossRef]**

75. Luo, S.; Zhang, L.; Fan, Y. Echtzeitplanung für dynamische, teilweise wartefreie, multiobjektive, flexible Job-Shops von Deep Bestärkendes Lernen. IEEE Trans. Autom. Sci. Eng. **2022,** 19, **3020–3038. [CrossRef]**

Machine Translated by Google

76. **Hu, J.; Wang, H.; Tang, H.-K.; Kanazawa, T.; Gupta, C.; Farahat, A. Wissensgestütztes bestärkendes Lernen für die integrierte Produktions- und Wartungsplanung mehrerer Maschinen.** Comput. Ind. Eng. **2023,** 185, **109631.** [CrossRef]
77. **Du, Y.; Li, JQ; Chen, XL; Duan, PY; Pan, QK Wissensbasiertes Verstärkungslernen und Schätzung des Verteilungsalgorithmus für flexible Job-Shop-Planungsprobleme.** IEEE Trans. Emerg. Top. Comput. Intell. **2023, 7, 1036–1050.** [CrossRef]
78. **Li, Y.; Gu, W.; Yuan, M.; Tang, Y. Datengesteuerte dynamische Planung in Echtzeit für flexible Werkstätten mit unzureichenden Transportmöglichkeiten Ressourcen mithilfe eines hybriden Deep-Q-Netzwerks.** Robot. Comput. Integr. Manuf. **2022,** 74, **102283.** [CrossRef]
79. **Zhou, T.; Zhu, H.; Tang, D.; Liu, C.; Cai, Q.; Shi, W.; Gui, Y. Reinforcement Learning zur Online-Optimierung der Werkstattplanung in einer intelligenten Fertigungsfabrik.** Adv. Mech. Eng. **2022,** 14, **16878132221086120.** [CrossRef]
80. **Wang, L.; Hu, X.; Wang, Y.; Xu, S.; Ma, S.; Yang, K.; Wang, W. Dynamische Job-Shop-Planung in der intelligenten Fertigung mit Deep bestärkendes Lernen.** Comput. Netw. **2021,** 190, **107969.** [CrossRef]
81. **Wang, Y.; Usher, JM Anwendung von Reinforcement Learning für agentenbasierte Produktionsplanung.** Eng. Appl. Artif. Intell. **2005,** 18, **73–82.** [CrossRef]
82. **Cancino, CA; Merigó, JM; Coronado, FC Eine bibliometrische Analyse führender Universitäten in der Innovationsforschung.** J. Innov. Knowl. **2017, 2, 106–124.** [CrossRef]
83. **Varin, C.; Cattelan, M.; Firth, D. Statistische Modellierung des Zitationsaustauschs zwischen Statistikzeitschriften.** JR Stat. Soc. Ser. A Stat. Chem . Soc. **2016,** 179, **1–63.** [PubMed]
84. **Moral-Muñoz, JA; Herrera-Viedma, E.; Santisteban-Espejo, A.; Cobo, MJ Softwaretools zur Durchführung bibliometrischer Analysen in Wissenschaft: Eine aktuelle Übersicht.** Prof. De La Inf./Inf. Prof. **2020,** 29, **e290103.** [CrossRef]
85. **Curry, S. Lassen wir die Rhetorik hinter uns: Es ist Zeit, die Art und Weise zu ändern, wie wir Forschung beurteilen.** Nature **2018,** 554, **147–148.** [CrossRef] [PubMed]
86. **Al-Hoorie, A.; Vitta, JP Die sieben Sünden der L2-Forschung: Eine Überprüfung der statistischen Qualität von 30 Zeitschriften und ihres CiteScore, SJR, SNIP, JCR Impact Factors.** Lang. Teach. Res. **2019,** 23, **727–744.** [CrossRef]
87. **Waltman, L.; Van Eck, NJ; Noyons, ECM Ein einheitlicher Ansatz zur Kartierung und Clusterung bibliometrischer Netzwerke.** J. Informetr. **2010, 4, 629–635.** [CrossRef]
88. **Cheng, F.; Huang, Y.; Tanpure, B.; Sawalani, P.; Cheng, L.; Liu, C. Kostenbewusste Jobplanung für Cloud-Instanzen mit Deep bestärkendes Lernen.** Clust. Comput. **2022,** 25, **619–631.** [CrossRef]
89. **Thaipisutikul, T.; Chen, Y.-C.; Hui, L.; Chen, S.-C.; Mongkolwat, P.; Shih, TK Die Bedeutung von Deep Reinforcement Learning für praktische KI-Anwendungen. In Proceedings der 12. Internationalen Konferenz zu Ubi-Media Computing, Bali, Indonesien, 5.–8. August 2019; S. 24–29.**
90. **Yan, J.; Huang, Y.; Gupta, A.; Gupta, A.; Liu, C.; Li, J.; Cheng, L. Energiebewusste Systeme für die Echtzeit-Jobplanung in Cloud-Daten Zentren: Ein Ansatz des tiefen bestärkenden Lernens.** Comput. Electr. Eng. **2022,** 99, **107688.** [CrossRef]
91. **Piller, FT Mass Customization: Überlegungen zum aktuellen Stand des Konzepts.** Int. J. Flex. Manuf. Syst. **2004,** 16, **313–334.** [CrossRef]
92. **Suzi´c, N.; Forza, C.; Trentin, A.; Aniši´c, Z. Implementierungsrichtlinien für Mass Customization: Aktuelle Merkmale und Verbesserungsvorschläge.** Prod. Plan. Control **2018,** 29, **856–871.** [CrossRef]
93. **Altenmüller, T.; Stüker, T.; Waschneck, B.; Kuhnle, A.; Lanza, G. Reinforcement learning für eine intelligente und autonome Produktionssteuerung komplexer Werkstätten unter Zeitdruck.** Prod. Eng. **2020,** 14, **319–328.** [CrossRef]
94. **Zhao, Y.; Zhang, H. Anwendung von maschinellem Lernen und Regelplanung in einem Job-Shop-Produktionssteuerungssystem.** Int. J. Simul. Modell **2021,** 20, **410–421.** [CrossRef]
95. **Kuhnle, A.; Kaiser, JP; Theiß, F.; Stricker, N.; Lanza, G. Designing an adaptive production control system using reinforcement learning.** J. Intell. Manuf. **2021,** 32, **855–876.** [CrossRef]
96. **Panzer, M.; Bender, B.; Gronau, N. Deep Reinforcement Learning in der Produktionsplanung und -steuerung: Eine systematische Literaturübersicht . In Proceedings der Conference on Production Systems and Logistics, Online, 10.–11. August 2021.**
97. **Wojakowski, P.; Warzoÿek, D. Die Klassifizierung von Planungsproblemen unter Produktionsunsicherheit.** Res. Logist. Prod. **2014, 4, 245–256.**
98. **Blackstone, JH; Phillips, DT; Hogg, GL Eine aktuelle Untersuchung von Dispositionsregeln für Fertigungsbetriebe.** Int. J. Prod. Res. **1982,** 20, **27–45.** [CrossRef]
99. **Blazewicz, J.; Ecker, KH; Pesch, E.; Schmidt, G.; Weglarz, J.** Handbook on Scheduling: From Theory to Applications; **Springer: Berlin/Heidelberg, Deutschland, 2007.**
100. **Ivanov, D.; Dolgui, A.; Sokolov, B. Ein dynamischer Ansatz für die mehrstufige Werkstattplanung in einem flexiblen Montagesystem auf Basis von Industrie 4.0. In** Fortschritte bei Produktionsmanagementsystemen. Der Weg zur intelligenten, kollaborativen und nachhaltigen Fertigung: IFIP WG 5.7 Internationale Konferenz, APMS 2017, Hamburg, Deutschland, 3.–7. September 2017, Proceedings, Teil **I; Springer International Publishing: Cham, Schweiz; S. 475–482.**
101. **Modrak, V. (Hrsg.)** Mass Customized Manufacturing: Theoretische Konzepte und praktische Ansätze; **CRC Press: Boca Raton, FL, USA, 2017.**
102. **Komaki, GM; Sheikh, S.; Malakooti, B. Probleme bei der Flow-Shop-Planung bei Montagevorgängen: Ein Überblick und neue Trends.** Int. J. Prod. Res. **2019,** 57, **2926–2955.** [CrossRef]
103. **Yang, Y.; Li, X. Ein wissensbasierter konstruktiver heuristischer Algorithmus für das verteilte Montageblockierungs-Flow-Shop-Planungsproblem .** Expert Syst. Appl. **2022,** 202, **117269.** [CrossRef]

Machine Translated by Google

104. Nasiri, MM; Yazdanparast, R.; Jolai, F. Ein Simulationsoptimierungsansatz für die Echtzeitplanung in einer Open-Shop-Umgebung unter Verwendung einer zusammengesetzten Dispatching-Regel. Int. J. Comput. Integr. Manuf. **2017,** 30, **1239–1252.** [CrossRef]

105. Abdelmaguid, TF Bi-objektive dynamische Multiprozessor-Open-Shop-Planung für Wartung und Gesundheitsdiagnostik. Experte Syst. Appl. **2021,** 186, **115777.** [CrossRef]

106. Tremblet, D.; Thevenin, S.; Dolgui, A. Makespan-Schätzung in einer flexiblen Job-Shop-Planungsumgebung mithilfe von maschinellem Lernen. Int. J. Prod. Res. **2024,** 62, **3654–3670.** [CrossRef]

107. Fu, Y.; Hou, Y.; Wang, Z.; Wu, X.; Gao, K.; Wang, L. Verteilte Planungsprobleme in intelligenten Fertigungssystemen. Tsinghua Sci. Technol. **2021,** 26, **625–645.** [CrossRef]

108. Zhang, Y.; Zhang, X.; Ding, Y. Ein auf bestärkendem Lernen basierender Ansatz zur stochastischen Werkstattplanung. IEEE Trans. Autom. Sci. Eng. **2020,** 17, **72–83.**

109. Rinciog, A.; Meyer, A. Auf dem Weg zur Standardisierung von Reinforcement-Learning-Ansätzen für die stochastische Produktionsplanung. arXiv **2021, arXiv:2104.08196.**

110. Zeng, B.; Yang, Y. Ein hybrider Reinforcement Learning- und Fuzzy Logic-Ansatz für die Job Shop Scheduling. J. Intell. Manuf. **2017,** 28, **1189–1201.**

111. Zhang, G.; Huang, Y. Fuzzy-Verstärkungslernen für die mehrzielige dynamische Planung eines flexiblen Fertigungssystems. J. Intel. Hersteller **2005,** 16, **293–304.**

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultieren.

Machine Translated by Google

Algorithmen

MDPI

Artikel

Inversionsbasiertes Deblending im gemeinsamen Mittelpunktsbereich Verwendung von hochauflösendem Radon im Zeitbereich

Kai Zhuang [1], **Daniel Trad** [1,*] **und Amr Ibrahim** [2]

[1] Department of Earth, Energy, and Environment, Faculty of Science, University of Calgary, Calgary, AB T2N 1N4, Kanada; kzhuang@ucalgary.ca
[2] Department of Physics, Faculty of Science, Beni-Suef University, Beni-Suef 62521, Ägypten; amr.ibrahim@ucalgary.ca
* Korrespondenz: daniel.trad@ucalgary.ca

Zusammenfassung: Wir implementieren eine inversionsbasierte Deblending-Methode in den Common Midpoint Gathers (CMP) als Alternative zu den Standardmethoden der Common Receiver Gather (CRG)-Domäne. Der Hauptvorteil des Deblendings in der CMP-Domäne besteht darin, dass Reflexionen von abfallenden Schichten um Null-Offsets zentriert sind. Infolgedessen weisen CMP-Gathers eine einfachere Struktur auf als Common Receiver Gathers (CRGs), bei denen diese Reflexionen apexverschoben sind. Folglich können wir einen hyperbolischen Radon-Operator mit Null-Offset verwenden, um CMP-Gathers zu verarbeiten. Dieser Operator ist eine rechnerisch effizientere Alternative zum apexverschobenen hyperbolischen Radon, der für die Verarbeitung von CRG-Gathers erforderlich ist. Sparse- Transformationen wie die Radon-Transformation können Reflexionen stapeln und Sparse-Modelle erstellen, die gemischte Quellen trennen können. Wir verwenden den Radon-Operator, um ein inversionsbasiertes Deblending-Framework zu entwickeln, das eine Sparse-Modell-Beschränkung enthält. Die Einbeziehung einer Sparsity-Beschränkung in den Inversionsprozess verbessert die Fokussierung der Transformation und die Datenwiederherstellung. Durch inversionsbasiertes Deblending können wir alle beobachteten Daten berücksichtigen, indem wir den Blending-Operator in die Kostenfunktion integrieren. Unsere Beispiele aus synthetisc

Schlüsselwörter: Entmischung; spärliches Radon; Inversion

check for updates

Zitat: Zhuang, K.; Trad, D.; Ibrahim, A. Inversionsbasiert Deblending im gemeinsamen Mittelpunktsbereich mithilfe von hochauflösendem Radon im Zeitbereich. Algorithms **2024,** 17, **344. https://doi.org/10.3390/a17080344**

Wissenschaftlicher Redakteur: Frank Werner

Empfangen: 16. Mai 2024
Überarbeitet: 25. Juni 2024
Akzeptiert: 5. August 2024
Veröffentlicht: 7. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/4.0/).

1. Einleitung

Die Suche nach einer effizienten seismischen Datenerfassung ist eine bekannte Herausforderung bei geophysikalischen Erkundungen. Die simultane Erfassung, auch als gemischte Erfassung bekannt, bietet eine Lösung, da sie eine Methode bietet, die die Zeit, die im Feld verbracht wird, durch gleichzeitiges Auslösen mehrerer Quellen erheblich reduziert. Diese Technik, die ursprünglich von Garottu [1] für die Erfassung von Vibroseis vorgeschlagen und später von Beasley et al. [2], Beasley [3] für Meeresumgebungen erweitert wurde, ermöglicht die Interferenz von Quellen zum Zeitpunkt der Erfassung, die anschließend während der Verarbeitung getrennt werden. Trotz dieser zeitsparenden Vorteile sind bestehende Modelle und Algorithmen zum Entm

Ziel dieser Studie ist es, diese Einschränkungen zu beseitigen und die Genauigkeit und Effizienz der Entmischung seismischer Daten zu verbessern. Aktuelle Methoden, darunter Zufallszündung und Phasenkodierung, haben den Weg für neue Ansätze zur Datenerfassung geebnet. Der Entmischungsprozess bleibt jedoch ein erheblicher Engpass, wobei zwei Denkrichtungen vorherrschen : der auf Rauschunterdrückung basierende Ansatz [4], der sich auf die sprunghafte und zufällige Rauschdämpfung konzentriert , und der auf Inversion basierende Ansatz [5], der versucht, das Signal durch Umkehrung des Mischprozesses zu rekonstruieren. In diesem Artikel wird der Schwerpunkt hauptsächlich auf die Verarbeitung von Daten aus dem Zufallszündung

Die Einführung kleiner zufälliger Verzögerungen in die Quellenzündungszeiten, wie von Berkhout [6] vorgeschlagen, ist ein kritischer Schritt im Deblending-Prozess. Beim Pseudo-Deblending werden die gemischten Aufnahmen auf Nullzündungszeiten umgestellt und die Bilder entsprechend der Anzahl gleichzeitiger Quellen dupliziert. Diese Manipulation aufgrund der zufäl

Machine Translated by Google

die Interferenzen von benachbarten Quellen als zufälliges Rauschen in einigen der gemeinsamen Datenbereiche , so dass nur das Zielsignal zur Wiederherstellung als kohärentes Ereignis übrig bleibt. Pseudo-entmischte Daten weisen dann inkohärente Strukturen innerhalb der Bereiche Common Midpoint Gather (CMP), Common Receiver Gather (CRG) und Common Offset Gather (COF) auf, wodurch der Entmischprozess erleichtert wird. Es gibt verschiedene Methoden, um die Inkohärenz von Mischrauschen auszunutzen, wie die Trennungsarbeit von Lin und Herrmann [7] und nachfolgende Techniken wie Stolt-Radon-Rauschunterdrückung [8], SSA-Filterung [9], Rauschunterdrückung mit einem raumvariablen Medianfilter [10], Vorkonditionierung mit einem diskreten Curvelet-Filter [11] und Vorkonditionierung mit robuster Schwellenwertbildung im FK-Bereich [12], jede mit ihren eigenen Vor- und Nachteilen, und die oft entweder mit Problemen bei der effektiven Unterscheidung zwischen Signal und Rauschen (Quelleninterferenz) bei den Rauschunterdrückungsansätzen oder mit sehr teuren und komplexen Transformationsoperatoren bei den inversionsbasierten Ansätzen zu kämpfen haben .

In diesem Artikel wird ein inversionsbasierter Deblending-Ansatz unter Verwendung der CMP-Domäne vorgestellt, der eine Hochgeschwindigkeitslösung unter Verwendung eines Radon-Transformationsoperators verspricht. Die Verwendung der hyperbolischen Radon-Transformation zum Deblending dient mehreren Zwecken und trägt erheblich zur Wiederherstellung des seismischen Signals bei. Angesichts der intrinsischen Natur der seismischen Erfassung manifestieren sich die akustischen Reaktionen, die aus dem Erdinneren kommen, als hyperbolische Ankünfte an der Oberfläche. Die hyperbolische Radon-Transformation ist durch die Aggregation von Daten auf der Grundlage hyperbolischer Flugbahnen ideal für die Kartierung seismischer Daten geeignet. Die Entscheidung für die CMP-Domäne anstelle der herkömmlichen CRG-Domänenmethode ermöglicht die Verwendung einer rationalisierteren Version des Radon-Transformationsoperators,

2. Theorie

2.1. Blending-Operator

Der Blending-Operator ermöglicht uns den Übergang zwischen dem Blending-Bereich, in dem sich Quellen überlappen (Supershot genannt), und dem Unblending-Bereich, in dem Quellen zu ihren jeweiligen Shot-Aufzeichnungen gehören, wobei die Feuerzeiten für jeden Shot auf Null gesetzt werden. Dies kann als Blending-Operator ÿ dargestellt werden, der die Informationen zu Überlappung und Verzögerungszeit enthält . Wir können dann unseren Unblending-Datenwürfel als D und unsere Blending-Daten als b mit der folgenden Beziehung darstellen [6,13]:

$$\mathbf{b} = \ddot{y}\mathbf{D}. \quad (1)$$

Wenn wir den Adjunkt des Blending-Operators auf die gemischten Daten anwenden, erhalten wir die pseudo- deblendeten Daten im ungemischten Raum, wobei jede Aufnahme ihren eigenen Datenrahmen belegt, im Gegensatz zu mehreren Aufnahmen pro Datenrahmen. Dies kann wie folgt ausgedrückt werden [6,13]:

$$\tilde{\mathbf{D}} = \ddot{y}\,\mathbf{H}\mathbf{b}. \quad (2)$$

Hier stellt $\tilde{\mathbf{D}}$ die pseudo-entmischten Daten in Standard-Aufnahmebildern dar. Die Aufgabe der Mischmatrix besteht darin, Aufnahmen mit zufälligen Verzögerungszeiten zusammenzufügen, wodurch die Daten einzelner Aufnahmeaufzeichnungen effektiv übereinandergelegt werden. Dies bedeutet auch, dass die Pseudo-Entmischung dazu dient, die Daten auf die jeweiligen Bilder neu zu verteilen. Dadurch entsteht eine übervollständige Darstellung der Daten mit redundanten Informationen, die in jedem pseudo-entmischten Quellfeld gespeichert sind. Es ist wichtig zu beachten, dass der Adjungierte Operator die Quelleninterferenzen nicht aus den Daten entfernt, sondern sie lediglich kopiert.

Die erfolgreiche Entmischung von Daten mithilfe der Random-Firing-Technik hängt von der Einbeziehung zufälliger Verzögerungszeiten ab. Diese Verzögerungen sind entscheidend, um Quelleninterferenzen in etwas umzuwandeln, das zufälligem Rauschen ähnelt, wenn die Daten über den Adjungierten-Blending- Operator neu zugeordnet werden. Es ist wichtig, dass diese Verzögerungszeiten ausreichend randomisiert sind, sodass in den neu zugeordneten Datendomänen keine Muster entstehen, da kohärente Muster in den Verzögerungszeiten die Wirksamkeit unserer Kohärenzbeschränkung verringern. Darüber hinaus sollten die Verzögerungen ausreichend lang sein, um die Trennbarkeit von Ereignissen sicherzustellen, wobei die Mindestverzögerungslänge eingehalten werden muss, die durch eine Funktion der niedrigs

Machine Translated by Google

und Abma [14]. Gleichzeitig müssen sie kurz genug sein, um einen Vorteil bei der Verkürzung der Erfassungszeiten zu erzielen.

Zufällige Verzögerungen, die in den Mischplan und folglich in den Mischoperator eingeführt werden , sind entscheidend für die Vereinfachung des Deblendings. Wenn der adjungierte Mischoperator (Pseudo-Deblending) auf den gemischten Datensatz angewendet wird, werden die zufälligen Verzögerungen ents Das Mischen kann als eine Operation an standardmäßigen Erfassungsdaten konzipiert werden, die Quelleninterferenzen einführt.

Der Einfluss der Addition auf die zufälligen Verzögerungszeiten ist sichtbar, wenn die Daten nach der Pseudo-Deblendung in einen der anderen geometrischen Bereiche sortiert werden. Nach der Anwendung des Adjungierten-Blending-Operators sind neben den inkohärenten Ereignissen, die vor der Anwendung der Transformation sichtbar waren, auch kohärente Ereignisse sichtbar. Diese kohärenten Ereignisse wollen wir extrahieren, da sie den Teil der Daten darstellen, dessen Verzögerungszeiten richtig korrigiert wurden, während die spontanen Daten aufgrund störender Schüsse entfern Durch die Verwendung einer Inversionsmethode und einer Kohärenzbeschränkung wie der hochauflösenden Radon-Transformation können wir das Störsignal effektiv entfernen und unsere Daten entmischen.

2.2. Radon-Transformation

Um das Problem der Unterbestimmtheit zu lösen, das im Inversionsproblem auftritt, verwenden wir die hyperbolische Radon-Transformation (HRT), um das Problem weiter einzuschränken. Dieser zusammengesetzte Operator, der sowohl aus dem Mischoperator als auch dem Radon-Transformationsoperator besteht, formuliert die Inversion nun in ein spärlich eingeschränktes Problem um. Die HRT ist eine Variante der Radon-Transformation, die über hyperbolische Basisfunktionen summiert; in der Geophysik wird diese Gleichung als Funktion der Ankunftszeiten und des Quelle-Empfänger-Offsets und der Intervallgeschwindigkeit definiert , wie bei Thorson

$$m(p, \ddot{y}) = \int_{h1}^{h2} d(h, t = \ddot{y} \sqrt{{}^2 + p\, 2h\, 2)}\, dh. \quad (3)$$

Dabei stellt m(p, ÿ) die Radon-Raumdaten dar, p bezeichnet die Langsamkeit, $\ddot{y}$ ist die Hin- und Rücklaufzeit, h1 und h2 sind jeweils die unteren und oberen Offset-Grenzen und d ist der zu transformierende Datenraum. Die Langsamkeit p wird als Kehrwert der Geschwindigkeit 1/v definiert. Die Kartierung auf Grundlage hyperbolischer Ankünfte ermöglicht eine präzise und spärliche Darstellung seismischer Daten, da seismische Signale bei standardmäßigen seismischen Untersuchungen als hyperbolische Ereignisse erscheinen. Die Radon-Transformation wird in verschiedenen Anwendungen eingesetzt, darunter Geschwindigkeitsanalyse [15], Mehrfachunterdrückung [16], Interpolation im Frequenz- oder Zeitbereich [17] und Entmischung durch Rausc

Obwohl sich die HRT gut für seismische Daten eignet, weist sie mehrere Einschränkungen auf. In erster Linie erfordert die Transformation, dass die hyperbolischen Ankünfte symmetrisch und im Datenraum zentriert sind, was bei realen Daten nicht unbedingt der Fall ist. In der Realität führt das Einfallen unterirdischer Strukturen dazu, dass sich hyperbolische Ankünfte nach oben verschieben, was zu nicht zentrierten Ankünften führt. Die Symmetrie wird durch zwei Faktoren in Frage gestellt: Anisotropie und Vermessungsbedingungen. Oberflächensensor-Arrays werden im Allgemeinen nicht auf einer perfekt flachen Oberfläche platziert, was zu Abweichungen von hyperbolischen Ankünften führt. Dieses Problem lässt sich leicht durch eine Spurstatikkorrektur lösen, die Teil des Standardverarbeitungsablaufs ist. Für das Anisotropieproblem kann in einfachen Fällen eine modifizierte Gleichung verwendet werden, um die Kurvenstreckung zu berücksichtigen. Die Verschiebung von zentrierten Positionen stellt ein ziemlich komplexes Problem mit mehreren möglichen Lösungen dar. Die erste Lö Bei diesem Ansatz werden mehrere Radon-Panels für ein einzelnes Datenbild generiert, die jeweils einer Reihe von erwarteten verschobenen Scheitelpunktpositionen entsprechen. Diese Lösung ist zwar robust und lässt sich gut auf seismische Daten abbilden, erhöht jedoch die Laufzeiten erheblich, da es keine Abkürzung zur Generierung von Radon-Panels gibt und die Verarbeitungszeiten mit der Anzahl der gescannten Scheitelpunkte multipliziert werden. Die zweite Lösung für das Zentrierproblem besteht darin, die Daten in die CMP-Domäne zu transformieren, wo die Scheitelpunkte auch bei abfallenden Reflexionen z

Machine Translated by Google

Lösung hat einen großen Nachteil: Die Abtastung im CMP-Bereich ist im Allgemeinen schlechter als bei anderen Datenbereichen, was Aliasing zu einem erheblichen Problem macht.

2.3. Inversionsbasiertes Deblending

Angesichts unseres bereits definierten Transformationsoperators, des Abbildungsoperators ÿ , der den Datenraum b auf den Modellraum D in Gleichung (1) abbildet, können wir das Problem als Inversionsproblem formulieren, das die Operation umkehrt [19]. Um diese Operation umzukehren, müssen wir eine Zielfunktion formulieren, die die Nichtübereinstimmung zwischen den Daten und den modellierten Daten minimiert, mit anderen Worten die Modellinformationen minimiert, die nicht auf die Daten abgebildet werden können. Diese Zielfunktion wird oft wie folgt dargestellt [17]:

$$\begin{aligned} &\textbf{Minimieren Sie } ||D|| \\ &2\ \textbf{2 Vorbehaltlich } ||(b \ddot{y} \ddot{y}D)||2\ \textbf{2 .} \end{aligned} \quad \textbf{(4)}$$

Die Zielfunktion für die Inversion gemischter Daten wird wie folgt geschrieben:

$$\mathbf{J} = ||b \ddot{y} \ddot{y}D||, \quad \textbf{(5)}$$

Die direkte Inversion des Mischoperators ist schlecht gestellt; daher kann keine eindeutige Lösung ermittelt werden, indem nur der Mischoperator ÿ invertiert wird. Diese schlechte Stellung ist auf die Tatsache zurückzuführen, dass der Mischoperator nur dazu dient, Daten zu sammeln und zu verschieben, während der adjungierte Operator nur dazu dient, sie zu verteilen. Um dieses Problem zu umgehen, haben wir das Inversionsschema neu formuliert, indem wir den Radon-Operator als Kohärenzbeschränkung hinzug

$$\mathbf{J} = ||b \ddot{y} \ddot{y}\mathbf{Lm}|| + \mu||m||, \quad \textbf{(6)}$$

wobei L die Radon-Transformation ist. Indem wir das Problem auf diese Weise formulieren, gelangen wir von einem schlecht gestellten Problem zu einem mit spärlichen Beschränkungen.

2.4. Spärliche Regularisierung

Wir verwenden die spärliche Regularisierung für unser Inversionsschema als Möglichkeit, die ÿ1-Norm im Modellraum zu lösen, als robustere Lösung für die Radon-Transformation [17] und wiederum das Deblending-Schema. Zur Lösung der spärlichen Lösung in diesem Dokument verwenden wir die Methode der iterativ neu gewichteten Kleinstquadrate-Inversion (IRLS) [20]. Aufgrund ihrer Einfachheit ist IRLS eine beliebte Methode zur Lösung spärlicher und/oder robuster inverser Lösungen. Tatsächlich formuliert IRLS die Berechnung der Lösung mit dem kleinsten Betrag als Wiederholung von Kleinstquadrate-Iterationen unter Verwendung von Gewichten in den Daten- und Modellräumen neu. Diese Herangehensweise ist effizienter als die direkte Lösung mit dem kleinsten Betrag, da wir schnellere Kleinstquadrate-Löser wie den konjugierten Gradienten verwenden können, um die Rechenzeit zu verringern. Die Verwendung von IRLS ermöglicht auch Inversionen gemischter Normen; Für dieses Papier verwenden wir eine ÿ2-Norm im Datenraum und eine ÿ1-Norm im Modellraum.

Wie viele Forscher herausgefunden haben, bietet die Anwendung einer spärlichen Beschränkung auf die Radon-Transformation zahlreiche Vorteile und nur wenige Nachteile für die Kartierung seismischer Daten [17]. Diese Verbesserungen beheben Probleme wie fehlende Daten aufgrund von Aliasing und begrenzte Blendeninformationen bei der Abbildung und Rekonstruktion. Die Vorteile erstrecken sich auch auf die Verwendung der Radon-Transformation beim Deblending. Der Hauptvorteil der Durchsetzung einer spärlichen Einschränkung im Modellraum für unsere Methode besteht darin, dass sich das unregelmäßige und zufällige Rauschen aus den zufälligen Verzögerungszeiten als Pfadintegral schlecht im Radon-Raum abbilden lässt. Diese schlechte Abbildung führt zu Schmierereien mit geringer Amplitude im Radon-Raum, im Gegensatz zu den Brennpunkten mit hoher Amplitude, die bei kohärenten hyperbolischen Ankünften auftreten. Ergänzt wird dies durch die Tatsache, dass in den Aufnahmebildern, zu denen die Daten gehören, die Signale richtig summiert und fokussiert werden. Folglich konvergiert der Löser auf die spärlichste Lösung, indem er verschmiertes gemischtes Rauschen herunter- und kohärente Brennpunkte hochgewichtet. Dies ist auch die am stärksten entmischte Lösung, da die Daten in die Frames neu zugeordnet werden, in denen sie am kohärentesten erscheinen, und das Mischrauschen vollständig du

Machine Translated by Google

IRLS ist eine Möglichkeit, die spärliche Lösung eines Inversionsproblems indirekt zu lösen, indem es als mehrere Iterationen eines kleinsten Quadrate-Problems, in diesem Fall eines konjugierten Gradienten (CG), dargestellt wird, wobei die äußeren Iterationen Daten und Modellgewichte aktualisieren, um den CG-Algorithmus zu lösen und die Lösung auf die ÿ1-Norm zu bringen. Die Durchsetzung der ÿ1-Norm im Datenraum ermöglicht eine geringere Empfindlichkeit gegenüber Ausreißern, die sich andernfalls negativ auf die Inversion auswirken würden, was oft als robuste Inversion bezeichnet wird. Die Durchsetzung der ÿ1-Norm im Modellraum löst die spärlichste Lösung, hauptsächlich das Modell, das die Daten mit der geringsten Anzahl von Datenpunkten erklärt ; dies wird als spärliche Lösung bezeichnet. Die Verwendung von IRLS ermöglicht es uns, durch die Berechnung der Gewichte die Normen, die wir verwenden möchten, zu mischen und anzupassen und so eine Lösung mit gemischten Normen zu finden.

Durch Hinzufügen von Modell- und Datengewichten können wir dann den Algorithmus so steuern, dass er zur Lösung mit dem kleinsten Betrag (ÿ1) oder den kleinsten Quadraten (ÿ2) konvergiert . **Dazu formulieren wir die ursprüngliche Gleichung der kleinsten Quadrate durch Minimierung der geänderten Gleichung ne**

$$\text{Minimieren Sie } \|\mathbf{Wmm}\|_2^2 \quad \text{unter Berücksichtigung von } \|\mathbf{Wr}(\mathbf{b} - \ddot{y}\,\ddot{y}\mathrm{Lm})\|_2^2 . \tag{7}$$

wobei Wr der Datengewichtsoperator und Wm der Modellgewichtsoperator ist. Diese Gewichte ermöglichen dem Benutzer, die Präferenz entweder auf das Residuum oder das Modell einzustellen, wobei Wm geändert werden kann, um eine bestimmte Anpassung an entweder ein spärliches ÿ1- oder glattes ÿ2- Modell zu erhalten, und Wr geändert werden **kann , um nicht spontane Ereignisse zu bevorzugen oder alle Ereignisse in den Eingabedaten gleich zu „gewichten“ . Die Operatoren sind normalerweise Diagonalmatrizen. Beispielsweise kann Wr wie folgt konstruiert werden:**

$$\mathrm{diag}(\mathbf{Wr}) = \mathbf{|r|}\,(\mathbf{p\ddot{y}2})/2 \quad \text{wobei } \mathbf{r} = \ddot{y}\mathbf{Lm}\,\ddot{y}\,\mathbf{b}. \tag{8}$$

wobei p **entweder 1 oder 2 ist, um die Modellgröße unter Verwendung der Datennorm ÿ1 oder ÿ2** zu messen.

Wm wird normalerweise wie folgt konstruiert:

$$\mathrm{diag}(\mathbf{Wm}) = \mathbf{|m|}\,(2\ddot{y}p)/2. \tag{9}$$

wobei p **entweder 1 oder 2 ist, um basierend auf der Modellnorm ÿ1 oder ÿ2** invertiert zu werden. Da die **Datengewichte durch Division erstellt werden, wird ein Dämpfungsfaktor gewählt, um eine Division durch Null zu vermeiden. Der Dämpfungsfaktor wird als Perzentil der Daten bestimmt. Der Hauptvorteil des Inversionsansatzes für die Entmischung besteht darin, dass das gesamte Signal- und Mischrauschen in anderen Bereichen erklärt werden kann, in denen es kohärent erscheint, im Gegensatz zum Denoising- Ansatz, der versucht, das Mischrauschen Bild für**

Wir können die spärliche Inversion auch verwenden, um den Denoising-Ansatz zur Entmischung zu demonstrieren, bei dem die Radon-Transformation zum Entrauschen einzelner Empfänger-Sammlungen verwendet wird. Der Denoising- Ansatz wendet zunächst Pseudo-Deblending auf die gemischten Daten an, um einen Datensatz in Standard **-Aufnahmerahmen zurückzugeben ,** wie in Abbildung **1b zu sehen, der dann die Trennung der Aufnahmen durch Entrauschen in den anderen Bereichen ermöglicht. Der allgemeine Denoising-Prozess unter Verwendung von spärlichem/robustem Radon wendet zunächst den adjungierten Mischoperator auf die ge**

$$\tilde{\mathbf{D}} = \mathbf{b}\ddot{y}^{H}, \tag{10}$$

Anschließend werden die pseudo-entmischten Daten in der Inversion mit der folgenden Zielfunktion verwendet, um mithilfe der Radon-Transformation Rauschen zu beseitigen:

$$\mathbf{|Wr}(\tilde{\mathbf{D}} - \ddot{y}\,\mathrm{Rm})\|_p^p + \boldsymbol{\mu}\mathbf{|Wmm}\|_q^q, \tag{11}$$

wobei D˜ die pseudo-entmischten Daten, R der Radon-Transformationsoperator, m das Modell, μ der Kompromissparameter und p **und** q **die Daten- bzw. Modellnormen sind. Eine grafische Darstellung der Radon-Rauschunterdrückung ist in Abbildung 2 zu sehen. Dabei wird Standard**

Machine Translated by Google

hyperbolisches Radon, um spontane Ereignisse im Schussdatensatz herunterzugewichten und so das kohärente Signal wiederherzustellen.

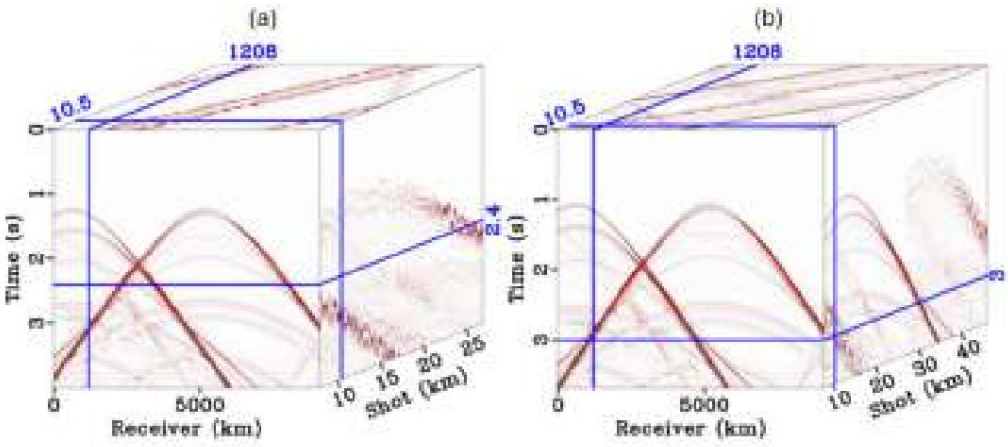


Abbildung 1. Anwendung des Adjungiertenoperators: (a,b) ist das Ergebnis; in anderen geometrischen Bereichen kann man sehen, dass unser Ziel kohärent ist, während Störsignale inkohärent bleiben.

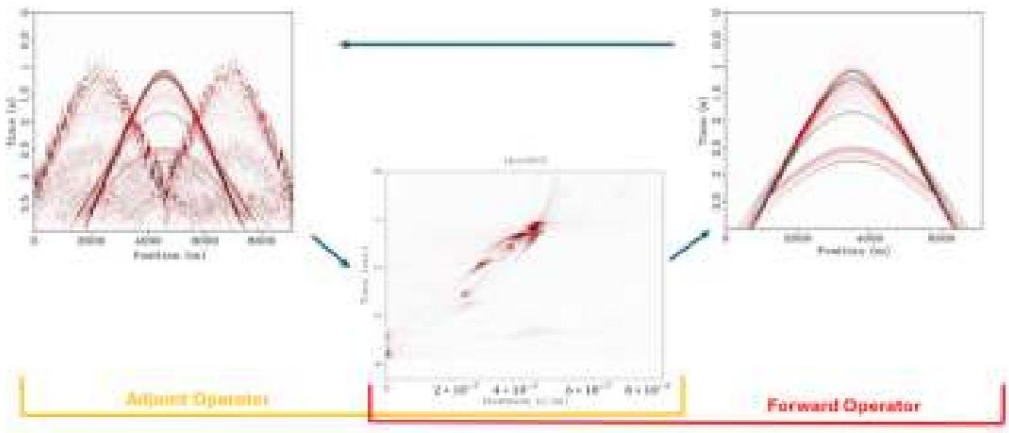


Abbildung 2. Denoising-Ansatz zum Deblending: Der Denoising-Ansatz zum Deblending verwendet spärliches/robustes Radon, um das Mischrauschen aus jedem pseudo-deblendeten Empfängerrahmen zu entfe

Da das Mischrauschen aus einer Aufnahme ein Signal aus einer anderen ist, hat unser sporadisches Mischrauschen die gleichen Amplituden wie die Aufnahme selbst. Dieses Problem bei hoher Mischfaltung manifestiert sich darin, dass das Rauschen mit hoher Amplitude Abschnitte mit Ereignissen mit niedriger Amplitude im Signal vollständig überwältigt, was herkömmliche filterbasierte Wiederherstellungsmethoden nahezu unmöglich macht . Die Verwendung der Inversion ermöglicht es uns jedoch, die **störenden Ereignisse mit hoher Amplitude zu erklären** und effektiv nur die störende Amplitude neu zuzuordnen, wodurch **wir die Signale mit niedrigerer Amplitude wiederherstellen können. Das Schema für den inversionsbasierten Ansatz ist in Abbildung 3 zu sehen. Dieser Ansatz hat jedoch seine eigenen Nachteile, hauptsächlich in Bezug auf den Speicherbedarf, da der gesamte Mischdatensatz plus das entsprechende Modell im Speicher gehalten werden müssen. Entgegen anfänglicher Erwartungen weist diese Methode keinen signifikanten Laufzeitunterschied im Vergleich zu einem ähnlichen inversionsbasierten Rauschminderungsansatz auf, da in beiden Fällen eine Inversion durchgeführt werden muss. Im Vergleich zu einem Radon-Rauschunterdrückungsansatz verwendet der Inversionsansatz eine andere Zie Wir können die Zielfunktion unter Verwendung der bereits besprochenen Gewichte sowie Gleichung (6) formulieren:**

$$J = \mathbf{||Wr(b} \ddot{y} \ddot{y}Lm)||p_P + \mathbf{\mu||Wmm||q}_q \quad \textbf{(12)}$$

Machine Translated by Google

Hier sind die Eingabedaten nicht die pseudo-entmischten Daten, sondern die gemischten Daten, und der Operator enthält sowohl die Misch- als auch die Radon-Operatoren.

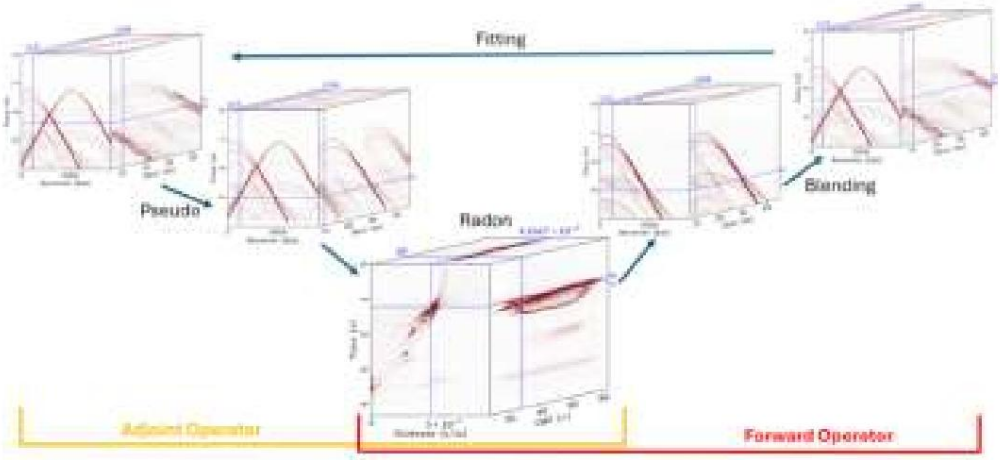


Abbildung 3. Schematische Darstellung der inversionsbasierten Entmischung: Kombination des Mischoperators mit dem Radon Transformation ermöglicht es uns, sporadische Mischstörungen wieder in das kohärente Signal in seiner Richtiger Ort, wodurch Störungen effektiv erklärt, anstatt sie zu beseitigen.

Durch die Implementierung von Deblending als inversionsbasierte Operation unter Verwendung der Blending- und Radon-Operatoren zusammen, wirkt das System wie ein übervollständiges Wörterbuch, in dem die Informationen werden in den pseudo-deblendeten Frames dargestellt und redundant dargestellt. Dies hilft bei Konvergenz, da die redundanten Informationen sich gegenseitig durch die Neuzuordnung von erklären Daten zurück zu ihren jeweiligen Aufnahmebildern.

3. Überlegungen zum High-Performance Computing (HPC)

Die spärliche hyperbolische Radon-Transformation wird in unserem Deblending-Algorithmus verwendet, um den schlecht gestellten Blending-Operator in ein Inversionsproblem umzuformatieren, das eine spärliche Regularisierung ver Dieser Radon-Operator ermöglicht uns, die Leistung eines hochgradig parallelisierbaren Operation mit einer Reihe paralleler Verarbeitungsmethoden. Um den HRT richtig zu codieren, Um effizient in openMP zu laufen, senden wir äußere Offsetschleifen an separate Threads, um die Transformation im Adjungierten-Operator; für den Vorwärts-Operator senden wir dann stattdessen jede äußere Langsamkeitsschleife. Der Pseudocode für die Radon-Transformation kann wie folgt geschrieben werden:

Algorithmus 1.

Algorithmus 1 Radon-Pseudocode

```
1: Funktion RADON
2: #PRAGMA OMP PARALLEL FOR für q =                    ÿ Insert zur Parallelisierung der Schleife
3:     Langsamkeit do
4:         für h = Offset do
5:             Auszug = h 2 ÿ q
6:             für es = 0 zu nt tun
7:                 Zeit = (it ÿ dt)  2 + Auszug
8:                 Modell[iqNt + it]+ = Daten[ihNt + INT(Zeit/dt)]
9:             Ende für
10:        Ende für
11: Ende für
12: Funktion beenden
```

Für unsere Implementierung in CUDA haben wir eine zweidimensionale Thread-Konfiguration eingerichtet mit jeweils einer Dimension für die Zeit- und Langsamkeitsdimensionen im Standardcode, dann berechnet die Transformation unabhängig für jede Ausgabezelle für jeden Thread. Zur Implementierung Mit dem HRT in openMPI senden wir die Daten zu jedem Schuss in der Transformation an separate

Machine Translated by Google

Knoten und berechnen dann die Transformation unabhängig auf jedem Knoten mithilfe der oben genannten CUDA- oder OpenMP-Methoden. Am Ende der Berechnung sammeln wir dann die Daten beim Master. Der CUDA-Pseudocode ist in Algorithmus 2 zu sehen.

Algorithmus 2 CUDA Radon Pseudocode 1:

```
__global__ 2:
Funktion CUDA_ RADON int it =
3:      blockIdx.x * blockDim.x + threadIdx.x int iq = blockIdx.y *
4:      blockDim.y + threadIdx.y wenn (iq >= nq || it >= nt) return
5:      int iqNt = iq ÿ nt double timemax = dt ÿ
6:      nt für h=Offset do int
7:      ihNt = ih ÿ nt double moveout = h
8:      2 ÿ q double time =
9:          (it ÿ dt) model[iqNt + it]
10:         + = data[ihNt + (INT)(time/dt)] 13:
11:         Ende für 14: Ende der Funktion 2 + Auszug
12:
```

Da eine beträchtliche Anzahl paralleler Anwendungen mehrdimensionale Daten verwendet, gruppiert die CUDA-API Threadblöcke in mehrere Dimensionen, bis zu drei. Dies ist für die Implementierung von Radon von Vorteil, da es im Allgemeinen drei Datendimensionen für die Transformation gibt, jeweils eine für Offset, Zeit und Geschwindigkeit. Indem wir jedoch jede unserer Variablen einer Dimension zuweisen, können wir ihre jeweiligen Schleifen eliminieren, aber dies ist nicht die effizienteste Art, den Algorithmus zu implementieren. Da das Modell eine Akkumulation über die Offsetdimension hinweg ist, würden wir auf einen Datenkonflikt stoßen, wenn die Offsetdimension h **parallel ausgeführt wird. Wenn der Offset in CUDA als dritte Threaddimension festgelegt würde, wäre ein atomarer Operator erforderlich; dieser wurde zunächst getestet und führte zu deutlich langsameren Ergebnissen als der oben angegebene Algorithmus. Indem wir nur die beiden Rasterdimensionen parallelisieren, vermeiden wir die Notwendigkeit, atomare Akkumulationsoperatoren zu verwenden, die zu erheblichen Verlangsamungen führen können, wobei jeder Thread in einer so genannten ausgabeorientierten Optimierung Arbeit für jede Rasterzelle leistet.**

Dieser Algorithmus ist insofern interessant, als dass er größtenteils parallel optimiert werden kann, aber eine ordnungsgemäße Parallelisierung birgt viele Herausforderungen. Die größte Herausforderung bei der Optimierung und Parallelisierung des Algorithmus liegt in den großen Speicherauslagerungen, die für jede Iteratio Diese Speicherprobleme entstehen durch die ständige Notwendigkeit, Domänen im Blending-Operator zu ändern, was zu einem groß angelegten, nicht sequentiellen Zugriff auf die Daten führt. Die andere Herausforderung, der wir uns stellen müssen, besteht darin, dass dieser Algorithmus mit openMPI nicht richtig über verteilte Speichersysteme (Server) parallelisiert werden kann, da die Gradientenberechnung bei jeder Iteration mit dem Blended-Operator berechnet werden muss, was eine Neuordnung aller Daten erfordert. Da verteilter Speicher in Clustern eine Aufteilung des Datensatzes auf Knoten erfordert, anstatt ihn vollständig im RAM des Masterknotens zu behalten, stellt der Neuordnungsprozess eine sehr schwierige Herausforderung für eine openMPI-Implementierung dar. Aufgrund dieser Anforderung müsste jeder Knoten in einem Cluster nach jeder Berechnung Daten zurücksenden, um den nächsten Schritt zu berechnen, und dann muss der Master die Ergebnisse erneut an jeden Knoten zurücksenden. Für unsere Implementierung in CUDA verwenden wir anstelle mehrerer Knoten eine einzelne Hochleistungsmaschine, auf der sowohl der Ra

4. Ergebnisse

Da es sich bei unseren Testdaten hauptsächlich um synthetische und Meeresdaten handelt, gehen wir davon aus, dass der Datenraum frei von Ausreißern ist. Daher haben wir ÿ2 im Datenraum für eine Inversion der gemischten Norm ÿ2-ÿ1 verwen **Für unsere Tests verwendeten wir drei Datensätze, zwei davon synthetisch durch finite Differenzen und einen realen Meeresdatensatz aus dem Golf von Mexiko, der numerisch gemischt wurde. Für die synthetischen Datensätze wurde eine finite Differenzenmodellierung mit 300 Schüssen und 300 Empfängern durchgeführt, die gleichmäßig auf der Oberfläche verteilt waren. Der erste Datensatz und der einfachste**

Machine Translated by Google

ist der Keilmodell-Datensatz aus Abbildung 4. Das Keilmodell besteht aus zwei Schichten, einem dreieckigen Keil und einer Reihe von harten Punkten, um die Rekonstruktion von Beugungen mit geringer Amplitude zu testen . Die Keildaten wurden mit zwei überlappenden Aufnahmen mit **zufälligen Verzögerungszeiten zwischen 0 und 400 Proben gemischt .**

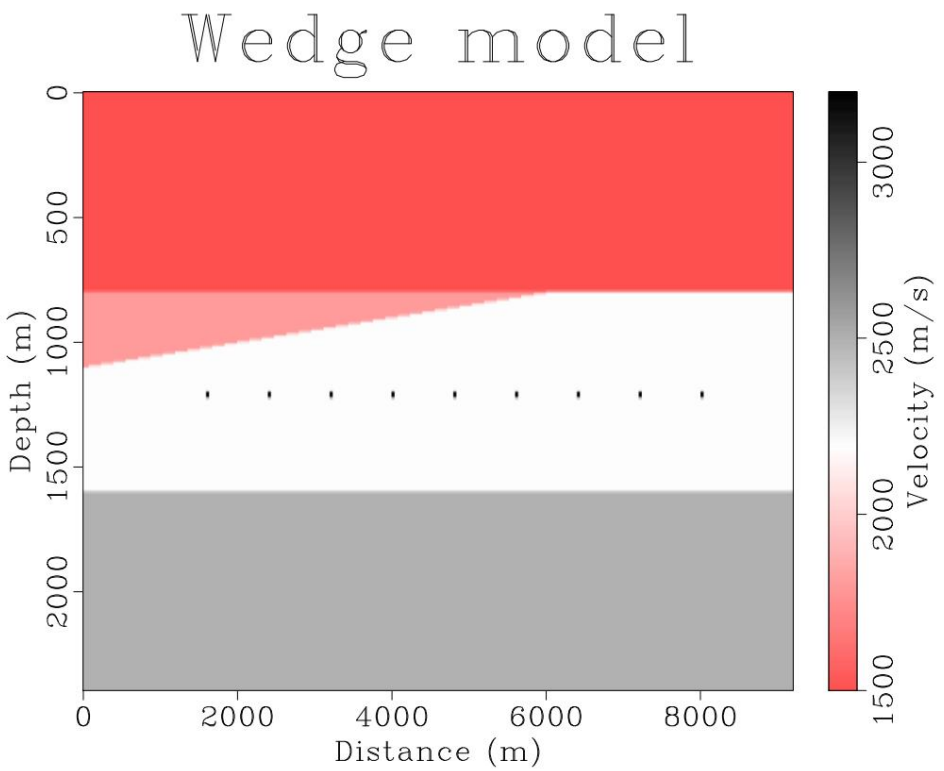


Abbildung 4. Geschwindigkeitsmodell 1: Keilmodell bestehend aus zwei Beugungspunkten der Schicht 9 und einem Keil.

Die Entmischung ist in Abbildung 5 zu sehen , wo das in den pseudo- **entmischten Daten in Abbildung** 5b **sichtbare Mischrauschen durch Inversion in Abbildung 5c entfernt wird . Anschließend wurden die Ergebnisse in den Shot-Bereich transformiert, wie in Abbildung 6 dargestellt.**

Trotz einiger Reststörungen im gemischten Schuss entfernte der Entmischungsprozess den sekundären Störschuss effektiv und behielt den Zielschuss, einschließlich der Beugungen, gut bei . Diese sind in Abbildung **6b und c oben z**

Einen Vergleich unserer Inversionsmethode und einer Rauschunterdrückungsmethode sehen Sie in Abbildungÿ7.

Die Rauschunterdrückungsmethode verwendet auch den spärlichen HRT als Rauschunterdrückungs-Engine, um das zufällige Mischrauschen in den pseudo-entmischten Sammlungen zu entfernen. Der Hauptunterschied zur Inversionsmethode besteht darin, dass der Rauschunterdrückungsansatz den Mischoperator nicht einbezieht. Im CMP-Bereich erschweren die hohen Amplituden des Mischrauschens die direkte Entfernung durch allgemeine Rauschunterdrückungsmethoden. Dies liegt daran, dass die Unterscheidung von Rauschen und Daten zunehmend schwieriger wird, da das Mischrauschen dieselbe Amplitude hat wie das Signal, das wir wiederherstellen möchten. Ohne die zusätzlichen Informationen, die der Mischoperator liefert, ist es aufgrund dieser Ähnlichkeit schwierig zu erkennen, welche Teile der Daten Rauschen und welche kohärente Signale sind. Wie in Abbildung 7 zu sehen ist , hat das Rauschunterdrückungsergebnis die Daten erfolgreich entmischt, jedoch auf Kosten der Daten mit niedrigerer Amplitude der Beugungsankünfte, die zusammen mit

Machine Translated by Google

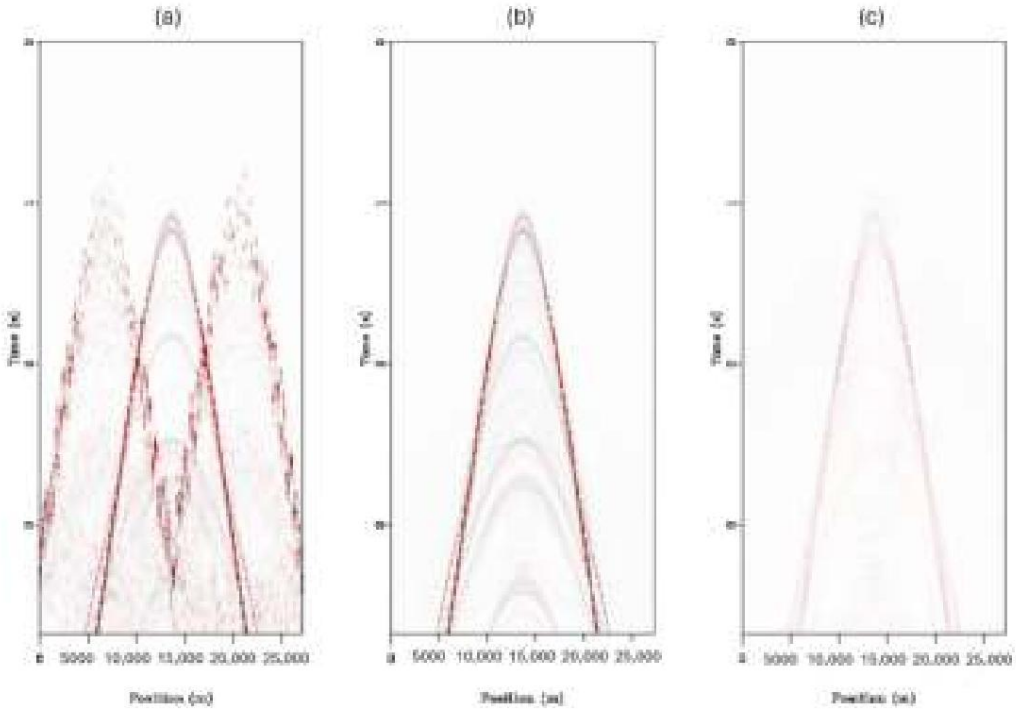


Abbildung 5. Ergebnisse des Keilmodell-Tests (CMP): Deblending-Test 1 im CMP-Bereich mit (a) den pseudo- deblendeten Daten, (b) den deblendeten Daten und (c) dem angezeigten Rekonstruktionsfehler.

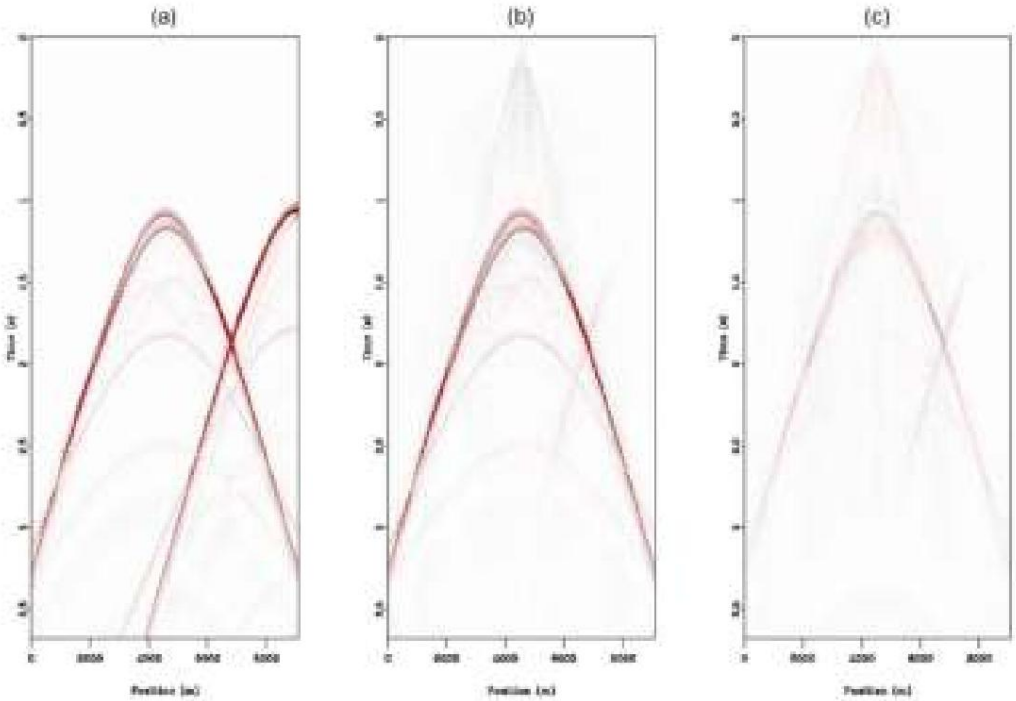


Abbildung 6. Ergebnisse des Keilmodelltests (Shot Domain): Deblending-Test 1 in Shot Domain mit (a) dem (b) die pseudo-entmischten Daten, (c) die entmischten Daten und (c) der gezeigte Rekonstruktionsfehler.

Machine Translated by Google

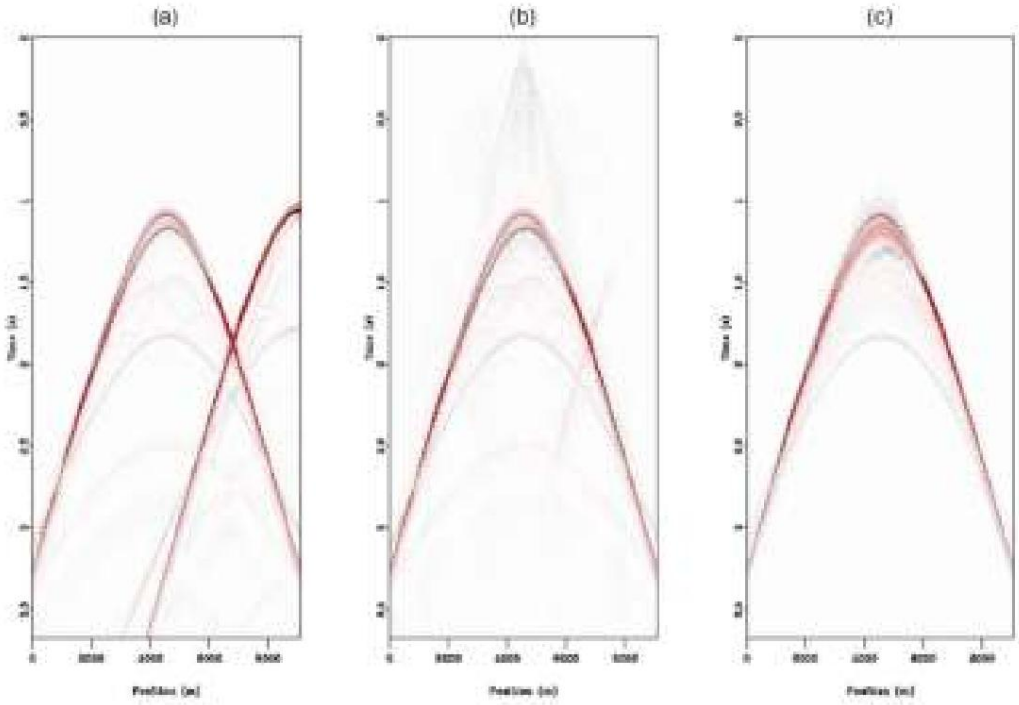


Abbildung 7. Vergleich der Ergebnisse des Keilmodells zwischen Inversions- und Rauschunterdrückungsansatz: Inversions- und Rauschunterdrückungsansätze im Shot-Bereich, mit (a) gemischt, (b) Inversion entmischt und (c) Rauschunterdrückung entmischt.

Der synthetische Test Nr. 2 wurde durch finite Differenzen unter Verwendung des Standard-Mar-Mousi-Modells [21] erzeugt. Da Test 1 zeigte, dass unsere Deblending-Engine unsere Daten effektiv entblendete, beschlossen wir, Test Nr. 2 eher zu einem Stresstest zu machen, bei dem eine nicht auf Inversion basierende Deblending-Methode Schwierigkeiten hätte, effektiv zu entblenden. Wir erhöhten den Schwierigkeitsgrad deutlich, indem wir den Bereich unserer zufälligen Verzögerungszeiten verringerten und auch die Anzahl der sich überlappenden Aufnahmen von zwei auf fünf erhöhten, Bedingungen, die wir derzeit in der realen Welt nicht erwarten. Abbildung 8a zeigt die gemischten Daten im CMP- Bereich; es lässt sich leicht vorstellen, wie schwierig es für einen normalen Denoiser ohne den Mischoperator wäre, die Daten von den Interferenzen in diesem Bild zu unterscheiden.

Bei der Analyse der Ergebnisse des zweiten Tests können wir in Abbildung 8 sehen , dass unsere Wiederherstellung von Reflexionen mit niedriger Amplitude effektiv ist, selbst wenn sie vollständig von Interferenzen mit hoher Amplitude aus anderen Aufnahmen überlagert werden. Wenn wir uns den Rekonstruktionsfehler für den Aufnahmebereich in Abbildung 9 ansehen, können wir sehen, dass die Hauptunterschiede reduzierte Amplituden der Ereignisse mit höherer Amplitude und die Einführung von Hintergrundrauschen mit sehr niedriger Amplitude sind. Wir glauben, dass der Amplitudenfehler möglicherweise dadurch entsteht, dass im Modellraum zu viel Spärlichkeit erzwungen wird, was diese Probleme im Allgemeinen bei hochauflösendem Radon verursacht, oder dadurch, dass die Mischaufnahmen nicht vollständig getrennt werden und die höheren Amplituden als Hintergrundrauschen über die Daten verschmiert

Unser letzter Test war ein realer Datensatz aus dem Golf von Mexiko, bekannt als BP Tiber WATS-Datensatz [8,22], der dann numerisch gemischt wurde. Das Mischschema war im Vergleich zu den anderen anders konfiguriert, da dies auch ein Test eines anderen Erfassungsplans war, der als kontinuierliches Abhören bekannt ist. Beim kontinuierlichen Abhören haben wir effektiv einen einzigen Superschuss, da die Daten aufgezeichnet werden, bis alle Schüsse abgefeuert sind und sich die Schüsse überlappen dürfen. Der reduzierte Datensatz bestand aus 90 Schüssen und 90 Empfängern, mit einer Schuss-zu-Schuss-Überlappung von 70 %, die zwischen 0 und 200 Proben randomisiert wurde. Die Ergebnisse der Entmischung des Datensatzes aus dem Golf von Mexiko sind in der CMP-Domäne in Abbildung 10 und der Schussdomäne in Abbildung 11 zu sehen. Bei der Analyse der Ergebnisse können wir sehen, dass der Entmischungsalgorithmus bei der Schusstrennung im realen Datensatz außergewöhnlich effektiv ist, mit einer guten Wiederherstellung aller in den Daten sichtbaren Ereignisse. Der Hauptfehler, den wir beobachten können

Machine Translated by Google

Domäne sind die Aufnahmen nicht perfekt entblendet, und eine gewisse Amplitude der nachfolgenden Aufnahme ist in der wiederhergestellten Aufnahme noch zu sehen. Dies ist bei etwa 5,2 s in Abbildung **11c zu beobachten .**

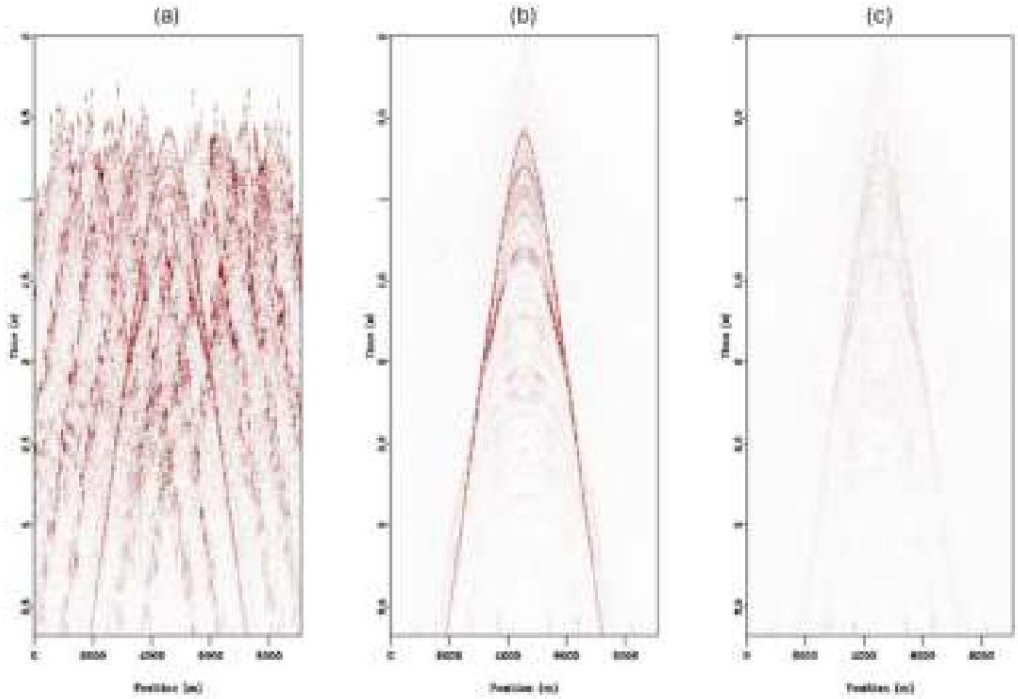


Abbildung 8. Testergebnisse des Marmousi-Modells (CMP): Deblending-Test 2 mit 5 Aufnahmen, mit (a) gemischt, (b) entblendet bzw. (c) Rekonstruktionsfehler.

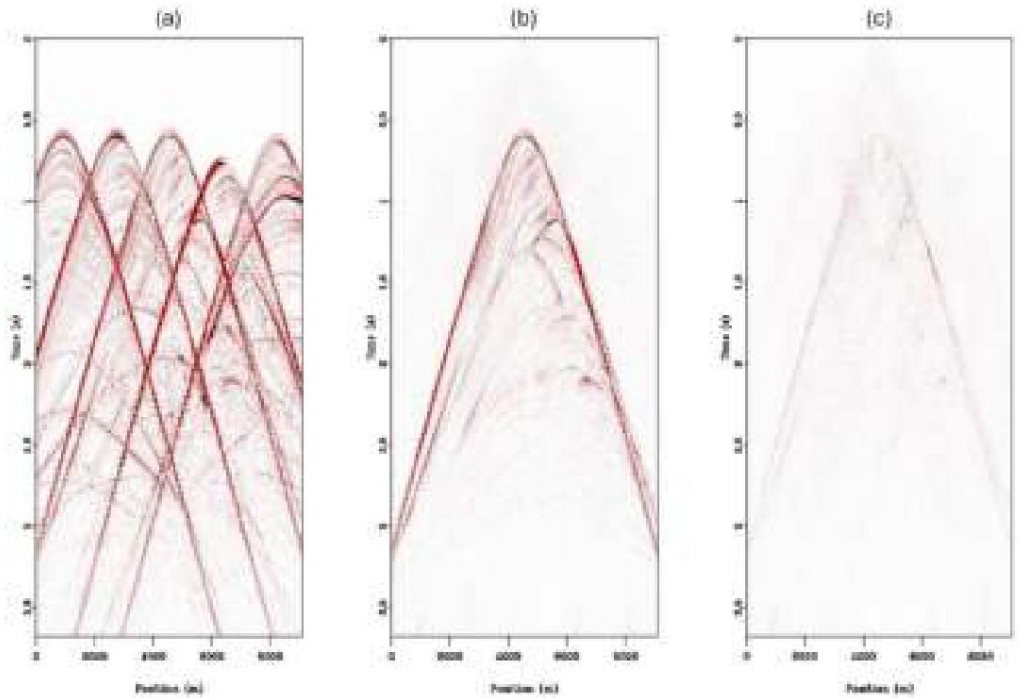


Abbildung 9. Testergebnisse des Marmousi-Modells (Aufnahmebereich): Deblending-Test 2 mit 5 Aufnahmen, mit (a) gemischt, (b) deblendet bzw. (c) Rekonstruktionsfehler.

Machine Translated by Google

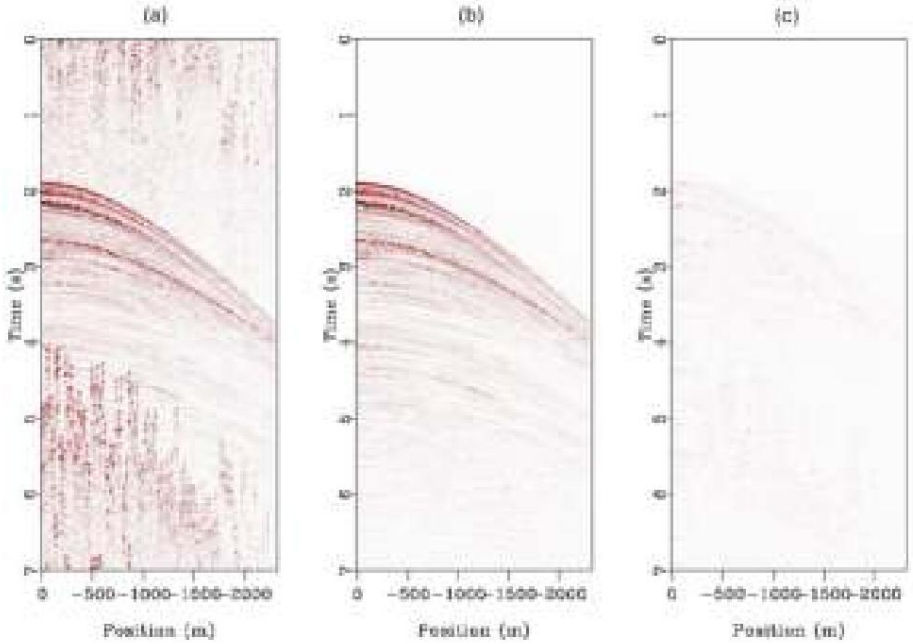


Abbildung 10. Daten zum Golf von Mexiko im CMP-Bereich: Entmischung unter Verwendung eines Meeresdatensatzes zum Golf von Mexiko mit Ergebnissen im CMP-Bereich, mit gemischten Daten in (a), entmischten Daten in (b) und dem Rekonstruktionsfehler in (c).

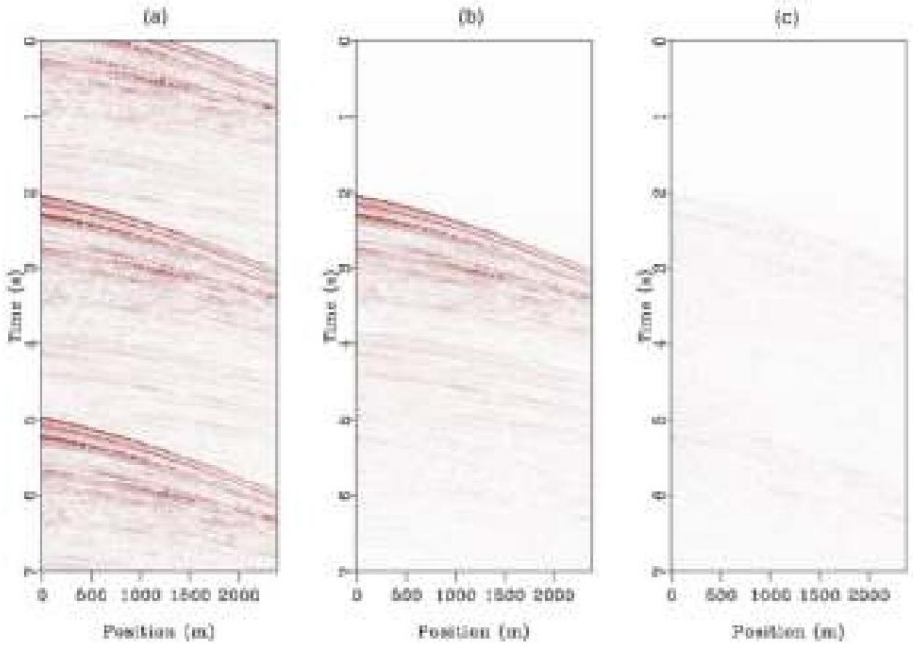


Abbildung 11. Daten aus dem Golf von Mexiko: Entmischung unter Verwendung eines Meeresdatensatzes aus dem Golf von Mexiko mit einer Aufnahmezeitinterferenz von 70ÿ%, mit gemischten Daten in (a), entmischten Daten in (b) und einem Rekonstruktionsfehler in

Die Einschränkungen dieses Ansatzes liegen in den Annahmen, die der Verwendung der hyperbolischen Radon-Transformation innewohnen. Die Hauptannahme dieses Ansatzes ist die hyperbolische Flugbahn aller seismischen Eingänge, die voraussichtlich im Radon-Raum genau dargestellt werden. Diese Annahme wird jedoch problematisch, wenn man die einzigartigen Eigenschaften seismischer Landdaten berücksichtigt . Die den Landdaten innewohnenden statischen Eigenschaften können vertikale Verschiebungen in den Proben verursachen, die von ihrer hyperbolischen Ausrichtung abweichen. Solche Diskrepanzen führen zu einer ineffektiven Abbildung im Radon-

Machine Translated by Google

Wir haben sowohl den Deblending-Algorithmus als auch den Radon-Transformationsalgorithmus auf einer Reihe von unterschiedliche Hardware, um die Leistung der Algorithmen mit den verschiedenen Parallelisierungsmethoden unter verschiedenen Bedingungen zu demonstrieren: ein Laptop, zwei Workstations Servercluster. Der Laptop nutzt eine Intel i7 8750H 6-Core-CPU mit einer Nvidia RTX 2060 MAX-Q GPU, die erste Workstation ist mit einem AMD Threadripper 3960× 24- ausgestattet Core-CPU mit einer Nvidia RTX 3080Ti, und die zweite Workstation verfügt über einen Intel i7 9800× 8-Kern-CPU mit Nvidia RTX 2060-Super-GPU ausgestattet. Diese Ergebnisse können sich sehen lassen in Tabelle 1. Die Laufzeit zwischen der Radon-Transformation und dem Deblending-Algorithmus können nicht direkt verglichen werden, da sie unterschiedliche Aufgaben ausführen; der relative Unterschied in der Laufzeit zwischen unterschiedlicher Hardware zeigt die Auswirkung der Sortierung auf die verschiedene Implementierungen.

Tabelle 1. Tabelle der durchschnittlichen Laufzeiten für Radon und Kleinstquadrate-Deblending auf verschiedenen Maschinen.

Prozessor (API)	Radon-Umwandlungszeit (s)	LS-Entmischungszeit (s)
i7 6800k 6-Kern (openMP) i5	4462	2590
8600 6-Kern (openMP) i7	3773	1748
8750H 6-Kern (openMP) i7	4044	2203
9800× 8-Kern (openMP)	2715	1390
TR 3960× 24-Kern (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	—

Es ist ersichtlich, dass es einen erheblichen rechnerischen Vorteil gibt, wenn man von CPU-Verarbeitung (openMP) zu GPU-basierter Verarbeitung (CUDA). Die openMPI-Ergebnisse zeigen eine deutliche Beschleunigung im Vergleich zu den openMP-Ergebnissen und die Schließung der Lücke zu den CUDA Ergebnisse; diese Beschleunigung kann hauptsächlich auf die gröberkörnige Natur des openMPI zurückgeführt werden. Die openMPI-Implementierung teilt weniger Ressourcen zwischen Threads; daher insgesamt weniger Die Speicherverwaltung muss vom System durchgeführt werden, auf Kosten des gesamten RAM-Speicherplatzes je mehr Kopien der Daten erstellt werden. Bei Standard-Radon-Transformationstests wird die CUDA-basierte Code reduzierte die Rechenzeit um mehr als eine Größenordnung im Vergleich zum OpenMP Version; nach der Einführung des Algorithmus der kleinsten Quadrate verringerte sich der Vorteil um eine große Menge. Durch die Verteilung aller Arbeiten durch openMPI, einschließlich der LS-Optimierung über die Knoten hinweg kann die LS-Optimierung parallel ausgeführt werden, so dass weder die weder openMP noch CUDA-Code können dies, was zu einer insgesamt schnelleren Rechenzeit führt als openMP und Ansätze zur CUDA-Implementierung. OpenMPI-Deblending jedoch kann nicht richtig implementiert werden, da der Hostknoten alle Daten vom anderen sammeln muss Knoten, die die LS-Berechnung bei jeder Iteration durchführen; dies führt zu großen Speichertransfers über das Netzwerk, was zu erheblichen Latenz- und Übertragungsrateneinbußen führt. Der Overhead mit der openMPI-Implementierung verbunden ist, ist so groß, dass wir keinen Vorteil feststellen konnten in unserem Prototyping des Programms im Vergleich zu Single-Thread-Laufzeiten. Es Es muss jedoch beachtet werden, dass openMPI in dieser Übung die größte Menge an RAM verwendet hat. etwa 122 GB, im Vergleich zu den OpenMP- und CUDA-Implementierungen, die 8 GB verwendeten RAM bzw. 8 GB + 8 GB (gespiegelter) VRAM. Der wichtigste Kompromiss zwischen einem CUDA/GPU-basierter Ansatz und ein OpenMP-Ansatz wären der von Geschwindigkeit versus Portabilität; der CUDA-Ansatz ist deutlich schneller, erfordert aber eine Nvidia-GPU, während Der OpenMP-Ansatz funktioniert auf jedem Gerät, das OpenMP-Parallelisierung unterstützt. der Compiler, bei dem es sich um jede moderne CPU auf dem Markt handelt.

5. Schlussfolgerungen

Wir haben eine Methode zur Entmischung vorgeschlagen, die eine lineare Inversion eines zusammengesetzten Transformationsoperator im CMP-Bereich. Diese Inversion, wenn sie mit dem Blending-

Machine Translated by Google

Operator, ermöglicht es uns, störende Schussinformationen vorherzusagen und neu zuzuordnen, anstatt sie zu eliminieren. Dies wiederum erleichtert die genaue Wiederherstellung von Ereignissen mit geringer Amplitude, selbst bei stark vermischten Daten. Die Implementierung einer spärlichen Einschränkung des Inversionsalgorithmus ermöglicht es uns, eine einzigartige Lösung zu finden, die die Kohärenz des Signals bei hyperbolischen Ankünften verbessert. Darüber hinaus bietet uns die Verwendung hyperbolischer Radon-Transformationen im CMP-Bereich einen Operator, der effizient und gut auf unsere Daten abgestimmt ist, die Kohärenz bei unserer Wiederherstellung fördert und Aliasing-Probleme umgeht. Daher sind wir zuversichtlich, dass unser Inversionsansatz eine effektive Strategie ist, um gleichzeitig gewonnene Daten zu entmischen und gleichzeitig Ereignisse mit geringer Amplitude zu schützen. Wir haben auch zusätzliche Methoden untersucht, um dieses Framework durch Anwendungen sowohl auf der CPU als auch auf der GPU zu beschleunigen, jede mit ihre

Beiträge der Autoren: Konzeptualisierung, KZ und DT; Methodik, KZ und DT; Software, KZ; Validierung, KZ, DT und AI; formale Analyse, KZ; Untersuchung, KZ; Datenkuratierung, KZ und AI; Schreiben – Vorbereitung des Originalentwurfs, KZ; Schreiben – Überprüfung und Bearbeitung, AI und DT; Überwachung, DT. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu.

Finanzierung: Wir danken den Sponsoren des Consortium for Research in Elastic Wave Exploration Seismology (CREWES) für ihre anhaltende Unterstützung. Diese Arbeit wurde von den industriellen **Sponsoren von CREWES und dem NSERC (Natural Science and Engineering Research Council of Canada) durch das** Stipendium CRDPJ 543578-19 **finanziert .**

Datenverfügbarkeitserklärung: In dieser Studie wurden keine neuen Daten erstellt oder analysiert. Die Datenfreigabe ist für diesen Artikel nicht anwendbar.

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Abkürzungen

In diesem Manuskript werden die folgenden Abkürzungen verwendet:

COF Gemeinsamer Offset
CMP Gemeinsamer Mittelpunkt
CRG Gemeinsamer Empfänger sammeln
HRT Hyperbolische Radontransformation
ASHRT Apex-verschobene hyperbolische Radon-Transformation
CG Konjugierter Gradient
IRLS Iterative neugewichtete kleinste Quadrate
CUDA Compute Unified Device Architecture (Nvidia-Programmier-API)
API Anwendungsprogrammierschnittstelle

Verweise

1. **Garottu, R. Gleichzeitige Aufzeichnung mehrerer seismischer Vibroseis-Linien. In** SEG Technical Program Expanded Abstracts 1983; **Society of Exploration Geophysicists: Houston, TX, USA, 1983; Seiten 308–310. [CrossRef]**
2. **Beasley, CJ; Chambers, RE; Jiang, Z. Ein neuer Blick auf gleichzeitige Quellen. In** SEG Technical Program Expanded Abstracts 1998; **Society of Exploration Geophysicists: Houston, TX, USA, 1998; S. 133–135. [CrossRef]**
3. **Beasley, CJ Ein neuer Blick auf marine Simultanquellen.** Lead. Edge **2008,** 27, **914–917. [CrossRef]**
4. **Mahdad, A.; Doulgeris, P.; Blacquiere, G. Trennung gemischter Daten durch iterative Schätzung und Subtraktion der gemischten Interferenzrauschen.** Geophysics **2011,** 76, **Q9–Q17. [CrossRef]**
5. **Abma, R.; Howe, D.; Foster, M.; Ahmed, I.; Tanis, M.; Zhang, Q.; Arogunmati, A.; Alexander, G. Unabhängige gleichzeitige Quellenerfassung und -verarbeitung.** Geophysics **2015,** 80, **WD37–WD44. [CrossRef]**
6. **Berkhout, AJ Die Denkweise bei der Erfassung seismischer Daten ändern.** Lead. Edge **2008,** 27, **924–938. [CrossRef]** 7. **Lin, T.; Herrmann, F. Designing Simultaneous Acquisitions with Compressive Sensing. In Proceedings der 71. EAGE-Konferenz und Ausstellung mit SPE EUROPEC, Amsterdam, Niederlande, 8.–11. Juni 2009; S. cp-127-00269. [Querverweis]**
8. **Ibrahim, A. Trennung gleichzeitiger seismischer Quellen durch robuste Inversion von Radon- und Migrationsoperatoren. Dissertation, University of Alberta, Edmonton, AB, Kanada, 2015.**
9. **Liu, B.; Sacchi, M. Minimal gewichtete Norminterpolation seismischer Aufzeichnungen.** Geophysics **2004,** 69, **1560–1568. [CrossRef]**
10. **Chen, Y. Deblending mit einem raumvariablen Medianfilter.** Explor. Geophys. **2015,** 46, **332–341. [CrossRef]**

Machine Translated by Google

11. Mahmoudian, F.; Curtis, J.; De Meersman, K.; Kennedy, R.; Staples, G. Deblending of Simultaneous Source Data. Online verfügbar: https://geoconvention.com/wp-content/uploads/abstracts/2023/91390-deblending-of-simultaneous-source-.pdf (abgerufen am 25. Juni 2024).

12. Stanton, A.; Wilkinson, K. Robuste Entmischung seismischer Daten gleichzeitiger Quellen. arXiv **2018, arXiv:1812.06040.**

13. Berkhout, AJ; Blacquière, G.; Verschuur, DJ Das Konzept der Doppelblendung: Die Kombination inkohärenter Aufnahmen mit inkohärenter Sensorik. Geophysik **2009,** 74, **A59–A62. [CrossRef]**

14. Jiang, Z.; Abma, R. Eine Analyse der gleichzeitigen Bildgebung gleichzeitiger Quelldaten. In SEG Technical Program Expanded Abstracts 2010; **Society of Exploration Geophysicists: Houston, TX, USA, 2010; Seiten 3115–3119. [CrossRef]**

15. Thorson, JR; Claerbout, JF Stochastische Inversion von Geschwindigkeitsstapeln und Schrägstapeln. Geophysics **1985,** 50, **2727–2741. [CrossRef]**

16. Foster, DJ; Mosher, CC Unterdrückung von Mehrfachreflexionen mit der Radon-Transformation. Geophysics **1992,** 57, **386–395. [CrossRef]**

17. Trad, D.; Ulrych, T.; Sacchi, M. Neueste Ansichten der spärlichen Radon-Transformation. Geophysics **2003,** 68, **386–399. [CrossRef]**

18. Trad, D.; Hargreaves, N.; verWest, B.; Wombell, R. Mehrfachdämpfung durch Verwendung einer Apex-Shift-Radon-Transformation. In SEG Expanded Zusammenfassungen; **Society of Exploration Geophysicists: Houston, TX, USA, 2004. [CrossRef]**

19. Claerbout, JF Erdsondierungsanalyse: Verarbeitung versus Inversion; **Blackwell Scientific Publications: London, Großbritannien, 1992.**

20. Scales, JA; Gersztenkorn, A.; Treitel, S. Schnelle lp-Lösung großer, dünn besetzter linearer Systeme: Anwendung in der seismischen Laufzeittomographie . J. Comput. Phys. **1988,** 75, **314–333. [CrossRef]**

21. Brougois, A.; Bourget, M.; Lailly, P.; Poulet, M.; Ricarte, P.; Versteeg, R. Marmousi, Modell und Daten. In Proceedings of the EAEG Workshop – Praktische Aspekte der Inversion seismischer Daten, Kopenhagen, Dänemark, 28. Mai–1. Juni 1990.

22. Naranjo, JC; Ekstrand, EJ; Etgen, J.; Hargrove, KL; Brewton, JG; Garcia, O.; Astvatsaturov, G.; Hite, D.; Howieson, B.; Bai, B.; et al. Untersuchungsdesign und Implementierung einer Strategie zur seismischen Erfassung mit mehreren weitwinkligen Schleppstreamern bei der Tiber Discovery; tiefes Wasser im Golf von Mexiko, USA. In SEG Technical Program Expanded Abstracts 2011; **Society of Exploration Geophysicists: Houston, TX, USA, 2011; S. 132–136. [CrossRef]**

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Artikel

Algorithmus zur Erstellung von Präzedenztabellen für CFGs, unabhängig davon, ob es sich um OPGs handelt

Leonardo Lizcano, Eduardo Angulo und José Márquez *

Fakultät für Informatik und Ingenieurwesen, Universidad del Norte, Barranquilla 081007, Kolumbien; ldlizcano@uninorte.edu.co (LL); edangulo@uninorte.edu.co (EA)
*** Korrespondenz: jmarquez@uninorte.edu.co**

Zusammenfassung: Operatorpräzedenzgrammatiken (OPG) sind kontextfreie Grammatiken (CFG), die durch das Fehlen von zwei benachbarten Nichtterminalsymbolen im Hauptteil jeder Produktion (rechte Seite) gekennzeichnet sind. Operatorpräzedenzsprachen (OPL) sind deterministisch und kontextfrei. Für diese Sprachen werden drei mögliche Präzedenzbeziehungen zwischen Paaren von Terminalsymbolen festgelegt. Viele CFGs sind keine OPGs, da die Operatorpräzedenz nicht auf sie angewendet werden kann, da sie **die Grundregel nicht erfüllen .** Um dieses Problem zu lösen, haben wir eine gründliche Neudefinition der linken **und rechten Terminalsätze durchgeführt, die die Grundlage für die Berechnung der Präzedenzbeziehungen bilden, und einen neuen ganz linken Satz definiert. Die Algorithmen zu ihrer Berechnung werden ebenfalls ausführlich beschrieben. Der wichtigste Beitrag unserer Arbeit besteht darin, dass wir Präzedenzbeziehungen zwischen Terminals festlegen, indem wir die Grundregel überwinden, dass keine zwei aufeinanderfolgenden Nichtterminals vorhanden sein dürfen, indem wir einen Algorithmus verwenden, der den Aufbau der Operatorpräzedenztabelle für eine CFG ermöglicht, unabhängig davon, ob es sich um eine OPG handelt oder nicht. Das Dokument zeigt die Komplexität der vorgeschlagenen Algorithmen und mögliche Ausnahmen von den vorgeschlagenen Regeln. Wir präsentieren Beispiele, indem wir einen OPG und zwei Nicht-OPGs verwenden, um die Funktionsweise der vorgeschlagenen Algorithmen zu veranschaulichen. Mit diesen wird die Operatorpräzedenztabelle erstellt und die Bottom**

Schlüsselwörter: Operatorpräzedenzsprache; Operatorpräzedenzgrammatik; Präzedenztabelle; Präzedenzbeziehungen; Algorithmus

check for updates

Zitat: Lizcano, L.; Angulo, E.; Márquez, J. Algorithmus zur Konstruktion von Präzedenztabellen für CFGs, unabhängig davon, ob sie OPGs sind. Algorithms **2024,** 17, **345. https://doi.org/ 10.3390/a17080345**

Wissenschaftlicher Redakteur: Dimitris Fotakis

Empfangen: 24. Juni 2024
Überarbeitet: 29. Juli 2024
Akzeptiert: 30. Juli 2024
Veröffentlicht: 7. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/ 4.0/).

1. Einleitung

Dieser Aufsatz befasst sich mit der von **R. Floyd** [1] **entwickelten Familie von Operator Precedence Languages (OPL) , die für effizientes Bottom-Up-Parsing verwendet wird [2,3].** Eine wichtige Erkenntnis **des Autors ist die Inspiration durch die Struktur arithmetischer Ausdrücke, in denen multiplikativen Operatoren (ÿ und /) Vorrang vor additiven Operatoren** (+ **und ÿ) eingeräumt wird. Durch den Prozess des Bottom-Up-Parsings ersetzt die linke Seite (Nicht-Terminalsymbol) die identifizierte Produktion auf der rechten Seite (Body) in einer kontextfreien Grammatik (CFG) vollständig und lässt keinen Raum für Mehrdeutigkeiten. Die Operatorpräzedenz wurde auf Grammatiken erweitert, die aus Terminalen auf der rechten Seite bestehen, selbst wenn diese durch Nicht-Terminale getrennt sind. Eine CFG wird im Allgemeinen als Operator Precedence Grammar (OPG) betrachtet, wenn sie keine Produktionen mit angrenzenden Nicht-Terminalsymbolen auf il**

OPLs haben sehr dabei geholfen, kontextfreie Grammatiken abzuleiten. Ein Beispiel wird in [4–6] beschrieben, wo eine Methode zur Ableitung von Operatorpräzedenz-freien Grammatiken aus eingeklammerten Stichproben vorgestellt wird. Die Methode basiert auf der Konstruktion von Parsebäumen und der Verwendung von Funktionen, die die Grammatiken erzeugen. Präzedenzregeln werden verwendet, um die Klammern zu identifizieren, die die Endelemente einer Zeichenfolge in der Stichprobe umgeben und die Reihenfolge angeben, in der die Zeichenfolge ausgewertet werden soll.

Die Forschung über eingabegesteuerte formale Sprachen (IDL) [7–9] (später umbenannt in Visibly Pushdown Languages (VPL) [10]) hat ergeben [11] , dass OP-Abschlusseigenschaften ID-Abschlusseigenschaften implizieren und dass ID-Sprachen ein spezieller Typ von OP-Sprachen sind, der durch eingeschränkte OP-Relationen gekennzeichnet ist **[12].**

Machine Translated by Google

Seit den Fortschritten in der Parallelverarbeitung werden lokale Parsability-Eigenschaften abgeleitet von OPL wurden verwendet, um schnelle parallele Parser zu erzeugen [13].

Der Hauptbeitrag dieses Papiers ist sein neuartiger Ansatz für den Aufbau von Vorrangbeziehungen zwischen Terminalsymbolen in Nicht-OPGs. Dieser Ansatz ermöglicht die Generierung von Operatorvorrangtabellen und ein Bottom-up-Parsing eines vom Nicht-OPG abgeleiteten Strings mit Unterstützung der erhaltenen Tabelle. Die Neuartigkeit dieser Arbeit ist ein wichtiger Aspekt, der sie von früherer Forschung auf diesem Gebiet abhebt. • Die in

früheren Arbeiten [1,11,14] zum Aufbau von Vorrangbeziehungen zwischen den Terminals eines OPG vorgeschlagenen Sätze von linken und rechten Terminals wurden neu definiert. • Ein neuer Satz von Terminals namens „Leftmost“ wurde definiert, um das Lösen von Vorrangbeziehungen zu unterstützen, wenn sich auf der rechten Seite der Produktionen zwei oder mehr benachbarte Terminals befinden. Dieser Satz ist bedeutsam, da er die Vorrangbeziehungen in derart komplexen Szenarien systematisch behandelt und den Bottom-up-Parsing-Prozess mit Nicht-OPGs erweitert.

- Es wurde ein Algorithmus zum Erstellen der Vorrangtabelle mit Unterstützung der drei vorgeschlagenen Terminalsätze entwickelt.
- Schließlich ermöglicht die Anwendung der vorgeschlagenen Algorithmen, aus einem Nicht- OPG die Relationen und Operatorpräzedenztabellen zu erhalten, die bisher nur auf OPGs beschränkt waren. Darüber hinaus könnte ein Bottom-up-Operatorpräzedenzparser verwendet werden, um jede von einem Nicht-OPG generierte Zeichenfolge zu analysieren, eine Technik, die nur auf OPGs angewendet wer

Wie in dieser Arbeit gezeigt wird, ist die vorgeschlagene Lösung zum Erstellen der Operatorpräzedenztabelle nicht nur ein theoretisches Konzept, sondern ein praktisches Werkzeug, das ohne Einschränkungen auf OPGs angewendet werden kann. Am Ende der Arbeit wird die Funktionalität demonstriert, indem sie auf ein OPG und zwei Nicht-OPGs angewendet und das jeweilige Bottom-up-Parsing durchgeführt wird . Diese praktische Demonstration soll Vertrauen in die Wirksamkeit der vorgeschlagenen Lösung schaffen und ihre praktische Anwendbarkeit sicherstellen.

Das Dokument ist wie folgt aufgebaut: Abschnitt 2 enthält die notwendigen Definitionen zum Verständnis dieser Arbeit. Abschnitt 3 präsentiert die bisherige Arbeit im Bereich CFG. Abschnitt 4 stellt unseren Ansatz vor und erläutert das Problem, das wir lösen möchten, sowie die Neudefinitionen der Konzepte früherer Arbeiten, die zur Entwicklung unserer Algorithmen verwendet wurden. Abschnitt 5 präsentiert Beispiele für die Anwendung des vorgeschlagenen Algorithmus auf ein OPG und zwei Nicht-OPGs und demonstriert die Anwendung der Bottom-up-Analyse mit den Präzedenztabellen, um deren Verwendung zu testen. Abschnitt 6 beschreibt mehrere Ausnahmen, bei denen die in den vorgeschlagenen Algorithmen beschriebenen Regeln nicht gelten. Abschließend enthält Abschnitt 7 Schlussfolgerungen zu unserer A

2. Grundlegende Definitionen

2.1. Kontextfreie Grammatiken und Sprachen

Innerhalb der Hierarchie von Chomsky [15] spielen kontextfreie Grammatiken eine bedeutende Rolle bei der Anwendung von Programmier- und Kompilierungssprachen, indem sie die syntaktische Struktur von Programmiersprachen adressieren.

Eine kontextfreie Grammatik (CFG) G generiert eine Sprache L(G), die allgemein als kontextfreie Sprache (CFL) bezeichnet wird und aus den von der CFG generierten Zeichenfolgen besteht. Die CFG ist definiert als ein 4-Tupel G(S) =< T, N, S, P >,

wobei: • T die Menge der Terminalsymbole darstellt, die die Zeichenfolgen innerhalb der von der CFG generierten
CFL erstellen. • N die Menge der Nicht-Terminalsymbole oder syntaktischen Variablen darstellt, die die Zeichenfolgen bestimmen, die zur
CFL-Generierung führen. • S ist das anfängliche Nicht-Terminalsymbol, aus dem alle von der GFC definierte generiert.
• P ist die Menge der Produktionen oder Regeln, die Terminale und Nichtterminale gruppieren, um die Zeichenfolgen zu erzeugen. Eine Produktion hat die Form A ÿÿ ÿ, wobei A ein Nichtterminal darstellt und ÿ eine Kombination aus Terminalen und Nichtterminalen ist. Das ÿÿ wird „produzieren“ ausgesprochen . Daher wird die Produktion gelesen als „A produziert Alpha“.

Machine Translated by Google

Es wird die folgende Notation verwendet: Kleinbuchstaben vom Anfang des Alphabets stellen Terminalsymbole dar (a, **b, c, . . . ÿ** T); **Großbuchstaben vom Anfang des Alphabets stellen Nicht-Terminalsymbole dar** (A, **B, C, . . . ÿ** N); **Kleinbuchstaben am Ende des Alphabets stellen Terminalzeichenfolgen dar** (u, **v, w, x, y,** z **ÿ** T +); **griechische Kleinbuchstaben stellen im Allgemeinen Terminal- und/oder Nicht-Terminalzeichenfolgen dar** *(ÿ,* **ÿ, ÿ, . . . ÿ** (T **ÿ** N) **ÿ) und die** leere Zeichenfolge wird durch *ÿ* **dargestellt (dieser griechische Buchstabe stellt die Ausnahme von der obigen Konvention dar).**

2.2. Grammatiken mit Operatorpräzedenz (Grundregel)

Eine kontextgebundene Grammatik wird als Operatorgrammatik betrachtet, wenn keine der Produktionen in P **auf ihrer rechten Seite benachbarte Nichtterminalsymbole hat .** Mit anderen Worten: Für eine Produktion A **ÿÿ** *ÿ* **in G nimmt** *ÿ* nicht die Form uBCv an.

2.3. Ableitungen

Eine Zeichenkette *ÿ* **ÿ** (T **ÿ** N) **ÿ** kann in k **Schritten aus einem Nichtterminal** A **ÿ** N **abgeleitet werden , wenn es eine Folge von Zeichenketten der Form ÿ1, ÿ2, . . . ,** ÿk gibt . **Mit dem Symbol =ÿ,** das eine einstufige Ableitung **bedeutet ,** wird die Folge von Zeichenketten in der Form **ÿ1 =ÿ ÿ2 =ÿ . . . =ÿ** ÿk **angeordnet .**

In dieser Folge entspricht ÿ1 A **und** ÿk **ÿ . Jedes Zwischenprodukt** ÿi **in der Folge hat die Form** *ÿBÿ,* **wobei** B **ÿ** N und es eine Produktion B **ÿÿ ÿ gibt, so dass durch Einsetzen auf der rechten Seite ÿi+1(=** *ÿÿÿ)* **erhalten wird. Die** Herleitung in einem oder mehreren Schritten wird **mit dem Operator** +=ÿ **gekennzeichnet . Allgemein wird gesagt, dass jede Folge von Terminalsymbolen** w **ÿ** L(G) als S +=ÿ w **oder durch Erweiterung der Herleitung in einem oder mehreren Schritten erstellt werden kann :** S = **ÿ1 =ÿ ÿ2 =ÿ . . . =ÿ** ÿk = **w. Für eine Grammatik G gilt die Satzform** *ÿ* **ÿ** (T **ÿ** N) **ÿ =ÿ ÿ, was bedeutet, dass** *ÿ* **mit null oder mehr Ableitungen aus S erhalten werden kann . Zeichenfolge, die** S **Wenn zwei Grammatiken ist ein identische Sprachen erzeugen, werden sie als** äquivalent betrachtet.

2.4. Ableitungsarten **Nach**

[16] ist S = **ÿ1 =ÿ ÿ2 =ÿ . . . =ÿ** ÿk = w **eine ganz links stehende Ableitung der Zeichenkette** w **ÿ** T , **wobei jedes** ÿi , **2 ÿ** i **ÿ** k **ÿ 1, die Form** xiAiÿi **mit** xi **ÿ** T Ai **ÿ** N **und** ÿi **ÿ** (T **ÿ** N) **ÿ** hat . **Weiterhin** gehört Ai **ÿÿ** ÿi **, zu P, und dann wird in jedem** ÿi **das Symbol** Ai durch ÿi **ersetzt , was die Satzform** ÿi+1 ergibt . **Jedes** ÿi wird als **links-satzförmige Form bezeichnet .**

Wenn jedes ÿi , **2 ÿ** i **ÿ** k **ÿ 1, die Form von** ÿiAixi **mit** xi **ÿ** T ÿi **ÿ** (T **ÿ** N) **ÿ , Ai ÿ N, und rechtssatzförmige , dann würde eine Ableitung ganz rechts gebildet. Jedes** ÿi **heißt Form annimmt.**

2.5. Bäume analysieren

Gemäß Referenz [16] ist ein Ableitungsbaum für eine CFG G(S) =< **T, N, S,** P > ein beschrifteter und geordneter Baum, in dem jeder Knoten eine Beschriftung erhält, die einem Symbol **aus der Menge** N **ÿ** T **ÿ {ÿ} entspricht. Wenn ein Nicht-Blattknoten mit** A beschriftet ist **und seine unmittelbaren Nachkommen** mit **X1, X2, . . . ,** Xn **beschriftet** sind, **dann** ist A **ÿÿ** X1X2 . . . Xn **eine Produktion in P. Ein beschrifteter geordneter Baum** D **ist ein Ableitungsbaum für eine CFG** G(A) =< **T, N, A,** P > , **wenn**

1. Die Grundnote von D **wird mit A** bezeichnet.

2. **Wenn D1, ... ,** Dk **die Teilbäume der direkten Nachkommen der Wurzel sind und die Wurzel von** Di mit Xi **bezeichnet ist , dann** liegt A **ÿÿ X1 ...** Xk **in P.**

3. Di **ist ein Ableitungsbaum für** G(Xi) =< **T, N,** Xi , P > **wenn** Xi **in** N liegt **oder** Di **ein einzelner Knoten ist mit der Bezeichnung** Xi , **wenn** Xi in T **liegt , oder** Di **ein einzelner Knoten mit der Bezeichnung ÿ ist.**

Analysebäume können aus jedem Ableitungstyp erstellt werden.

Machine Translated by Google

2.6. Bottom-Up-Analyse

Die Bottom-Up-Parsing-Syntaxanalyse [2], auch als Shift-Reduce-Parsing bekannt, versucht, einen Parsing-Baum für eine Eingabezeichenkette aufzubauen, der in den Blättern (Baumboden) beginnt und bewegt sich zur Wurzel (Baumspitze). Dieser Vorgang wird als Reduktion der Zeichenfolge w **auf das Anfangssymbol** S **einer CFG. Bei jedem Schritt der Parsing-Reduktion wird ein bestimmter Teilstring der Satzform, die der rechten Seite einer Produktion entspricht, wird durch das Nichtterminal ersetzt Symbol der linken Seite dieser Produktion. Wenn bei jedem Schritt die Teilzeichenfolge richtig gewählt wird, Eine Ableitung ganz rechts wird invers nachgezeichnet.**

2.7. Griff

Ein Griff [2,17] einer rechtssatzförmigen Form *ÿ* **ist eine Produktion** A **ÿÿ** *ÿ* **und eine Position von** *ÿ* **, wo die Zeichenfolge** *ÿ* gefunden und durch A **ersetzt werden könnte, um die vorherige rechtssatzförmige Form in einer ganz rechts stehenden Ableitung von ÿ zu erzeugen. Im Allgemeinen, wenn** S **=ÿ** ÿAw **=ÿ** ÿÿw, **dann** A **ÿÿ** *ÿ* **, wenn die Position neben** *ÿ* **ein Handle von** ÿÿw ist. **Wenn die Grammatik mehrdeutig ist, ein Handle wird erhalten. Wenn die Grammatik eindeutig ist, dann ist jeder rechtssatzförmige Form hat genau einen Handle. Normalerweise wird der String** *ÿ* **als Handle von** *ÿÿw* bezeichnet , **wenn der Die Voraussetzungen hierfür sind klar.**

2.8. Implementierung eines Bottom-Up-Parsings

Das Bottom-Up-Parsing durch Verschiebung und Reduktion verwendet einen Stapel zum Speichern von Grammatiksymbolen und ein Puffer zur Verarbeitung der zu analysierenden Eingabezeichenfolge. Das Symbol $ wird auch zur Abgrenzung verwendet unten im Stapel und rechts im Eingabepuffer. Das Erkennungsmodellformat beginnt mit dem Stapel bei $ und der Zeichenfolge w$ in der Eingabe, wie in Tabelle 1 gezeigt.

Tabelle 1. Bottom-Up-Analysetabelle.

Stapel	Eingang	Aktion
$	**w$**	
…	…	**Verschieben/Verkleinern**
…	…	**Verschieben/Verkleinern**
$S	**$**	**Akzeptieren**

Bei der Bottom-Up-Analyse werden null oder mehr Eingabesymbole auf den Stapel verschoben, bis ein Handle oben. Das Parsing reduziert dann den Handle auf die linke Seite der entsprechenden Produktion um die Satzform zu erhalten, die dem vorherigen Schritt der ganz rechts stehenden Ableitung entspricht. Beim Reduzieren erkennt das Parsing, dass sich das rechte Ende des Griffs oben im Stapel. Daher wird das Parsen aktiv, um das linke Ende im Stapel zu finden und das Nichtterminal, das es entsprechend der rechten Seite einer Produktion ersetzen wird, wo es mit dem Handle übereinstimmt. Diese Schritte werden wiederholt, bis ein Fehler erkannt wird oder bis der Stapel enthält $S **und die Eingabe hat das Symbol $. An diesem Punkt endet die Analyse und die Die Eingabezeichenfolge wird für die CFG als gültig betrachtet.**

3. Bisherige Arbeiten

Nach [1,11,14] sind **für eine kontextgebundene Gruppe G die linken Terminalmengen** LG(A) **und die rechten Terminalmengen Die Mengen** RG(A) **werden wie folgt definiert:**

$$LG(A) = \mathbf{a|A} \quad \overset{ÿ}{=}\mathbf{ÿ}\ ÿaÿ,\ ÿ\ \mathbf{ÿ}\ N\ \mathbf{ÿ}\ \{ÿ\}\ .$$

$$RG(A) = \mathbf{a|A} \quad \overset{ÿ}{=}\mathbf{ÿ}\ ÿaÿ,\ ÿ\ \mathbf{ÿ}\ N\ \mathbf{ÿ}\ \{ÿ\}\ .$$

Vorrangsbeziehungen entstehen durch die Herstellung binärer Beziehungen zwischen aufeinanderfolgenden Terminals oder solchen, die nach einer Bottom-Up-Analyse sequentiell werden, durch den Analysebaum. **Prozess hin zum nicht-terminalen Symbol S.**

Machine Translated by Google

Die Vorrangbeziehungen werden wie folgt hergestellt: a .= **b, genau**

1. **dann, wenn** ÿA **ÿÿ** *ÿaÿbÿ* **ÿ P,** *ÿ* **ÿ** N **ÿ {ÿ}.**
2. a ÿ **b, genau dann wenn** ÿA **ÿÿ** ÿaDÿ **ÿ P,** b **ÿ** LG(D). **3.** a ÿ **b, genau**
dann wenn ÿA **ÿÿ** ÿDbÿ **ÿ P,** a **ÿ** RG(D).

Gemäß [2] werden Vorrangsrelationen verwendet, um einen Handle in einer rechtssatzförmigen Form abzugrenzen. Das heißt, in der satzförmigen Form ÿÿw ist **der Teilstring** *ÿ* **ein Handle, falls es eine Produktion** A **ÿÿ** *ÿ* **in der CFG gibt; er kann also durch die Vorrangsrelationen abgegrenzt werden, wodurch** *ÿ* ÿ *ÿ* ÿ **w erhalten wird. Die Relation .= zwischen den Symbolen wird verwendet, wenn der Handle mehr als ein aufeinanderfolgendes Terminalsymbol hat. Das heißt, wenn** *ÿ* = a1a2 . . . an, **dann sollte die Relation** ai .= ÿi = ai+1 , **1, . . . ,** n **ÿ 1 hergestellt werden.Angenommen, eine rechtssatzförmige Form ist** A0a1A1a2 . . . anAn, **wobei jedes** Ai **ein Nichtterminal ist.** In diesem Fall können die Beziehungen als ai ÿ Aiai+1 **hergestellt werden , mit maximal einem Nichtterminal (Operatorgrammatikprinzip) zwischen den Terminalen. Nichtterminale könnten in der richtigen Satzform eliminiert werden, sodass nur die Beziehungen zwischen den Terminalen übrig bleiben.**

Kurz gesagt, jedes Mal, wenn ein Handle erhalten wird, kann es zwischen den Symbolen von ÿ **und ÿ eingeschlossen werden. Darüber hinaus können sich zwei aufeinanderfolgende Terminalsymbole innerhalb eines Handles befinden (Nichtterminale in der Mitte), und die Beziehung muss .=** hergestellt werden. **Schließlich können Nichtterminalsymbole aus der rechtssatzförmigen Form entfernt werden, wenn ein Handle gefunden und zwischen den Vorrangsbeziehungen gesperrt wird.**

Gemäß [2], den in Abschnitt **2.8 diskutierten Konzepten und den Vorrangbeziehungen** ist **der Vorranganalysealgorithmus 1 wie folgt aufgebaut:**

Algorithmus 1 Algorithmus zur Vorranganalyse Erfordert: Satz von

```
Vorrangbeziehungen. Prozedur PrecedenceParsing(w)
    Stapel ÿ [$] Ende ÿ falsch pos ÿ 0 mache a ÿ
        Stapel[oben] b
        ÿ w$[pos] wenn a
        = $ und b
        = $
            dann


                print(„Aktion: Akzeptieren“) end ÿ
                true
            anders
                wenn a ÿ b oder a .= b dann
                    pos + +
                    push(stack, b)
                    print(“Aktion: Shift”) sonst
                wenn
                    a ÿ b dann
                        Tun
                            e ÿ pull(stack) while
                        not (stack[top] ÿ e) print(„Aktion:
                        Reduzieren“) else print(„Aktion:


                    Fehler“) end ÿ true end if end if end if
                        while not end
                    end
                procedure
```

Machine Translated by Google

4. Vorgehensweise

Diese Arbeit schlägt eine Neudefinition der linken **und** rechten **Terminalmengen vor,** die LG(A) **bzw.** RG(A) **entsprechen . Zusätzlich wird die Definition der neuen** ganz links stehenden **Menge vorgestellt. Die neuen Regeln zur Bestimmung von Vorrangsbeziehungen werden demonstriert, auch wenn die CFG nicht der Grundregel der Operatorgrammatik entspricht, was bedeutet, dass die CFG mehr als ein benachbartes Nicht-Terminalsymbol auf der rechten Seite ihrer Produktionen haben kann.**

Wir präsentieren die Algorithmen zum Erstellen der drei vorgeschlagenen Sets: Links, Ganz links **und** Rechts. **Darüber hinaus demonstrieren wir den Algorithmus, der die neu festgelegten Vorrangregeln einbezieht.**

4.1. Definitionen

Es gibt folgende Symbole:

• Eine Folge von Terminalen x **ÿ** T **+. •**
Eine Folge von Nicht-Terminalen *ÿ* **ÿ Nÿ . •**
Eine Satzform *ÿ* **ÿ** (T **ÿ** N) **ÿ Die folgenden ·**

Mengen sind definiert:

• Die Linke, bezeichnet als L(A), **ist die Menge der Terminalsymbole, die auf der linken Seite jeder Satzform als Produkt einer ganz rechts stehenden Ableitung von A** erscheinen können.

$$L(A) = \{a \ \ddot{y} \ T \mid A \Rightarrow^{\ddot{y}} \ddot{y}a\ddot{y}\}.$$

• Das äußerste linke Element, bezeichnet als Lm(A), **ist die Menge der Terminalsymbole, die am Ende der linken Seite in jeder Ableitung von A erscheinen können, in Abwesenheit anderer äußerster linker grammatischer Symbole (oder die in deren Abwesenheit zu** *ÿ* abgeleitet werden).

$$Lm(A) = \{a \ \ddot{y} \ T \mid A \Rightarrow^{\ddot{y}} a\ddot{y}\}.$$

• Die rechte Seite, bezeichnet als R(A), **ist die Menge der Terminalsymbole, die am rechten Ende der am weitesten rechts stehenden Ableitungen von** A oder in jeder **Produktion von A weiter rechts erscheinen** können.

$$R(A) = \{a \ \ddot{y} \ T \mid A \Rightarrow^{\ddot{y}} \ddot{y}a \ \ddot{y} \ A \Rightarrow^{\ddot{y}} \ddot{y}\ddot{y}a\ddot{y}\}.$$

Demnach sind folgende Vorrangregeln definiert: b **ÿ** L(B), **dann** a ÿ **b.**

- **Wenn** C **ÿÿ** *ÿ1aÿBÿ2* ,
- **ÿÿ** *ÿ1AÿBÿ2* , **=ÿ ÿ, dann** aÿb(A) **ÿ** b **ÿ** Lm(B) **ÿ** *ÿ* **Wenn** C $\Rightarrow^{\ddot{y}}$
- **Wenn** C **ÿÿ** *ÿ1Aÿbÿ2* , a **ÿ** R(A) **ÿ** *ÿ* **=ÿ ÿ, dann** $\Rightarrow^{\ddot{y}}$ a ÿ **b.**
- **Wenn** C **ÿÿ** ÿ1aÿbÿ2, **dann** a .= **b.**

4.2. Problemanalyse

4.2.1. Grammatik absoluter Operatoren

Gegeben sei die äußerste rechte Ableitungsreihe der Form:

$$S \Rightarrow^{\ddot{y}} \ddot{y}1Ax1 \Rightarrow^{\ddot{y}} \ddot{y}1\ddot{y}1x1, \text{ wobei } A \Rightarrow^{\ddot{y}\ddot{y}} \ddot{y}1 \text{ und } \ddot{y}1 \Rightarrow^{*} \ddot{y}2B\ddot{y}3. \quad (1)$$

Es wird darauf hingewiesen, dass ÿ1 gemäß der vorherigen Definition in Abschnitt 2.7 ein Griff ist. Die rechtssatzförmige Form von (1) kann wie folgt umgeschrieben werden:

$$\ddot{y}1\ddot{y}1x1 \Rightarrow^{*} \ddot{y}1\ddot{y}2B\ddot{y}3x1 \Rightarrow^{\ddot{y}} \ddot{y}1\ddot{y}2Bx2x1 \Rightarrow^{\ddot{y}} \ddot{y}1\ddot{y}2\ddot{y}2x2x1, \text{ wobei } B \Rightarrow^{\ddot{y}\ddot{y}} \ddot{y}2. \quad (2)$$

Aus der letzten Satzform in (2) folgt, dass ÿ2 ein Handle gemäß der Definition in Abschnitt 2.7 ist.

Machine Translated by Google

Aus diesem Grund konzentriert sich die Analyse auf die Produktionen von A, die mindestens ein Nichtterminal B **auf der rechten Seite haben, wie in (1) zu sehen. Die Produktionen in** A **werden durch die Grammatik durchlaufen, um die Vorrangsbeziehungen des Nichtterminals** B **auf der rechten Seite zu finden.**

Im Kontext von ÿ1, wobei B **ÿÿ ÿ2 gilt und ÿ2 ein Griff ist, bilden sich folgende Formen kann aus seiner Definition in (1) abgeleitet**

werden : (a) ÿ1 = ÿ4aB, **daraus folgt** a ÿ L(B). **Hier ist ÿ2** = **ÿ4a und ÿ3** = **ÿ. (b) ÿ1** = *Bbÿ5,* **daraus folgt** R(B) ÿ **b. Hier** ist **ÿ2** = *ÿ* **und ÿ3** = **bÿ5. (c) ÿ1** = ÿ4aBbÿ5, **daraus folgt** a ÿ L(B) **und** R(B) ÿ **b. Hier ist ÿ2** = **ÿ4a und ÿ3** = **bÿ5.**

Für das nicht-terminale Startsymbol S wird eine anfängliche Dummy-Produktion E **ÿÿ** S **definiert und Regel (c) angewendet.**

Für eine Produktion A **ÿÿ** ÿ1aXbÿ2, X **ÿ** N **ÿ {ÿ}, folgt, dass** a .= **b.**

4.2.2. Allgemeine Grammatik mit zwei oder mehr aufeinanderfolgenden nicht-terminalen

Symbolen. Neudefinition von Links, Rechts und Definition von „ganz links“.

Links

(B): Gegeben sei die Ableitung ganz rechts:

$$ÿ1aBx1 \Rightarrow ÿ\ ÿ1aÿ1x1. \quad (3)$$

Es wird beobachtet, dass ÿ1 ein Handle innerhalb einer Bottom-Up-Analyse ist (inverse Ableitung ganz rechts), da B **=ÿ ÿ1;** daher **kann aus** (3) abgeleitet werden, dass **ÿ1a** ÿ **ÿ1** ÿ **x1. Darüber hinaus** kann **ÿ1 die folgenden Formen annehmen: (a) ÿ1** = *ÿ1bÿ2.* **(b) ÿ1** = *ÿ1Cÿ3.*

Wenn wir mit der ganz rechts stehenden Ableitung in (3) fortfahren **und jede der Optionen von ÿ1 berücksichtigen, erhalten wir:**

(a) ÿ1aBx1 **=ÿ** ÿ1aÿ1bÿ2x1, **da ÿ1 ein Griff ist, folglich ÿ1a** ÿ ÿ1bÿ2 ÿ **x1; daher daher** a ÿ b , **wobei** B **ÿÿ** ÿ1bÿ2 **und** b **ÿ** L(B). **(b)**
ÿ1aBx1 **=ÿ** ÿ1aÿ1Cÿÿÿÿÿ3x1 **=ÿ** ÿ1aÿ1Cx2x1 **=ÿ** ÿ1aÿ1ÿ2Dx3x2x1 **=ÿ ÿ1aÿ1ÿ2ÿ3bÿ4 x3x2x1. Es wird beobachtet, dass** C +=ÿ ÿ2Dx3 **und** D **ÿÿ** *ÿ3bÿ4;* **daher ist** ÿ3bÿ4 ein **Griff und folglich** *ÿ1aÿ1ÿ2* ÿ ÿ3bÿ4 ÿ **x3x2x1 und** a ÿ **b.**

Kurz gesagt, ÿ1aBx1 +=ÿ ÿ1aÿ1ÿ2ÿ3bÿ4x3x2x1; wenn *ÿ* = **ÿ1ÿ2ÿ3 und ÿ5** = **ÿ4x3x2 gesetzt ist , erhalten wir die zusammenfassende Ableitung** ÿ1aBx1 +=ÿ *ÿ1aÿbÿ5x1,* **wobei** B +=ÿ *ÿbÿ5* **und** b **ÿ** L(B).

Rechts

(B): Gegeben sei die Ableitung ganz rechts:

$$ÿ1Bax1 \Rightarrow ÿ\ ÿ1ÿ1ax1. \quad (4)$$

Man beobachtet, dass ÿ1 ein Handle innerhalb einer Bottom-Up-Analyse ist (inverse Ableitung ganz rechts), da B **=ÿ ÿ1; daher kann aus (4) abgeleitet werden, dass ÿ1** ÿ **ÿ1** ÿ **ax1. Außerdem** kann **ÿ1 die folgenden Formen annehmen:**

(a) ÿ1 = ÿ2bÿ. **(b) ÿ1** = ÿÿÿÿÿ

Machine Translated by Google

ÿ1ÿ3ÿ4 ÿ ÿ5bÿ2 ÿ ax1 **und** b ÿ **a.**

Kurz gesagt, ÿ1Bax1 +=ÿ **ÿ1ÿ3ÿ4ÿ5bÿ2ax1;** wenn wir **ÿ6** = ÿ3ÿ4ÿ5 **setzen , erhalten wir die zusammenfassende Ableitung** ÿ1Bax1 +=ÿ *ÿ1ÿ6bÿ2ax1,* **wobei** B +=ÿ ÿ6bÿ2 **und** b **ÿ** R(B).

Ganz links (B):

Gegeben sei die Produktion mit zwei nicht-terminalen Symbolen auf der rechten Seite (Hauptteil):

S **ÿÿ** AB **(5)**

A **ÿÿ ÿ1b1ÿ1** **(6)**

Die ganz rechte Ableitung von S **in (5) lautet:**

S **=ÿ** AB + =ÿ Aaÿ + = ÿ Aax1 **=ÿ** ÿ1b1ÿ1ax1 = ÿ2Cax1 + =ÿ **ÿ2ÿ3ÿ4b2ÿ2ax1.**
(ii) (iv) ab (vi) (vii)

Zusammenfassend:

- **In** (v) kann **die Beziehung** ÿÿ1b1ÿ1 ÿ ax1 **hergestellt werden, da der rechte Teil der Produktion (6) ist ein Handle. Daher** gilt **b1** ÿ a **und b1 ÿ** R(A).
- **Aus** (vi) **ist bekannt, dass ÿ1** = C0C1 . . . CnC, **wobei** Ci **ÿ** N **und** C jeweils **das Nichtterminal ist , das zuerst von rechts abgeleitet wurde. Es wird darauf hingewiesen, dass ÿ2** = **ÿ1b1C0 . . .** Cn. •

Aus (vii) ist C +=ÿ ÿ3ÿ4b2ÿ2 , **wobei ÿ4b2ÿ2 das Ergebnis der zuletzt durchgeführten Ableitung ganz rechts ist und somit ein Handle; daher ist** die Beziehung *ÿ2ÿ3* ÿ **ÿ4b2ÿ2** ÿ ax1 **hergestellt, da ÿ4b2ÿ2 ein Handle ist, also b2** ÿ **a.** • **Aus** (iv) folgt Aax1 +=ÿ **ÿ2ÿ3ÿ4b2ÿ2ax1.** Wenn **ÿ5** =

ÿ2ÿ3ÿ4 gilt , dann ist A +=ÿ **ÿ5b2ÿ2 und daher b2 ÿ** R(A). • **Aus** (ii) folgt B +=ÿ *aÿ ,* **woraus wir die neue ganz linke Menge von** B **als** Lm(B) =

{a **ÿ** T **|** B +=ÿ **aÿ} definieren.**

4.2.3. Eine intuitive Methode zur Beobachtung der Elemente jedes Sets

Vision of Left(·):
Gegeben sei die
kontextgebundene Formel: A **ÿÿ** CDE
E **ÿÿ** eh D
ÿÿ d C **ÿÿ**
cF F **ÿÿ** f

Die

äußerste rechte Ableitung der Zeichenfolge cf deh **bietet die folgende Möglichkeit zur Erstellung von** Le f t(A) **oder** L(A).

A **=ÿ** CDE **=ÿ CDeh =ÿ** Cdeh **=ÿ** cFdeh **=ÿ** cfdeh.
ÿ t D c

Die Terminalsymbole e, d und c **werden von links betrachtet und in** L(A) aufgezeichnet. **Das Symbol** e **ist das erste von links gesehene und wird in** L(A) aufgezeichnet, **und das Symbol** h **wird** durch e **blockiert, da beide erscheinen, wenn** E **ersetzt wird. Dann** wird D **durch d ersetzt , was beobachtet und in L(A) aufgezeichnet wird . Wenn** cF C **ersetzt , beobachten und zeichnen wir** c **in** L(A) auf.
Wenn schließlich die Ersetzung von F **durch** f **durchgeführt wird, kann sie nicht mehr beobachtet (durch c blockiert) und nicht registriert werden. Folglich:**

Die Gleichung lautet: L(A) = {e, **d, c}.**

Vision des Rechts(·):
Angesichts der CFG:
A **ÿÿ** abBC
B **ÿÿ** eF

Machine Translated by Google

C ÿÿ cD D
ÿÿ d F ÿÿ

f **Die am**

weitesten rechts stehende Ableitung der Zeichenfolge abe f cd **führt zu der folgenden Möglichkeit,** Right(A) **oder** R(A) zu erstellen.

A **=ÿ** abBC **=ÿ** abBcD **=ÿ** abBcd **=ÿ** abeFcd **=ÿ abefcd.**
B C D ÿ

Die Terminalsymbole b, c und d **werden von rechts betrachtet und in** R(A) aufgezeichnet. **Das Symbol** b **ist das erste, das in** R(A) beobachtet und aufgezeichnet wird. **Das Symbol** a wird durch **b blockiert ,** da beide erscheinen, wenn A **abgeleitet wird.** C **wird dann durch** cD ersetzt **und** c **wird in R(A) beobachtet und aufgezeichnet . Wenn** D für **d eingesetzt wird ,** beobachten und zeichnen wir d **in** R(A) **Dann wird** B durch eF abgeleitet , **aber** e **kann nicht aufgezeichnet werden, da die vorherige Einbeziehung von** c **und** d **in** R(A) **es unsichtbar macht oder dessen Beobachtung von rechts blockiert. Schließlich kann es, wenn** man f **durch** F **ersetzt , auch von rechts nicht beobachtet werden (blockiert durch** c **und d) und kann nicht aufgezeichnet werden. Folglich gilt:**

Die Gleichung lautet: R(A) = {b, **c, d}.**

Vision von ganz links(·):
Angesichts der CFG:
S ÿÿ dAB

A ÿÿ fa
B ÿÿ Cb
C ÿÿ c

Die rechteste Ableitung der Zeichenfolge df acb **ergibt die folgende Möglichkeit zur Erzeugung der** Ganz links (B) **oder** Lm (B), **wie in Abbildung 1 dargestellt.**

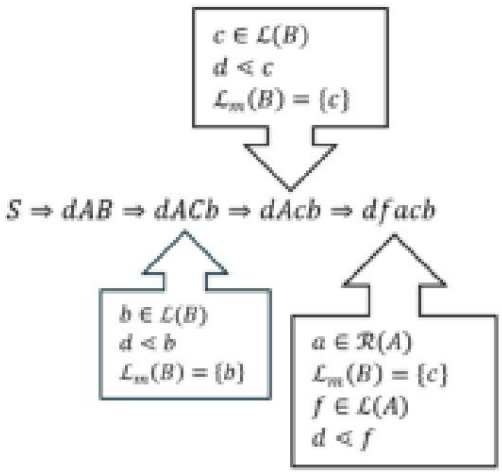


Abbildung 1. Herleitung der df acb- **Kette und** der Leftmost(B) **-Set-Erstellungssequenz.**

Man sieht, dass L(B) = {b, **c}. Wenn** B Cb **ableitet , beobachten wir, dass** Lm(B) = {b} , **weil bis dahin keine weiteren Terminalsymbole zu sehen sind. In der folgenden Ableitung** wird C **durch c ersetzt ; daher** befindet sich c **weiter links von der Teilzeichenkette** cb, sodass Lm(B) = {c} **bleibt .**

Machine Translated by Google

4.2.4. Zusammenfassung der Beziehungen zwischen grammatischen Symbolen aus Produktionen und Mengen

Tabelle 2 zeigt die Beziehungen zwischen den grammatikalischen Symbolen der Produktionen und den Terminalmengen.

Tabelle 2. Allgemeine Produktionen und erhaltene Vorrangbeziehungen.

Produktion	Beziehung
A **ÿÿ** *aÿb*	a .= b
A **ÿÿ** CÿB, *ÿ* =ÿ *ÿ*	R(C) ÿ ba
A **ÿÿ** *aÿC*	ÿ L(C)
A **ÿÿ** BÿC, *ÿ* =ÿ *ÿ*	R(B) ÿ Lm(C)

4.3. Algorithmen

Die folgenden Algorithmen wurden auf Grundlage der in Abschnitt 4.1 vorgestellten Definitionen entwickelt.

Algorithmus 2 empfängt ein Nicht-Terminal A , **dessen Produktionen bekannt sind. Für jedes dieser Produkte wird überprüft, ob sie der Form** *ÿBÿ entsprechen.* **Wenn dies der Fall ist, wird Algorithmus 2** erneut **ausgeführt , wobei das Nicht-Terminal** B **als Parameter gesendet wird (vorausgesetzt,** B ÿ = **A). Seine Antwort** wird in die Menge L(A) **aufgenommen .**

Anschließend wird überprüft, ob die Produktion der Form *ÿaÿ entspricht,* **und wenn ja, wird das Terminal** a **zur Menge** L(A) hinzugefügt.

Sobald die Überprüfung aller Produktionen von A abgeschlossen ist, **wird die Menge** L(A) als **Ausgabe von Algorithmus 2 zurückgegeben.**

Algorithmus 2 Linker **Algorithmus eines Nichtterminals** A

```
Funktion Links(A)
    L(A) ÿ ÿ für
    jede Produktion von A gilt,
        wenn A ÿÿ ÿBÿ dann
            L(A) ÿ Left(B) // L(A) enthält ein Left(B) -Ende, wenn
        wenn
        A ÿÿ ÿaÿ dann a ÿ
            L(A) // a ist in L(A) -Ende,
        wenn
    Ende für
    Rückgabewert L(A)
Endfunktion
```

Algorithmus 3 empfängt ein Nicht-Terminal A , **dessen Produktionen bekannt sind. Für jedes dieser Produkte wird überprüft, ob sie der Form** ÿBÿ **und** *ÿ* **=ÿ** ÿ entsprechen. Wenn dies der Fall ist, **wird Algorithmus 3** erneut ausgeführt, wobei das Nicht-Terminal B **als Parameter gesendet wird** (vorausgesetzt, B ÿ= **A). Seine Antwort wird in die Menge Lm(A) aufgenommen .**

Anschließend wird überprüft, ob die Produktion der Form ÿaÿ **und** *ÿ* entspricht . **Wenn** =ÿ ÿ. **ja, wird das Terminal** a **zur Menge** Lm(A) hinzugefügt.

Sobald die Überprüfung aller Produktionen von A **abgeschlossen ist, wird die Menge** Lm(A) **zurückgegeben als Ausgabe von Algorithmus 3.**

Machine Translated by Google

Algorithmus 3 Ganz links stehender **Algorithmus eines Nichtterminals** A

```
Funktion Ganz links(A)
    Lm(A) ÿ ÿ für
    jede Produktion von A gilt, wenn
        A ÿÿ ÿBÿ und ÿ =ÿ ÿ dann
            Lm(A) ÿ Leftmost(B) Ende
        wenn
        wenn A ÿÿ ÿaÿ und ÿ =ÿ ÿ dann a ÿ
            Lm(A) Ende
        wenn
    Ende für
    return Lm(A)
Ende Funktion
```

Algorithmus 4 empfängt ein Nicht-Terminal A , **dessen Produktionen bekannt sind. Für jedes dieser Produkte wird überprüft, ob sie der Form** ÿBÿ **und** ÿ **=ÿ** ÿ entsprechen. Wenn dies der Fall ist, **wird Algorithmus 4** erneut ausgeführt, wobei das Nicht-Terminal B **als Parameter gesendet wird** (vorausgesetzt, B ÿ= **A). Seine Antwort wird in die Menge** R(A) aufgenommen.

Anschließend wird überprüft, ob die Produktion der Form *ÿaÿ entspricht,* **und wenn ja, wird das Terminal** a **zur Menge** R(A) hinzugefügt.

Sobald alle Produktionen von A **verifiziert wurden, wird die Menge** R(A) **als Ausgabe von Algorithmus 4 zurückgegeben.**

Algorithmus 4 Richtiger **Algorithmus eines Nichtterminals** A

```
Funktion Rechts(A)
    R(A) ÿ ÿ für
    jede Produktion von A gilt, wenn
        A ÿÿ ÿBÿ und ÿ =ÿ ÿ dann
            R(A) ÿ Rechts(B) Ende
        wenn
        wenn A ÿÿ ÿaÿ dann
            a ÿ R(A)
        Ende
    wenn
    Ende für return
R(A) Ende Funktion
```

Algorithmus 5 erhält eine Grammatik G(S) =< **T, N, S,** P > **und die Algorithmen 2–4. Zunächst werden Vorrangbeziehungen zwischen dem Begrenzungssymbol $ und den** linken **und** rechten **Mengen des nicht-terminalen Symbols des Starts S hinzugefügt, gemäß den Regeln** (A.1)

Algorithmus 5 arbeitet iterativ. Er durchläuft alle Produktionen der Menge P und initialisiert für jede Iteration zwei temporäre Variablen, u **und l, auf null bzw. eine leere Liste. Anschließend führt er eine Reihe von Prozessen für jedes Paar zusammenhängender grammatischer Symbole** X **und** Y **in einer Produktion aus.**

1. **Der Algorithmus prüft, ob** XY **die Form** ab hat; **wenn ja, fügt er die entsprechende Präzedenz hinzu Beziehung zwischen** a **und b gemäß Regel** (B).
2. **Der Algorithmus prüft, ob** XY **die Form** aB hat; **wenn ja, weist er** u **den Wert von** a zu **und fügt die entsprechenden Vorrangsbeziehungen zwischen** a **und den Terminals in** Left(B) **gemäß Regel** (C) hinzu.
3. **Der Algorithmus prüft, ob** XY **die Form Ab hat ;** wenn ja, fügt er **gemäß Regel** (D.1) **die entsprechenden Vorrangsbeziehungen zwischen** b **und den Terminals in** Rechts(A) hinzu. **Dann prüft er, ob** u ÿ = null ist, und wenn ja, fügt er **gemäß Regel** (D.2) **die entsprechende Vorrangsbeziehung zwischen** u **und** b hinzu **und setzt** u **auf** null zurück. **Schließlich prüft er, ob Liste** l **Elemente hat,** und wenn ja, fügt er **gemäß Regel** (D.3) **die entsprechenden Vorrangsbeziehungen zwischen** b **und den Terminals in** Rechts(B) hinzu **und setzt Liste** l **auf leer zurück.**

Machine Translated by Google

4. Der Algorithmus prüft, ob XY **die Form** AB hat; **wenn ja, fügt er die entsprechenden Vorrangsbeziehungen zwischen den Terminals in** Rechts(A) **und den Terminals in** Ganz links(B) hinzu, **gemäß Regel** (E.1). **Dann prüft er, ob** u ÿ = null ist, **und wenn ja, fügt er die entsprechende Vorrangsbeziehung zwischen** u **und den Terminals in** Links(B) hinzu, **gemäß Regel** (E.2). **Schließlich prüft er, ob die Liste** l **Elemente hat, und wenn ja, fügt er die entsprechenden Vorrangsbeziehungen zwischen den Terminals in** Rechts(C), **mit** C **ÿ** l , **und den Terminals in** Ganz links(B) hinzu, **gemäß Regel** (E.3), **und setzt die Liste** l zurück , **falls nötig.**

Algorithmus 5 Algorithmus zum Erhalten der Operatorpräzedenztabelle einer G- **Grammatik**

Erfordert: G(S) =< **T, N, S,** P > **Prozedur**
Tabelle(G, Links(·), Rechts(·), Ganz links(·))
Berechne Links(A), Rechts(A), Ganz links(A), ÿA **ÿ** N **und addiere $**
ÿ **a,** ÿa **ÿ** Links(S). **Addiere** a ÿ (A.1)
$, ÿa **ÿ** Rechts(S) **für jedes** A **ÿÿ** (A.2)
ÿ **in** P **und mache** u **ÿ** null.

l **ÿ** [ ] // **leere Liste für**
jedes XY, **das in** *ÿ* aneinandergrenzt,
mache wenn XY = ab , **dann**
füge a .= b hinzu (B)
Ende, wenn
wenn XY = aB **dann**
u **ÿ** ein
addiere a ÿ **b,** ÿb **ÿ** Links(B) (C)
Ende
wenn wenn XY = Ab
dann addiere a ÿ **b,** ÿa **ÿ** (D.1)
Rechts(A) **wenn** u ÿ=
null **dann** (D.2)
addiere
u .= bu
ÿ null **Ende wenn**
wenn l ÿ= [ ]
dann für B **ÿ** l **addiere** a ÿ **b,** ÿa (D.3)
ÿ

Rechts(B) **Ende für** l **ÿ** [ ] **Ende wenn Ende wenn wenn** XY = AB **dann**
addiere a ÿ **b,** ÿa **ÿ** Rechts(A) **ÿ** ÿb **ÿ** Ganz links(B) **wenn** u (E.1)
ÿ= null **dann addiere**
u ÿ **b,** ÿb **ÿ** Links(B) (E.2)
Ende, wenn
wenn l ÿ= [ ] **dann**
für C **ÿ** l
addiere c ÿ **b,** ÿc **ÿ** Rechts(C) **ÿ** ÿb **ÿ** Ganz links(B) **Ende für** (E.3)

Ende, wenn
wenn $B^{ÿ}$**=ÿ** *ÿ* **dann**
addiere A **zu**
l
sonst l **ÿ** [ ]
Ende
wenn
Ende
wenn
Ende für Ende für Ende Prozedur

Machine Translated by Google

4.4. Komplexitätsanalyse **Die**

folgenden Variablen werden für die Berechnung der Komplexität von die vier vorgeschlagenen Algorithmen:

- t : **Kardinalität der Menge der Terminalsymbole.** p : **Kardinalität der Menge der Produktionen.** g : **Anzahl der grammatischen Symbole in den Produktionen.**
- n : **Kardinalität der Menge der nicht-terminalen Symbole.**

4.4.1. Komplexität von Algorithmus 2 (links)

Die Funktion Left **prüft, in welchen Produktionen das Symbol** A **auf der linken Seite vorkommt. Im ersten if-Block wird ein rekursiver Aufruf von** Left **beobachtet, was bedeutet, dass alle nicht-terminalen Symbole von** G **am Ende der Ausführung auf der Suche nach** Left(A) durchlaufen werden könnten. **Daher würden alle Worst-Case-Produktionen durchlaufen, was eine Rechenzeit von** O(p) ergibt. **Der zweite if-Block hat eine Rechenzeit von O(1), die nicht zur Gesamtrechenzeit** O(p) beiträgt.

4.4.2. Komplexität von Algorithmus 3 (ganz links)

Die Funktion Leftmost **hat eine ähnliche Struktur wie die vorherige. Daher ist die Die Rechenausführungszeit beträgt** O(p).

4.4.3. Komplexität von Algorithmus 4 (rechts)

Die Funktion Right **weist eine ähnliche Struktur wie die Funktion** Left auf. **Daher beträgt die Rechenausführungszeit** O(p).

4.4.4. Komplexität von Algorithmus 5 (Tabelle)

Gemäß dem ersten For-Zyklus beträgt die Anzahl der Ausführungen p. Unmittelbar im Anschluss an den vorherigen wird der innere For-Zyklus im schlimmsten Fall g- **mal ausgeführt, wobei davon ausgegangen wird, dass die Gesamtzahl der grammatikalischen Symbole in jeder Produktion Daher beträgt der Rechenaufwand der beiden Zyklen** O(p × g).

Die Rechenkosten für die Konstruktion von Left(A), Right(A) **und** Leftmost(A), ÿA **ÿ N, sind bereits berechnet. Bei der Berechnung der Komplexität von** Table **wird die Länge jedes generierten Satzes berücksichtigt, die im schlimmsten Fall t beträgt.**

Die Analyse der vier if-Blöcke innerhalb des zweiten for-Zyklus ist wie folgt: • Im ersten if-Block wird nach der Überprüfung die ausgeführte Anweisung in konstanter Zeit und wird mit O(1) bezeichnet.

- **Im zweiten if-Block wird nach der Überprüfung die erste Anweisung in einer konstanten Zeit O(1) und die zweite in einer Rechenzeit** O(t) ausgeführt; **daher beträgt die Rechenzeit des if-Blocks** O(t).

- **Im dritten if-Block werden nach der Überprüfung drei Unterblöcke unterschieden; der erste entspricht der Bewegung von** Right(A), **die in einer Rechenzeit** O(t) ausgeführt wird. **Nach der Überprüfung** von u ÿ= null **stellt der zweite Unterblock zwei Anweisungen dar, die in einer konstanten Zeit O(1) ausgeführt werden. Der dritte Unterblock stellt nach der Überprüfung von** l ÿ= [ ] **einen Rechenaufwand** O(n × t) dar, der durch den im schlimmsten Fall O(n) Mal **ausgeführten for-Zyklus bestimmt wird , entsprechend der Anzahl der nicht-terminalen Symbole in** G **und der Bewegung ,** die dieser von Right(B) **mit einem Aufwand** O(t) **macht . Zusammenfassend beträgt der Rechenaufwand des dritten if-Blocks** O(n × t), **was dem schlimmsten Fall entspricht.**
- **Im vierten if-Block werden nach der Überprüfung vier Unterblöcke unterschieden. Der erste entspricht den Kombinationen, die durch das Durchlaufen von** Right(A) **und** Leftmost(B) gebildet werden , **um die Vorrangsbeziehungen** a ÿ **b zu erhalten. Da die maximale Größe jedes Satzes t ist, beträgt der Rechenaufwand dieser** t^2**). Nach der Überprüfung von** u ÿ= **null , Kombinationen** O(t) . **Der zweite Unterblock stellt eine Anweisung dar, bei der** Left(B) **durchlaufen wird; daher beträgt der Rechenaufwand** O(t). **Im dritten Unterblock wird nach der Überprüfung von** l ÿ= [ ] **ein for-Zyklus dargestellt, der höchstens** n- **mal die Kombinatio**

Machine Translated by Google

Durchlaufen von Right(C) **und** Leftmost(B), **um die Vorrangsbeziehungen** c ÿ **b zu erhalten. Diese Kombinationen werden maximal in** O(t) **Rechenzeit durchlaufen, da** t **die Größe beider Mengen ist und folglich der Rechenaufwand des dritten Unterblocks** O(n × t) **beträgt. Der letzte Unterblock hat einen Aufwand von O(1). Zusammenfassend beträgt der Rechenaufwand des vierten Blocks** O(n × t). **Zusammenfassend beträgt der** **), was dem schlimmsten Fall entspricht.**

Rechenaufwand der vier if-Blöcke innerhalb des zweiten for-Zyklus von Algorithmus 5 ist O(n × t. **), was der schlimmste Fall ist; daher beträgt der Rechenaufwand für Der Hauptteil von Algorithmus 5 , der aus den beiden verschachtelten For-Zyklen und den vier If-Blöcken besteht, ist** O(p × g × n), **was größer ist als die Kosten für die Berechnung eines der drei Sätze** × t) **an Terminalsymbolen** Left(A), Right(A) **und** Leftmost(A), ÿA **ÿ N, beim Starten des Algorithmus und entsprechend** O(p).

5. Beispiele

5.1. Beispiel 1

5.1.1. Grammatik

Gegeben sei folgende OPG: E
ÿÿ E + T **|** TT **ÿÿ** T
ÿ F **|** FF **ÿÿ** (E) **|** id.
Die Mengen
werden schrittweise gemäß den Algorithmen aufgebaut.

5.1.2. Links

Die linken **Sätze werden wie in Tabelle 3 beschrieben berechnet .**

Tabelle 3. Berechnung der linken **Mengen für die CFG in Beispiel 1.**

Links (E)			
Für E **ÿÿ** E + T			
1. Auswerten der ersten Bedingung von Algorithmus 2.	EIN = E. *ÿ* = *ÿ* B = E *ÿ* = +T L(E) **ÿ** L(E)	**2. Auswerten der zweiten Bedingung von Algorithmus 2.**	EIN = E. *ÿ* = E a = + *ÿ* = T L(E) **ÿ** {+}
Für E **ÿÿ** T			
3. Auswerten der ersten Bedingung von Algorithmus 2.	EIN = E. *ÿ* = *ÿ* B = T *ÿ* = *ÿ* L(E) = {+} **ÿ** L(T)		

Links (T)			
Für T **ÿÿ** T **ÿ** F			
1. Auswerten der ersten Bedingung von Algorithmus 2.	A = T *ÿ* = *ÿ* B = T *ÿ* = ÿF L(T) **ÿ** L(T)	**2. Auswerten der zweiten Bedingung von Algorithmus 2.**	A = T *ÿ* = T ein = ÿ *ÿ* = F L(T) **ÿ {ÿ}**
Für T **ÿÿ** F			
3. Auswerten der ersten Bedingung von Algorithmus 2.	A = T *ÿ* = *ÿ* B = F *ÿ* = *ÿ* L(T) = **{ÿ} ÿ** L(F)		

Machine Translated by Google

Tabelle 3. Fortsetzung

Links (F)			
Für F **ÿÿ** (E)		**Für** F **ÿÿ** id	
1. Auswerten der zweiten Bedingung von Algorithmus 2.	A = F *ÿ* = *ÿ* a = (ÿ = E) L(F) **ÿ** {(}	**2. Auswerten der zweiten Bedingung von Algorithmus 2.**	A = F *ÿ* = *ÿ* a = id *ÿ* = *ÿ* L(F) = {(, id}

Zusammenfassung der linken **Mengen von Nicht-**

Terminalen: • Links(E) = {+} **ÿ** L(T).
• Links(T) = **{ÿ} ÿ** L(F). • Links(F) =
{(, **id}.**

Ersetzen der

Mengen: • Links(E) = {+,
ÿ,(, **id}.** • Links(T) = **{ÿ,**
(, id}. • Links(F) = {(, **id}.**

5.1.3. Rechts

Die rechten Sätze werden wie in Tabelle **4 beschrieben berechnet .**

Tabelle 4. Berechnung der rechten **Sätze für die CFG in Beispiel 1.**

Rechts (E)			
Für E **ÿÿ** E + T			
1. Auswerten der ersten Bedingung von Algorithmus 4.	EIN = E. *ÿ* = E+ B = T *ÿ* = *ÿ* R(E) **ÿ** R(T)	**2. Auswerten der zweiten Bedingung von Algorithmus 4.**	EIN = E. *ÿ* = E a = + *ÿ* = T R(E) **ÿ** R(T) **ÿ** {+}
Für E **ÿÿ** T			
3. Auswerten der ersten Bedingung von Algorithmus 4.	EIN = E. *ÿ* = *ÿ* B = T *ÿ* = *ÿ* R(E) = R(T) **ÿ** {+}		

Rechts (T)			
Für T **ÿÿ** T **ÿ** F			
1. Auswerten der ersten Bedingung von Algorithmus 4.	A = T *ÿ* = **Tÿ** B = F *ÿ* = *ÿ* R(T) **ÿ** R(F)	**2. Auswerten der zweiten Bedingung von Algorithmus 4.**	A = T *ÿ* = T ein = ÿ *ÿ* = F R(T) **ÿ** R(F) **ÿ {ÿ}**
Für T **ÿÿ** F			
3. Auswerten der ersten Bedingung von Algorithmus 4.	A = T *ÿ* = *ÿ* B = F *ÿ* = *ÿ* R(T) = R(F) **ÿ {ÿ}**		

Machine Translated by Google

Tabelle 4. Fortsetzung

Rechts (F)			
Für F ÿÿ (E)		**Für** F ÿÿ id	
1. Auswertung der zweiten Bedingung von Algorithmus 2.	A = F ÿ = (E a =) ÿ = ÿ R(F) **ÿ** {)}	**2. Auswertung der zweiten Bedingung von Algorithmus 2.**	A = F ÿ = ÿ a = Ich würde ÿ = ÿ R(F) = {), id}

Zusammenfassung der rechten **Mengen von Nichtterminals:**

- Rechts(E) = R(T) **ÿ {+}.**
- Rechts(T) = R(F) **ÿ {ÿ}.**
- Rechts(F) = {), **id}.**

Ersetzen der Sets:

- Rechts(E) = {), id, **ÿ, +}.**
- Rechts(T) = {), id, **ÿ}.**
- Rechts(F) = {), **id}.**

5.1.4. Ganz links

Die ganz linken Werte werden wie in Tabelle **5 beschrieben berechnet .**

Tabelle 5. Berechnung der ganz linken **Mengen für die CFG in Beispiel 1.**

Ganz links (E)			
Für E ÿÿ E + T		**Für** E ÿÿ T	
1. Auswertung der ersten Bedingung von Algorithmus 3.	EIN = E. ÿ = ÿ B = E ÿ = +T Lm(E) **ÿ** Lm(E)	**2. Auswertung der ersten Bedingung von Algorithmus 3.**	EIN = E. ÿ = ÿ B = T ÿ = ÿ Lm(E) = Lm(T)

Ganz links (T)			
Für T ÿÿ T **ÿ** F		**Für** T ÿÿ F	
1. Auswertung der ersten Bedingung von Algorithmus 3.	A = T ÿ = ÿ B = T ÿ = ÿF Lm(T) **ÿ** Lm(T)	**2. Auswertung der ersten Bedingung von Algorithmus 3.**	A = T ÿ = ÿ B = F ÿ = ÿ Lm(T) = Lm(F)

Ganz links (F)			
Für F ÿÿ (E)		**Für** F ÿÿ id	
1. Auswertung der zweiten Bedingung von Algorithmus 3.	A = F ÿ = ÿ a = (ÿ = E) Lm(F) **ÿ** {(}	**2. Auswertung der zweiten Bedingung von Algorithmus 3.**	A = F ÿ = ÿ a = Ich würde ÿ = ÿ Lm(F) = {(, id}

Zusammenfassung der äußersten linken **Mengen von Nichtterminals:**

- Ganz links (E) = Lm(T).

Machine Translated by Google

• Ganz links (T) = Lm (F).
• Ganz links(F) = {(, **id}.**

Ersetzen der Sets:

• Ganz links(E) = {(, **id}.**
• Ganz links(T) = {(, **id}.**
• Ganz links(F) = {(, **id}.**

5.1.5. Fallstudien

Für Algorithmus 5 wird die Anwendung einiger Vorrangbeziehungen gezeigt.

1. Für das Trennzeichen $:

(a) $ ÿ L(E) **ÿÿ Regel** (A.1).

- **$** ÿ +
- **$** ÿ **ÿ**
- **$** ÿ (
- **$** ÿ Ich würde

(b) R(E) ÿ **$ ÿÿ Regel** (A.2).

-) ÿ **$**
- Ich würde ÿ **$**
- **ÿ** ÿ **$**
- + ÿ **$**

2. Für E **ÿÿ** E + T

(a) X = **E,** Y = + **ÿÿ Regel** (D.1).

R(E) ÿ +

-) ÿ +
- Ich würde ÿ +
- **ÿ** ÿ +
- + ÿ +

u = null, l = [ ]

(b) X = **+,** Y = T **ÿÿ Regel** (C).

+ ÿ L(T)

- + ÿ **ÿ**
- + ÿ (
- + ÿ Ich würde

u = **+,** l = [ ]

3. Für F **ÿÿ** (E)

(a) X = **(,** Y = E **ÿÿ Regel** (C).

(ÿ L(E)

- (ÿ +
- (ÿ **ÿ**
- (ÿ (
- (ÿ Ich würde

u = **(,** l = [ ]

(b) X = **E,** Y =) **ÿÿ Regel** (D.1).

R(E) ÿ)

-) ÿ)
- Ich würde ÿ)
- **ÿ** ÿ)
- + ÿ)

u ÿ= null **ÿÿ** u .= b **ÿÿ** (.=) **ÿÿ Regel** (D.2).

u = null, l = [ ]

Machine Translated by Google

4. In folgenden Fällen ist die Form XY **im rechten Teil der Produktion nicht erfüllt , deshalb kann keine Regel angewendet werden:**

(A) E **ÿÿ** T

(b) T **ÿÿ** F

(c) F **ÿÿ** id

Die verbleibenden Vorrangbeziehungen werden mit der gleichen Methode wie in den zuvor gezeigten Fällen abgeleitet.

5.1.6. Operator-Präzedenztabelle

Am Ende der Algorithmusausführung wird die Operatorpräzedenztabelle 6 abgerufen.

Tabelle 6. Vorrangtabelle von Beispiel 1.

	+	ÿ	(	)	Auswerts	**$**
+	ÿ	ÿ	ÿ	ÿ	ÿ	ÿ
ÿ	ÿ	ÿ	ÿ	ÿ	ÿ	ÿ
	ÿ	ÿ	ÿ	.=	ÿ	
	ÿ	ÿ		ÿ		ÿ
( ) Ich	ÿ	ÿ		ÿ		ÿ
würde $	ÿ	ÿ	ÿ		ÿ	

5.1.7. Bottom-Up-Analyse

Es sei id + ((id + id) **ÿ** (id)) **ÿ** id **ein String, der durch die Ableitung ganz rechts erhalten wird. Dann führt die Anwendung von Algorithmus 1 zu der in Tabelle 7 beschriebenen Bottom-Up-Analyse.**

Tabelle 7. Bottom-Up-Parsing von Beispiel 1.

Stapel	Eingabe-	Aktion
$	ID + ((ID + ID) **ÿ** (ID)) **ÿ** id$ +((ID + ID)	**Schicht**
$id	**ÿ** (ID)) **ÿ** id$ +((ID + ID) **ÿ** (ID)) **ÿ**	**Reduzieren**
$	id$ ((ID + ID) **ÿ** (ID)) **ÿ** id$ (ID + ID)	**Schicht**
$+	**ÿ** (ID)) **ÿ** id$ id + ID) **ÿ** (ID)) **ÿ**	**Schicht**
$ +	id$ +ID) **ÿ** (ID)) **ÿ** id$ +ID) **ÿ**	**Schicht**
($ +	(ID)) **ÿ** id$ id) **ÿ** (ID)) **ÿ** id$) **ÿ**	**Schicht**
(($ +	(ID)) **ÿ** id$) **ÿ** (ID)) **ÿ** id$)	**Reduzieren**
((id	**ÿ** (ID)) **ÿ** id$ ÿ(id)) **ÿ** id$	**Schicht**
$ + (($	ÿ(id)) **ÿ** id$ (id)) **ÿ** id$	**Schicht**
+ ((+ **$** +	id)) **ÿ** id $)) **ÿ** id$))	**Reduzieren**
((+id **$**	**ÿ** id$	**Reduzieren**
+ ((+		**Schicht**
$ +		**Reduzieren**
(($		**Schicht**
+ (()		**Schicht**
$ + ($		**Schicht**
+ **(ÿ $** +		**Reduzieren**
(ÿ($** + (ÿ(id **$ + **(ÿ(**		**Schicht**

Machine Translated by Google

Tabelle 7. Fortsetzung

Stapel		Aktion
$ + (ÿ()		**Reduzieren**
$ + **(ÿ**		**Reduzieren**
$ +	**Eingabe) ÿ** id$) **ÿ** id$) **ÿ** id$	**Schicht**
($ + ()	ÿid$	**Reduzieren**
$+	ÿid$	**Schicht**
$ + **ÿ**	Ich würde$	**Schicht**
$ + ÿid	**$**	**Reduzieren**
$ + **ÿ**	**$**	**Reduzieren**
$+	**$**	**Reduzieren**
$	**$**	**Akzeptieren**

5.2. Beispiel 2

5.2.1. Grammatik

Angesichts des folgenden Nicht-OPG:

S **ÿÿ** SD; **| D;**
D **ÿÿ** Tid(L)
T **ÿÿ** T ÿ **|**
L *ÿ* **ÿ** I **|**
Ich **ÿÿ** T **| T,** Ich

Die Sets werden Schritt für Schritt gemäß den Algorithmen aufgebaut.

5.2.2. Links

Die linken **Sätze werden wie in Tabelle 8 beschrieben berechnet .**

Tabelle 8. Berechnung der linken **Mengen für die CFG in Beispiel 2.**

Links (S)			
Für S **ÿÿ** SD;			
1. Auswertung der ersten Bedingung von Algorithmus 2.	EIN = S. *ÿ* = *ÿ* B = S *ÿ* = **D;** L(S) **ÿ** L(S)	**2. Auswertung der ersten Bedingung von Algorithmus 2.**	EIN = S. *ÿ* = S B = D *ÿ* **=;** L(S) **ÿ** L(D)
3. Auswertung der zweiten Bedingung von Algorithmus 2.	EIN = S. *ÿ* = SD ein **=;** *ÿ* = *ÿ* L(S) **ÿ** L(D) **ÿ {; }**		
Für S **ÿÿ D;**			
4. Auswertung der ersten Bedingung von Algorithmus 2.	EIN = S. *ÿ* = *ÿ* B = D *ÿ* **=;** L(S) **ÿ** L(D) **ÿ {; } ÿ** L(D)	**5. Auswertung der zweiten Bedingung von Algorithmus 2.**	EIN = S. *ÿ* = D ein **=;** *ÿ* = *ÿ* L(S) = L(D) **ÿ {; }**

Machine Translated by Google

Tabelle 8. Fortsetzung

Links (D)			
Für D **ÿÿ** Tid(L)			
1. Auswerten der ersten Bedingung von Algorithmus 2.	EIN = D. *ÿ* = *ÿ* B = T *ÿ* = id(L) L(D) **ÿ** L(T)	**2. Auswerten der zweiten Bedingung von Algorithmus 2.**	EIN = D. *ÿ* = T a = id *ÿ* = (L) L(D) = L(T) **ÿ** {id}

Links (T)			
Für T **ÿÿ Tÿ**			
1. Auswerten der ersten Bedingung von Algorithmus 2.	A = T *ÿ* = *ÿ* B = T *ÿ* = **ÿ** L(T) **ÿ** L(T)	**2. Auswerten der zweiten Bedingung von Algorithmus 2.**	A = T *ÿ* = T ein = *ÿ* *ÿ* = *ÿ* L(T) **ÿ {ÿ}**
Für T **ÿÿ** int			
3. Auswerten der zweiten Bedingung von Algorithmus 2.	A = T *ÿ* = *ÿ* a = int *ÿ* = *ÿ* L(T) = **{ÿ,** int}		

Links (L)			
Für L **ÿÿ** I			
1. Auswerten der ersten Bedingung von Algorithmus 2.	A = L *ÿ* = *ÿ* B = I *ÿ* = *ÿ* L(L) = L(Ich)		

Links (I)			
Für I **ÿÿ** T			
1. Auswerten der ersten Bedingung von Algorithmus 2.	A = ich *ÿ* = *ÿ* B = T *ÿ* = *ÿ* L(I) **ÿ** L(T)		
Für I **ÿÿ T,** I			
2. Auswerten der ersten Bedingung von Algorithmus 2.	A = Ich *ÿ* = *ÿ* B = T *ÿ* **=,** ich L(I) **ÿ** L(T)	**3. Auswerten der zweiten Bedingung von Algorithmus 2.**	A = Ich *ÿ* = T a **=,** *ÿ* = I L(I) = L(T) **ÿ {, }**

Zusammenfassung der linken **Mengen von Nicht-Terminalen:** • Links(S) = L(D) **ÿ {; }.** • Links(D) = L(T) **ÿ** {id}. • Links(T) = **{ÿ,** int}. • Links(L) = L(I). • Links(I) = L(T) **ÿ {, }.**

Machine Translated by Google

Ersetzen der Mengen: •

Links(S) = **{; ,** id, **ÿ,** int}. • Links(D) = {id, **ÿ,** int}. • Links(T) = **{ÿ,** int}. • Links(L) = **{, , ÿ,** int}. • Links(I) = **{, , ÿ,** int}.

5.2.3. Rechts

Die rechten **Sätze werden wie in Tabelle 9 beschrieben berechnet .**

Tabelle 9. Berechnung der rechten **Sätze für die CFG in Beispiel 2.**

Rechte)			
Für S **ÿÿ** SD;		**Für** S **ÿÿ D;**	
1. Auswerten der zweiten Bedingung von Algorithmus 4.	EIN = S. *ÿ* = SD ein **=;** *ÿ* = *ÿ* R(S) **ÿ {; }**	**2. Auswerten der zweiten Bedingung von Algorithmus 4.**	EIN = S. *ÿ* = D ein **=;** *ÿ* = *ÿ* R(S) = **{; }**

Rechts (D)	
Für D **ÿÿ** Tid(L)	
1. Auswerten der zweiten Bedingung von Algorithmus 4.	EIN = D. *ÿ* = Tid(L a =) *ÿ* = *ÿ* R(D) = {)}

Rechts (T)			
Für T **ÿÿ Tÿ**		**Für** T **ÿÿ** int	
1. Auswerten der zweiten Bedingung von Algorithmus 4.	A = T *ÿ* = T ein = ÿ *ÿ* = *ÿ* R(T) **ÿ {ÿ}**	**2. Auswerten der zweiten Bedingung von Algorithmus 4.**	A = T *ÿ* = *ÿ* a = int *ÿ* = *ÿ* R(T) = **{ÿ,** int}

Rechts (L)	
Für L **ÿÿ** I	
1. Auswerten der ersten Bedingung von Algorithmus 4.	EIN = L. *ÿ* = *ÿ* B = I *ÿ* = *ÿ* R(L) = R(Ich)

Machine Translated by Google

Tabelle 9. Fortsetzung

Rechts (I)			
Für I **ÿÿ** T			
1. Auswerten der ersten Bedingung von Algorithmus 4.	A = Ich *ÿ* = *ÿ* B = T *ÿ* = *ÿ* R(ich) **ÿ** R(T)		
Für I **ÿÿ T,** I			
2. Auswerten der ersten Bedingung von Algorithmus 4.	A = Ich *ÿ* = **T,** B = I *ÿ* = *ÿ* R(I) **ÿ** R(T) **ÿ** R(I)	**3. Auswerten der zweiten Bedingung von Algorithmus 4.**	A = Ich *ÿ* = T ein **=,** *ÿ* = ich R(I) = R(T) **ÿ {, }**

Zusammenfassung der rechten **Mengen von Nicht-**

Terminalen: • Rechts(S) = **{; }.** • Rechts(D) = **{)}.** • Rechts(T) = **{ÿ,** int}. • Rechts(L) = R(I). • Rechts(I) = R(T) **ÿ {, }.**

Ersetzen der Mengen: • Rechts(S) = **{; }.** • Rechts(D) = **{)}.** • Rechts(T) = **{ÿ,** int}. • Rechts(L) = **{, , ÿ,** int}. • Rechts(I) = **{, , ÿ,** int}.

5.2.4. Ganz links

Die ganz linken Sätze werden wie in Tabelle **10** **beschrieben berechnet .**

Tabelle 10. Berechnung der ganz linken **Mengen für die CFG in Beispiel 2.**

Ganz links (S)			
Für S **ÿÿ** SD;		**Für** S **ÿÿ D;**	
1. Auswerten der ersten Bedingung von Algorithmus 3.	EIN = S. *ÿ* = *ÿ* B = S *ÿ* = **D;** Lm(S) **ÿ** Lm(S)	**2. Auswerten der ersten Bedingung von Algorithmus 3.**	EIN = S. *ÿ* = *ÿ* B = D *ÿ* **=;** Lm(S) = Lm(D)

Ganz links (D)	
Für D **ÿÿ** Tid(L)	
1. Auswerten der ersten Bedingung von Algorithmus 3.	EIN = D. *ÿ* = *ÿ* B = T *ÿ* = id(L) Lm(D) = Lm(T)

Machine Translated by Google

Tabelle 10. Fortsetzung

Ganz links (T)			
Für T **ÿÿ Tÿ**		**Für** T **ÿÿ** int	
1. Auswertung der ersten Bedingung von Algorithmus 3.	A = T *ÿ* = *ÿ* B = T *ÿ* = **ÿ** Lm(T) **ÿ** Lm(T)	**2. Auswertung der zweiten Bedingung von Algorithmus 3.**	A = T *ÿ* = *ÿ* a = int *ÿ* = *ÿ* Lm(T) = {int}

Ganz links (L)	
Für L **ÿÿ** I	
1. Auswertung der ersten Bedingung von Algorithmus 3.	EIN = L. *ÿ* = *ÿ* B = I *ÿ* = *ÿ* Lm(L) = Lm(I)

Ganz links (I)			
Für I **ÿÿ** T		**Für** I **ÿÿ T,** I	
1. Auswertung der ersten Bedingung von Algorithmus 3.	A = Ich *ÿ* = *ÿ* B = T *ÿ* = *ÿ* Lm(I) **ÿ** Lm(T)	**2. Auswertung der ersten Bedingung von Algorithmus 3.**	A = Ich *ÿ* = *ÿ* B = T *ÿ* **=,** ich Lm(I) = Lm(T)

Zusammenfassung der äußersten linken **Mengen von Nichtterminals:**

- Ganz links (S) = Lm(D).
- Ganz links (D) = Lm(T).
- Ganz links(T) = {int}.
- Ganz links (L) = Lm (I).
- Ganz links(I) = Lm(T).

Ersetzen der Sets:

- Ganz links(S) = {int}.
- Ganz links(D) = {int}.
- Ganz links(T) = {int}.
- Ganz links (L) = {int}.
- Ganz links(I) = {int}.

5.2.5. Fallstudien

Für Algorithmus 5 wird die Anwendung einiger Vorrangbeziehungen gezeigt.

1. Für das Trennzeichen $:

(a) $ ÿ L(S) **ÿÿ Regel** (A.1).

- **$** ÿ ;
- $ ÿ Ich würde
- **$** ÿ **ÿ**
- **$** ÿ int

(b) R(S) ÿ **$ ÿÿ Regel** (A.2).

- **;** ÿ **$**

2. Für S **ÿÿ** SD;

Machine Translated by Google

(a) X = **S,** Y = D **ÿÿ Regel** (E.1).

R(S) ÿ Lm(D)

- **;** ÿ int

u = null, l = [ ]

(b) X = **D,** Y = ; **ÿÿ Regel** (D.1).

R(D) ÿ ;

-) ÿ ;

u = null, l = [ ]

3. Für D **ÿÿ** Tid(L)

(a) X = id, Y = (ÿÿ **Regel** (B).

- Ich würde .= (

u = null, l = [ ]

(b) X = **(,** Y = L **ÿÿ Regel** (C).

(ÿ L(L)

- (ÿ **,**
- (ÿ **ÿ**
- (ÿ int

u = **(,** l = [ ]

(c) X = **L,** Y =) **ÿÿ Regel** (D.1).

R(L) ÿ)

- **,** ÿ)
- **ÿ** ÿ)
- int ÿ)

u ÿ= null **ÿÿ** u .= b **ÿÿ** (.=) **ÿÿ Regel** (D.2).

u = null, l = [ ]

4. In folgenden Fällen ist die Form XY **im rechten Teil der Produktion nicht erfüllt , deshalb kann keine Regel angewendet werden:**

(a) T **ÿÿ** int

(B) L **ÿÿ** Ich

(C) L **ÿÿ** *ÿ*

(D) Ich **ÿÿ** T

Die verbleibenden Vorrangbeziehungen werden mit der gleichen Methode wie in den zuvor gezeigten Fällen abgeleitet.

5.2.6. Operator-Präzedenztabelle

Am Ende der Algorithmusausführung wird die Operatorpräzedenztabelle 11 abgerufen.

Tabelle 11. Vorrangtabelle von Beispiel 2.

	;	Auswais	(	)	ÿ	int	**,**	**$**
:						ÿ		ÿ
			.=					
Auswais (				.=	ÿ	ÿ	ÿ	
)	ÿ							
ÿ		ÿ		ÿ	ÿ		ÿ	
int		ÿ		ÿ	ÿ		ÿ	
,				ÿ	ÿ	ÿ	ÿ	
$	ÿ	ÿ			ÿ	ÿ		

Machine Translated by Google

5.2.7. Bottom-Up-Analyse

Sei int id(); int id(int, int); **ein String, der durch Ableitung von rechts erhalten wurde, und dann Algorithmus 1 führt zur Bottom-Up-Analyse, die in Tabelle 12 detailliert beschrieben wird.**

Tabelle 12. Bottom-Up-Analyse von Beispiel 2.

Stapel	Eingabe	Aktion
$	int id(); int id(int, int); **$** id(); int	**Schicht**
$int	id(int, int); **$** id(); int id(int,	**Reduzieren**
$	int); **$** (); int id(int, int); **$);**	**Schicht**
$id	int id(int, int); $; **int** id	**Schicht**
$ich	(int, int); **$** int id(int, int);	**Schicht**
würde($ich	**$** int id(int, int); **$** id(int,	**Reduzieren**
würde() $	int); **$** id(int, int); **$** (int,	**Schicht**
$;	int); **$** int, int) ; **$,**	**Reduzieren**
$	**int);** $, int); **$** int) ;	**Schicht**
$int	**$) ; $) ; $; $**	**Reduzieren**
$		**Schicht**
$id		**Schicht**
		Schicht
$id($id(int		**Reduzieren**
		Schicht
		Schicht
$id($id(,		**Reduzieren**
$id(,		**Reduzieren**
int		**Schicht**
$id(, $id($id()		**Reduzieren**
$	**; $**	**Schicht**
$;	**$**	**Reduzieren**
$	**$**	**Akzeptieren**

5.3. Beispiel 3

5.3.1. Grammatik

Angesichts des folgenden Nicht-OPG:
S **ÿÿ** ABC

A **ÿÿ** aA **|** ein

B *ÿ* **ÿ** bB **|** b **|**

C **ÿÿ** CDc **|** c
D **ÿÿ** d

Die Sets werden Schritt für Schritt gemäß den Algorithmen aufgebaut.

5.3.2. Links

Die linken **Sätze werden wie in Tabelle 13 beschrieben berechnet .**

Machine Translated by Google

Tabelle 13. Berechnung der linken **Mengen für die CFG in Beispiel 3.**

Links (S)			
Für S **ÿÿ** ABC			
1. Auswerten der ersten Bedingung von Algorithmus 2.	A = S *ÿ* = *ÿ* B = A *ÿ* = BC L(S) **ÿ** L(A)	**2. Auswerten der ersten Bedingung von Algorithmus 2.**	A = S *ÿ* = A B = B *ÿ* = C L(S) **ÿ** L(A) **ÿ** L(B)
3. Auswerten der ersten Bedingung von Algorithmus 2.	EIN = S. *ÿ* = AB B = C *ÿ* = *ÿ* L(S) = L(A) **ÿ** L(B) **ÿ** L(C)		

Links (A)			
Für A **ÿÿ** aA			
1. Auswerten der zweiten Bedingung von Algorithmus 2.	A = A *ÿ* = *ÿ* ein = ein *ÿ* = A L(A) **ÿ** {a}	**2. Auswerten der ersten Bedingung von Algorithmus 2.**	A = A *ÿ* = ein B = A *ÿ* = *ÿ* L(A) **ÿ** {a} **ÿ** L(A)
Für A **ÿÿ** a			
3. Auswerten der zweiten Bedingung von Algorithmus 2.	A = A *ÿ* = *ÿ* ein = ein *ÿ* = *ÿ* L. (EIN) = {ein}		

Links (B)			
Für B **ÿÿ** bB			
1. Auswerten der zweiten Bedingung von Algorithmus 2.	A = B *ÿ* = *ÿ* a = b *ÿ* = B L(B) **ÿ** {b}	**2. Auswerten der ersten Bedingung von Algorithmus 2.**	A = B *ÿ* = b B = B *ÿ* = *ÿ* L(B) **ÿ** {b} **ÿ** L(B)
Für B **ÿÿ** b			
3. Auswerten der zweiten Bedingung von Algorithmus 2.	A = B *ÿ* = *ÿ* a = b *ÿ* = *ÿ* L(B) = {b}		

Machine Translated by Google

Tabelle 13. Fortsetzung

Links (C)			
Für C **ÿÿ** CDc			
1. Auswerten der ersten Bedingung von Algorithmus 2.	A = C *ÿ* = *ÿ* B = C *ÿ* = Dc L(C) **ÿ** L(C)	**2. Auswerten der ersten Bedingung von Algorithmus 2.**	A = C *ÿ* = C B = D *ÿ* = c L(C) **ÿ** L(D)
3. Auswerten der zweiten Bedingung von Algorithmus 2.	A = C *ÿ* = CD a = c *ÿ* = *ÿ* L(C) **ÿ** L(D) **ÿ** {c}		
Für C **ÿÿ** c			
4. Auswerten der zweiten Bedingung von Algorithmus 2.	A = C *ÿ* = *ÿ* a = c *ÿ* = *ÿ* L(C) = L(D) **ÿ** {c}		

Links (D)	
Für D **ÿÿ** d	
1. Auswerten der zweiten Bedingung von Algorithmus 2.	A = D *ÿ* = *ÿ* a = d *ÿ* = *ÿ* L. (D.) = {d}

Zusammenfassung der linken **Mengen von Nicht-Terminalen:** • Links(S) = L(A) **ÿ** L(B) **ÿ** L(C). • Links(A) = **{a}.** • Links(B) = **{b}.** • Links(C) = L(D) **ÿ {c}.** • Links(D) = **{d}.**

Ersetzen der Mengen: • Links(S) = {a, **b, d, c}.** • Links(A) = **{a}.** • Links(B) = **{b}.** • Links(C) = {d, **c}.** • Links(D) = **{d}.**

5.3.3. Rechts

Die rechten Sätze werden wie in Tabelle **14** **beschrieben berechnet .**

Machine Translated by Google

Tabelle 14. Berechnung der rechten **Sätze für die CFG in Beispiel 3.**

Rechte)			
Für S **ÿÿ** ABC			
1. Auswerten der ersten Bedingung von Algorithmus 4.	A = S *ÿ* = AB B = C *ÿ* = *ÿ* R(S) = R(C)		

Rechts (A)			
Für A **ÿÿ** aA			
1. Auswerten der ersten Bedingung von Algorithmus 4.	A = A *ÿ* = ein B = A *ÿ* = *ÿ* R(A) **ÿ** R(A)	**2. Auswerten der zweiten Bedingung von Algorithmus 4.**	A = A *ÿ* = *ÿ* ein = ein *ÿ* = A R(A) **ÿ** {a}
Für A **ÿÿ** a			
3. Auswerten der zweiten Bedingung von Algorithmus 4.	A = A *ÿ* = *ÿ* ein = ein *ÿ* = *ÿ* R(A) = {a}		

Rechts (B)			
Für B **ÿÿ** bB			
1. Auswerten der ersten Bedingung von Algorithmus 4.	A = B *ÿ* = b B = B *ÿ* = *ÿ* R(B) **ÿ** R(B)	**2. Auswerten der zweiten Bedingung von Algorithmus 4.**	A = B *ÿ* = *ÿ* a = b *ÿ* = B R(B) **ÿ** {b}
Für B **ÿÿ** b			
3. Auswerten der zweiten Bedingung von Algorithmus 4.	A = B *ÿ* = *ÿ* a = b *ÿ* = *ÿ* R(b) = {b}		

Rechts (C)			
Für C **ÿÿ** CDc		**Für** C **ÿÿ** c	
1. Auswerten der zweiten Bedingung von Algorithmus 4.	A = C *ÿ* = CD a = c *ÿ* = *ÿ* R(C) **ÿ** {c}	**2. Auswerten der zweiten Bedingung von Algorithmus 4.**	A = C *ÿ* = *ÿ* a = c *ÿ* = *ÿ* R(C) = {c}

Machine Translated by Google

Tabelle 14. Fortsetzung

Rechts (D)	
Für D **ÿÿ** d	
1. Auswerten der zweiten Bedingung von Algorithmus 4.	EIN = D. *ÿ* = *ÿ* a = d *ÿ* = *ÿ* R(D) = {d}

Zusammenfassung der rechten **Mengen von Nichtterminals:** • Rechts(S) = R(C). • Rechts(A) = **{a}.** • Rechts(B) = **{b}.** • Rechts(C) = **{c}.** • Rechts(D) = **{d}.**

Ersetzen der Mengen: • Rechts(S) = **{c}.** • Rechts(A) = **{a}.** • Rechts(B) = **{b}.** • Rechts(C) = **{c}.** • Rechts(D) = **{d}.**

5.3.4. Ganz links

Die ganz linken Sätze werden wie in Tabelle **15** **beschrieben berechnet .**

Tabelle 15. Berechnung der ganz linken **Mengen für die CFG in Beispiel 3.**

Ganz links (S)	
Für S **ÿÿ** ABC	
1. Auswerten der ersten Bedingung von Algorithmus 3.	A = S *ÿ* = *ÿ* B = A *ÿ* = BC Lm(S) = Lm(A)

Ganz links (A)			
Für A **ÿÿ** aA		**Für** A **ÿÿ** a	
1. Auswerten der zweiten Bedingung von Algorithmus 3.	A = A *ÿ* = *ÿ* ein = ein *ÿ* = A Lm(A) **ÿ** {a}	**2. Auswerten der zweiten Bedingung von Algorithmus 3.**	A = A *ÿ* = *ÿ* ein = ein *ÿ* = *ÿ* Lm(A) = {a}

Ganz links (B)			
Für B **ÿÿ** bB		**Für** B **ÿÿ** b	
1. Auswerten der zweiten Bedingung von Algorithmus 3.	A = B *ÿ* = *ÿ* a = b *ÿ* = B Lm(B) **ÿ** {b}	**2. Auswerten der zweiten Bedingung von Algorithmus 3.**	A = B *ÿ* = *ÿ* a = b *ÿ* = *ÿ* Lm(B) = {b}

Machine Translated by Google

Tabelle 15. Fortsetzung

Ganz links (C)			
Für C **ÿÿ** CDc		**Für** C **ÿÿ** c	
1. Auswerten der ersten Bedingung von Algorithmus 3.	A = C *ÿ* = *ÿ* B = C *ÿ* = Dc Lm(C) **ÿ** Lm(C)	**2. Auswerten der zweiten Bedingung von Algorithmus 3.**	A = C *ÿ* = *ÿ* a = c *ÿ* = *ÿ* Lm(C) = {c}

Ganz links (D)	
Für D **ÿÿ** d	
1. Auswerten der zweiten Bedingung von Algorithmus 3.	EIN = D. *ÿ* = *ÿ* a = d *ÿ* = *ÿ* Lm(D) = {d}

Zusammenfassung der ganz links stehenden

Mengen von Nicht-Terminalen:

• Ganz links(S) = Lm(A). • Ganz links(A) = **{a}.** • Ganz links(B) = **{b}.** • Ganz links(C) = **{c}.** • Ganz links(D) = **{d}.**

Ersetzen der Mengen: •

Ganz links (S) = **{a}.** • Ganz links (A) = **{a}.** • Ganz links (B) = **{b}.** • Ganz links (C) = **{c}.** • Ganz links (D) = **{d}.**

5.3.5. Fallstudien

Für Algorithmus 5 wird die Anwendung einiger Vorrangbeziehungen gezeigt.

1. Für das Trennzeichen $:

(a) $ ÿ L(S) **ÿÿ Regel** (A.1). **$** ÿ a **$** ÿ
- b **$** ÿ d
- **$** ÿ c
-
-

(b) R(S) ÿ **$ ÿÿ Regel** (A.2). c ÿ **$**
-

2. Für S **ÿÿ** ABC

(a) X = **A,** Y = B **ÿÿ Regel** (E.1).
R(A) ÿ Lm(B) a ÿ b
-

u = null, l = [A] **(b)** X
= **B,** Y = C **ÿÿ Regel** (E.1).
R(B) ÿ Lm(C) b ÿ c
-

Machine Translated by Google

A **ÿ** l ÿ= [ ] **ÿÿ Regel** (E.3).

R(A) ÿ Lm(C)

- a ÿ c

u = null, l = [ ]

3. Für A **ÿÿ** aA

(a) X = **a,** Y = A **ÿÿ Regel** (C).

a ÿ L(A)

- ein ÿ ein

u = **a,** l = [ ]

4. Für B **ÿÿ** bB

(a) X = **b,** Y = B **ÿÿ Regel** (C).

b ÿ L(B)

- b ÿ b

u = **b,** l = [ ]

5. Für C **ÿÿ** CDc

(a) X = **C,** Y = D **ÿÿ Regel** (E.1).

R(C) ÿ Lm(D)

- c ÿ d

u = null, l = [ ]

(b) X = **D,** Y = c **ÿÿ Regel** (D.1).

R(D) ÿ c

• d ÿ c

u = null, l = [ ]

6. In folgenden Fällen ist die Form XY **im rechten Teil der Produktion nicht erfüllt , deshalb kann keine Regel angewendet werden:**

a) A **ÿÿ** a
(B) B **ÿÿ** b
(C) B **ÿÿ** *ÿ*
(d) C **ÿÿ** c
(e) D **ÿÿ** d

Die verbleibenden Vorrangbeziehungen werden mit der gleichen Methode wie in den zuvor gezeigten Fällen abgeleitet.

5.3.6. Operator-Präzedenztabelle

Am Ende der Algorithmusausführung wird die Operatorpräzedenztabelle 16 abgerufen.

Tabelle 16. Vorrangtabelle von Beispiel 3.

	A	B	C	D	**$**
A	ÿ	ÿ	ÿ		
B		ÿ	ÿ		
C				ÿ	ÿ
d			ÿ		
$	ÿ	ÿ	ÿ	ÿ	

5.3.7. Bottom-Up-Analyse

Sei aabbcdc **eine Zeichenfolge, die durch Ableitung ganz rechts erhalten wurde. Dann wird Algorithmus 1** angewendet **Ergebnisse der Bottom-Up-Analyse, detailliert in Tabelle 17.**

Machine Translated by Google

Tabelle 17. Bottom-Up-Parsing von Beispiel 3.

Stapel	Eingang	Aktion
$	aabbcdc$	**Schicht**
$ein	abbcdc$	**Schicht**
$aa	bbcdc$	**Reduzieren**
$ein	bbcdc$	**Reduzieren**
$	bbcdc$	**Schicht**
$b	bcdc$	**Schicht**

Tabelle 17. Fortsetzung

Stapel	Eingabe	Aktion
$bb	cdc$	**Reduzieren**
$b	CDC$	**Reduzieren**
$	CDC$	**Schicht**
$c	dc$	**Reduzieren**
$	dc$	**Schicht**
$d	**c$**	**Reduzieren**
$	**c$**	**Schicht**
$c	**$**	**Reduzieren**
$	**$**	**Akzeptieren**

5.4. Diskussion

In den Beispielen werden drei CFGs betrachtet: ein OPG und zwei Nicht-OPGs.
ersten CFG wird festgestellt, dass es der Grundregel entspricht, während die folgenden
nicht, da sie Produktionen mit aufeinanderfolgenden nichtterminalen Symbolen auf der rechten Seite darstellen
Seite. Unabhängig von der Art der CFG spielen die vorgeschlagenen Algorithmen eine entscheidende Rolle bei
Abrufen der Operatorpräzedenztabelle. Mit Algorithmus 1, angewendet auf eine bestimmte Zeichenfolge für
In jedem CFG wird eine Bottom-Up-Analyse durchgeführt, um zu überprüfen, ob die Zeichenfolgen generiert werden von
ihre jeweiligen Grammatiken. Dieser Prozess unterstreicht die Funktionalität des Vorschlags,
Führen Sie eine Bottom-Up-Analyse basierend auf der Operatorpriorität durch, auch wenn die CFGs dies nicht tun.
erfüllen die Grundvoraussetzung der OPGs.

6. Ausnahmen

Wenn die CFG eine der folgenden Formen hat, gelten die in den Algorithmen dieser
Papier würde nicht gelten:

1. A **ÿÿ** aBb

B **ÿÿ** ÿa

Die Sets für B **in der zweiten Produktion sind:**
: L(B) = **{a}.**
: R(B) = **{a}.**

Die auf den Abschnitten von Algorithmus **5 basierenden Vorrangbeziehungen weisen Folgendes auf:**

- **Für die Regel** (C): a ÿ a **und** u = **a.**
- **Für die Regel** (D.1): a ÿ **b.**
- **Für die Regel** (D.2): a .= **b.**

Die Anwendung der Regeln (D.1) **und** (D.2) **führt zum Konflikt.**

Machine Translated by Google

2. A **ÿÿ** aBb B **ÿÿ** *aÿ*

Sei *ÿ* **=ÿ** *ÿ*

ÿ *ÿ* **=ÿ** ÿa.

Gemäß den festgelegten Bedingungen sind die Mengen für B **in der zweiten Produktion** L(B) = **{a}.**

: R(B) = **{a}.**

Die auf den Abschnitten von Algorithmus **5 basierenden Vorrangbeziehungen ergeben:**

• Für die Regel (C): a ÿ a **und** u = **a. • Für die**

Regel (D.1): a ÿ **b. • Für die Regel**

(D.2): a .= **b.**

Die Anwendung der Regeln (D.1) **und** (D.2) **führt zum Konflikt.**

Wenn R(B) **kein a enthält , dann tritt der Konflikt nicht auf.**

3. A **ÿÿ** bBa B **ÿÿ** *aÿ*

Die

Mengen für B **in der zweiten Produktion sind:** L(B)

= **{a}.**

: R(B) = **{a}.**

Die auf den Abschnitten von Algorithmus **5 basierenden Vorrangbeziehungen ergeben:**

• Für die Regel (C): b ÿ a **und** u = **b. • Für die**

Regel (D.1): a ÿ **a. • Für die Regel**

(D.2): b .= **a.**

Die Anwendung der Regeln (C) **und** (D.2) **führt zum Konflikt.**

4. A **ÿÿ** bBa

B **ÿÿ** ÿa

Sei *ÿ* **=ÿ** *ÿ* **ÿ** *ÿ* **=ÿ** *aÿ.*

Gemäß den festgelegten Bedingungen lauten die Mengen für B **in der zweiten Produktion:** L(B) = **{a}.**

: R(B) = **{a}.**

Die auf den Abschnitten von Algorithmus **5 basierenden Vorrangbeziehungen ergeben:**

• Für die Regel (C): b ÿ a **und** u = **b. • Für die**

Regel (D.1): a ÿ **a. • Für die Regel**

(D.2): b .= **a.**

Die Anwendung der Regeln (C) **und** (D.2) **führt zum Konflikt.**

Wenn L(B) **kein a enthält , dann tritt der Konflikt nicht auf.**

Wenn CFG-Produktionen die folgende Form annehmen, gibt es keine Konflikte:
A **ÿÿ** aBb

B **ÿÿ** *ÿ* ,

wobei *ÿ* weder a **am Ende der rechten Seite noch** b **am Ende der linken Seite hat .**

7. Schlussfolgerungen

Frühere Arbeiten befassten sich mit der Anwendung von OPGs zur Lösung von Problemen im Zusammenhang mit der Analyse verschiedener Sprachformen. Dabei wurde allerdings weder viel Wert auf die Definition der linken und rechten Mengen noch auf die Generierung eines Algorithmus zum Ermitteln der Operatorpräzedenztabelle gelegt.

In dieser Arbeit haben wir die zuvor von anderen Autoren vorgestellten Mengen Right und Left neu definiert und eine neue Menge namens Leftmost eingeführt. Diese Mengen bilden die Grundlage für einen Algorithmus, der die Operatorpräzedenztabelle erstellt. Jede Menge wird mithilfe eines Algorithmus erstellt, der ihre Beschaffung erleichtert und der bestehenden Forschung eine neue Dimension hinzufügt.

Der vorgeschlagene Algorithmus zum Erstellen der Präzedenztabelle bricht zwar (mit einigen Ausnahmen) die grundlegende Definition eines OPG, eröffnet aber neue Möglichkeiten. Er ermöglicht das Erstellen einer Operatorpräzedenztabelle aus einer Nicht-OPG-CFG und ebnet damit den Weg für einen Bottom-up-Analysealgorithmus zum Erkennen einer von der CFG generierten Zeichenfolge.

Machine Translated by Google

Es werden drei Beispiele für CFGs gezeigt: ein OPG und zwei Nicht-OPGs. Die vorgeschlagenen Algorithmen werden systematisch Schritt für Schritt auf diese Beispiele angewendet, bis die Operatorpräzedenztabellen vorliegen, unabhängig von der Art des CFGs. Darüber hinaus wird der Bottom-up-Parsing-Algorithmus auf bestimmte Zeichenfolgen für jede Grammatik angewendet, um zu verifizieren , dass sie von ihnen generiert wurden.

Es ist wichtig zu beachten, dass der Algorithmus und die Definition der neuen Terminalsymbolsätze zwar bedeutende Fortschritte darstellen, aber auch Einschränkungen haben. Es ist nicht immer möglich, alle Fälle zu lösen, in denen sich zwei oder mehr benachbarte Terminalsymbole auf der rechten Seite befinden. Es gibt Ausnahmen, und wir müssen uns dieser bewuss

Beiträge der Autoren: Konzeptualisierung, LL und JM; Methodik, JM; Software, LL; Validierung, LL und JM; formale Analyse, LL, EA und JM; Untersuchung, LL und JM; Ressourcen, JM; Datenkuratierung , EA; Schreiben – Vorbereitung des Originalentwurfs, EA und JM; Schreiben – Überprüfung und Bearbeitung, EA und JM; Visualisierung, EA; Überwachung, JM; Projektverwaltung, JM; Mittelbeschaffung, JM; Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ih

Finanzierung: Diese Forschung erhielt keine externe Finanzierung und das APC wurde von der Universidad del Norte finanziert.

Datenverfügbarkeitserklärung: Die Datenfreigabe trifft auf diesen Artikel nicht zu.

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte vorliegen.

Verweise

1. **Floyd, RW Syntaktische Analyse und Operatorpräzedenz.** J. ACM (JACM) **1963,** 10, **316–333.** [CrossRef] **2.**
   Alfred, V.; Monica, S.; Ravi, S.; Jeffrey, DU -Compiler – Prinzipien und Techniken; **Pearson Education: London, Großbritannien, 2007.**
3. **Grune, D.; Jacobs, CJH** Parsing Techniques (Monographien in Informatik); **Springer: Berlin, Heidelberg, 2006.**
4. **Crespi-Reghizzi, S. Ein effektives Modell für Grammatik-Inferenz. In** Information Processing 71; **Gilchrist, B., Hrsg.; Elsevier: North- Holland, New York, 1972; S. 524–529.**
5. **Crespi-Reghizzi, S. Reduktion der Aufzählung beim Grammatikerwerb. In Proceedings der 2. Internationalen Gemeinsamen Konferenz on Artificial Intelligence, San Francisco, CA, USA, 1.–3. September 1971; S. 546–552.**
6. **Crespi-Reghizzi, S.; Melkanoff, MA; Lichten, L. Die Verwendung grammatischer Inferenz für die Entwicklung von Programmiersprachen.** Commun. ACM **1973,** 16, **83–90.** [CrossRef]
7. **von Braunmühl, B.; Verbeek, R. Input Driven Languages are Recognized in log n Space. In** Topics in the Theory of Computation; **North-Holland Mathematics Studies; Karplnski, M., van Leeuwen, J., Hrsg.; Elsevier: North-Holland, Niederlande, 1985; Band 102, Seiten 1–19.**
8. **Mehlhorn, K. Pebbling-Gebirgszüge und ihre Anwendung zur DCFL-Erkennung. In** Automata, Languages and Programming; **de Bakker, J., van Leeuwen, J., Hrsg.; Springer: Berlin/Heidelberg, Deutschland, 1980; S. 422–435.**
9. Mandrioli, D.; Pradella, **M. Generalisierung eingabegesteuerter Sprachen: Theoretische und praktische Vorteile .** 27, **61–87.** [CrossRef]
10. **Alur, R.; Madhusudan, P. Hinzufügen einer Verschachtelungsstruktur zu Wörtern.** J. ACM **2009,** 56, **1–43.** [CrossRef]
11. **Crespi Reghizzi, S.; Mandrioli, D. Operatorpräzedenz und die sichtbare Pushdown-Eigenschaft.** J. Comput. Syst. Sci. **2012,** 78, **1837–1867.** [CrossRef]
12. **Crespi Reghizzi, S.; Pradella, M. Mehr als Operatorpräzedenzgrammatiken und -sprachen.** J. Comput. Syst. Sci. **2020,** 113, **18–41.** [Querverweis]
13. **Barenghi, A.; Crespi Reghizzi, S.; Mandrioli, D.; Panella, F.; Pradella, M. Paralleles Parsen praktisch gemacht.** Wissenschaft. Berechnen. Programm. **2015,** 112, **195–226.** [CrossRef]
14. **Crespi-Reghizzi, S.; Mandrioli, D.; Martin, DF Algebraische Eigenschaften von Operatorpräzedenzsprachen.** Inf. Control **1978,** 37, **115–133.** [CrossRef]
15. **Chomsky, N. Drei Modelle zur Beschreibung von Sprache.** IRE Trans. Inf. Theory **1956, 2, 113–124.** [CrossRef]
16. **Aho, AV; Ullman, JD** Die Theorie des Parsing, Übersetzens und Kompilierens; **Prentice-Hall Englewood: Cliffs, NJ, USA, 1973; Band 1.**
17. **Louden Kenneth, C. Compilerkonstruktion: Prinzipien und Praxis. In** Course Technology; **PWS Publishing Co.: Montgomery, IL, USA, 1997.**

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Artikel

EEG-Kanalauswahl für die Rehabilitation von Schlaganfallpatienten mit BAT-Optimierer

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] und Christoph Guger [6]

[1] Forschungszentrum für künstliche Intelligenz (AIRC), College of Engineering and Information Technology, Ajman University, Ajman PO Box 346, Vereinigte Arabische Emirate; sharif.makhadmeh@uop.edu.jo
[2] Forschungs- und Entwicklungszentrum für Informationstechnologie (ITRDC), Universität Kufa, Nadschaf 54001, Irak; nabeel@uokufa.edu.iq
[3] College of Engineering, Universität Warith Al-Anbiyaa, Karbala 56001, Irak Abteilung
[4] für Biomedizintechnik, AL-Khwarizmi College of Engineering, Universität Bagdad, Bagdad 47146, Irak; noorbme@kecbu.uobaghdad.edu.iq Abteilung für
[5] Datenwissenschaft und künstliche Intelligenz, Fakultät für Informationstechnologie, Universität Petra, Amman 1196, Jordanien G.Tec Medical
[6] Engineering GmbH, 4521 Schiedlberg, Österreich; guger@gtec.at * Korrespondenz: m.albetar@ajman.ac.ae (MAA-B.); zaid.alyasseri@uokufa.edu.iq (ZAA)

Zusammenfassung: Schlaganfälle sind weltweit eine der häufigsten Todesursachen. Sie stören den zerebralen Blutfluss und führen zu schweren Hirnschäden. Hemiplegie, eine häufige Folge, führt zu einem Verlust der motorischen Funktion auf einer Körperseite . Viele Schlaganfallpatienten leiden unter langfristigen motorischen Beeinträchtigungen und benötigen eine umfa Elektroenzephalogramme (EEGs) bieten eine nichtinvasive Methode zur Überwachung der Gehirnaktivität und werden in Gehirn-Computer-Schnittstellen (BCIs) zur Unterstützung der Rehabilitation eingesetzt. Motorische Imagery-Aufgaben (MI), die durch EEGs erfasst werden, sind von entscheidender Bedeutung für die Entwicklung von BCIs, die Patienten dabei helfen, ihre motorischen Fähigke Die Interpretation von EEG-Signalen für MI-Aufgaben bleibt jedoch aufgrund ihrer Komplexität und ihres niedrigen Signal-Rausch-Verhältnisses eine Herausforderung. Das Hauptziel dieser Studie ist die Optimierung der Kanalauswahl in EEG-basierten BCIs speziell für die Schlaganfallrehabilitation. Die Bestimmung der informativsten EEG- Kanäle ist entscheidend für die Erfassung der neuronalen Signale im Zusammenhang mit motorischen Beeinträchtigungen bei Schlaganfallpatienten. In diesem Artikel wird eine auf dem binären Bat-Algorithmus (BA) basierende Optimierungsmethode vorgeschlagen, um die relevantesten Kanäle auszuwählen, die auf die einzigartigen neurophysiologischen Veränderungen bei Schlaganfallpatienten zugeschnitten sind. Dieser Ansatz kann die BCI-Leistung verbessern, indem er die Klassifizierungsgenauigkeit verbessert und die Datendimensionalität reduziert. Wir verwenden Zeit-Entropie-Frequenz-Attribute (TEF), die durch eine automatisierte unabhängige Komponentenanalyse mit Wavelet-Transformation (AICA-WT)-Rauschunterdrückung verarbeitet werden, um die Signalklarheit zu verbessern . Die ausgewählten Kanäle und Funktionen werden durch einen K-Nearest-Neighbor-Klassifikator (KNN) unter Verwendung öffentlicher BCI-Datensätze nachgewiesen, was eine verbesserte Klassifizierung von MI-Aufgaben und das Potenzial für bessere Rehabilitationsergebnisse demonstriert.

Schlüsselwörter: EEG; Merkmalsextraktion; Kanalauswahl; Schlaganfallpatienten; BAT-Algorithmus

check for updates

Zitat: Al-Betar, MA; Alyasseri, ZAA; Al-Qazzaz, NK; Makhadmeh, SN; Ali, NS; Guger, C. EEG-Kanal Auswahl für Schlaganfallpatienten Rehabilitation mit BAT Optimizer. Algorithmen **2024,** 17, **346. https://doi.org/10.3390/a17080346**

Wissenschaftliche Redakteurin: Gabriella Trucco

Empfangen: 21. Mai 2024
Überarbeitet: 14. Juli 2024
Akzeptiert: 15. Juli 2024
Veröffentlicht: 8. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/4.0/).

1. Einleitung

Die Weltgesundheitsorganisation (WHO) nennt Schlaganfälle als zweithäufigste Todesursache weltweit, da sie durch Störungen der Hirndurchblutung Schäden am Gehirn verursachen **[1,2]. Eine der schwerwiegendsten Formen von Nervenschädigungen ist die Hemiplegie, die eine teilweise oder vollständige Lähmung einer Körperhälfte einschließlich Arm, Bein, Fuß und Hand beschreibt. Der ischämische Schlaganfall, die häufigste Art von Schlaganfall, führt zu einer Unterbrechung der Hirndurchblutung, was zu einem raschen Verlust der Hirnfunktion führt [3,4]. 60 Prozent oder mehr der Schlaganfallpatienten benötigen aufgrund dauerhafter motorischer Funktionsstörungen eine Rehabilitation [5]. Schlaganfallpatienten leiden jedoch unter einer Vielzahl von Behinderungen, darunter Seh- und kog**

Machine Translated by Google

selbst die einfachsten Aufgaben zu erledigen. Deshalb steht die Forschung zur effizienten Behandlung und Rehabilitation von Schlaganfallpatienten schon seit vielen Jahren im Fokus [6].

Ein Elektroenzephalogramm (EEG) ist eine nichtinvasive Methode, die die elektrische Aktivität des Gehirns mit guter Zeitauflösung schätzt, um die Ergebnisse einer Vielzahl von mentalen Aufgaben zu definieren. Indem sie EEG-Signale des Gehirns in Steueranweisungen übersetzen, helfen Gehirn-Computer-Schnittstellen (BCIs) Menschen mit körperlichen Beeinträchtigungen, mit der sensorischen Außenwelt zu interagieren [7]. Motorische Imagery (MI) ist die mentale Vorstellung von Körperbewegungen ohne echte Muskelbewegung, und die damit einhergehenden rhythmischen Aktivitäten des Gehirns können aufgezeichnet und als Eingangssignale von BCI-Systemen verwendet werden. Ereignisbezogene Desynchronisation /Synchronisation (ERD/ERS) bezieht sich auf die Erkennung der rhythmischen Leistungsänderungen im sensorischen Motorbereich innerhalb der rolandischen Frequenzbänder mu (9–13 Hz) und beta (13–30 Hz) , die zur Unterscheidung zwischen verschiedenen Arten von MI-Aufgaben verwendet werden können [8]. Da MI keine äußeren Reize erfordert, haben auf MI basierende BCIs eine breite Palette potenzieller Anwendungen [9].

Die Volatilität und Komplexität von EEG-Signalen bereiten jedoch Schwierigkeiten bei der konventionellen MI-BCI-Dekodierung [7]. Entscheidende und schwierige Probleme verbleiben bei der Verbesserung der EEG- Dekodierkapazität, der Extraktion diskriminierender Informationen aus EEGs mit niedrigem Signal-Rausch-Verhältnis und der Realisierung einer zuverlässigen Klassifizierung verschiedener MI-Aufgabe

Die Kanalauswahl bei EEG-basierten Untersuchungen ist wichtig, um die Teilmenge an Kanälen zu bestimmen, die im Hinblick auf die Erfassung der richtigen neuronalen Signale für die gewählte Anwendung am aussagekräftigsten ist [10]. Dennoch können solche Überlegungen und die damit verbundenen Herausforderungen in anderen Anwendungskontexten anders sein, wie bei einem Vergleich der Schlaganfallrehabilitation mit der Emotionsverarbeitung und Demenzerkennung anhand eines EEG-basierten Datensatzes deutlich wird [11]. Beispielsweise ist die Schlaganfallrehabilitation dadurch gekennzeichnet, dass es bei neurologischen Defekten infolge des Schlaganfallereignisses zu bestimmten Veränderungen der Gehirnaktivität und der Funktion der neuronalen Netzwerke kommt. EEG-Daten können zudem durch bestimmte Merkmale charakterisiert sein, die sie von denen gesunder Probanden oder anderer Patienten unterscheiden, wie z. B. Änderungen in der topologischen Organisation der Netzwerke, abweichende Aktivität und das Auftreten neuartiger Regulierungsstrategien [12].

Daher muss der Ansatz zur Auswahl der wichtigsten EEG-Kanäle für die Schlaganfallrehabilitation möglicherweise selektiver und spezifischer sein. Die Idee besteht darin, herauszufinden, welcher der Kanäle die neuronalen Eigenschaften der jeweiligen motorischen, kognitiven oder Verhaltensbeeinträchtigung, die ein Schlaganfallpatient aufweisen kann, am besten aufzeichnet, und es werden nicht dieselben Kanäle sein, die für andere Anwendungen ideal wären [13]. Es gibt jedoch unterschiedliche Ansätze wie die Verarbeitung von Emotionen und die Erkennung von Demenz, die unterschiedliche neurophysiologische Mechanismen und/oder spektrotemporale Muster verwenden können. Beispielsweise kann bei der Verarbeitung von Emotionen das mit der Verarbeitung von Emotionen verbundene Frequenzband über verschiedene Bereiche des Kortex verteilt sein, und das Ziel der Kanalauswahl könnte darin bestehen, diese verteilte Frequenzaktivität zu erfassen. In ähnlicher Weise könnte im Fall der Demenzerkennung die Veränderung der Gehirnstruktur und -funktion, die die kognitive Beeinträchtigung definiert, eine andere räumliche Belastungsverteilung darstellen, die andere Regeln für die Auswahl der Kanäle erfordert [14]. Mit einem direkten Fokus auf das Problem der Kanalauswahl im Zusammenhang mit der Schlaganfallrehabilitation soll die vorgeschlagene Studie eine wirksame Lösung liefern, die die Leistung eines BCI und damit den Rehabilitationsprozess verbessern kann.

Die Kanalauswahl ist der Hauptschwerpunkt dieser Studie. Viele Elektroden werden in medizinischen und diagnostischen Behandlungen eingesetzt. Für praktische BCI-Anwendungen werden viele Elektroden für eine genaue Klassifizierung benötigt; das Platzieren vieler Elektroden auf der Kopfhaut ist jedoch zeitaufwändig und kann zu einer Überanpassung des EEG-Signals führen. Um diese Probleme zu beheben, müssen die Forscher die Elektroden identifizieren, die nicht klassifiziert werden können. Die Elektrodenauswahl kann das neurologische Wissen von Personen aufgrund unterschiedlicher Reaktionen und der vom Patienten abhängigen idealen Elektrodenplatzierung bewerten [15]. Die benutzerabhängige Klassifizierung erfordert eine automatische Bestimmung der Elektrodenrelevanz. Diese Herausforderung kann überwunden werden, indem ein breites Spektrum an Elektroden berücksichtigt und viele Methoden verwendet werden, um die geeigneten Kanäle für jeden Patienten zu identifizieren. Eine auf Partikel-BA basierende

Machine Translated by Google

EEG-Kanäle für die Rehabilitation von Schlaganfallpatienten. Jeder EEG-Datenkanal bietet einzigartige Klassifizierungsfunktionen. Die Kanalauswahl reduziert die Datenspeicherung und -verarbeitung, beschleunigt das Klassifikatortraining durch die Verwendung einfacherer Daten und reduziert den „Fluch der Dimensionalität“.

In vielen Studien wurde die Kanalauswahl bei gesunden Menschen untersucht [16,17]. Die vorliegende Studie konzentriert sich jedoch auf die Schlaganfallrehabilitation, bei der die zugrunde liegenden neurophysiologischen Anomalien und Einschränkungen von Schlaganfallpatienten eine individuelle Strategie zur Kanalauswahl erfordern können.

Die Auswahl von EEG-Kanälen für die Verarbeitung von Emotionen, die kognitive Beurteilung und die klinische Diagnose erfordert häufig Kenntnisse über neuronale Mechanismen und Gehirnregionen, die mit den gewünschten Aktivitäten verbunden sind. Übliche Gehirnaktivitätsmuster und die optimale Kanalauswahl sind bei gesunden Menschen gut bekannt. Schlaganfälle können jedoch die Struktur, Konnektivität und Funktion des Gehirns verändern. Diese Veränderungen können abnormale Gehirnmuster und Kompensationsprozesse verursachen [18,19]. Daher können sich die besten EEG-Kanäle für die Schlaganfallrehabilitation von denen unterscheiden, die für gesunde Menschen empfohlen werden.

Um die neuronalen Korrelate motorischer, kognitiver und Verhaltensstörungen bei Schlaganfallpatienten zu erfassen, muss die Kanalauswahl aufgrund der durch den Schlaganfall verursachten Neuroplastizität spezialisierter und personalisierter sein. Die gezielte Behandlung von Schlaganfallpatienten mit einer maßgeschneiderten Lösung, die neuronale Fingerabdrücke und Gehirnaktivitätsmuster richtig erkennen kann, soll dazu beitragen, schlaganfallbedingte Defizite zu erkennen. Diese Phase ist entscheidend für die Entwicklung von BCI-basierten Schlaganfalltherapien, die die Rehabilitationsergebnisse verbessern [20].

Die Literatur zur Kanalauswahl bei gesunden Menschen kann hilfreich sein, aber diese Studie erkennt die Notwendigkeit an, die besonderen Probleme und Überlegungen von Schlaganfallpatienten zu berücksichtigen. Wissenschaftler untersuchten die Unterschiede bei der Kanalauswahl zwischen Schlaganfallpatienten und gesunden Menschen, um neurophysiologische Veränderungen und optimale BCI-Einstellungen für die Schlaganfallrehabili

Daher wird in dieser Studie eine auf dem binären Bat-Algorithmus (BA) basierende Optimierungsmethode verwendet, um die wichtigsten EEG-Kanäle für die Schlaganfallrehabilitation zu identifizieren. Somit ist diese Methode auf die Bedürfnisse von Schlaganfallpatienten zugeschnitten, indem sie die BCI-Technologie und die Effektivität der R

Um die oben genannten Probleme zu lösen, werden Roh-EEG-Daten segmentiert und mithilfe von Standardfiltern und der automatisierten unabhängigen Komponentenanalyse mit Wavelet-Transformation (AICA-WT) zur Rauschunterdrückung gefiltert. Anschließend **werden Zeit-Entropie-Frequenz-Attribute (TEF) erstellt, indem die effektiven Merkmale aus den Bereichen Zeit, Entropie und Frequenz kombiniert werden. Infolgedessen wird die BA-basierte Optimierungsmethode verwendet, um die TEF-Eigenschaften zu optimieren und die effizienten Kanäle auszuwählen, die die Stationarität und Belastbarkeit des Systems verbessern. Nachdem wir eine KNN-Klassifizierungstechnik trainiert haben, führen wir eine Reihe von Experimenten durch, um unser vorgeschlagenes Framework zu bewerten. Dabei verwenden wir 25 MI-basierte BCI-Sitzungen mit anschließenden Beurteilungsbesuchen, um die funktionellen Änderungen vor und nach der EEG- Neurorehabilitation anhand öffentlicher Datensätze aus dem BCI-Wettbewerb zu untersuchen. Die im AICA-WT-TEF-Chs-Framework für die MI-BCI-Klassifizierung verwendeten hybriden Zeit -Entropie-Frequenz-Attribute (TEF) sollen die zugrunde liegenden Informationen der Bereiche Zeit, Entropie und Frequenz effizient auf die Klassifizierungsleistung auswirken.**

Der in diesem Dokument vorgeschlagene Ansatz leistet folgende Beiträge: Zunächst wird untersucht, wie sich Zeit-Entropie-Frequenz-Variablen (TEF) auf die Klassifizierungsleistung auswirken.
Zweitens werden die effektiven EEG-Kanäle verwendet, die von einer BA-basierten Optimierungsmethode generiert werden, um Änderungen sowohl vor als auch nach der Rehabilitation zu erkennen. Drittens wird die Wirksamkeit des AICA-WT-TEF-Chs-Frameworks unter Verwendung verschiedener Klassifizierungsmodelle und modernster Techniken getestet. Durch die Eliminierung irrelevanter Dimensionen verringert die effektive Kanalauswahl die Datenmenge, die gespeichert und verarbeitet werden muss, beschleunigt das Klassifikatortraining durch die Verwendung einfacherer Daten und verhindert eine Überanpassung des Klassifikators durch die Verringerung der Auswirkungen des „Fluchs der Dimensionalität“.

2. Verwandte Werke

Insbesondere EEG-basierte BCIs weisen eine gute zeitliche Auflösung auf [21], aber die aufgezeichnete Wellenaktivität kann, abhängig von den aufgetretenen Artefakten, auf unterschiedliche Art und Weise verfälscht sein, sodass die Vorverarbeitung der Roh-EEG-Signale bei der Klassifizierung von MI-basierten Signalen von entscheidend

Machine Translated by Google

Meistens können diese Artefakte das abnormale Verhalten des Gehirns imitieren oder überlagern. Studien haben die Auswirkungen von Artefakten auf EEG-Signale untersucht, darunter Augenblinzeln, Augenbewegungen , Herzartefakte und Muskelaktivität. Auch Rauschen von Stromleitungen kann die Frequenzen eines EEGs überlagern [22,23]. Daher kann die Beurteilung der Effizienz von EEG-Signalen bei Hintergrundrauschen schwierig sein. Räumliche Filter können in dieser Situation ein lokalisierteres Signal für jede Elektrode erzeugen, da insbesondere EEG-basierte BCIs eine schlechte räumliche Auflösung haben.

Um die besten Charakteristiken für eine BCI-MI-basierte EEG-Analyse auszuwählen, wird für jeden aufgezeichneten Kanal in mehreren Domänen eine Merkmalsextraktion durchgeführt. Die gängigste Technik besteht darin, gemeinsame räumliche Muster (CSPs) zu verwenden, um diskriminante Merkmale aus hochdimensionalen EEG-Signalen zu erhalten, die verwendet werden können, um die räumlichen Charakteristiken im Zusammenhang mit ERD zu identifizieren, die durch verschiedene MI-Aufgaben erzeugt werden [24]. Die Effizienz von CSPs wird jedoch leicht durch Rauschen beeinträchtigt und reagiert äußerst empfindlich auf die Wahl der Frequenzbänder [16]. Um das Problem der Wirksamkeit zu lösen, wurden mehrere Techniken zur Merkmalsextraktion entwickelt, um die EEG-Daten in den Domänen Zeit, Frequenz und Zeit-Frequenz zu analysieren [25]. Zusätzlich wurden in [26] Wavelet-Kohärenz -Bispektrum-Charakteristiken (WC) zur Unterscheidung zwischen Rechts- und Linksherzinfarkt vorgeschlagen .

Die Größen der aus EEG-Signalen abgeleiteten Merkmale sind für jeden Kanal normalerweise ziemlich groß und werden größer, wenn man zum nächsten Kanal übergeht. Die Klassifizierung mehrerer Merkmale erfordert zusätzliche Berechnungen und Zeit. Um dieses Problem zu lösen, muss der effektivste Merkmalssatz durch Auswahl einer Teilmenge von EEG-Kanälen erhalten werden, die stärker mit geistiger Arbeit verbunden sind als andere. Die vier Bewertungsmethoden Filtern, Umhüllen, Einbetten und Hybrid können verwendet werden, um Merkmalsauswahlalgorithmen zu kategorisieren [27]. Basierend auf bestimmten statistischen Kriterien wird die Filtermethode berücksichtigt, die für motorische Absichten basierende EEG-Aktivitäten verwendet wird. Beispielsweise entwickelten He et al. [28] vor der Klassifizierung der motorischen Absichtsaktivität die statistische Methode der Bhattacharyya- Grenze für die Kanalauswahl. Die beste Kandidatenteilmenge wurde mithilfe dieser Methode ausgewählt, die auch eine CSP- und eine sequentielle Suchstrategie verwendete. Tam et al. [29] schlugen eine andere statistische Methode basierend auf dem CSP-Rang vor, um Filterkoeffizienten mit absoluten Werten zu sortieren und dann die Merkmale mit dem größten nachfolgenden Koeffizienten auszuwählen. Obwohl die Filtertechnik eine geringe Genauigkeit aufweist, arbeitet sie schnell und wird von der Auswahl des Subjekts oder Klassifikators nicht beeinflusst.

Andererseits sind der Klassifikator und das Subjekt im Wrapper- Auswahlprozess recht wichtig [30]. Auf diese Weise werden die Teilmengenkandidaten anhand der Klassifizierungsgenauigkeit bewertet und können somit genauere Vorhersagen treffen als Filtertechniken. Infolgedessen ist die Wrapper-Strategie rechnerisch aufwändiger als die Filtertechnik und unterliegt dem Überanpassungsproblem [31], das durch den Einsatz von Kreuzvalidierungsmaßnahmen zur Vorhersage vermieden werden kann. Zur Auswahl des besten Kanals verwenden die meisten Wrapper-Strategien die sequentielle Vorwärtssuche, die Rückwärtseliminierungsstrategie [32] und die heuristische Suchmethode [33,34]. Ohne eine Abbruchbedingung zu erwähnen, wurden die Filter- und Wrapping-Verfahren auch zu Auswahlzwecken gekoppelt. Diese Hybridstrategie wurde entwickelt, um mit riesigen Datensätzen umzugehen. Gaur et al. beispielsweise [35] boten die bestmögliche Kanalauswahltechnik basierend auf CSPs an. Bevor sie alle Kanäle nach ihren Punktzahlen ordneten und die Klassifizierungsgenauigkeit zur Bewertung der gewählten idealen Kanalkombination verwendeten, wendeten sie zunächst die L1-Norm eines CSP an, um die Kanalbeitragspunktzahlen zu gruppieren. Abschließend wurden mithilfe der definierten Kriterien für den Lernprozess eines bestimmten Klassifikators mithilfe des eingebetteten Ansatzes signifikante Kanäle ausgewählt.

Um die Kanäle mit den besten Klassifizierungsergebnissen auszuwählen, verwendeten Shi et al. [36] Merkmalsauswahltechniken, rekursive Merkmalsentfernung und Nullnormoptimierung, die mit Support Vector Machines trainiert wurden. Der Einsatz dieser Techniken ermöglichte eine deutliche Reduzierung der EEG-Kanäle [37].

Der entscheidende Schritt bei der Klassifizierung von EEG-Daten ist die Verwendung relevanter Merkmale, die die Fähigkeit des Systems zur effektiven Klassifizierung von Daten erheblich beeinflussen [38–40]. Um die beste Klassifizierungsleistung zu erzielen, ist es daher notwendig, gültige Merkmale aus EEG-Signale

Machine Translated by Google

Darüber hinaus können globale Optimierungstechniken und maschinelles Lernen kombiniert werden, um solche Probleme im BCI-Bereich zu lösen, einschließlich der Merkmalsextraktion und -auswahl [42]. Mehrere Algorithmen haben aufgrund ihrer Anwendung in der Merkmalsauswahl großes Interesse geweckt , beispielsweise Differential Evolution (DE), Simulated Annealing (SA), Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC) und Ant

Es wurde eine Untersuchung zur Kanalauswahl für verschiedene BCI-Paradigmen durchgeführt, darunter die rekursive Kanaleliminierung für Aufgaben zur motorischen Vorstellungskraft [44]. In einer Studie von Rakotomamonjy et al. [45] verwendeten die Forscher eine Kanalauswahltechnik , die in das Training eines SVM-Klassifikators integriert war. Der Merkmalsraum wurde mithilfe von GA in den Referenzen [46,47] durchsucht **und die Fitnessfunktion war die gewichtete lineare Kombination aus der Anzahl der Merkmale und der Genauigkeit des SVM. Udhaya et al. schlugen in [48] einen PSO-basierten Ansatz zur Auswahl von Rough-Set-Merkmalen vor, um die beste Teilmenge an Merkmalen zu bestimmen, und die Genauigkeit eines Nachbarschaftsklassifikators wurde als Bewertungskriterium für die Merkmalsteilmenge verwendet. Baig et al. [49]** verwendeten DE und **Atyabi et al. [50] PSO und GA, um jeweils die beste Merkmalsteilmenge zu finden. In [51]** wurden **bei der Auswahl von Merkmalen für die EEG-basierte Emotionserkennung Ameisenkolonieoptimierung , simuliertes Abkühlen, genetische Algorithmen, Partikelschwarmoptimierung und differenzielle Evolution eingesetzt. Die Leistung dieser EA-basierten Ansätze zur Merkmalsauswahl ist recht ermutigend.**

Die Größe des Suchbereichs ist bei diesen Algorithmen jedoch vorgegeben. Daher ist es wichtig, bei jeder Iteration eine Entscheidung darüber zu treffen, ob jedes Merkmal ausgewählt werden soll oder nicht. Unter diesen Umständen wird der Rechenaufwand für einige Merkmale verschwendet, die überflüssig, unnötig oder trivialer Natur sind. Darüber hinaus konvergieren evolutionäre Algorithmen (EAs) aufgrund des breiten Suchraums sowie der Interferenz redundanter, irrelevanter und trivialer Merkmale leicht zu einem lokalen Optimum. In dieser Studie entwickeln wir EA-basierte Merkmalsauswahlmethoden für motorische Bildgebung BCI durch Reduzierung der Dimensionalität.

Die Leistung eines Klassifikators nimmt mit zunehmender Merkmalsanzahl dramatisch ab. Grund dafür ist der Dimensionsfluch, der auftritt, wenn die Anzahl der Merkmale einen bestimmten Schwellenwert überschreitet. Darüber hinaus dauert das Training länger, je höher die Merkmalsdimensionalität wird [46].

Um eine große Anzahl vorhandener Kanäle zu lokalisieren, ist eine Merkmalsauswahl mit einer gründlichen Suche nach der Extraktion verwandter Kanäle erforderlich **[10,52]. Forscher haben sich auf Probleme der Merkmalsauswahl konzentriert und Ansätze auf Basis der vollständigen Suche, der gierigen Suche und der zufälligen Suche** vorgeschlagen **[11]. Hohe Rechenkosten und die Konvergenz zum lokalen Optimum bleiben bei diesen Auswahltechniken Probleme [53].**

3. BAT-Algorithmus

In diesem Abschnitt wird der Fledermaus-Optimierungsalgorithmus vorgestellt und im Hinblick auf seine Inspiration, die Fledermäuse und ihre Lebensweise in der Natur veranschaulicht , zusammen mit dem allgemeinen **Optimierungsverfahren und der mathematischen Formulierung der Methode .**

3.1. Inspiration für BAT **Der**

Fledermaus-Optimierungsalgorithmus wurde erstmals 2010 von Xin-She Yang vorgeschlagen [54]. Xin - She Yang schlug den Fledermaus-Algorithmus basierend auf den Eigenschaften der Fledermäuse vor: Echoortung, die zur Jagd auf Beute verwendet wird, und Bewegung im Dunkeln. Eine Fledermaus erzeugt extrem laute Schallimpulse und hört deren Echo von Objekten und Beute um sie herum. Anhand des Echos können Fledermäuse die Art der Beute und die Jagdstrategie bestimmen. Wenn Fledermäuse nach Beute suchen, erzeugen sie normalerweise 10 bis 20 Impulse pro Sekunde. Sobald sie sich der Beute nähern , erhöht sich die Anzahl der Impulse auf 200 Impulse pro Sekunde und der Schallimpuls wird leiser.

In der Natur haben Microbats ein kollektives Jagdverhalten, bei dem sie gemeinsam mit einer bestimmten Geschwindigkeit fliegen, um gleichzeitig Beute zu jagen, indem sie die Echoortungseigenschaft nutzen, wie in Abbildung 1 dargestellt. Die Abbildung zeigt die Bewegung von Microb

Machine Translated by Google

tionsmechanismus, bei dem die Microbats ihre Standorte aktualisieren und sich auf der Grundlage des Echos und des Standorts anderer Microbats, der durch Interaktion untereinander ermittelt wird, auf **die Beute zubewegen. Darüber hinaus wird die Geschwindigkeit der Microbats höher und die Impulse langsamer, wenn sich die Microbats in der Nähe des Standorts der Beute befinden.**

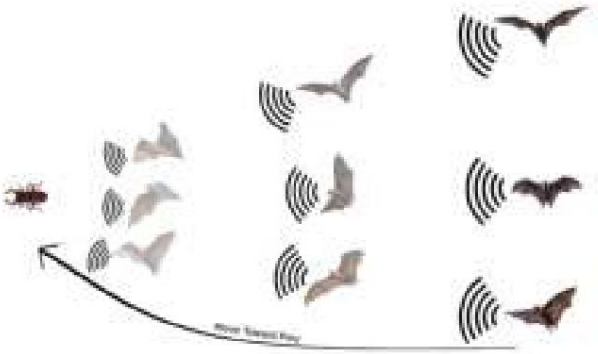


Abbildung 1. Fledermausbewegung in Richtung Beute.

Im BAT-Algorithmus wird jede Fledermaus als Lösung in der Population betrachtet, die Position der Beute ist die optimale Lösung für ein bestimmtes Optimierungsproblem, der Standort wird durch einen Vektor x **mit einem Bereich zwischen** fmin **und** fmax dargestellt, **die Geschwindigkeit der Fledermäuse wird durch einen Vektor v dargestellt und die Wellenlänge ist** *ÿ* **mit Pulsfrequenz** r ÿ **[0,** 1] **und Lautstärkebereich** A, wobei der größte Wert als Ao **und der niedrigste a**

3.2. Vorgehensweise

bei BAT **Dieser Abschnitt stellt die allgemeine Optimierungsprozedur des BAT-Algorithmus vor. Diese Optimierungsprozedur umfasst sechs Hauptschritte, die in Abbildung 2** dargestellt **und im Folgenden ausführlich erläutert werden:**

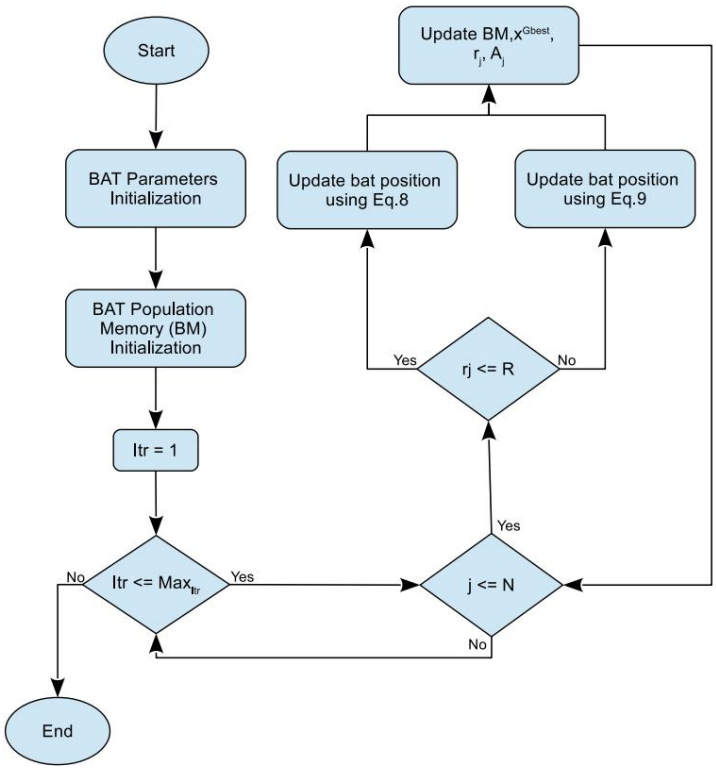


Abbildung 2. Flussdiagramm des Bat-Algorithmus.

- **Schritt 1: Initialisierung der BAT-Algorithmusparameter. In diesem Schritt werden die BAT-Parameter mit Anfangswerten initialisiert. Diese Parameter sind die Anzahl der Microbats**

Machine Translated by Google

(Lösungen) (N), maximale Anzahl von Iterationen (MaxItr), **maximaler** (fmax) **und minimaler** (fmin) **Frequenzbereich, Geschwindigkeitsvektor (v), Lautstärkerate (A), Pulsrate (r), anfängliche Pulsrate** (r-^{0}), **zwei Parameter** *(ÿ* **und ÿ) mit konstanten Werten im Bereich [0, 1] und Bandbreite Bereich** *(ÿ)* **im Bereich [ÿ1, 1].**

- **Schritt 2: Initialisierung des BAT-Populationsspeichers. Die BAT-Lösungen werden in diesem Schritt mit Hilfe von Gleichung (1) zufällig generiert, wobei der BAT-Algorithmus und besondere Parameter und Einschränkungen des Optimierungsproblems berücksichtigt werden.**

$$x_{ij} = lbi + (ubi \ddot{y} lbi) \times \mathbf{R}, \quad (1)$$

wobei ÿi = **1, 2, . . . , d,** ÿj = **1, 2, . . . , N und** R **eine Zufallszahl zwischen 0 und 1 ist.**

Die erzeugten Lösungen bilden zusammen die Population, wie in Gleichung (2) dargestellt, und werden auf der Grundlage ihrer Fitness Gbest **in aufsteigender Reihenfolge im BAT-Speicher (BM) abgelegt** und in den beste Lösung mit den fittesten Werten wird der ersten Position x **des BM zugewiesen.** Werten gespeichert. Die

$$BM = \begin{matrix} X^{1} & X^{1}_{2} & \cdots & X^{1}_{D} \\ \ddot{y}^{1}_{x\,1} & x^{2}_{2} & d\,2 \cdots & x\,d\,\ddot{y} \\ \vdots & \vdots & \cdots & \vdots \\ \ddot{y}x^{N}_{1} & x^{N}_{2} & \cdots & x^{N}_{D}\,\ddot{y} \end{matrix}, \quad (2)$$

- **Schritt 3: Intensivierung der Fledermauspopulation. Jetzt fliegen alle Microbats** x **und ändern ihre Position unter Berücksichtigung einer Geschwindigkeit** v , die **durch eine zufällig generierte** Frequenz f definiert ist, **wie in den Gleichungen (3) und (4) gezeigt. Dementsprechend werden die Fledermauspositionen gemäß Gleichung (5) aktualisiert.**

$$fi = fmin + (fmin \ddot{y} fmax) \times \mathbf{R}, \quad (3)$$

$$v^{'J}_{ich} = v^{jj}_{ich} + (x_{_} \ddot{y} x\,Gbest) \times fi, \quad (4)$$

$$x^{'J}_{ich} = xii, ^{J}_{j} + v^{\ddot{y}} \quad (5)$$

Es ist bemerkenswert, dass die Microbats ihre Standorte aktualisieren, um näher am globalen Bestwert (x Gbest) zu sein. **Somit verstärkt die neue Position der Microbats die Position in Richtung** Gbest **in Richtung** x **.**

- **Schritt 4: Diversifizierung der Fledermauspopulation.**

In diesem Schritt werden die Positionen der Microbats anhand von Zufallsparametern aktualisiert, um bessere globale Lösungen zu finden. Aus dem BM wird anhand einer Auswahlmethode eine neue Lösung ausgewählt und x Gbest **zugewiesen . Anschließend wird die aktualisiert mit** x ^{Gbest} **wie folgt:**

$$x^{'J}_{ich} = x^{Gbest} + \ddot{y} \times Aj, ^{\ddot{y}} \quad (6)$$

wobei A $^{\ddot{y}}$ **ist der Lautstärkedurchschnitt. Da die Schritte 3 und 4 widersprüchlich sind, wählt der BAT-Algorithmus jeweils einen davon zur Ausführung aus. Diese Auswahl basiert auf der folgenden Gleichung:**

$$x^{\ddot{y}j} = \begin{cases} x^{Gbest} + \ddot{y} \times A\,j^{\ddot{y}} & \text{wenn } rj \ddot{y} R \\ Xj + v\,\ddot{y}j & \text{ansonsten} \end{cases}, \quad (7)$$

- **Schritt 5: BM-Update. Die Position der Microbats im BM wird in diesem** Schritt aktualisiert , **wenn der neue Standort fitter ist als der alte und** R ÿ Aj . **Außerdem gilt** x

Machine Translated by Google

aktualisiert, wenn das neue BM eine Lösung mit einem besseren Fitnesswert enthält. Anschließend werden die rj- **und** Aj **-Werte gemäß den Gleichungen (8) und (9) aktualisiert.**

$$r_j = r_j^0 \left(1 - e^{(-\gamma \times itr)}\right), \quad (8)$$

$$A_j = \alpha \times A_j, \quad (9)$$

wobei itr **die aktuelle Iterationsnummer bezeichnet.**

- **Schritt 6: Abbruchkriterium. Die Schritte 3, 4 und 5 werden wiederholt, bis der Algorithmus das Abbruchkriterium.**

4. Vorgeschlagene

Methode Dieser Abschnitt stellt den Hauptbeitrag dieses Papiers dar. Die vorgeschlagene Methode umfasst fünf Phasen, die in Abbildung **3 dargestellt und im Folgenden ausführlich erläutert werden.**

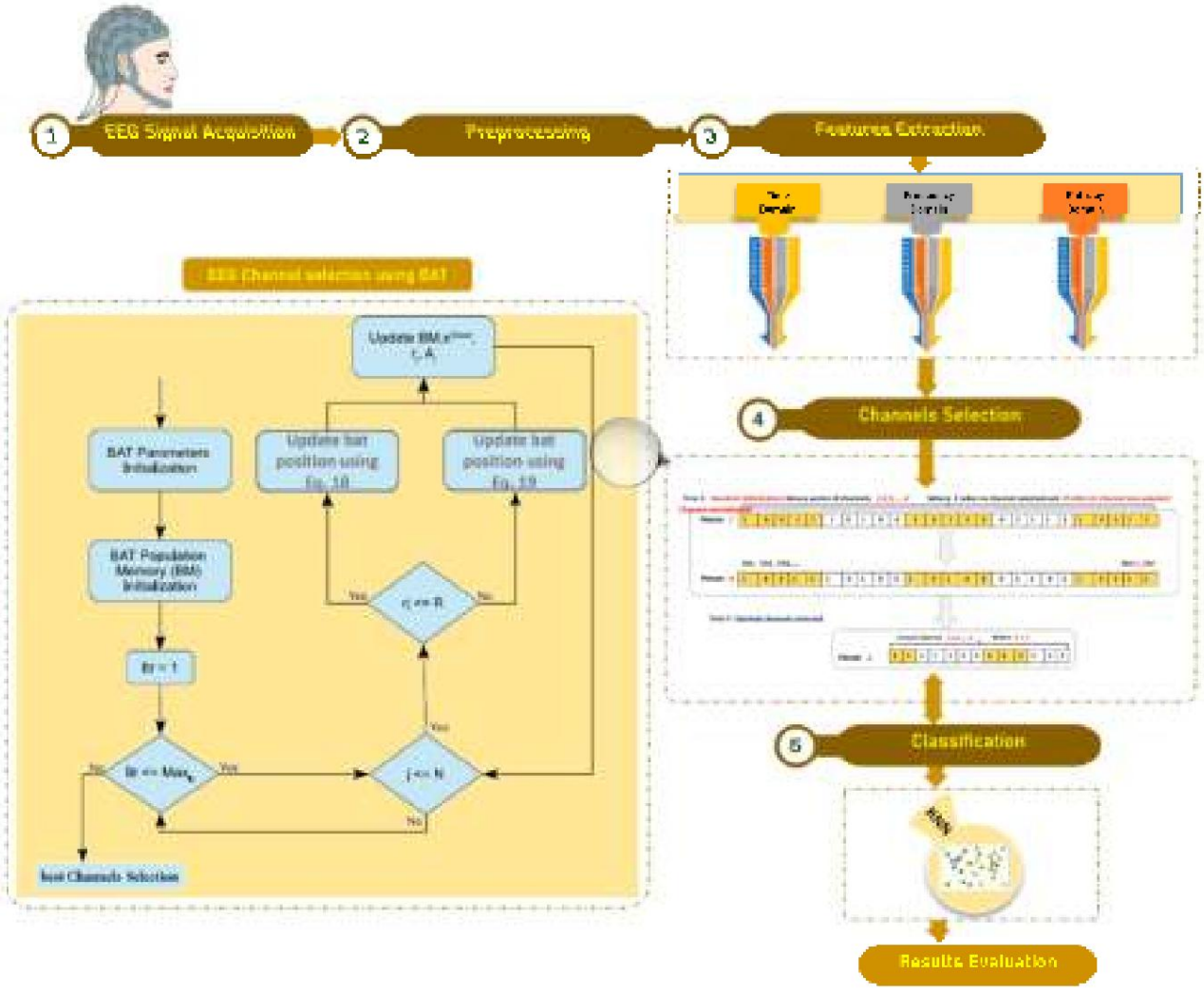


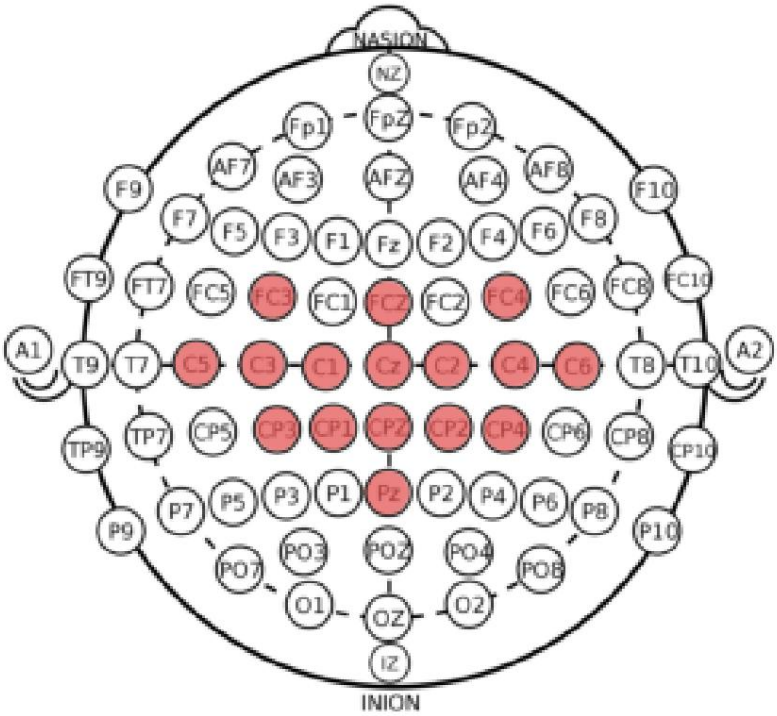
Abbildung 3. Eine vorgeschlagene Methode zur Auswahl von Elektroenzephalogramm-Kanälen.

4.1. Phase I: EEG-Signalerfassung **In**

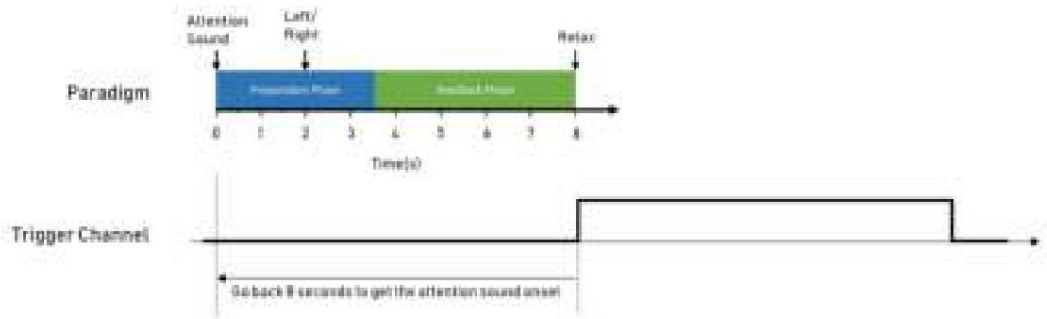
dieser Studie stammt der Datensatz von g.tec Medical Engineering GmbH, ein Standard-EEG-Datensatz; Einzelheiten zu diesem Datensatz werden in [8] erläutert. Die EEG-Daten von Patienten nach einem Schlaganfall mit Hemiparese der oberen Extremitäten wurden untersucht. An dieser Studie nahmen 8 Patienten nach einem Schlaganfall teil, die mit dem RecoveriX-System (g.tec medical engine

Machine Translated by Google

GmbH), mit einem Durchschnittsalter von 22 Jahren (SD = **4,582). Jeder Teilnehmer erhielt drei Monate lang ein BCI-basiertes MI-Training mit 2 Trainingseinheiten pro Woche (insgesamt 25 Trainingseinheiten). Das Studienteam führte zwei Bewertungen durch und analysierte sie (vor und nach d Die Vortrainingsbewertung war 30 bis 35 Tage vor der Intervention geplant, und die Nachtrainingsbewertung wurde einige Tage nach der Intervention durchgeführt (siehe Abbildung 4). Die Ethikkommission des Landes Oberösterreich in Österreich (#D-42-17) genehmigte dieses Studienprotokoll , und jeder Patient unterzeichnete vor der Voruntersuchung eine Schließlich wird dieser Datensatz mit einer Abtastrate von 256 Hz erfasst.**



(A)



(B)

Abbildung 4. (a) EEG-Elektrodenverteilungen basierend auf einem 10–20-System; (b) Schematische Darstellung des EEG- Aufzeichnungsprotokolls.

Gemäß den Systemanweisungen für eine MI-Gedankenaufgabe wurden die Patienten angewiesen, die dorsale Handgelenksbewegung zu visualisieren. Vor und nach der EEG-Neurorehabilitation nahmen die Patienten an 25 MI-basierten BCI-Sitzungen mit Nachuntersuchungen zur Messung der funktionellen Veränderungen teil. Jede Sitzung bestand aus 240 MI-Wiederholungen mit beiden Händen, aufgeteilt in drei Durchläufe mit je 80 Versuchen. Jede Sitzung dauerte ungefähr eine Stunde, einschließlich der Zeit für Vorbereitung und Aufräumen. Die MI-basierten BCI-Aufgaben wurden mit **zufälligen Intervallen zwischen den Versuchen in pseudozufällige**

Machine Translated by Google

Um alle neuronalen Prozesse, die während der Schlaganfallrehabilitation stattfinden – kognitive, affektive und motorische Komponenten – zu erfassen, wurden aktive EEG-Kappen mit 16 Kanälen eingesetzt. Die Kappen wurden von der Firma g.tec medical engineering GmbH in Österreich hergestellt.

Gemäß dem internationalen 10–20-System wurden die EEG-Elektroden wie folgt positioniert: (FC5, **FC1,** FCz, **FC2, FC6, C5, C3, C1,** Cz, **C2, C4, C6, CP5, CP1, CP2 und CP6).**
Dem Probanden wurde eine Erdungselektrode an der FPz angelegt **und eine Referenzelektrode am rechten Ohrläppchen befestigt.**

Diese Kanäle sind mit verschiedenen kognitiven und sensorisch-motorischen Prozessen verbunden, die bei Patienten nach einem Schlaganfall in gewisser Weise beeinträchtigt sind. Insbesondere die frontozentralen Kanäle (FC5, **FC1,** FCz, **FC2, FC6) sind mit der Emotionsregulation, dem Arbeitsgedächtnis und dem sekundären Motorbereich verbunden. Darüber hinaus ist es wichtig zu beachten, dass die zentralen Kanäle (C5, C3, C1,** Cz, **C2, C4 und C6) mit der sensorisch-motorischen Integration und der Darstellung von Körperbewegungen in Zusammenhang stehen . Darüber hinaus wird beschrieben, dass die zentroparietalen Kanäle** (CP5, **CP1, CP2, CP6)** am Prozess der Integration der exterozeptiven und propriozeptiven **Eingänge sowie an der Regulierung der willkürlic**

4.2. Phase II: Vorverarbeitung

Jeder Kanal des aufgezeichneten EEG-Datensatzes verwendete zunächst zwei Standardfilter. Erstens wurde ein Bandpassfilter (BPF) **mit Frequenzen zwischen (8 und 30 Hz) verwendet, um das Band der aufgezeichneten EEG-Daten einzuschränken, und zweitens wurde ein Butterworth** (BW) **-Kerbfilter bei (50 Hz) verwendet, um das Rauschen durch Netzstörungen zu reduzieren.** (ICs) s(t) = [s1(t), . . . ,sn(t)], **unter Verwendung des** von **[55]** vorgeschlagenen FastICA **-Algorithmus . Der Satz** s(t) **von** n **unbekannten Komponenten , die linear innerhalb der Matrix A gemischt wurden, und der Satz** x(t) **von** n **Beobachtungen, wobei** x(t) = [x1(t), . . . , xn(t)] **[56–58]** die EEGs darstellen und mit s(t) **in Beziehung stehen , wobei t die Zeit ist; daher lautet die ICA-Gleichung**

$$x(t) = As(t) \quad (10)$$

Anschließend wurden drei Metriken verwendet, um die Artefaktkomponenten (ICs) zu bewerten: **Kurtosis** (Kurt), Schiefe **(Skw) und Probenentropie** (SampEn). **Wenn diese Parameter den pm1.2-Schwellenwert für jeden** IC überschritten, wurde **der** IC **als kritisch markiert und mithilfe von WT entraus Der praktische Wert des Schwellenwerts wurde durch Versuch und Irrtum sowie frühere Untersuchungen** ermittelt **[59–61]. Der Schwellenwert von ±1,2 ist kein Nachteil der AICA-WT-Technik, da die künstlichen** ICs **nicht verworfen, sondern mit der WT-Technik entrauscht wurden. Infolgedessen entrauschte WT die markierten** ICs , **bevor die verstärkten Komponenten in den ursprünglichen EEG-Datensatz** zurückgeführt wurden **[22,23].**

4.3. Phase III: EEG-Merkmalsextraktion

In dieser Arbeit werden drei verschiedene Techniken zur Merkmalsextraktion verwendet. Diese Techniken umfassen Merkmale im Zeitbereich, Entropiemerkmale und Merkmale im Frequenzbereich, wie sie in [8] extrahiert wurden.

4.4. Phase IV: EEG-Kanalauswahl mit BAT Optimizer

Diese Phase ist der Hauptbeitrag und umfasst mehrere Schritte, um die optimale Teilmenge der EEG-Kanäle zu erreichen, die die höchste Genauigkeitsrate liefern kann. Die folgenden Schritte zeigen, wie wir den BAT-Optimierer für das Problem der EEG-Kanalauswahl anpassen: • Schritt 1: Initialisierung der BAT-Algorithmusparameter. In diesem Schritt werden die BAT-Parameter mit Anfangswerten initialisiert. Diese Parameter sind die Anzahl der Microbats (Lösungen) (N), die maximale Anzahl von Iterationen (MaxItr), **der maximale** (fmax) **und minimale** (fmin) **Frequenzbereich, der Geschwindigkeitsvektor (v), die Lautstärkerate (A), die Pulsrate** (r), der **Bereich der anfänglichen Pulsrate** (r) *(ÿ)* **im Bereich [ÿ1, ^{0}), zwei Parameter** *(ÿ* **und ÿ) mit konstanten Werten im Bereich [0, 1] und Bandbreite 1], wie in Tabelle 1 gezeigt.**

Machine Translated by Google

Tabelle 1. Metaheuristische Parameter.

Algorithmus	Parameter	Bevölkerungsgröße	Läuft
ES	*ÿ* **= 1,5,** Pa = **0,25 c1**	**20**	**30**
PSO	= **c2** = **2** A **=**	**20**	**30**
SCHLÄGER	**0,5,** r **= 0,5** *ÿ* **= 0,5,**	**20**	**30**
FFA	*ÿ* **= 1, ÿ0 = 0,2**	**20**	**30**
GWO	**-**	**20**	**30**

- **Schritt 2: Initialisierung des BAT-Populationsspeichers. Die BAT-Lösungen werden in diesem Schritt zufällig generiert, indem Gleichung (11) verwendet wird, wobei der BAT-Algorithmus und bestimmte Parameter und Einschränkungen des Optimierungsproblems.**

$$J = \text{lbi} + (\text{ubi} \ddot{y} \text{lbi}) \times \mathbf{R}, x \quad \textbf{(11)}$$

wobei ÿi = **1, 2, . . . , d,** ÿj = **1, 2, . . . , N und** R **eine Zufallszahl zwischen 0 und 1 ist.**

Die erzeugten Lösungen ergeben zusammen die Population, wie in Gleichung (12) gezeigt, und werden im BAT-Speicher (BM) in aufsteigender Reihenfolge nach ihrer Fitness gespeichert Werte. Die beste Lösung mit den fittesten Werten wird x **zugewiesen** Gbest **und gespeichert im erste Position des BM.**

$$\text{BM} = \begin{matrix} X^1_1 & X^1_2 & \cdots X & 1 \\ \ddot{y} X^2_1 & X^2_2 & \cdots X & 2\ \ddot{y} \\ \vdots & \vdots & \cdots & \vdots \\ \ddot{y} X^N_1 & X^N_2 & \cdots X & N\ \ddot{y} \end{matrix} \quad \begin{matrix} \\ D \\ D \\ \\ D \end{matrix}, \quad \textbf{(12)}$$

- **Schritt 3: Intensivierung der Fledermauspopulation. Jetzt fliegen alle Microbats** x **und ändern ihre Position unter Berücksichtigung einer Geschwindigkeit** v **, die durch eine zufällig erzeugte** Frequenz f **definiert ist , wie in den Gleichungen (13) und (14) gezeigt. Dementsprechend werden die Schlägerpositionen aktualisiert mit Gleichung (15).**

$$\text{fi} = \text{fmin} + (\text{fmin} \ddot{y} \text{fmax}) \times \mathbf{R}, \quad \textbf{(13)}$$

$$'J = v \ J + (x \ J \ \ddot{y} x \ G \text{ am besten}) \times \text{fi}, \quad \textbf{(14)}$$

$$X'^J_{\text{iii}} = x \ J + v \ 'J \quad \textbf{(15)}$$

Es ist bemerkenswert, dass die Microbats ihre Standorte aktualisieren, um näher an den globalen Besten zu sein (x Gbest). **Die neue Position der Fledermäuse verstärkt also die Position in eine Richtung in Richtung** x^{Gbest} **.**

- **Schritt 4: Diversifizierung der BAT-Population.**

In diesem Schritt werden die Positionen der Fledermäuse anhand von Zufallsparametern aktualisiert, um versuchen, bessere globale Lösungen zu finden. Eine neue Lösung wird aus den BM-basierten auf eine Auswahlmethode und wird x am besten zugewiesen. **Anschließend wird die aktuelle Lösung aktualisiert mit** x best, **wie folgt:**

$$X'^J_{\text{ich}} = x \ \text{am besten} + \ddot{y} \times A \ \ddot{y}_{\text{j.}} \quad \textbf{(16)}$$

wobei A $\ddot{y}$ **ist der Lautstärkedurchschnitt. Da die Schritte 3 und 4 widersprüchliche Schritte sind, ist der BAT Algorithmus wählt jeweils einen davon zur Ausführung aus. Diese Auswahl basiert auf der folgende Gleichung:**

$$x \ \ddot{y}j = \begin{cases} x \ \text{bestes} + \ddot{y} \times A \ \ddot{y} \ J & \textbf{wenn } rj \ \ddot{y} \ R \\ Xj + v \ 'J & \textbf{ansonsten} \end{cases}, \quad \textbf{(17)}$$

Machine Translated by Google

• **Schritt 5: BM-Update. Die Position der Microbats im BM wird in diesem** Schritt aktualisiert , **wenn der neue Standort fitter ist als der alte und** R ÿ Aj . **Außerdem** wird X **aktualisiert, wenn das neue BM eine Lösung mit einem besseren Fitnesswert enthält. Anschließend** werden die rj- **und** Aj **-Werte gemäß den Gleichungen (18) und (19) aktualisiert.**

$$rj = r j^{0} (1 \ddot{y} e^{(\ddot{y}\ddot{y}\times itr)}, \quad (18)$$

$$Aj = \ddot{y} \times Aj , \quad (19)$$

wobei itr **die aktuelle Iterationsnummer bezeichnet.**

• **Schritt 6: Abbruchkriterium. Die Schritte 3, 4 und 5 werden wiederholt, bis der Algorithmus das Abbruchkriterium.**

4.5. Phase V: Klassifizierung **In**

dieser Arbeit wird ein KNN-Klassifikator verwendet. Bei dieser vorgeschlagenen Methode wird der Datensatz in 60 % Training, 20 % Validierung und 20 % Test aufgeteilt.

5. Experimente und Ergebnisse

In diesem Abschnitt wird die Leistung der vorgeschlagenen Methode (BAT) von EEG-Kanälen für die Rehabilitation von Schlaganfallpatienten ausführlich erläutert. Da der vorgeschlagene Ansatz nicht deterministisch ist, wird die mittlere Genauigkeitsrate über 20 Durchläufe berechnet, um verzerrte Ergebni Für die Experimente wird ein Ideapad 310 Lenovo PC, China, verwendet. Intel Core i7® 2,59 GHz Prozessor, 16 GB RAM und offizielles Windows 10. Die Leistung der vorgeschlagenen Methode wird anhand von zwei Maßstäben bewertet, nämlich Genauigkeit und Anzahl der ausgewählten EEG-Kanäle. Dabei wird die Stabilität der Algorithmen durch Berechnung der besten, schlechtesten und mittleren Genauigkeit untersucht. Darüber hinaus wird ein statistischer Test basierend auf der Klassifizierungsgenauigkeit des Motor-/Bildgebungs-EEG-Datensatzes unter Verwendung der Summe der Ränge Die vorgeschlagene Methode wird anhand von drei Arten von EEG-Merkmalen getestet: Zeitbereichs-Merkmale, Frequenzbereichs-Merkmale und Entropiebereichs-Merkmale. Darüber hinaus werden die Leistungsergebnisse der vorgeschlagenen Methode anhand eines statistischen Tests ausgewertet, um die Signifikanz **der vorgeschlagenen Methode im Vergleich zu anderen metaheuristischen Algorithmen aufzuzeigen .**

Tabelle 1 zeigt die Parameter der ausgewählten metaheuristischen Algorithmen, die in dieser Arbeit verwendet werden.

5.1. Ergebnisse im Zeitbereich

Tabelle 2 präsentiert die Ergebnisse von sieben Merkmalen im Zeitbereich: Higuchis fraktale Dimension (HFD); die Hjorth-Parameter Hjorth-Aktivität (HjAc), Hjorth-Komplexität (Hj-Comp) und Hjorth-Mobilität (HjMo); Hurst (Hur); Kurtosis (Kurt); und Schiefe (Skw).

Die Genauigkeit und die Anzahl der ausgewählten EEG-Kanäle sind zwei Messungen zur Bewertung der sieben für den Motor-/Bildgebungs-EEG-Datensatz ausgewählten Merkmale. Die Bewertungsmetriken werden in drei Kategorien eingeteilt: beste, schlechteste und mittlere Genauigkeit und Anzahl der ausgewählten Kanäle zur Bewertung der Stabilität der Algorithmen.

Laut Tabelle 2 erreichten die Leistungsergebnisse für die metaheuristischen Algorithmen PSO, GWO, FFA, CS und BAT mit mehreren Features unterschiedliche und bemerkenswerte Genauigkeiten hinsichtlich der besten, schlechtesten und durchschnittlichen Genauigkeit und der Anzahl der ausgewählten Kanäle. Beim HFD-Feature erreichte der BAT-Algorithmus bessere Genauigkeiten von 99, 97,5 und 92,8 für die beste, mittlere und schlechteste Genauigkeit bzw. Kanalauswahl. Beim HjAc-Feature erreichte der BAT-Algorithmus im besten Fall eine hohe Genauigkeit von 97,7 für Genauigkeit und Anzahl der ausgewählten Kanäle , während PSO im durchschnittlichen und schlechtesten Fall wertvolle Verhältnisse von 94,4 und 85,3 bei Genauigkeit und Anzahl der ausgewählten Kanäle erreichte. Beim HjComp-Feature erreichte der PSO-Algorithmus im mittleren und schlechtesten Fall eine Genauigkeit von 88,6 und 84,7, während der BAT-Algorithmus im besten Fall die höchste Genauigkeit von 93,8 erreichte. Für das HjMo-Feature erreichte BAT 97,8, 95,4 und 88,6 für die besten, mittleren und schlechtesten Werte unter anderen metaheuristischen Algorithmen. Für das Hur-Feature erreichte PSO hohe Genauigkeiten von 72,8, 65,2 und 56,6 für die besten, mittleren und schlechtesten Werte. Fü

Machine Translated by Google

Algorithmus erzielte wertvolle Genauigkeiten von 73,2, 68,8 und 63,1 für die beste, mittlere und schlechteste Genauigkeit bzw. Anzahl der ausgewählten Kanäle.

Darüber hinaus erreichte BAT für Skw hohe Genauigkeiten von 79,2 und 72,5 für den besten und mittleren Wert. Im Vergleich zu PSO, das im schlimmsten Fall eine Genauigkeit von 58 erreichte der Genauigkeit. Abbildung 4 zeigt die Ergebnisse der Zeitbereichsfunktionen für die Konvergenz Rate und Kanalverteilung, Vergleich der BAT- und PSO-Algorithmen, die erreichten die höchste Genauigkeit im Vergleich zu den anderen metaheuristischen Algorithmen, d.ÿh. GWO, FFA und CS. In Abbildung 4a zeigt die Konvergenzrate die Kanäle und die Anzahl der Iterationen für die BAT- und PSO-Algorithmen. Darüber hinaus zeigt Abbildung 4b die effektiven und emotionalen Kanäle, wenn Kanäle in den BAT- und PSO-Algorithmen verteilt werden.

Tabelle 2. Die Leistung von metaheuristischen **Algorithmen mit mehreren Merkmalsextraktionsansätzen für Motorischer/bildgebender EEG-Datensatz.**

Funktion/Algorithmus			PSO	GWO	FFA	CS	SCHLÄGER
HFD	Am besten	Genauigkeit	97,9	90,7	66,7	54,9	99
		Anzahl Kanäle	8	3	1	1	6
	Am schlimmsten	Genauigkeit	81,5	67,4	45.1	43.2	92,8
		Anzahl Kanäle	12	6	3	2	13
	Bedeuten	Genauigkeit	94,4	76,2	56,7	50,6	97,5
		Anzahl Kanäle	9	4	2	2	10
HjAc	Am besten	Genauigkeit	97,1	88,4	68,8	55.1	97,7
		Anzahl Kanäle	5	3	1	1	4
	Am schlimmsten	Genauigkeit	85,3	70,4	41,7	42.1	77,9
		Anzahl Kanäle	9	5	3	2	14
	Bedeuten	Genauigkeit	94,4	82.1	60,7	48,7	90,8
		Anzahl Kanäle	8	4	3	2	8
HjComp	Am besten	Genauigkeit	91,2	78,9	58,8	48,7	93,8
		Anzahl Kanäle	7	3	2	1	5
	Am schlimmsten	Genauigkeit	84,7	58.1	47,4	40,9	74,7
		Anzahl Kanäle	15	6	3	2	13
	Bedeuten	Genauigkeit	88,6	69,9	53,9	44.2	88,2
		Anzahl Kanäle	9	4	3	1	9
HjMo	Am besten	Genauigkeit	96,5	89,3	67,9	53,5	97,9
		Anzahl Kanäle	6	2	2	1	5
	Am schlimmsten	Genauigkeit	86,4	51,8	53,9	43,8	88,6
		Anzahl Kanäle	11	6	3	2	12
	Bedeuten	Genauigkeit	92,3	80,6	61,2	47,9	95,4
		Anzahl Kanäle	9	4	3	1	9
Hur	Am besten	Genauigkeit	72,8	57,9	47,2	44.1	70.1
		Anzahl Kanäle	4	2	1	1	5
	Am schlimmsten	Genauigkeit	56,6	44.1	41,5	41.3	53,9
		Anzahl Kanäle	10	6	3	2	11
	Bedeuten	Genauigkeit	65,2	51,7	44,7	42,6	64,9
		Anzahl Kanäle	8	4	3	2	9

Machine Translated by Google

Tabelle 2. Fortsetzung

Funktion/Algorithmus			PSO	GWO	FFA	ES	SCHLÄGER
Kurt	Am besten	Genauigkeit	71,2	61	48,4	43.4	73,2
		Anzahl Kanäle	7	3	2	1	7
	Am schlimmsten	Genauigkeit	56,5	48,2	43.3	40,3	63.1
		Anzahl Kanäle	12	6	3	2	12
	Bedeuten	Genauigkeit	65,6	54,9	45,4	42	68,8
		Anzahl Kanäle	9	5	3	2	10
Skw	Am besten	Genauigkeit	78,7	63,7	50	44,4	79,2
		Anzahl Kanäle	4	1	2	1	3
	Am schlimmsten	Genauigkeit	58	41	43.2	40,4	48,2
		Anzahl Kanäle	11	6	3	2	13
	Bedeuten	Genauigkeit	69,6	53,3	45,9	42.3	72,5
		Anzahl Kanäle	7	4	2	2	9

5.2. Ergebnisse im Entropiebereich

Tabelle 3 zeigt die Ergebnisse der Leistungsgenauigkeit unter Verwendung von sieben Entropiemerkmalen. nämlich ConFuzEn, FuzEn, impe, MFEmu, RCMFEmu, SampEn und TsEn. Basierend auf der Genauigkeit und Anzahl der ausgewählten Kanäle für den besten, schlechtesten und mittleren Fall, diese Entropie Merkmale werden verwendet, um die fünf metaheuristischen Algorithmen PSO, GWO, FFA, CS und BAT bzw.. Im Vergleich zu den anderen sieben Entropiemerkmalen ist die beste Klassifizierung Die Leistung erreichte die höchste Genauigkeit mit ConFuzEn mit dem PSO-Algorithmus im Vergleich zu den Algorithmen GWO, FFA, CS und BAT. Andere Entropiemerkmale führten zu hö Genauigkeitsergebnissen mit dem BAT-Algorithmus als mit den Algorithmen PSO, GWO, FFA und CS. Der Der BAT-Algorithmus erreichte Genauigkeiten von 95,5, 95,6, 88,1, 87,2, 91,4 und 84,6 unter Verwendung von FuzEn. impe, MFEmu, RCMFEmu, SampEn und TsEn. Darüber hinaus zeigt Abbildung 5 **die Ergebnisse der Leistungsbewertung zwischen den BAT- und PSO-Algorithmen in Bezug auf die Konvergenzrate und Kanalverteilung.**

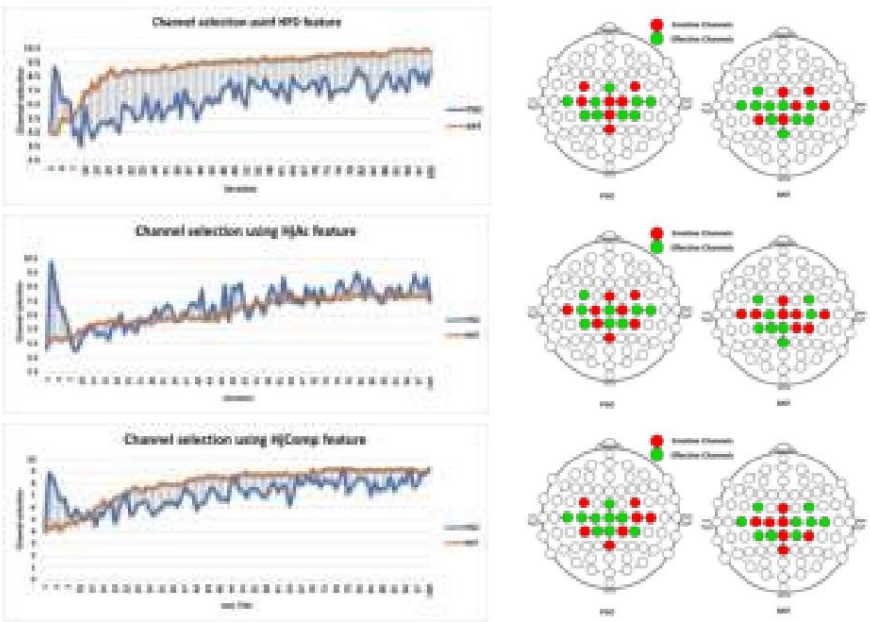


Abbildung 5. Fortsetzung

Machine Translated by Google

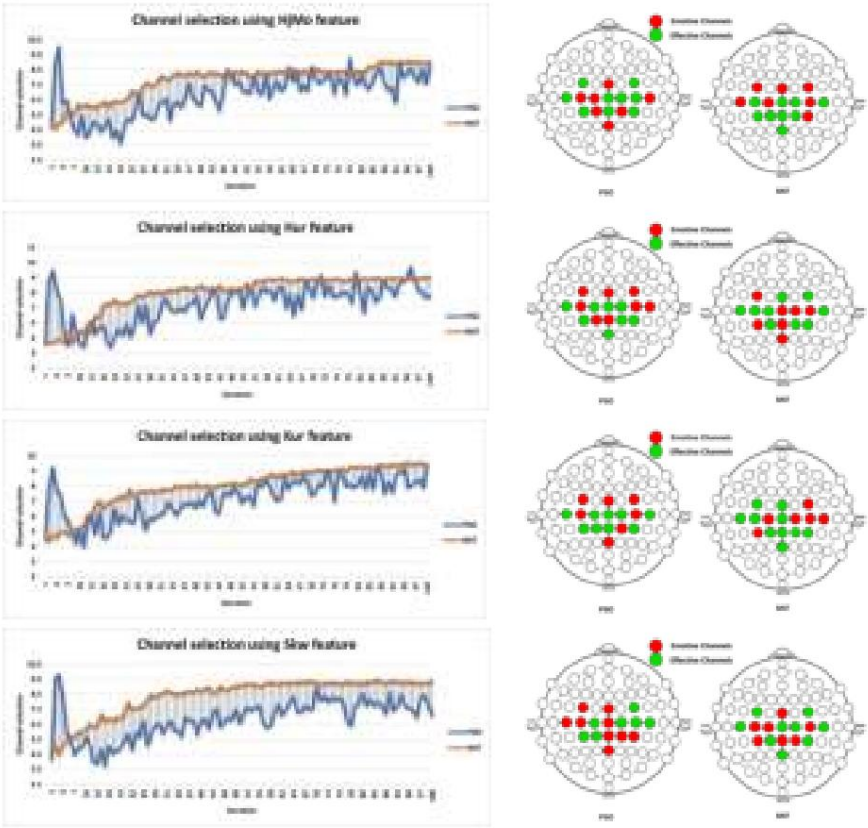


Abbildung 5. Konvergenzrate und Kanalverteilung.

Tabelle 3. Die Leistung von metaheuristischen **Algorithmen mit mehreren Merkmalsextraktionsansätzen für Motorischer/bildgebender EEG-Datensatz.**

Funktion/Algorithmus			PSO	GWO	FFA	ES	SCHLÄGER
ConvFuzEn	Am besten	Genauigkeit	96,6	82,7	61,3	52	96,5
		Anzahl Kanäle	3	3	1	1	7
	Am schlimmsten	Genauigkeit	76,4	64,7	45,5	42,5	87,1
		Anzahl Kanäle	11	6	3	2	14
	Bedeuten	Genauigkeit	90,2	73,6	54,9	45,8	93,2
		Anzahl Kanäle	8	5	2	1	11
FuzEn	Am besten	Genauigkeit	93,6	81,6	60,6	49.1	95,5
		Anzahl Kanäle	2	4	2	1	8
	Am schlimmsten	Genauigkeit	82,2	68,7	50,5	41,8	88,9
		Anzahl Kanäle	10	6	3	2	12
	Bedeuten	Genauigkeit	88,6	74,4	55,7	45,5	92,1
		Anzahl Kanäle	7	5	3	1	10
impfen	Am besten	Genauigkeit	94,7	78,6	59,3	49,8	95,6
		Anzahl Kanäle	5	3	2	1	7
	Am schlimmsten	Genauigkeit	75,8	56.1	47,9	44.3	84,9
		Anzahl Kanäle	10	6	3	2	12
	Bedeuten	Genauigkeit	85,3	69,8	52,3	46,6	90,8
		Anzahl Kanäle	8	5	2	2	10

Machine Translated by Google

Tabelle 3. Fortsetzung

Funktion/Algorithmus			PSO	GWO	FFA	CS	SCHLÄGER
MFEmu	Am besten	Genauigkeit	85,2	73,6	54,3	46,8	88,1
		Anzahl Kanäle	7	3	2	1	7
	Am schlimmsten	Genauigkeit	70,8	55,9	45.1	40,5	75,5
		Anzahl Kanäle	12	6	3	2	12
	Bedeuten	Genauigkeit	80,7	66.1	49,4	44,8	84
		Anzahl Kanäle	9	5	2	2	9
RCMFEmu	Am besten	Genauigkeit	84,3	67,7	51,9	46,6	87,2
		Anzahl Kanäle	4	2	2	1	8
	Schlimmste	Genauigkeit	63,7	47	44,9	41.2	75,3
		Anzahl Kanäle	12	6	3	2	13
	Bedeuten	Genauigkeit	77,3	56,5	47,7	43	80,3
		Anzahl Kanäle	8	4	2	1	10
SampEn	Am besten	Genauigkeit	87,9	71,5	52,9	51,9	91,4
		Anzahl Kanäle	6	3	1	1	7
	Am schlimmsten	Genauigkeit	75,4	53.2	42,6	40,4	75,9
		Anzahl Kanäle	10	6	3	3	14
	Bedeuten	Genauigkeit	81,4	63,7	48,2	44.3	83,6
		Anzahl Kanäle	8	5	2	2	10
TsEn	Am besten	Genauigkeit	83,9	72,5	52,3	46,6	84,6
		Anzahl Kanäle	7	3	1	1	4
	Am schlimmsten	Genauigkeit	65,9	54,4	41.2	40.2	61,3
		Anzahl Kanäle	12	7	3	2	13
	Bedeuten	Genauigkeit	78,7	64,5	47,2	42,8	79,8
		Anzahl Kanäle	10	5	2	1	10

5.3. Ergebnisse im Frequenzbereich

Basierend auf Tabelle 4 erhielten die Frequenzbereichsmerkmale von meanF und medF hohe Genauigkeiten von 96,7 bzw. 83,4 für die PSO- und BAT-Algorithmen. Im besten, schlechtesten, und mittlere Fälle, meanF gewann eine signifikante Genauigkeit im Vergleich zu medF Merkmal auf der Gesamtgenauigkeitsergebnisse für die PSO- und BAT-Algorithmen. Die beste Genauigkeit von medF wurde Die mit PSO erzielte Genauigkeit betrug 96,7, während die schlechtesten und mittleren Genauigkeiten 90,3 und 93,8 betrugen. mit dem BAT-Algorithmus. Darüber hinaus erzielte der BAT-Algorithmus eine hohe Genauigkeit Klassifizierungsergebnisse von 83,4 und 78,4 im Verhältnis zur besten und schlechtesten Genauigkeit mit dem Merkmal von medF, während PSO eine Genauigkeit von 69,5 erreichte. Die Konvergenzrate und die Kanalverteilung sind in Abbildung 6 dargestellt.



Abbildung 6. Fortsetzung

Machine Translated by Google

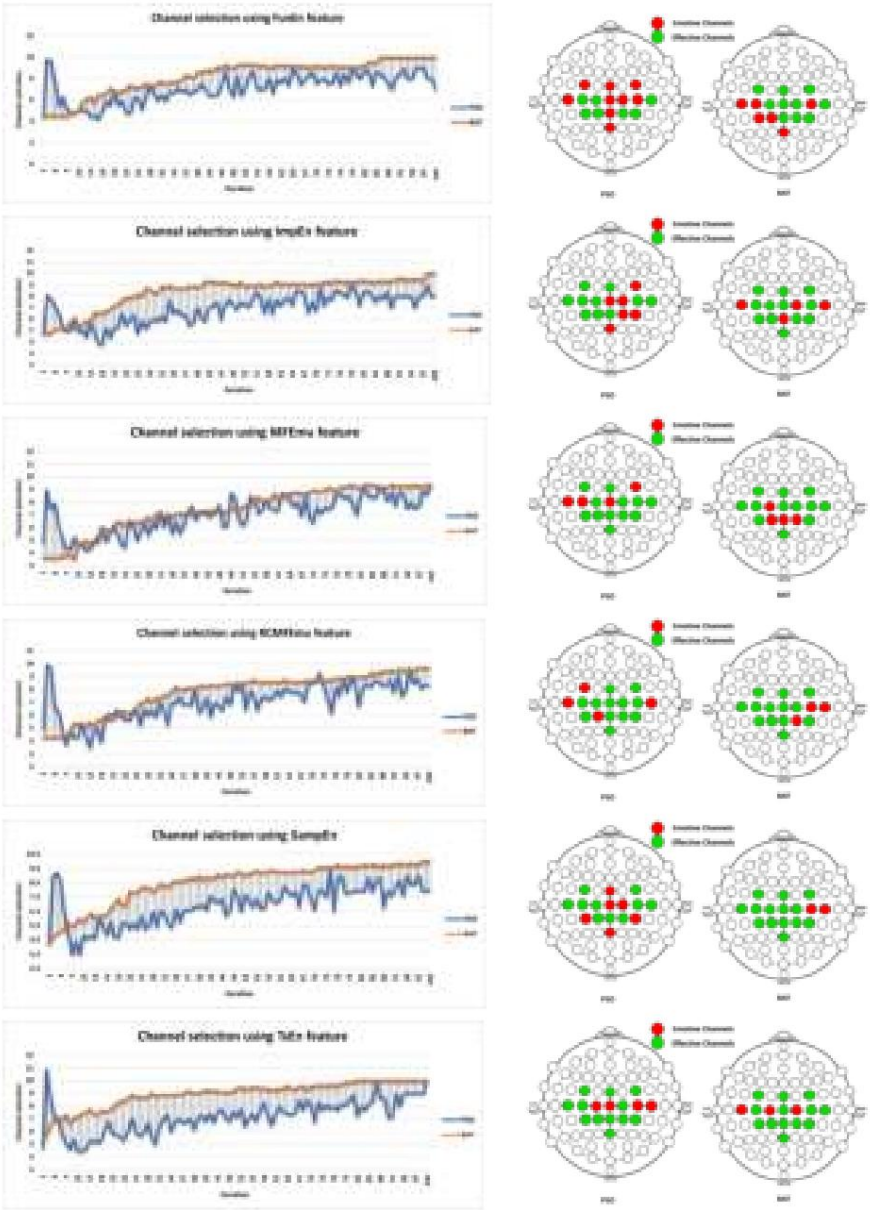


Abbildung 6. Konvergenzrate und Kanalverteilung.

Wie erwähnt, zeigt Tabelle 4 die Leistung der vorgeschlagenen Methode (BAT) im Vergleich zu den PSO-, GWO-, FFA- und CS-Algorithmen unter Verwendung von Frequenzbereichsme Der BAT-Algorithmus erzielte bessere Ergebnisse als alle anderen Algorithmen, außer in einigen Fällen , in denen der PSO-Algorithmus ihn leicht übertraf, wie z. B. bei meanF, **wo die beste Genauigkeit 96,7 für PSO und 96,3 für BAT betrug. Die Gesamtdurchläufe des BAT-Algorithmus erzielten jedoch bessere Ergebnisse mit Genauigkeitswerten von 93,8, 91,2, 76,4, 54 und 50 für die Algorithmen BAT, PSO, GWO, FFA und CS. Die Konvergenzrate und die Kanalverteilung sind in Abbildung 7 dargestellt.**



Abbildung 7. Fortsetzung

Machine Translated by Google



Abbildung 7. Konvergenzrate und Kanalverteilung.

Tabelle 4. Die Leistung des BAT-Algorithmus im Vergleich zu mehreren metaheuristischen Algorithmen unter Verwendung Frequenzbereichsfunktionen.

Funktion/Algorithmus			PSO	GWO	FFA	ES	SCHLÄGER
MittelwertF	Am besten	Genauigkeit	96,7	87,3	64.1	52,5	96,3
		Anzahl Kanäle	5	3	2	1	7
	Am schlimmsten	Genauigkeit	65,4	66	50,2	41,9	90,3
		Anzahl Kanäle	10	6	3	2	12
	Bedeuten	Genauigkeit	91,2	76,4	54	50	93,8
		Anzahl Kanäle	8	4	2	2	10
MedF	Am besten	Genauigkeit	82,7	69,2	53,6	46,6	83,4
		Anzahl Kanäle	6	3	2	1	6
	Am schlimmsten	Genauigkeit	69,5	55,5	41	20	68,7
		Anzahl Kanäle	10	6	3	2	14
	Bedeuten	Genauigkeit	76,3	60,6	46.1	36,5	78,4
		Anzahl Kanäle	9	4	2	2	10

5.4. Ergebnisse der statistischen Analyse

Bevor wir unsere Ergebnisse zur Genauigkeit und Anzahl der ausgewählten Kanäle präsentieren Um die Klassifizierungsleistung der Zeit-, Entropie- und Frequenzmerkmalsanalyse von Schlaganfall-EEG-Signalen zu verbessern , möchten wir den Einfluss der Klassifizierungsgenauigkeit u im Motor/Bildgebung EEG-Datensatz durch die Übernahme der Summe der Ränge der Metaheuristik Algorithmen. Auf diese Weise wurde die Funktion, die die Ranggenauigkeit ausgewählter Kanäle verbesserte, identifiziert werden können.

Wie in Tabelle 5 dargestellt, haben wir die Summe der Ränge von PSO, GWO, FFA, CS, und BAT-Metaheuristiken je nach individueller Merkmalsdomäne. Im Im Zeitbereich testeten wir sieben Merkmale. Der BAT-Algorithmus verbesserte die Genauigkeitsrang, insbesondere in HFD und Hjac, für einen Genauigkeitsrang von 99 bzw. 97,7. Darüber hinaus sind weitere Funktionen im Zeitbereich wie HjComp, HjMo, Kurt und Skw verfügbar. zeigte einen einflussreichen Einfluss auf die Klassifikationsgenauigkeit Ergebnisse im Verhältnis zur Summe des Rangs des BAT-Algorithmus mit Ausnahme der Hur-Funktion, die einen hohen Einfluss auf der PSO-Algorithmus. In der statistischen Testphase hatte der BAT-Algorithmus jedoch niedrige Summation der Rangergebnisse unter anderen metaheuristischen Algorithmen.

Andererseits zeigte die Summe der Ränge für den BAT-Algorithmus eine bedeutsame Auswirkungen auf die Klassifizierungsgenauigkeit Ergebnisse relativ zu den sieben Merkmalen der Entropie Domäne mit Ausnahme der ConvFuzEn-Funktion, die einen bemerkenswerten Effekt auf die PSO hatte Algorithmus mit einem Unterschied von 0,1 gegenüber BAT bezüglich der Summe der Rangergebnisse in der statistischen Analyse. Der BAT-Algorithmus hatte jedoch niedrige Summationsergebnisse in der statistischen Testphase unter den anderen metaheuristischen Algorithmen.

Die Genauigkeit der Klassifizierung ergibt sich jedoch aus den Merkmalen im Frequenzbereich. meanF zeigte einen deutlichen verbessernden Effekt auf die Genauigkeit mit dem PSO-Algorithmus relativ zum Genauigkeitsrang 96,7 im Vergleich zur medF-Funktion, die eine Genauigkeit von 83,4 hatte

Machine Translated by Google

Rang mit dem BAT-Algorithmus. Darüber hinaus hatten die BAT- und PSO-Algorithmen im Vergleich zu anderen metaheuristischen Algorithmen niedrige Rangsummenergebnisse.

Tabelle 5. Die Summe der Ränge der metaheuristischen **Algorithmen basierend auf der Klassifizierungsgenauigkeit des motorischen /bildgebenden EEG-Datensatzes.**

Funktion/Algorithmus			PSO	GWO	FFA	CS	SCHLÄGER
TDF HjMo	HFD	Genauigkeit	97,9	90,7	66,7	54,9	99
		Rang	2	3	4	5	1
	HjAc	Genauigkeit	97,1	88,4	68,8	55.1	97,7
		Rang	2	3	4	5	1
	HjComp	Genauigkeit	91,2	78,9	58,8	48,7	93,8
		Rang	2	3	4	5	1
		Genauigkeit	96,5	89,3	67,9	53,5	97,9
		Rang	2	3	4	5	1
	Hur	Genauigkeit	72,8	57,9	47,2	44.1	70.1
		Rang	1	3	4	5	2
	Kur	Genauigkeit	71,2	61	48,4	43.4	73,2
		Rang	2	3	4	5	1
	Skw	Genauigkeit	78,7	63,7	50	44,4	79,2
		Rang	2	3	4	5	1
	Summierung der Ränge		13	21	28	35	8
FDF MedF	MittelwertF	Genauigkeit	96,7	87,3	64.1	52,5	96,3
		Rang	1	3	4	5	2
		Genauigkeit	82,7	69,2	53,6	46,6	83,4
		Rang	2	3	4	5	1
	Summierung der Ränge		3	6	8	10	3
EDF MFEmu	ConvFuzEn	Genauigkeit	96,6	82,7	61,3	52	96,5
		Rang	1	3	4	5	2
	FuzEn	Genauigkeit	93,6	81,6	60,6	49.1	95,5
		Rang	2	3	4	5	1
	impfen	Genauigkeit	94,7	78,6	59,3	49,8	95,6
		Rang	2	3	4	5	1
		Genauigkeit	85,2	73,6	54,3	46,8	88,1
		Rang	2	3	4	5	1
	RCMFEmu	Genauigkeit	84,3	67,7	51,9	46,6	87,2
		Rang	2	3	4	5	1
	SampEn	Genauigkeit	87,9	71,5	52,9	51,9	91,4
		Rang	2	3	4	5	1
	TsEn	Genauigkeit	83,9	72,5	52,3	46,6	84,6
		Rang	2	3	4	5	1
	Summierung der Ränge		13	21	28	35	8

Abbildung 8 zeigt die Ergebnisse der Geschwindigkeit metaheuristischer Algorithmen in Sekunden.

Machine Translated by Google

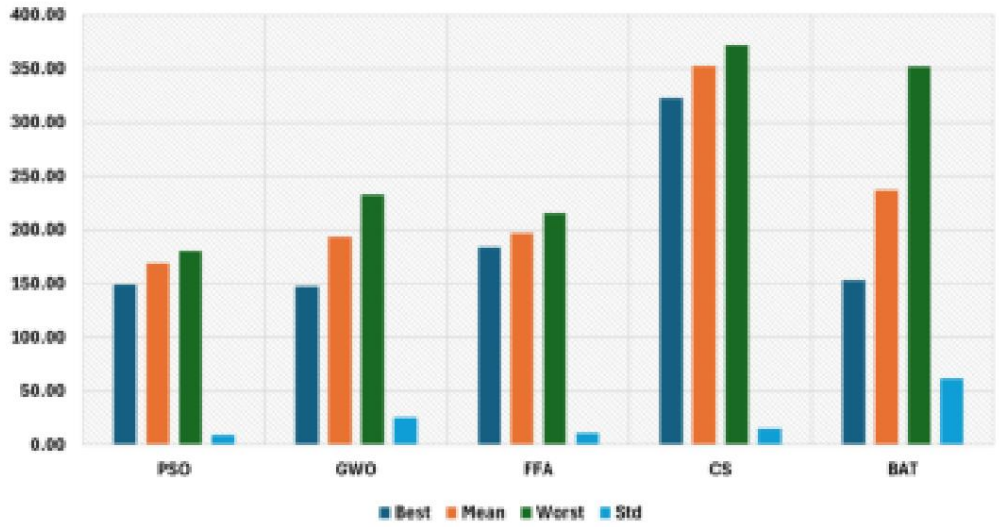


Abbildung 8. Geschwindigkeit metaheuristischer Algorithmen in Sekunden.

5.5. Diskussion der Ergebnisse

Ziel dieser Forschung ist es, die relevantesten EEG-Kanäle auszuwählen, die eine effektive Genauigkeitsrate für die Rehabilitation von Schlaganfallpatienten liefern, indem eine auf einem binären Bat-Algorithmus basierende Optimierungstechnik zur Optimierung der TEF-Eigenschaften verwendet wird. Um die Klassifizierungsleistung von Schlaganfall-EEG-Signalen weiter zu verbessern, wurden drei Arten von Merkmalen (nämlich Zeit-, Entropie- und Frequenzbereichsmerkmale) als Grundlage für die Klassifizierungsaufgabe verwendet. Es wurden zwei Messungen durchgeführt, um die Leistung der vorgeschlagenen Methode, die Genauigkeit und die Anzahl der ausgewählten EEG-Kanäle zu bewerten. Darüber hinaus haben wir einen statistischen Test mit der Summe der Ränge für den Motor-/Bildgebungs-EEG-Datensatz durchgeführt, um den Einfluss der Klassifizierungsgenauigkeit und die Signifikanz der vorgeschlagenen Methode im Vergleich zu hochmodernen metaheuristischen Algorithmen wie PSO, GWO, FFA und CS zu untersuchen. Zunächst wurden die Genauigkeit und die Anzahl der ausgewählten EEG-Kanäle bewertet, indem die besten, schlechtesten und mittleren Genauigkeitswerte berechnet wurden, um die Stabilität des implementierten Algorithmus zu erfassen. Im Zeitbereich wurden sieben Funktionen getestet, und die HFD-Funktion hatte unter anderen Optimierungsalgorithmen den größten Einfluss auf die Verbesserung der Klassifizierungsaufgabe im BAT-Optimierer für die besten, durchschnittlichen und schlechtesten Genauigkeitsfälle bzw. die Anzahl der ausgewählten Kanäle, basierend auf einem kNN-Klassifikator. Darüber hinaus übertrafen die BAT-Ergebnisse bei den Testwerten für den Fall der besten Genauigkeit die von PSO, das die besten Ergebnisse im Vergleich zu den Algorithmen GWO, FFA und CS bei den Funktionen HjAc und HjComp erzielte. Darüber hinaus verzeichneten die Ergebnisse für den BAT-Optimierer hohe beste, durchschnittliche und schlechteste Genauigkeitswerte bzw. die Anzahl der ausgewählten Kanäle mit den Funktionen HFD, HjMo und Kurt, wä Bei den Best-Case-Werten übertraf der BAT-Optimierer PSO und die anderen metaheuristischen Algorithmen mit den sechs Merkmalen HFD, HjAc, HjComp, HjMo, Kurt und Skw bei den wertvollen und höchsten Werten (siehe Tabelle 2). Darüber hinaus wurde gemäß den Ergebnissen der Leistungsgenauigkeit unter Verwendung von sieben Entropiemerkmalen basierend auf dem KNN- Klassifikator die beste Klassifizierungsleistung mit der höchsten Genauigkeit unter Verwendung von ConFuzEn mit dem PSO-Algorithmus im Vergle

Andere Entropiefunktionen führten mit dem BAT-Algorithmus zu höheren Genauigkeitsergebnissen als die Algorithmen PSO, GWO, FFA und CS, und zwar je nach den besten, mittleren und schlechtesten Genauigkeitswerten und der Anzahl der ausgewählten Kanäle. Der BAT-Algorithmus verzeichnete unter Verwendung der Funktionen FuzEn, impe, MFEmu, RCMFEmu, SampEn und TsEn jeweils Genauigkeitswerte von 95,5, 95,6, 88,1, 87,2, 91,4 und 84,6 , während PSO BAT mit Genauigkeitsergebniswerten für die Funktion ConvFuzEn für den KNN-Klassifikator übertraf (siehe Tabelle 3).

Machine Translated by Google

Die BAT- und PSO-Algorithmen erzielten im Vergleich zu anderen Algorithmen in der Literatur für die Frequenzbereichsmerkmale meanF und medF hohe Genauigkeiten von 96,7 bzw. 83,4 . **Der BAT-Algorithmus erzielte hochgenaue Klassifizierungsergebnisse von 83,4 bzw. 78,4 im Verhältnis zur besten und schlechtesten Genauigkeit mit dem Merkmal medF, während PSO eine Genauigkeit von 69,5 erreichte; darüber hinaus gewannen die PSO- und BAT-Algorithmen im Verhältnis zum Merkmal medF bei den Gesamtgenauigkeitsergebnissen in den Fällen der besten, schlechtesten und mittleren Genauigkeit erheblich an Genauigkeit .** Basierend auf den Klassifizierungsergebnissen auf Grundlage der Merkmale aus den drei verschiedenen Domänentypen übten die BAT- und PSO-Algorithmen im Vergleich zu den Ergebnissen der **Optimierungsalgorithmen GWO, FFA und CS hinsichtlich der Genauigkeit bzw. der Anzahl der ausgewählten Kanäle erhebliche Auswirkungen auf die Klassifizierungsleistung aus . Sie weisen daher eine deutliche Verbesserung der durchschnittlichen Genauigkeit im Verhältnis zum Klassifizierungsergebnis auf Basis der einzelnen Merkmalsdomänen auf, wie in Abbildung 5 zu sehen ist, die den Vergleich zwischen der Konvergenzrate und der Kanalverteilung für die BAT- und PSO-Algorithmen erklärt.**

Zweitens untersuchten wir den Einfluss der Klassifizierungsgenauigkeit auf den Motor-/Bildgebungs-EEG-Datensatz mithilfe der angenommenen Rangsumme metaheuristischer Algorithmen. Somit kann das Merkmal identifiziert werden, das die Ranggenauigkeit der ausgewählten Kanäle verbessert hat. Der BAT-Algorithmus verbesserte den Genauigkeitsrang deutlich, insbesondere bei den Merkmalen HFD und Hjac, mit einem Genauigkeitsrang von 99 bzw. 97,7 im Vergleich zu den Algorithmen PSO, GWO, FFA und CS gemäß den Rangsummenergebnissen in Bezug auf den Zeitbereich.
Darüber hinaus zeigten Funktionen wie HjComp, HjMo, Kurt und Skw einen starken Einfluss auf die Ergebnisse der Klassifizierungsgenauigkeit im Verhältnis zur Summe der Ränge des BAT-Algorithmus mit Ausnahme der Funktion Hur, die einen großen Einfluss auf den PSO-Algorithmus hatte. In der statistischen Testphase hatte der BAT-Algorithmus jedoch im Vergleich zu den anderen metaheuristischen Algorithmen niedrige Ergebnisse bei der Rangsumme. Des Weiteren zeigte der BAT-Algorithmus einen bedeutenden Einfluss auf die Ergebnisse der Klassifizierungsgenauigkeit für die Summe der Ränge im Verhältnis zu den sieben Funktionen des Entropiebereichs mit Ausnahme der Funktion ConvFuzEn, die einen bemerkenswerten Einfluss auf den PSO-Algorithmus mit einem Unterschied von 0,1 im Vergleich zu BAT hinsichtlich der Ergebnisse der Rangsumme in der statistischen Analyse hatte. In der statistischen Testphase hatte der BAT-Algorithmus jedoch im Vergleich zu den anderen metaheuristischen Algorithmen niedrige Ergebnisse bei der Rangsumme.
Darüber hinaus erzielten die BAT- und PSO-Algorithmen im Vergleich zu den anderen metaheuristischen Algorithmen in den Frequenzbereichsfunktionen niedrige Rangsummenergebnisse. Die meanF-Funktion zeigte einen deutlichen Verbesserungseffekt auf die Genauigkeit mit dem PSO-Algorithmus im Verhältnis zum Genauigkeitsrang von 96,7 im Vergleich zur medF-Funktion, die **mit dem BAT-Algorithmus einen Genauigkeitsrang von 83,4 aufwies .**

6. Schlussfolgerungen und zukünftige Arbeiten

In diesem Artikel wird eine neue Methode zur Auswahl von EEG-Kanälen vorgeschlagen, die auf einem Optimierungsalgorithmus namens BAT-inspirierter Algorithmus basiert. Der Hauptzweck des vorgeschlagenen Algorithmus besteht darin, die relevantesten EEG-Kanäle auszuwählen, die eine höhere Genauigkeitsrate für **die Rehabilitation von Schlaganfallpatienten bieten können .**

Die vorgeschlagene Methode wurde anhand eines Standard-EEG-Datensatzes getestet, der von acht Schlaganfallpatienten mit Hemiparese der oberen Extremitäten gesammelt wurde. Die EEG-Daten wurden mithilfe von EEG-Kappen mit 16 aktiven Elektroden von g.Nautilus PRO, g.tec medical engineering GmbH, Österreich, erfasst.

Die vorgeschlagene Methode verwendete mehrere EEG-Merkmalsextraktionsmethoden wie den Zeitbereich , den Frequenzbereich und den Entropiebereich. Darüber hinaus wurde die vorgeschlagene Methode (BAT) mit vier metaheuristischen Algorithmen verglichen, nämlich Partikelschwarmoptimierung (PSO), Grey Wolf Optimizer (GWO), Cuckoo Search (CS) und Firefly-Algorithmus (FFA). Die Leistung der vorgeschlagenen Methode wurde anhand der Genauigkeitsrate und der Anzahl der ausgewählten Kanäle bewertet. Die vorgeschlagene Methode erzielte die besten Ergebnisse aller Merkmalsextraktionsmethoden und zeigte eine signifikante Verbesserung bei Verwendung eines standardmäßigen statistischen Analysetests, dem sogenannten Rangsummentest.

Machine Translated by Google

Schließlich gelang es mit der vorgeschlagenen Methode, die Anzahl der EEG-Kanäle unter Beibehaltung der Genauigkeitsrate auf weniger als die Hälfte zu reduzieren.

Im Hinblick auf zukünftige Arbeiten weist die vorgeschlagene Methode einige Probleme auf. Beispielsweise ist die vorgeschlagene Methode nicht in der Lage, neue Lösungen zu generieren, die die Genauigkeit erhöhen (d. h. sie bleibt in lokalen Minima stecken). Daher muss die vorgeschlagene Methode verbessert werden, um dieses Problem zu überwinden, indem der Mechanismus des BAT-Algorithmus geändert oder der BAT-Algorithmus mit einem anderen metaheuristischen Algorithmus hybridisiert wird, um die optimale Lösung finden zu können . Im Hinblick auf das zweite Problem konzentriert sich die aktuelle Version des BAT-Algorithmus darauf, die optimale Lösung zu finden, die die höchste Genauigkeitsrate bietet, und berücksichtigt nicht die Anzahl der EEG-Kanäle in der Zielfunktion. Daher werden wir in zukünftigen Arbeiten an der vorgeschlagenen multiobjektiven BAT-Version arbeiten, die **gleichzeitig die höchste Genauigkeit und die ni**

Beiträge der Autoren: Konzeptualisierung, MAA-B., ZAAA und NKA-Q.; Methodik, MAA-B., ZAAA und NKA-Q.; Software, MAA-B. und ZAAA; Validierung, MAA-B. und SNM; Untersuchung, MAA-B. und ZAAA; Ressourcen, MAA-B., ZAAA und CG; Schreiben – Originalentwurf , MAA-B., ZAAA, NKA-Q., SNM, NSA und CG; Visualisierung, MAA-B., ZAAA, SNM, NSA und CG. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu.

Finanzierung: Diese Arbeit wurde vom Dekanat für Forschung und Graduiertenstudien (DRG) der Ajman University, Ajman, VAE (Zuschuss-Nr. 2022-IRG-ENIT-7) unterstützt.

Datenverfügbarkeitserklärung: Die Daten sind im Artikel enthalten.

Interessenkonflikte: Autor Christoph Guger war bei der g.tec medical engineering GmbH angestellt. Die übrigen Autoren erklären, dass bei der Durchführung der Forschung keine kommerziellen oder finanziellen Beziehungen bestanden , die als potenzieller Interessenkonflikt ausgelegt werden könnten.

Verweise

1. **Li, F.; Fan, Y.; Zhang, X.; Wang, C.; Hu, F.; Jia, W.; Hui, H. Multi-Feature-Fusion-Methode basierend auf EEG-Signalen und ihre Anwendung in der Schlaganfallklassifizierung.** J. Med. Syst. **2020,** 44, **39.** [CrossRef]
2. **Khairunizam, W.; Yean, CW; Murugappan, M.; Junoh, AK; Razlan, ZM; Shahriman, A.; Mustafa, WAW; Ibrahim, Z.; Nurhafizah, S. Ein experimenteller Rahmen zur Beurteilung der Emotionen von Schlaganfallpatienten mithilfe eines Elektroenzephalogramms (E** Conf. Ser. **2020,** 1529, **052072.** [CrossRef]
3. **Carino-Escobar, RI; Carrillo-Mora, P.; Valdés-Cristerna, R.; Rodriguez-Barragan, MA; Hernandez-Arenas, C.; Quinzaños-Fresnedo, J.; Galicia-Alvarado, MA; Cantillo-Negrete, J. Längsschnittanalyse der Gehirnrhythmen von Schlaganfallpatienten während einer Intervention mit einer Gehirn-Computer-Schnittstelle.** Neural Plast. **2019,** 2019. [CrossRef] [PubMed]
4. **Grefkes, C.; Fink, GR Konnektivitätsbasierte Ansätze bei Schlaganfall und Funktionswiederherstellung.** Lancet Neurol. **2014,** 13, **206–216.** [Querverweis]
5. **Hachinski, V.; Donnan, GA; Gorelick, PB; Hacke, W.; Cramer, SC; Kaste, M.; Fisher, M.; Brainin, M.; Buchan, AM; Lo, EH; et al. Stroke: Arbeiten auf eine priorisierte Weltagenda hin.** Stroke **2010,** 41, **1084–1099.** [CrossRef] [PubMed]
6. **Gao, W.; Cui, Z.; Yu, Y.; Mao, J.; Xu, J.; Ji, L.; Kan, X.; Shen, X.; Li, X.; Zhu, S.; et al. Anwendung eines Brain-Computer-Interface- Systems mit visuellem und motorischem Feedback in der funktionellen Rehabilitation von Gliedmaßen und Gehirn nach einem Schlaganfall: Fallbericht.** Brain Sci. **2022,** 12, **1083.** [CrossRef]
7. **Jin, J.; Qu, T.; Xu, R.; Wang, X.; Cichocki, A. Motor Imagery EEG-Klassifizierung basierend auf Riemannscher Sparse-Optimierung und Dempster-Shafer-Fusion von Multi-Time-Frequency-Mustern.** IEEE Trans. Neural Syst. Rehabil. Eng. **2022,** 31, **58–67.** [CrossRef]
8. **Al-Qazzaz, NK; Alyasseri, ZAA; Abdulkareem, KH; Ali, NS; Al-Mhiqani, MN; Guger, C. EEG-Merkmalsfusion für motorische Bilder: Ein neuer robuster Rahmen für die Rehabilitation von Schlaganfallpatienten.** Berechnen. Biol. Med. **2021,** 137, **104799.** [CrossRef]
9. **Al-Timemy, AH; Bugmann, G.; Outram, N.; Escudero, J. Reduzierung von Klassifizierungsfehlern bei der myoelektrischen Steuerung von Handbewegungen durch unabhängige Komponentenanalyse. In Proceedings der 5. Internationalen Konferenz für Informationstechnologie , ICIT, Bali, Indonesien, 26.–28. Oktober 2011; Band 2011.**
10. **Al-Qazzaz, NK; Ali, SHBM; Ahmad, SA; Escudero, J. Optimale EEG-Kanalauswahl zur Identifizierung vaskulärer Demenz mithilfe eines verbesserten binären Gravitationssuchalgorithmus. In Proceedings der International Conference for Innovation in Biomedical Engineering and Life Sciences, Penang, Malaysia, 10.–13. Dezember 2017; S. 125–130.**
11. **Al-Qazzaz, NK; Sabir, MK; Ali, SHBM; Ahmad, SA; Grammer, K. Mehrkanaloptimierung mit hybriden Spektralentropiemarkern zur Verbesserung der Geschlechtsidentifikation von emotionsbasierten EEGs.** IEEE Access **2021, 9, 107059–107078.** [CrossRef]
12. **Velasco, I.; Sipols, A.; De Blas, CS; Pastor, L.; Bayona, S. Motorische Imagery EEG-Signalklassifizierung mit einem multivariaten Zeitreihenansatz .** Biomed. Eng. OnLine **2023,** 22, **29.** [CrossRef]

Machine Translated by Google

13. **Jouzizadeh, M. EEG-bewertete Netzwerk- und Signalvariabilitätsmerkmale bei Männern und Frauen während einer visuell-räumlichen Aufgabe. Ph.D. Dissertation, Université d'Ottawa/University of Ottawa, Ottawa, ON, Kanada, 2024.**
14. **Al-Qazzaz, NK; Sabir, MK; Ali, S.; Ahmad, SA; Grammer, K. Effektive EEG-Kanäle zur Emotionserkennung über die Gehirnregionen unter Verwendung eines Differentialevolutionsalgorithmus. In Proceedings der 41. jährlichen internationalen Konferenz der IEEE Engineering in Medicine and Biology Society (EMBC) 2019, Berlin, Deutschland, 23.–27. Juli 2019; S. 4703–4706.**
15. **Dornhege, G.; Millán, JdR; Hinterberger, T.; McFarland, DJ; Muller, KR** Toward Brain-Computer Interfacing; **MIT Press: Cambridge, MA, USA, 2007; Band 63.**
16. **Kingphai, K.; Moshfeghi, Y. Zur Kanalauswahl für die EEG-basierte Klassifizierung der mentalen Arbeitsbelastung. In Proceedings der International Conference on Machine Learning, Optimization, and Data Science, Grasmere, Großbritannien, 22.–26. September 2023; S. 403–417.**
17. **Liu, T.; Wu, Y.; Ye, A.; Cao, L.; Cao, Y. Zweistufiger spärlicher, multiobjektiver Evolutionsalgorithmus zur Optimierung der Kanalauswahl in BCIs.** Vorderseite. Hum. Neurosci. **2024,** 18, **1400077. [CrossRef] [PubMed]**
18. **Soleimani, B.; Dallasta, I.; Das, P.; Kulasingham, JP; Girgenti, S.; Simon, JZ; Babadi, B.; Marsh, EB Veränderte gerichtete funktionelle Konnektivität liegt der kognitiven Erholung nach einem Schlaganfall zugrunde.** Brain Commun. **2023, 5, fcad149. [CrossRef] [PubMed]**
19. **Seguin, C.; Sporns, O.; Zalesky, A. Kommunikation im Gehirnnetzwerk: Konzepte, Modelle und Anwendungen.** Nat. Rev. Neurosci. **2023,** 24, **557–574. [CrossRef] [PubMed]**
20. **Mang, J.; Xu, Z.; Qi, Y.; Zhang, T. Förderung des kognitiv-motorischen Prozesses im geschlossenen Kreislauf der BCI-vermittelten Wiederherstellung der motorischen Funktion nach einem Schlaganfall: Herausforderungen und Ansätze.** Vorderseite. Neurorobotics **2023,** 17, **1271967. [CrossRef] [PubMed]**
21. **Nunez, PL; Srinivasan, R.; Westdorp, AF; Wijesinghe, RS; Tucker, DM; Silberstein, RB; Cadusch, PJ EEG-Kohärenz: I: Statistik, Referenzelektrode, Volumenleitung, Laplace-Operatoren, kortikale Bildgebung und Interpretation auf mehreren Skalen.** Elektroenzephalograph. Clin. Neurophysiol. **1997,** 103, **499–515. [CrossRef] [PubMed]**
22. **Al-Qazzaz, NK; Hamid Bin Mohd Ali, S.; Ahmad, SA; Islam, MS; Escudero, J. Automatische Artefaktentfernung im EEG normaler und demente Personen, die ICA–WT während Arbeitsgedächtnisaufgaben verwenden.** Sensors **2017,** 17, **1326. [CrossRef] [PubMed]**
23. **Al-Qazzaz, NK; Ali, SHBM; Ahmad, SA; Islam, MS; Escudero, J. Unterscheidung von Schlaganfall-bedingten leichten kognitiven Beeinträchtigungen und vaskulärer Demenz mittels EEG-Signalanalyse.** Med. Biol. Eng. Comput. **2018,** 56, **137–157. [CrossRef] [PubMed]**
24. **Zhang, Y.; Nam, CS; Zhou, G.; Jin, J.; Wang, X.; Cichocki, A. Zeitlich beschränkte spärliche Gruppenraummuster für motorische Bildgebung BCI.** IEEE Trans. Cybern. **2018,** 49, **3322–3332. [CrossRef] [PubMed]**
25. **Chaudhary, S.; Taran, S.; Bajaj, V.; Siuly, S. Ein flexibler analytischer Ansatz auf Basis der Wavelet-Transformation für motorische Imagery-Aufgaben Klassifizierung in BCI-Anwendungen.** Comput. Methods Programs Biomed. **2020,** 187, **105325. [CrossRef]**
26. **Duan, L.; Hongxin, Z.; Khan, MS; Fang, M. Erkennung von motorischen Vorstellungsaufgaben für BCI unter Verwendung von CSP und chaotischem PSO-Zwillings-SVM.** J. China Univ. Posts Telecommun. **2017,** 24, **83–90. [CrossRef]**
27. **Alotaiby, T.; Abd El-Samie, FE; Alshebeili, SA; Ahmad, I. Ein Überblick über Kanalauswahlalgorithmen für die EEG-Signalverarbeitung.** EURASIP J. Adv. Signal Process. **2015,** 2015, **66. [CrossRef]**
28. **He, L.; Yu, Z.; Gu, Z.; Li, Y. Bhattacharyya-Grenzenbasierte Kanalauswahl zur Klassifizierung von Motorbildern in EEG-Signalen. In Proceedings der 2009 Chinese Control and Decision Conference, Guilin, China, 17.–19. Juni 2009; S. 2353–2356.**
29. **Tam, WK; Ke, Z.; Tong, KY Leistung eines gemeinsamen räumlichen Musters unter einem kleineren Satz EEG-Elektroden in der Gehirn-Computer-Schnittstelle bei Patienten mit chronischem Schlaganfall: Eine Studie mit mehreren Sitzungen und Datensätzen. In Proceedings der jährlichen internationalen Konferenz 2011 der IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30. August–3. September 2011; S. 6344–6347.**
30. **Gómez-López, JC; Escobar, JJ; Díaz, AF; Damas, M.; Gil-Montoya, F.; González, J. Verbesserung der Konvergenz eines GA-basierten Wrappers für Merkmalsauswahlprobleme bei hochdimensionalen Daten. In Proceedings of the Genetic and Evolutionary Computation Conference Companion, Boston, MA, USA, 9.–13. Juli 2022; S. 431–434.**
31. **Shen, C.; Zhang, K. Zweistufig verbesserter Grey Wolf-Optimierungsalgorithmus für die Merkmalsauswahl bei hochdimensionaler Klassifizierung.** Complex Intell. Syst. **2022, 8, 2769–2789. [CrossRef]**
32. **Tiwari, A.; Chaturvedi, A. Automatische EEG-Kanalauswahl für die Mehrklassen-Klassifizierung von Gehirn-Computer-Schnittstellen mittels Multiobjektiver verbesserter Firefly-Algorithmus.** Multimed. Tools Appl. **2023,** 82, **5405–5433. [CrossRef]**
33. **Millán, JdR; Franzé, M.; Mouriño, J.; Cincotti, F.; Babiloni, F. Relevante EEG-Merkmale für die Klassifizierung von spontanen motorische Aufgaben.** Biol. Cybern. **2002,** 86, **89–95. [CrossRef] [PubMed]**
34. **Martínez-Cagigal, V.; Santamaría-Vázquez, E.; Hornero, R. Optimierung der Kanalauswahl an der Schnittstelle zwischen Gehirn und Computer mithilfe von Metaheuristiken und evolutionären Algorithmen.** Appl. Soft Comput. **2022,** 115, **108176. [CrossRef]**
35. **Gaur, P.; Gupta, H.; Chowdhury, A.; McCreadie, K.; Pachori, RB; Wang, H. Ein gleitendes Fenster mit gemeinsamem räumlichen Muster zur Verbesserung der Klassifizierung motorischer Vorstellungen in EEG-BCI.** IEEE Trans. Instrum. Meas. **2021,** 70, **4002709. [CrossRef]**
36. **Shi, B.; Wang, Q.; Yin, S.; Yue, Z.; Huai, Y.; Wang, J. Ein binärer Harmoniesuchalgorithmus als Kanalauswahlmethode für Motor Bildbasiertes BCI.** Neurocomputing **2021,** 443, **12–25. [CrossRef]**
37. **Ghaemi, A.; Rashedi, E.; Pourrahimi, AM; Kamandar, M.; Rahdari, F. Automatische Kanalauswahl in EEG-Signalen zur Klassifizierung der Links- oder Rechtshandbewegung in Brain Computer Interfaces unter Verwendung eines verbesserten binären Gravitationssuchalgorithmus.** Biomed. Signal Process. Control **2017,** 33, **109–118. [CrossRef]**
38. **Lotte, F.; Congedo, M.; Lécuyer, A.; Lamarche, F.; Arnaldi, B. Eine Überprüfung der Klassifizierungsalgorithmen für EEG-basierte Gehirn-Computer Schnittstellen.** J. Neural Eng. **2007, 4, R1. [CrossRef]**
39. **Hassan, MM; Alam, MGR; Uddin, MZ; Huda, S.; Almogren, A.; Fortino, G. Erkennung menschlicher Emotionen mithilfe der Architektur eines Deep Belief Network.** Inf. Fusion **2019,** 51, **10–18. [CrossRef]**

Machine Translated by Google

40. Khosla, A.; Khandnor, P.; Chand, T. Eine vergleichende Analyse von Signalverarbeitungs- und Klassifizierungsmethoden für verschiedene Anwendungen basierend auf EEG-Signalen. Biocybern. Biomed. Eng. **2020,** 40, **649–690.** [CrossRef]

41. Altan, G.; Kutlu, Y.; Allahverdi, N. Auf tiefen Glaubensnetzwerken basierende Klassifizierung der Gehirnaktivität mittels EEG aus langsamen kortikalen Potentialen bei Schlaganfall. Int. J. Appl. Math. Electron. Comput. **2016, 4, 205–210.** [CrossRef]

42. García-Laencina, PJ; Rodríguez-Bermudez, G.; Roca-Dorda, J. Erforschung der Dimensionsreduktion von EEG-Merkmalen im Motor Klassifizierung von Bildgebungsaufgaben. Expert Syst. Appl. **2014,** 41, **5285–5295.** [CrossRef]

43. Aler, R.; Galván, IM Optimierung der Anzahl von Elektroden und Raumfiltern für Brain–Computer Interfaces mittels evolutionärer multiobjektiver Ansatz. Expert Syst. Appl. **2015,** 42, **6215–6223.** [CrossRef]

44. Lal, TN; Schroder, M.; Hinterberger, T.; Weston, J.; Bogdan, M.; Birbaumer, N.; Scholkopf, B. Support-Vektor-Kanalauswahl in BCI. IEEE Trans. Biomed. Eng. **2004,** 51, **1003–1010.** [CrossRef] [PubMed]

45. Rakotomamonjy, A.; Guigue, V. BCI-Wettbewerb III: Datensatz II-Ensemble von SVMs für BCI P300-Rechtschreibprüfung. IEEE Trans. Biomed. Eng. **2008,** 55, **1147–1154.** [CrossRef] [PubMed]

46. Tan, P.; Wang, X.; Wang, Y. Dimensionsreduktion in der auf evolutionären Algorithmen basierenden Merkmalsauswahl für motorische Vorstellungen Gehirn-Computer-Schnittstelle. Swarm Evol. Comput. **2020,** 52, **100597.** [CrossRef]

47. Tsai, CF; Eberle, W.; Chu, CY Genetische Algorithmen in der Merkmals- und Instanzauswahl. Knowl.-Based Syst. **2013,** 39, **240–247.** [Querverweis]

48. Udhaya Kumar, S.; Hannah Inbarani, H. PSO-basierte Merkmalsauswahl und auf Nachbarschafts-Rough-Sets basierende Klassifizierung für BCI Mehrklassen-Motorbildaufgabe. Neural Comput. Appl. **2017,** 28, **3239–3258.** [CrossRef]

49. Baig, MZ; Aslam, N.; Shum, HP; Zhang, L. Differentialevolutionsalgorithmus als Werkzeug zur optimalen Merkmalsauswahl in Motorische Imagery EEG. Expert Syst. Appl. **2017,** 90, **184–195.** [CrossRef]

50. Atyabi, A.; Luerssen, M.; Fitzgibbon, S.; Powers, DM Evolutionäre Merkmalsauswahl und Elektrodenreduktion für die EEG-Klassifizierung. In Proceedings des 2012 IEEE Congress on Evolutionary Computation, Brisbane, QLD, Australien, 10.–15. Juni 2012; S. 1–8.

51. Nakisa, B.; Rastgoo, MN; Tjondronegoro, D.; Chandran, V. Evolutionäre Berechnungsalgorithmen für die Merkmalsauswahl von EEG-basierte Emotionserkennung mit mobilen Sensoren. Expert Syst. Appl. **2018,** 93, **143–155.** [CrossRef]

52. Al-Qazzaz, NK; Ali, SHM; Ahmad, SA Differential Evolution Based Channel Selection Algorithm on EEG Signal for Early Detection of Vascular Dementia among Stroke Survivors. In Proceedings der 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Sarawak, Malaysia, 3.–6. Dezember 2018; S. 239–244.

53. Alyasseri, ZAA; Khader, AT; Al-Betar, MA; Joao, PP; Osama, AA EEG-basierte Personenauthentifizierung mithilfe eines multiobjektiven Blütenbestäubungsalgorithmus. In Proceedings of the Evolutionary Computation (CEC), Rio de Janeiro, Brasilien, 8.–13. Juli 2018; S. 15–22.

54. Yang, XS Ein neuer metaheuristischer, von Fledermäusen inspirierter Algorithmus. In Nature Inspired Cooperative Strategies for Optimization (NICSO 2010); **Springer: Berlin/Heidelberg, Deutschland, 2010; S. 65–74.**

55. Hyvarinen, A. Schnelle und robuste Fixpunktalgorithmen für die unabhängige Komponentenanalyse. IEEE Trans. Neural Netw. **1999,** 10, **626–634.** [CrossRef] [PubMed]

56. Escudero, J.; Hornero, R.; Abásolo, D.; Fernández, A. Blinde Quellentrennung zur Verbesserung spektraler und nichtlinearer Merkmale von Magnetoenzephalogramm-Aufzeichnungen. Anwendung bei der Alzheimer-Krankheit. Med. Eng. Phys. **2009,** 31, **872–879.** [CrossRef] [Pub

57. James, CJ; Hesse, CW Unabhängige Komponentenanalyse für biomedizinische Signale. Physiol. Meas. **2005,** 26, **R15.** [CrossRef] [PubMed]

58. Al-Qazzaz, NK; Hamid Bin Mohd Ali, S.; Ahmad, SA; Islam, MS; Escudero, J. Auswahl von Mutter-Wavelet-Funktionen für die mehrkanalige EEG-Signalanalyse während einer Arbeitsgedächtnisaufgabe. Sensors **2015,** 15, **29015–29035.** [CrossRef] [PubMed]

59. Escudero, J.; Hornero, R.; Abásolo, D.; Fernández, A. Quantitative Bewertung der Artefaktentfernung im realen Magnetenzephalogramm Signale mit blinder Quellentrennung. Ann. Biomed. Eng. **2011,** 39, **2274–2286.** [CrossRef] [PubMed]

60. Barbati, G.; Porcaro, C.; Zappasodi, F.; Rossini, PM; Tecchio, F. Optimierung eines unabhängigen Komponentenanalyseansatzes zur Artefaktidentifizierung und -entfernung in magnetoenzephalographischen Signalen. Clin. Neurophysiol. **2004,** 115, **1220–1232.** [CrossRef]

61. Mammone, N.; La Foresta, F.; Morabito, FC Automatische Artefaktunterdrückung aus mehrkanaligem Kopfhaut-EEG durch Wavelet-ICA. IEEE Sens. J. **2011,** 12, **533–542.** [CrossRef]

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultieren.

Machine Translated by Google

Algorithmen

MDPI

Artikel

Klassifizierung und Regression von Nadelstichkorrosionen an Rohrleitunge

Basierend auf magnetischen Streuflusssignalen mittels Faltungsalgorithmus Neuronale Netze

Yufei Shen und Wenxing Zhou *

Fakultät für Bau- und Umweltingenieurwesen, University of Western Ontario, London, ON N6A 5B9, Kanada *

Korrespondenz: wzhou@eng.uwo.ca

Zusammenfassung: Nadelstichkorrosionen an Öl- und Gaspipelines sind schwer zu erkennen und zu dimensionieren und stellen daher eine erhebliche Herausforderung für die Praxis des Integritätsmanagements von Pipelines dar. Diese Studie entwickelt zwei Convolutional Neural Network (CNN)-Modelle zur Erkennung von Nadelstichen und zur Vorhersage der Größe und Lage der Nadelstichkorrosionen anhand der magnetischen Flusslecksignale, die mithilfe der magnetostatischen Finite-Elemente-Analyse generiert werden. Um die beiden CNN-Modelle zu trainieren und zu validieren, werden umfangreiche dreidimensionale parametrische Finite-Elemente-Analysefälle generiert. Darüber hinaus wird die Modellleistung durch eine umfassende Algorithmenanalyse bewertet, die Einblicke in die praktische Anwendung von CNN-Modellen im Integritätsmanagement von Pipelines bietet. Das vorgeschlagene Klassifizierungs-CNN-Modell ist bei der Klassifizierung von Nadelstichen und Nadelstichkorrosionsdefekten im Allgemeinen äußerst genau. Das vorgeschlagene Regressions-CNN-Modell ist bei der Vorhersage der Lage des Nadelstichs äußerst genau und erreicht eine relativ hohe Genauigkeit bei der Schätzung der Tiefe und des Durchmessers des Nadelstichs, selbst bei Vorhandensein von Messrauschen. Diese Studie zeigt die Wirksamkeit des Einsatzes von Deep

Schlüsselwörter: Deep-Learning-Algorithmus; Convolutional Neural Network; Pipeline; Korrosion; Lochblende; Finite-Elemente-Analyse; magnetisches Streuflusssignal

check for updates

Zitat: Shen, Y.; Zhou, W. Klassifizierung und Regression von Nadelstichkorrosionen an Rohrleitungen auf Basis von magnetischen Streuflusssignalen unter Verwendung von Convolutional Neural Networks. Algorithmen **2024,** 17, **347. https://doi.org/10.3390/a17080347**

Wissenschaftliche Herausgeber: Chih-Lung Lin, Bor-Jiunn Hwang, Shaou-Gang Miaou und Chi-Hung Chuang

Empfangen: 24. Juni 2024
Überarbeitet: 1. August 2024
Akzeptiert: 6. August 2024
Veröffentlicht: 8. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/ 4.0/).

1. Einleitung

Pipelines sind ein kostengünstiges und sicheres Transportmittel für große Mengen an Öl- und Gasprodukten über weite Entfernungen [1]. Allerdings können Pipeline-Ausfälle schwerwiegende Folgen haben, beispielsweise Todesfälle, wirtschaftliche Verluste und Umweltschä Korrosion ist eine der größten Bedrohungen für die strukturelle Integrität von Pipelines unter den verschiedenen Ausfallmechanismen [3]. Das In-Line-Inspektionswerkzeug (ILI) wird häufig zum Erkennen und Messen von Korrosion eingesetzt, um die Integrität von Pipelines zu beurteilen [4], und die magnetische Flussleckage (MFL) ist die vorherrschende ILI-Technologie zum Erkennen von inneren und äußeren Korrosionsfehlern in Flüssigkeits- und Gaspipelines [5]. Das Pipeline Operators Forum (POF) schlägt vor, eine einzelne Korrosionsanomalie auf der Innen- oder Außenfläche einer Pipeline auf Grundlage der vom ILI gemeldeten Korrosionslänge und -breite in sieben Klassen einzuteilen, wie in Abbildung 1 dargestellt [6], d. h. axiale Schlitzung, Umfangsschlitzung, axiale Rillenbildung, Umfangsrillenbildung , allgemein, Nadellöcher und Lochfraß, um die Genauigkeit von ILI-Werkzeugen besser zu quantifizieren und die Auswahl geeigneter Modelle zur Bewertung der Berstkapazität von Pipelines mit

Die Schätzung der Korrosionsanomaliegrößen auf der Grundlage von MFL-Signalen fällt in den Bereich der inversen Modellierung. Die Pipeline-Industrie verwendet häufig das iterative inverse Modell aufgrund seiner bemerkenswerten Präzision [8–10]. Das iterative inverse Modell ist jedoch rechnerisch aufwändig. Inverse Modelle auf der Grundlage von maschinellem Lernen bieten vielversprechende Alternativen zur Größenbestimmung von Korrosionsfehlern aus MFL-Signalen mit hoher Genauigkeit, sind aber rechnerisch effizienter als die iterativen inversen Modelle [11–13]. Zum Beispiel

Machine Translated by Google

Kandroodi et al. [14] speisten die algorithmisch geschätzte Defektbreite und -länge sowie die extrahierten Peak-to-Peak-Werte aus den MFL-Signalen in ein neuronales Netzwerk mit radialer Basisfunktion und Gaußscher Verteilung ein, um die Korrosionstiefe vorherzusagen. Feng et al. [15] schlugen eine Fehlerkorrekturmethode zur Rekonstruktion der Korrosionsprofile aus MFL-Signalen auf der Grundlage eines neuronalen Netzwerks mit radialer Basisfunktion vor. Diese Modelle haben jedoch bestimmte Einschränkungen: Das Training dieser Modelle basiert auf manuell ausgewählten Merkmalen aus den MFL-Signalen, wodurch möglicherweise wertvolle Daten weggelassen werden, die für eine Darüber hinaus konzentriert sich die Mehrheit dieser Techniken lediglich auf eine MFL-Signaldimension und vernachlässigt die wichtigen Informationen, die in den anderen beiden Dimensionen vorhanden sind.

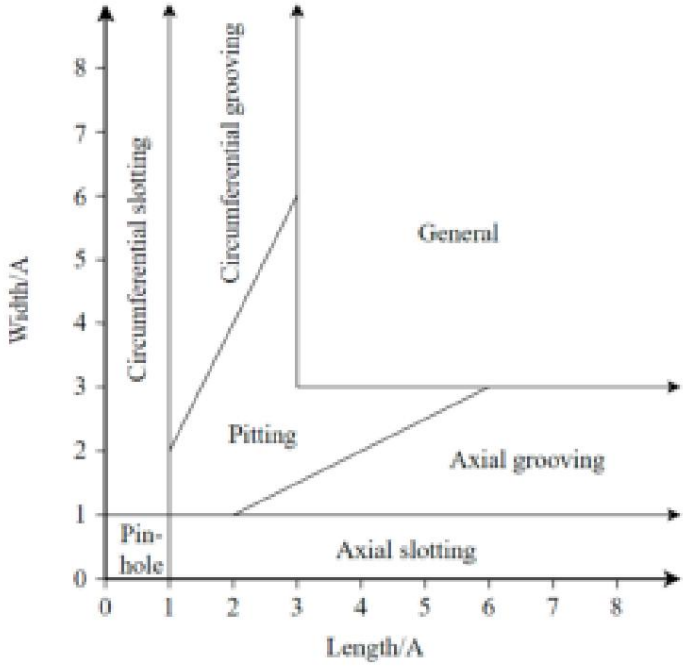


Abbildung 1. Sieben Korrosionsanomaliekategorien basieren auf der Länge und Breite der Anomalie. Hinweis: A = max. {10 mm, Rohrwandstärke}.

In den letzten zehn Jahren haben sich Convolutional Neural Networks (CNNs) als revolutionärer Deep-Learning-Algorithmus etabliert und bemerkenswerte Ergebnisse in verschiedenen Bereichen der Mustererkennung geliefert, darunter auch in der Bild- und Sprachverarbeitung [16,17]. Das CNN-Tool kann anstelle manuell entwickelter Merkmalsextraktoren automatisch tiefe Merkmale aus den Eingabedaten extrahieren und eignet sich daher gut zur Lösung der inversen Modellierung auf Basis von MFL- Signalen. Über die Anwendung von CNN in der inversen Analyse von MFL-Signalen wurde in der Literatur berichtet. Lu et al. [18] führten das visuelle Transformations-CNN ein, das Länge, Breite und Tiefe von Anomalien vorhersagt, indem es zwei Komponenten (d. h. die Umfangs- und Radialkomponente) des MFL-Signals berücksichtigt und die Längskomponente außer Acht lässt. Shen und Zhou [19] verwendeten CNN, um die Lage und Ausmaße von Korrosionen an Stahlpipelines anhand von drei Komponenten der MFL-Signale abzuschätzen. Wang und Chen [20] schlugen eine CNN-Architektur mit zwei Hauptmodulen vor: eines für die Anomalieklassifizierung und ein weiteres für die Anomaliegrößenregression. Die Eingabe des Klassifizierungsmoduls besteht aus allen drei Komponenten des MFL-Signals, und die Ausgaben sind sieben Anomaliekategorien, wie in Abbildung 1 dargestellt. Das Regressionsmodul umfasst sieben separate CNNs, von denen jedes auf einen bestimmten Anomalietyp zugeschnitten ist und Abmessungen wie Breite, Länge und Tiefe des Defekts vorhersagt.

Laut einem Bericht der European Gas Pipeline Incident Data Group (EGIG) [21] wird Nadelstichkorrosion als Hauptursache für Pipeline-Ausfälle identifiziert, da sie sich von einem anfänglich geringen Metallverlust zu einem größeren Metallverlust entwickeln kann, da die Erkennung durch Drucküberwachung schwierig ist, insbesondere wenn die ausgetretenen Produkte nicht über die Bodenebene hinausgehen [22,23]. Trotz der Auswirkungen von Nadelstichkorrosi

Machine Translated by Google

die sich speziell mit der Erkennung und Größenbestimmung von Nadelstichkorrosionen befassen, sind in der Literatur nur begrenzt vorhanden, sodass eine Wissenslücke entsteht. Darüber hinaus können sich kleine Korrosionsanomalien wie Nadelstichkorrosionen manchmal mit einem relativ großen Korrosionsbereich überlappen [23–25], wodurch ein komplexer Defekt entsteht, der in dieser Studie als Nadelstichkorrosionsdefekt (PIC) bezeichnet wird. Es gibt in der Literatur auch nur wenige Studien zur Erkennung

Das Ziel dieser Studie ist die Entwicklung von Deep-Learning-Algorithmen, d. h. CNN-Modellen, zur Klassifizierung und Größenbestimmung von Nadellöchern und PIC-Defekten auf der Grundlage einer großen Menge von MFL-Signalen. Das Klassifizierungs-CNN-Modell befasst sich mit sechs Szenarien, d. h. Nadellöchern auf der inneren oder äußeren Rohroberfläche, allgemeinen Korrosionen auf der inneren oder äußeren Oberfläche und PIC-Defekten auf der inneren oder äußeren Oberfläche. Die Signale für die vom Klassifizierungsmodell identifizierten Nadellöcher und PIC-Defekte werden dann in das Regressions-CNN-Modell eingespeist, um die Größen und Positionen der Nadellöcher und PIC-Defekte vorherzusagen. In der Praxis ist es nicht erforderlich, die beiden CNN- Modelle nacheinander anzuwenden. Abhängig von der spezifischen Anwendung können Daten direkt in das Regressionsmodell eingespeist werden, wenn die Klassifizierung bereits auf andere Weise durchgeführt wurde und das Hauptziel darin besteht, die Größe des Defekts vorherzusagen. Da echte MFL-Signale typischerweise proprietär und nicht öffentlich zugänglich sind, werden die in dieser Studie verwendeten MFL-Signale numerisch generiert, indem dreidimensionale (3D) magnetostatische Finite-Elemente-Analysen (FEA) mit der kommerziellen Software COMSOL (Version 5.6) angewendet werden. Die dreidimensionalen (d. h. longitudinalen, zirkulären und radialen) MFL-Signalkarten dienen als Eingaben für das Trainieren und Validieren von CNN-Modellen, um die Defekte zu klassifizieren und die Größe sowie die Umfangs- und Längspositionen von Nadellöchern in der Pipeline vorherzusagen. Die Auswirkungen der mit dem MFL-Signal verbundenen Messgeräusche auf die Genauigkeit des Regressions-CNN-Modells werden ebenfalls untersucht. Diese Forschung schließt eine Wissenslücke in der Literatur, indem CNN angewendet wird, um Nadellöcher und Nadelloch-in-Korrosionsdefekte zu erkennen, zu lokalisieren und zu dimensionieren. Es wird darauf hingewiesen, dass die genaue Identifizierung der Position des Nadellochs innerhalb der allgemeinen Korrosion von entscheidender Bedeutung ist, da die relative Position des Nadellochs innerhalb der allgemeinen Korrosion einen deutlichen Einfluss auf die Berstkapazität des PIC-Defekts hat und daher starke Auswirkungen auf die genaue Vorhersage der Berstkapazität hat. Diese Studie demonstriert die Wirksamkeit vo

Der Rest des Dokuments ist wie folgt strukturiert. Abschnitt 2 stellt die Prinzipien der MFL-Technologie und wesentliche Hintergrundinformationen zur Beziehung zwischen MFL-Signalen und Anomaliemerkmalen vor. Abschnitt 3 stellt das vorgeschlagene FEA-Modell und die Simulationsparameter von MFL-Signalen vor. Abschnitt 4 beschreibt die Struktur der entwickelten CNN-Modelle und ihre Anwendung zur Klassifizierung und Regression von Nadellöchern. Die entsprechende Vorhersagegenauigkeit der CNN-Modelle wird ebenfalls in Ab Abschnitt 5 fasst die abschließenden Bemerkungen der Studie zusammen.

2. Grundsätze der MFL-Technik

Die magnetostaktische FEA wird in der Studie zur numerischen Generierung der MFL-Signale eingesetzt. Die entsprechende aus den Maxwell-Gleichungen abgeleitete Grundgleichung ist in Gleichung (1) dargestellt:

$$\ddot{y} \times \quad \frac{1}{\mu} \ddot{y} \times \mathbf{A} = \mathbf{0} \quad (1)$$

wobei ÿ× die Curl-Operation bezeichnet; μ stellt die magnetische Permeabilität der ferromagnetischen Probe (N/A2) dar und A ist der magnetische Potentialvektor. Gleichung (1) wird unter den angegebenen Randbedingungen gelöst, um A zu erhalten, das dann verwendet wird, um die magnetische Flussdichte B wie folgt zu bestimmen [26,27]:

$$\mathbf{B} = \ddot{y} \times \mathbf{A} \quad (2)$$

Die MFL-Technik basiert auf dem Messprinzip, bei dem ein ferromagnetisches Material in der Nähe eines Defekts MFL aufweist, wie in Abbildung 2 dargestellt [18]. Das Messverfahren läuft wie folgt ab: Während des ILI-Prozesses wird das MFL-Werkzeug angetrieben

Machine Translated by Google

durch die Rohrleitung, und Magnete im Werkzeug erzeugen ein Magnetfeld. In einem perfekten Rohr bleibt der gesamte magnetische Fluss innerhalb der Rohrwand [28], was zu einer gleichmäßigen Magnetfeldverteilung führt [29]. Bei Defekten tritt jedoch ein Teil des magnetischen Flusses aus dem Rohr aus und bildet in der Nähe des Defekts ein magnetisches Streufeld [30]. Dieses Streufeld wird von Hall-Sensoren erkannt, die elektrische Signale erzeugen [29]. Stärke und Verteilung des Streuflusses variieren je nach Geometrie des Defekts auf der Rohrleitung. Durch Erfassen des MFL-Signals, das typischerweise aus einem dreidimensionalen Vektorfeld der magnetischen Flussdichte (B) besteht, das durch Bx **in Umfangsrichtung des Rohrs,** By **in radialer Richtung und** Bz **in Längsrichtung (axial) dargestellt wird, können Größe und Lage des Korrosionsdefekts geschätzt werden.**

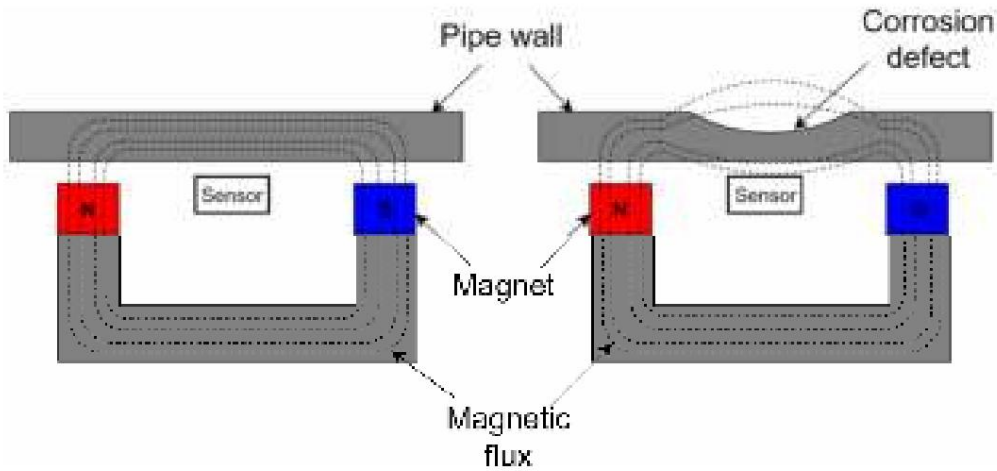


Abbildung 2. Prinzip der MFL-Technik.

Axiale (AMFL) und zirkuläre (CMFL) MFL-Werkzeuge sind die beiden Arten von MFL-Werkzeugen, die nach der Anordnung der Magnete im Inneren des MFL-Werkzeugs kategorisiert werden. Bei AMFL-Werkzeugen sind die Permanentmagnete parallel zur Längsachse des Rohrs ausgerichtet. Diese Werkzeuge eignen sich hervorragend zum Erkennen und Messen von Defekten in Umfangsrichtung des Rohrs, sind jedoch weniger effektiv bei der Größenbestimmung axial ausgerichteter Merkmale [31]. Im Gegensatz dazu sind bei CMFL-Werkzeugen die Magnete in Umfangsrichtung der Rohrleitung angeordnet. Sie sind **im Vergleich zu AMFL-Werkzeugen beim Messen von Merk**
Der Schwerpunkt dieser Studie liegt auf dem CMFL-Tool, da es bei der Größenbestimmung von längsorientierten Korrosionsanomalien, die senkrecht zur Umfangsspannung stehen, die sich aus dem Innendruck der Leitung ergibt, genauer ist als das AMFL-Tool und daher einen großen Einfluss auf die Berstfestigkeit der Rohrleitung haben kann.

3. Simulation von MFL-Signalen mit FEA 3.1.
Simulationsparameter

Das in dieser Studie betrachtete CMFL-Tool umfasst drei Paare Permanentmagnete.
Als magnetisches Material wird NdFeB mit einer hohen Koerzitivfeldstärke von 895.000 A/m, stabilen magnetischen Eigenschaften und hoher magnetischer Energie ausgewählt **[32]. Der zur Simulation des Rohrs verwendete Stahl wird als Güteklasse X52 angenommen, gekennzeichnet durch eine Remanentflussdichte** (Br) **von 1T, eine elektrische Leitfähigkeit von 5,882 × 106 S/m, eine Koerzitivfeldstärke** (Hc) **von 415 A/m und eine Sättigungsflussdichte** (Bs) **von 1,9T. Abbildung 3 zeigt schematisch einen Querschnitt des CMFL-Werkzeugs und des Rohrabschnitts, und die geometrischen und Materialeigenschaften der einzelnen Komponenten sind in Tabelle 1 zusammengefasst. Es sei darauf hingewiesen, dass der Lift-off-Wert mit 5 mm angenommen wird, was der Entfernung zwischen dem Sensor und der Innenfläche des Rohrs ents**

Machine Translated by Google

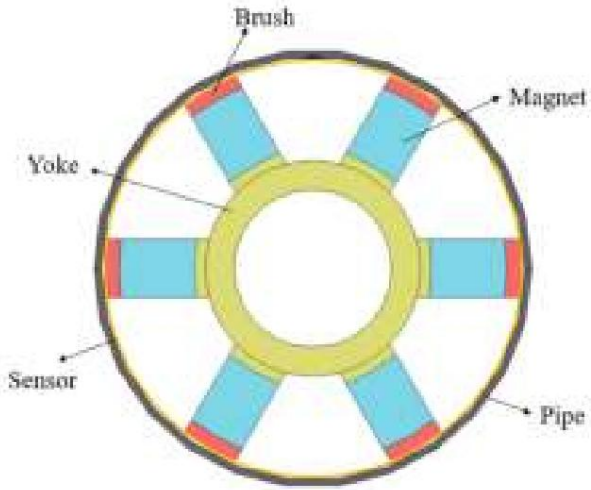


Abbildung 3. Abbildung des CMFL-Tools.

Tabelle 1. Geometrie jeder Komponente im vorgeschlagenen 3D-CMFL-Modell.

Elemente	Rohrabschnitt	Magnete	Pinsel	Joch
Geometrie (mm)	600 × 610 × 10 a Länge [A]	b 300 × 81 × 105	× 81 × 20	b 300 × 300 × 40 300 [A]

[A] × Außendurchmesser × Wandstärke. [B] Länge × Breite × Dicke.

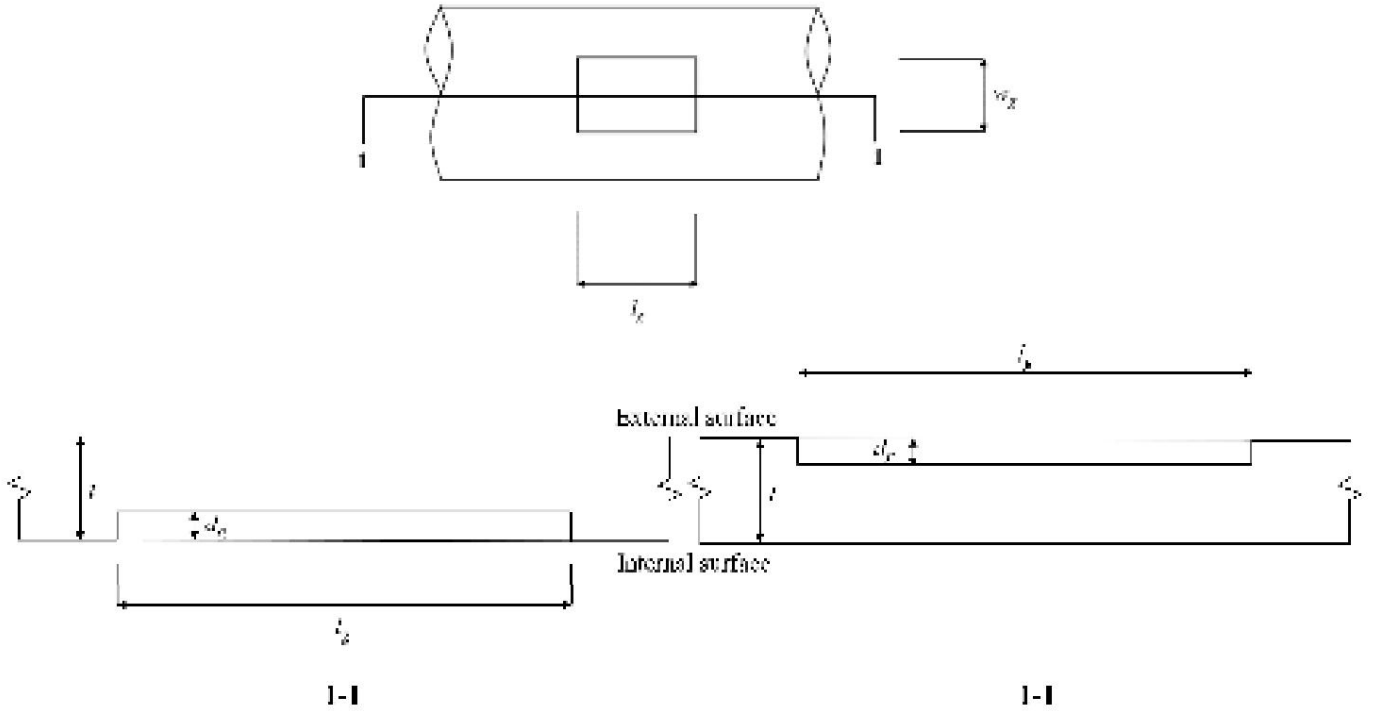
3.2. FEA-Fälle

Eine Reihe von FEA-Modellen wird mit den gleichen Magnetisierungsbedingungen und Rohrattributen erstellt und analysiert , sie unterscheiden sich jedoch in Korrosionsarten, geometrischen Par Standorte in dieser tt Studie. Die Simulation umfasst 9600 Analysefälle, wobei 1600 Fälle jedem der sechs Szenarien zugeordnet sind, wie in Abbildung 4 dargestellt, nämlich interne allgemeine Korrosio (Gin), äußere allgemeine Korrosion (Gex), inneres Nadelloch (Pin), äußeres Nadelloch (Pex), inneres PIC-Defekt (PICin) und externer PIC-Defekt (PICex). Die allgemeine Korrosion wird idealisiert als quaderförmig, definiert durch die Breite (wg), Tiefe (dg) und Länge (lg), und das Loch ist idealisiert als zylindrisch geformt, definiert durch den Radius (rp) und die Tiefe (dp). Beachten Sie das Zentrum Es wird angenommen, dass die Korrosionserscheinungen des Nadellochs in einem PIC-Defekt mit denen der allgemeinen Korrosion übereinstimmen. die durch zwei Koordinaten definiert ist, nämlich ihre Längsposition (h) und ihre Umfangsposition Ort (ÿ), innerhalb eines 160 × 160 mm großen Bereichs, wie in Abbildung 5 dargestellt. Dieser Bereich ist hellblau markiert, dessen Mittelpunkt durch die Koordinaten (r, ÿ = 0, h = 0) angegeben ist, wobei r stellt den Innenradius des Rohres dar. Diese spezielle 160 × 160 mm große quadratische Fläche auf dem Die innere Rohrwand wurde gewählt, da sie die größte Fläche einnimmt und eine gleichbleibende magnetisches Flussdichteprofil zwischen den benachbarten Magneten.

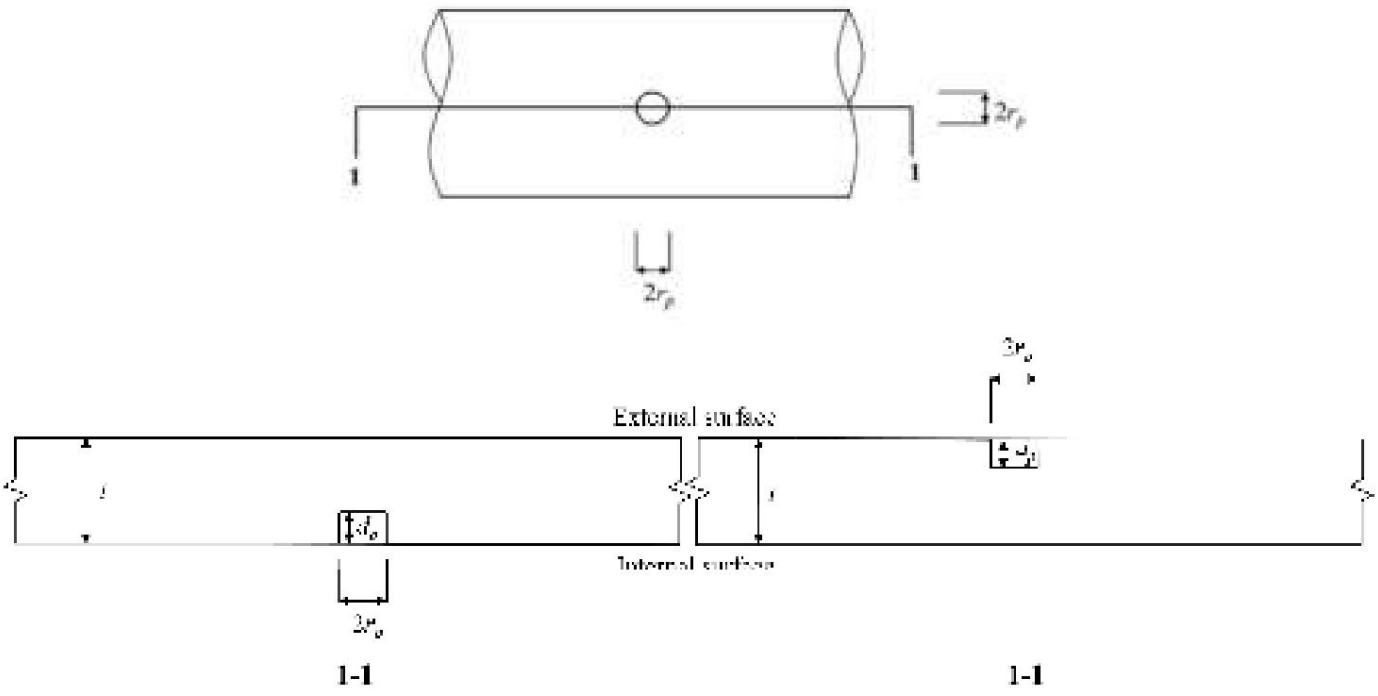
Die Analysefälle, die nur allgemeine Korrosion betreffen, d. h. Gin und Gex, umfassen vier Werte von wg, d. h. 30, 50, 70 und 90 mm, vier Werte von dg/t, d. h. 20, 40, 60 und 80 %, wobei t bezeichnet die Rohrwanddicke und vier Werte von lg, nämlich 30, 50, 70 und 90 mm. Daraus ergibt sich in insgesamt 64 allgemeinen Korrosionen mit unterschiedlichen Größen. In Bezug auf den Korrosionsort wurden fünf Werte von h (ÿ50, ÿ25, 0, 25 und 50 mm) und fünf Werte von ÿ (ÿ10, ÿ5, 0, 5 und 10 Grad) werden berücksichtigt. Die Permutationen von 64 Defekten und 25 Stellen führen zu 1600 Fällen für jeden von Gin und Gex.

Die Analysefälle, die nur Nadelstichkorrosion beinhalten, also Pin und Pex, umfassen acht Werte von rp, d. h. 1, 1,5, 2, 2,5, 3, 3,5, 4 und 4,5 mm, und acht Werte von dp/t, d. h. 10, 20, 30, 40, 50, 60, 70 und 80 %. Insgesamt ergeben sich 64 Nadelstichkorrosionen in unterschiedlichen Größen. In Bezug auf den Korrosionsort wurden fünf Werte von h (ÿ50, ÿ25, 0, 25 und 50 mm) und fünf Werte von ÿ (ÿ10, ÿ5, 0, 5 und 10 Grad) werden berücksichtigt. Die Permutation von 64 Defekten und 25 Standorte führen auch zu jeweils 1600 Fällen für Pin und Pex.

Machine Translated by Google



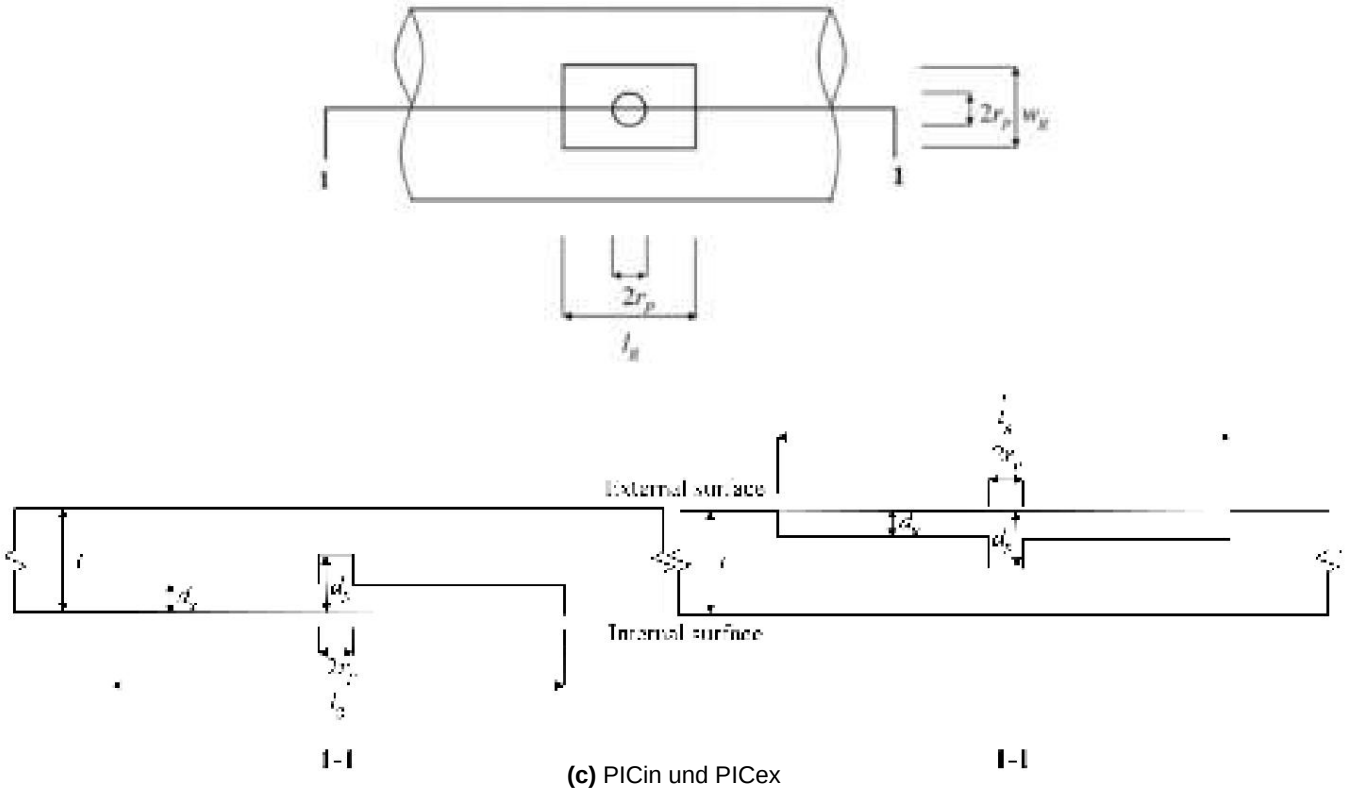
a) Gin und Gex



(b) Pin und Pex

Abbildung 4. Fortsetzung

Machine Translated by Google



(c) PICin und PICex

Abbildung 4. Sechs Arten von Korrosionssituationen: (a) Gin und Gex; (b) Pin und Pex; (c) PICin und PICex.

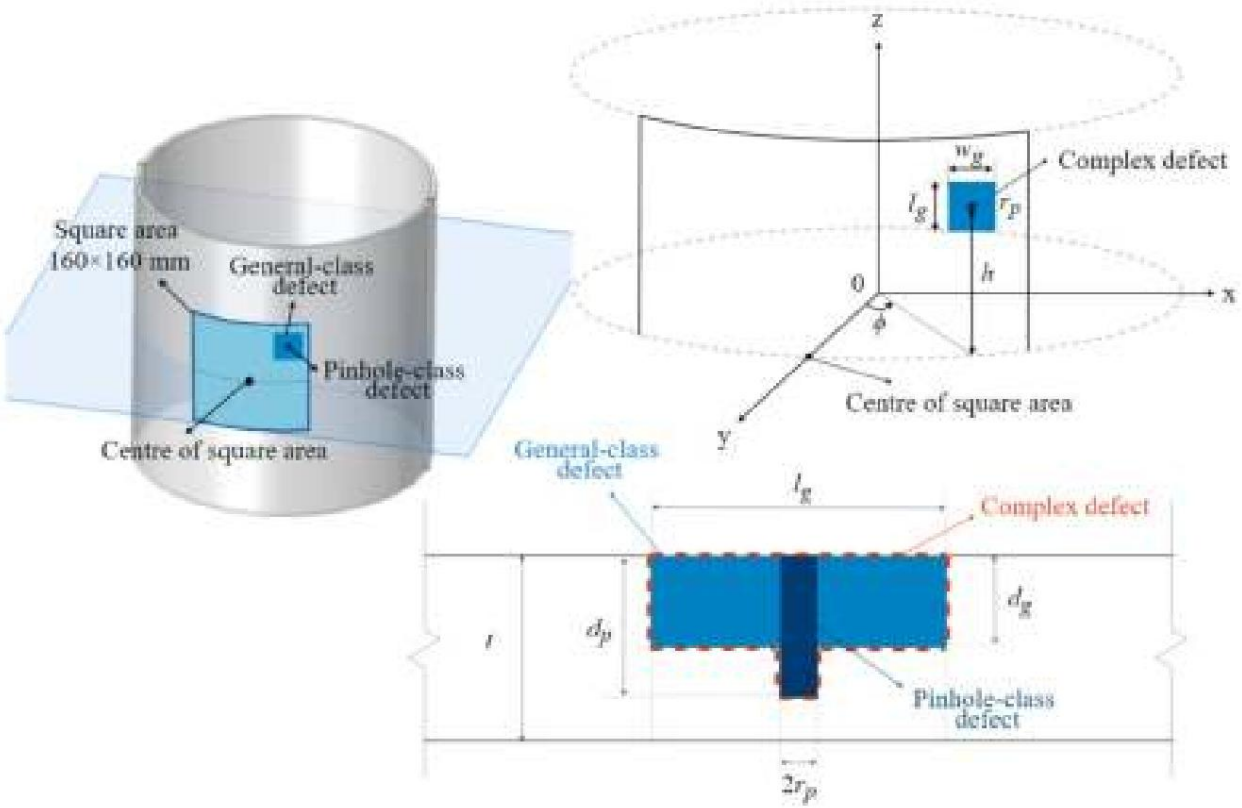


Abbildung 5. Zylindrisches Koordinatensystem, das die Lageparameter h **und ÿ definiert.**

ÿ ÿ ÿ

Machine Translated by Google

Analysefälle mit PIC-Defekten, d. h. PICin und PICex, umfassen wg **=** lg **= 50 mm und vier** rp-Werte, **d. h. 1, 2, 3 und 4 mm. In Bezug auf die Größen allgemeiner und nadelstichartiger Korrosion werden drei Gruppen berücksichtigt, um ein breiteres Spektrum relativer Beziehungen zwischen allgemeiner und nadelstichartiger Korrosion zu erfassen. Die erste umfasst** dg/t **gleich 20 % und dp/t gleich 30, 40, 50, 60, 70 oder 80 %. Die zweite Gruppe umfasst** dg/t **gleich 30 % und dp/t gleich 40, 50, 55, 60, 70 oder 80 %. Die letzte Gruppe umfasst** dg/t **gleich 40 % und dp/t gleich 50, 60, 70 oder 80 %. Es ist zu beachten, dass dp/t in** *ÿ ÿ* **allen drei Gruppen mindestens 10 % größer als** dg/t angenommen wird. Dies führt zu insgesamt 64 PIC-Defekten mit unterschiedlichen Größen. In Bezug auf den **Korrosionsort werden fünf Werte von** h **(ÿ50, ÿ25, 0, 25 und 50 mm) und fünf Werte von** *ÿ* **(ÿ10, ÿ5, 0, 5 und 10 Grad) berücksichtigt. Die Permutation von 64 Defekten und 25 Orten führt zu jeweils 1600 Fällen für PICin und PICex. Daraus folgt, dass in dieser Studie ein Datensatz bestehend aus 9600 Fällen (1600 × 6 = 9600) generiert wird. Dieser große Datensatz soll das Training und die Validierung des CNN-Modells erleichtern und eine umfassende Analyse und robuste Modellentwicklung ermöglichen.**

Das entwickelte FEA-Modell verwendet Maschen mit 4-Knoten-Tetraederelementen, wobei die Mindestgröße 4 mm beträgt und die maximale Maschengröße nach einer Konvergenzstudie 55 mm beträgt. Insbesondere sind die Gitter innerhalb der Fläche von 160 mm x **160 mm in 1-mm-Intervallen fein aufgelöst. An jedem Gitterpunkt wird der Wert von B durch Interpolation zwischen zwei Netzknoten berechnet. Das System nichtlinearer Gleichungen wird mit der Newton-Raphson-Methode gelöst. Zur Veranschaulichung zeigt Abbildung 6 die dreidimensionalen Informationen zur magnetischen Flussdichte (d. h.** Bx, By **und** Bz) **innerhalb des 160 mm × 160 mm großen Bereichs eines ausgewählten FE-Modells. Das vorgeschlagene FEA-Modell wurde validiert, indem die Simulationsergebnisse mit experimentellen Daten eines** von Ireland und Torres **[34] gemeldeten CMFL-Tools und eines von Azizzadeh und Safizadeh [35] gemeldeten AMFL-Tools verglichen wurden. Für Ei**

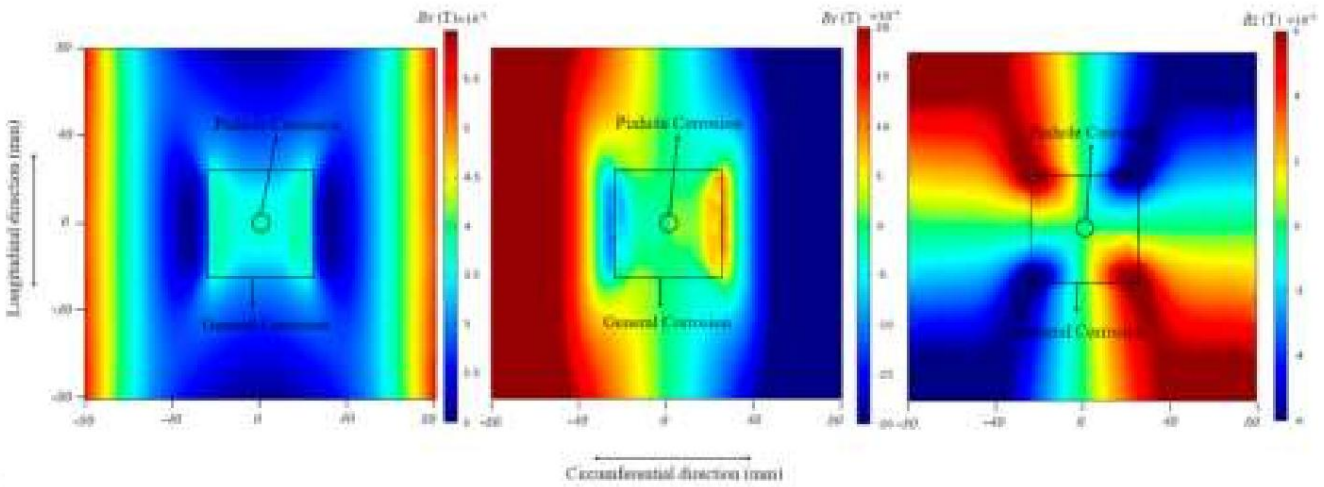


Abbildung 6. Durch FEA ermittelte Bx-, By- **und** Bz- **Werte entsprechend einem PIC-Defekt auf der inneren Rohroberfläche** (wg **= 50 mm,** lg **= 50 mm,** dg/t **= 40 %,** rp **= 4 mm, dp/t = 80 %,** h **= 0 mm und** *ÿ* **= 0 Grad).**

MFL-Signale in praktischen Anwendungen enthalten zwangsläufig Rauschen, das von Faktoren wie ungleichmäßiger Rohrwanddicke und Unterschieden in den Sensor-Abhebewerten herrührt **[36]. Dieses Rauschen kann die Genauigkeit des inversen Modells beeinträchtigen, sodass Rauschen in die FEA-simulierten MFL-Signale einbezogen werden muss. Das Signal-Rausch-Verhältnis (SNR) dient als wichtiger Parameter zur Quantifizierung des Rauschens im dreidimensionalen MFL-Signal. Gleichung (3) definiert das SNR in dieser Studie [37]:**

$$\text{SNR} = \mathbf{10log10}\ \frac{\sum_{n=1}^{160}\sum_{m=1}^{160}\left[Bx(n,m)^2 + Durch(n,m)^2 + Bz(n,m)^2\right]}{\sum_{n=1}^{160}\sum_{m=1}^{160}\left[wx(n,m)^2 + wy(n,m)^2 + wz(n,m)^2\right]} \quad \textbf{(3)}$$

Machine Translated by Google

wobei wx, wy **und** wz **die Geräusche entlang der drei Richtungen des Rohrs bezeichnen, nämlich Umfangs-, Radial- und Längsrichtung, und** (n, **m) (wobei n,** m **= 1, 2, ... , 160) die Koordinate jedes Gitters (mit 1 mm Auflösung) im Bereich von 160 mm** × **160 mm darstellt, mit insgesamt 25.600 Gittern. Unter Berücksichtigung eines repräsentativen SNR-Werts von 20** [38] werden **drei 160** × 160-Matrizen von weißem Rauschen mit Gauß-Verteilung, die als **unabhängig voneinander angenommen werden ,** simuliert und dann auf Bx, By **und** Bz angewendet .

Jede Matrix nimmt einen Mittelwert von Null an, wobei die Standardabweichungen durch Ausprobieren angepasst werden, bis das vordefinierte Signal-Rausch-Verhältnis erreicht ist. ÿn **bezeichnet die Standardabweichung von** wx; **die Standardabweichungen von** wy **und** wz **werden mit 0,75ÿn bzw. 0,25ÿn** angenommen . **Die Skalierung von** ÿn **basiert auf den relativen Größen von** Bx, By **und** Bz , die aus der Finite-Elemente-Analyse **abgeleitet wurden : Für einen bestimmten Analysefall** hat Bx **durchweg die höchste Größe, gefolgt von** By **bzw.** Bz . **Der endgültige verrauschte Eingang für das CNN-Modell besteht aus den drei Matrizen** Bx, By **und** Bz, die jeweils mit den jeweiligen verrauschten Matrizen wx, wy **und** wz **kombiniert werden , was zu verrauschten Darstellur**

Es wird betont, dass die verrauschten MFL-Signalmatrizen nur als erweiterter Teil in Abschnitt 4.2.4 in das Regressions-CNN-Modell eingespeist werden. Fan et al. [39] berichteten, dass der Effekt des Gaußschen Rauschens auf die Genauigkeit des CNN-Klassifizierungsmodells gering ist. Acharya et al. [40] gaben an, dass das Entfernen des Rauschens für die Bildklassifizierung mit Deep-Learning-Algorithmen nicht notwendig ist [41]. Daher berücksichtigen wir in der vorliegenden Studie nur den Einfluss des Rauschens auf die Genauigkeit des

4. Faltungsneuronales Netzwerk 4.1.

Klassifizierung CNN **4.1.1.**

Eingabeinformationen Die

Eingabe in ein CNN-Modell wird üblicherweise als RGB-Farbbild dargestellt, das drei Farbkanäle enthält , **die jeweils Grün, Rot und Blau entsprechen. Diese Eingabe ist mit den dreidimensionalen MFL-Signalen kompatibel. Für ein bestimmtes FE-Modell werden die 3D- Informationen zur magnetischen Flussdichte (d. h.** Bx, By **und** Bz) innerhalb des angegebenen quadratischen Bereichs **aus der COMSOL-Software in 160 mm** × **160 mm großen Flächenrastern mit einer Auflösung von 1 mm exportiert .**
Die drei generierten 160 × **160-Matrizen rauschfreier MFL-Signale (eine Matrix für jede Dimension) werden als Eingabe des vorgeschlagenen CNN-Modells bezeichnet, und die sechs Korrosionsarten (d.ÿh. Gin, Gex, Pin, Pex, PICin und PICex) sind die Ausgabe.**

4.1.2. Vorgeschlagene

Struktur Das in dieser Studie vorgeschlagene Klassifizierungs-CNN-Modell umfasst insgesamt 20 Schichten: Die Schichten 1 bis 12 sind der Merkmalsextraktion gewidmet, während sich die Schichten 13 bis 20 auf die Klassifizierung konzentrieren. Schicht 20 dient als Ausgabeschicht und besteht aus sechs Komponenten, die die sechs Korrosionsarten darstellen. Die detaillierten Informationen zu jeder Schicht si
Wie in Tabelle 2 gezeigt, wird die Größe des Filters in der Faltungsschicht auf (5 × **5) gewählt, was eine typische Größe für große Filter ist [42]. Die Größe des Filters in der Maxpooling-Schicht beträgt (2** × **2), was den in der vorhandenen Literatur üblicherweise verwendeten Größen entspricht [43,44].** Um eine Überanpassung zu vermeiden, sind **mehrere** Dropout-Schichten **in das Modell integriert, die während des Trainingsprozesses zufällig einen bestimmten Prozentsatz von Neuronen aus dem neuronalen Netzwerk entfernen .** Die verwendete **Dropout-Rate beträgt 0,2, d. h. 20 % der Neuronen im neuronal**
Abbildung 7 zeigt die Architektur des vorgeschlagenen Klassifizierungsmodells mit der Dimension jeder Schicht, wobei die Eingabe-, Faltungs- und Max-Pooling-Schichten jeweils in Rosa, Blau und Grün dargestellt sind. Beachten Sie, dass die Dropout-Schichten in Abbildung 7 nicht angezeigt werden , **da sie keinen Einfluss auf die Datendimensionalität haben. Die dicke weiße Linie stellt die abgeflachte Schicht dar, die vollständig verbundenen Schichten sind in Orange dargestellt und die Ausgabeschichten sind in Abbildung 7 in Gelb dargestellt.**

Machine Translated by Google

Tabelle 2. Informationen zur CNN-Klassifizierungsmodellschicht.

Teil	Schicht-Nr.	Ebenenname	Parameter
Merkmalsextraktion	1	Faltung	64 Filter mit einer Größe von (5 × 5)
	2	Maxpooling	Die Pooling-Größe beträgt (2 × 2)
	3	Ausfallen	Rate = 0,2
	4	Faltung	128 Filter mit einer Größe von (5 × 5)
	5	Maxpooling	Die Pooling-Größe beträgt (2 × 2)
	6	Ausfallen	Rate = 0,2
	7	Faltung	256 Filter mit einer Größe von (5 × 5)
	8	Maxpooling	Die Pooling-Größe beträgt (2 × 2)
	9	Ausfallen	Rate = 0,2
	10	Faltung	512 Filter mit einer Größe von (5 × 5)
	11	Maxpooling	Die Pooling-Größe beträgt (2 × 2)
	12	Ausfallen	Rate = 0,2
Einstufung	13	Ebene reduzieren	Die Einheitennummer beträgt 51.200
	14	Vollständig verbundene Schicht	Die Einheitennummer beträgt 128
	15	Ausfallen	Rate = 0,2
	16	Vollständig verbundene Schicht	Die Einheitennummer beträgt 64
	17	Ausfallen	Rate = 0,2
	18	Vollständig verbundene Schicht	Die Einheitennummer beträgt 32
	19	Ausfallen	Rate = 0,2
	20	Ausgabe 1	Die Einheitennummer ist gleich 1
		Ausgabe 2	Die Einheitennummer ist gleich 1
		Ausgabe 3	Die Einheitennummer ist gleich 1
		Ausgabe 4	Die Einheitennummer ist gleich 1
		Ausgabe 5	Die Einheitennummer ist gleich 1
		Ausgabe 6	Die Einheitennummer ist gleich 1

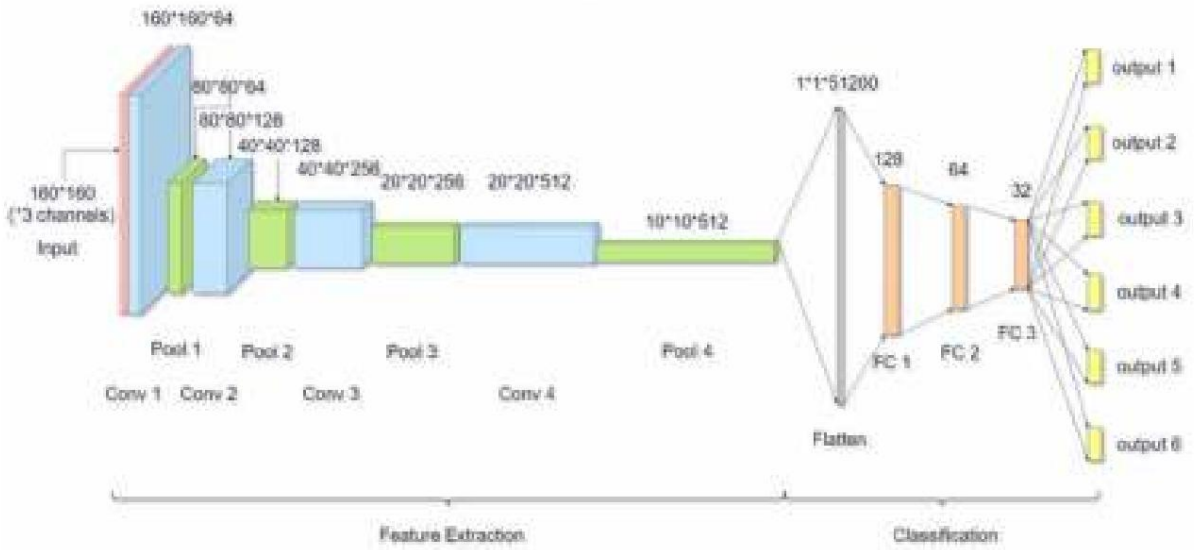


Abbildung 7. Vorgeschlagene Struktur des CNN-Klassifizierungsmodells.

Machine Translated by Google

4.1.3. Ergebnisse

Bei einem Verhältnis von 80–20 % Trainings-Testdaten gibt es 1280 Trainingsfälle und 320 Testfälle für jeden der sechs Korrosionstypen. Die Klassifizierungsergebnisse aus dem vorgeschlagenen CNN-Modells sind in Tabelle 3 zusammengefasst. Tabelle 3 zeigt, dass das vorgeschlagene Das Modell erreicht durchgängig Vorhersagegenauigkeiten von über 0,93 sowohl im Trainings- und Testdatensätze, wobei die niedrigste Genauigkeit von 0,934 im Trainingsdatensatz beobachtet wurde für Gex. Es wird darauf hingewiesen, dass das Modell im Testdatensatz besser abschneidet als im Trainingsdatensatz. mit Genauigkeitsraten für jeden Korrosionstyp im Testdatensatz, die höher oder gleich sind als Raten im Trainingsdatensatz. Dies kann an den Regularisierungstechniken liegen, wie z. B. Dropout, das im CNN-Modell übernommen wurde. Obwohl diese Techniken Überanpassung verhindern können, Sie können auch Lärm oder Unsicherheit während des Trainings verursachen, was vorübergehend die Leistung des Modells auf den Trainingsdaten verschlechtern und gleichzeitig die Generalisierung verbessern und Leistung auf den Testdaten [45].

Tabelle 3. Ergebnisse des Klassifizierungs-CNN-Modells.

Kategorie Nr.	Kategorie	Gesamt Zug		Prüfen	Falsche Vorhersage im Zug	Falsche Vorhersage im Test	Zuggenauigkeit	Testgenauigkeit
0	Gin	1600	1280	320	7	0	0,978	1.000
1	Gex	1600	1280	320	21	0	0,934	1.000
2	Stift	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

Insbesondere wurden 32 Fälle im Trainingsdatensatz und ein Fall im Testdatensatz falsch klassifiziert, was zu einer Fehlvorhersagerate von 0,34 % (33/9600 = 0,34 %) führt, was zeigt, die hervorragende Genauigkeit des entwickelten Modells. Abbildung 8 zeigt die Verteilung von Fehlvorhersagen im Trainingsdatensatz in verschiedenen Kategorien. Beispielsweise 65,6 % (d. h. 21) der falsch klassifizierten Trainingsfälle gehören zur Kategorie Gex . Unter ihnen sind vier Fälle werden fälschlicherweise als Pin klassifiziert und 17 Fälle werden fälschlicherweise als Pex klassifiziert. Darüber hinaus 33 Fälle haben falsch klassifizierte Standorte (Innen-/Außenfläche des Rohres), d.h. zwei Pex Fälle, die fälschlicherweise als Pin klassifiziert wurden, und vier Gex- Fälle, die fälschlicherweise als Pin klassifiziert wurden.

Beachten Sie, dass das Modell eine bessere Klassifizierungsgenauigkeit für Pin, Pex, PICin und PICex erreicht als für Gin und Gex. Die im gesamten Datensatz enthaltenen dg/t -Werte für alle sechs Arten von Korrosionen sind 20, 30, 40, 60 und 80 %, während die Werte von dp/t im gesamten Datensatz enthalten sind sind 20, 30, 40, 50, 55, 60, 70 und 80%. Der größere Bereich der Korrosionstiefe für Fälle mit Nadellöchern (z. B. Pin, Pex, PICin und PICex) im Vergleich zu den Fällen mit allgemeinen Korrosion (Gin und Gex) ermöglicht es dem Modell, eine größere Anzahl von Merkmalen und Variationen zu erlernen im Zusammenhang mit Lochkorrosion, wodurch das Modell besser unterscheiden kann zwischen verschiedene Arten und Schweregrade der Lochkorrosion. Alle 28 falsch klassifizierten Fälle in der Bei den Kategorien Gin und Gex handelt es sich um relativ flache Korrosionen, von denen 26 Fälle dg/t = 20% und zwei Fälle haben dg/t = 40%. Die Verteilung der 28 falsch klassifizierten Fälle von wg/lg ist in Abbildung 9 dargestellt. Abbildung 9 zeigt, dass das Klassifizierungs-CNN-Modell weniger genau für breite allgemeine Korrosionen mit wg/lg größer als 1,67. Es ist auch bemerkenswert, dass Das entwickelte Modell ist weniger genau für Fälle mit externer Korrosion (z. B. Gex, Pex, und PICex) als bei Fällen mit innerer Korrosion (Gin, Pin und PICin). Ein potenzielles Erklärung ist, dass der Modelleingang (d. h. die MFL-Signalkarte) für externe Defekte weniger ausgeprägte Merkmale, da die Variation in B aufgrund des Metallverlustdefekts auf der äußeren Rohroberfläche kleiner ist als die durch den inneren Metallverlustdefekt verursachte, da die äußere Der Defekt ist weiter vom Sensor entfernt als der interne Defekt.

Machine Translated by Google

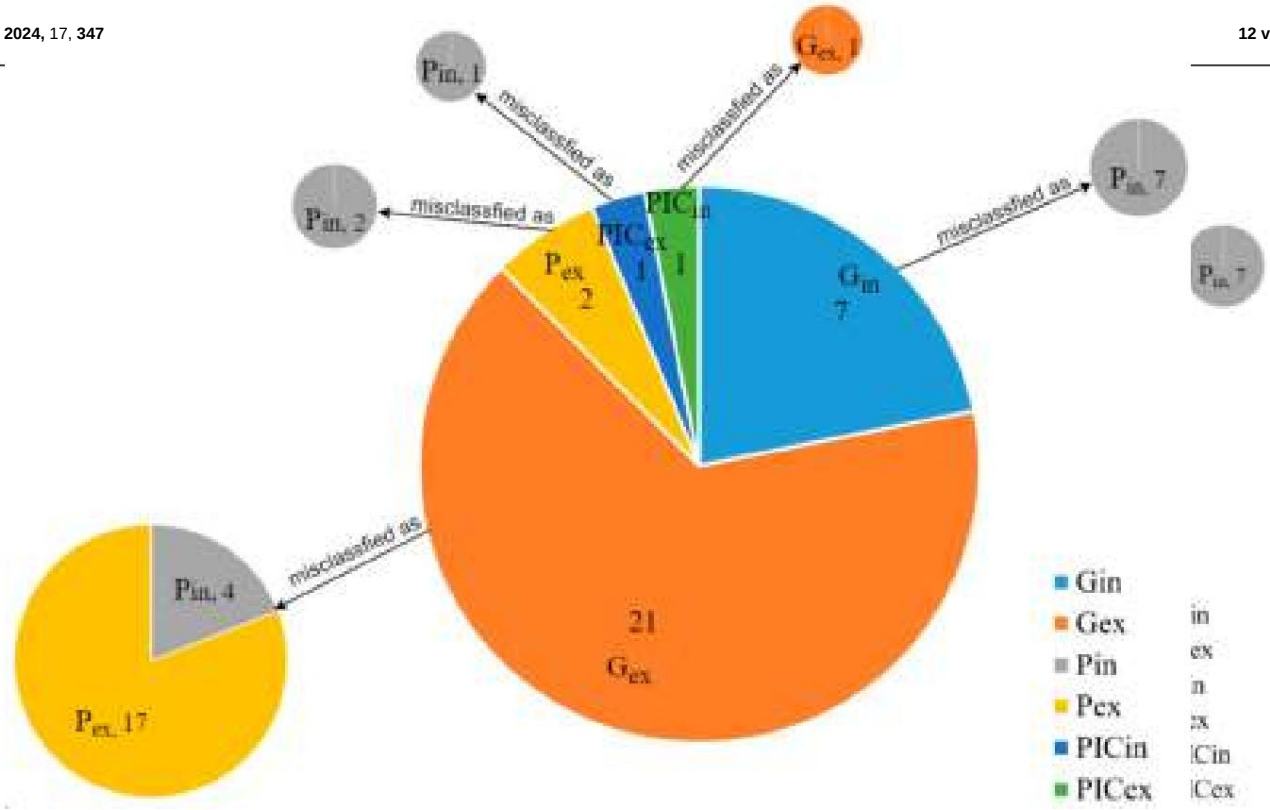


Abbildung 8. Verteilung der Fehlklassifizierungen im Trainingsdatensatz.

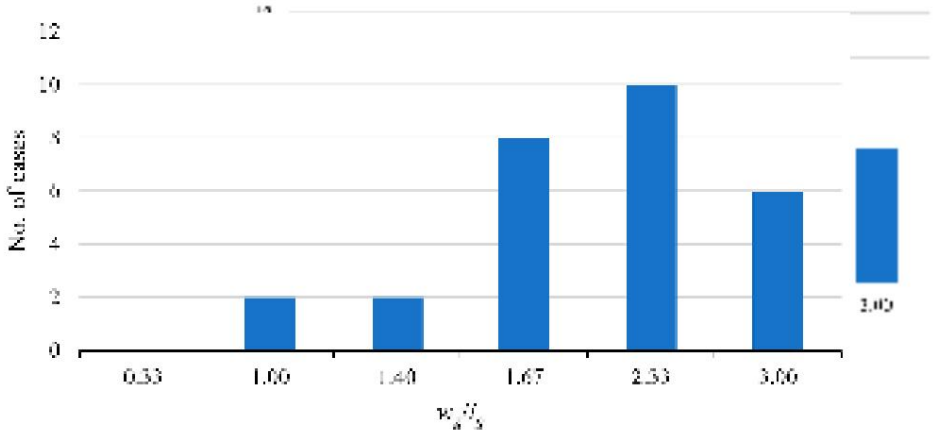


Abbildung 9. Verteilung der falsch klassifizierten Fälle in Gin und Gex nach wg/lg.

4.2. Regressions-CNN

4.2.1. Eingabeinformationen

Die Ergebnisse des Klassifikations-CNN-Modells werden verwendet, um korrekt klassifizierte Fälle für Pin, Pex, PICin und PICex zu extrahieren, wobei vier Fehlklassifizierungen im Trainingsdatensatz *ÿ* und eine Fehlklassifizierung im Testdatensatz ausgeschlossen werden, was zu 1600, 1598, 1599 und 1598 Eingabefällen für Pin, Pex, PICin und PICex führt, um das Regressions-CNN- Modell zu trainieren und zu validieren. Die drei generierten 160 × 160-Matrizen rauschfreier MFL-Signale für jeden der 6395 (=1600 + 1598 + 1599 + 1598) Fälle sind die Eingabe für das entwickelte Regressionsmodell und *ÿ* die vorhergesagten rp-, **dp/t-, h- und** *ÿ* **-Werte für jeden Fall sind die Ausgabe. Die korrekt klassifizierten Daten werden zum Trainieren des Regressions-CNN-Modells verwendet, da es problematisch ist, falsch klassifizierte Daten (wie eine allgemeine Korrosionsanomalie, die fälschlicherweise als Nadelstich klassifiziert wird) in das Training einzubeziehen, z. B. wenn eine Fehlerfunktion definiert werden muss, die auf alle Eingabedaten anwendbar ist, einschließlich falsch klassifizierter Daten. Alternativ kann das Training des Regressions-CNN-Modells vollständig vom Klassifizierungsmodell getrennt werden; das heißt, alle 6400 Fälle der parametrischen Finite-Elemente-Analyse mit Nadelstichen können zum Trainieren des Regressionsmodells verwendet werden. In dieser Studie wird der korrekt klassifizierte Datensatz, der aus**

Machine Translated by Google

4.2.2. Vorgeschlagene Struktur

Das in der vorliegenden Studie vorgeschlagene Regressions-CNN-Modell umfasst insgesamt 23 Schichten: Die Schichten 1 bis 16 sind der Merkmalsextraktion gewidmet, während die Schichten 17 bis 23 sich auf Regression. Schicht 23 ist die Ausgabeschicht, die vier Komponenten enthält, die die vorhergesagten rp-, dp/t-, h- und ÿ- Werte. Das detaillierte CNN-Regressionsmodell Informationen zu den Schichten finden Sie in Tabelle 4. Es wird darauf hingewiesen, dass zusätzlich zu den Dropout-Schichten In Abschnitt 4.1.2 wurden mehrere Batch-Normalisierungsebenen integriert . die typischen CNN-Schichten, um Überanpassung zu vermeiden [46–48]. Wie in Tabelle 4 gezeigt, ist die abgestimmte Hyperparameter im Merkmalsextraktionsprozess des Regressionsmodells, wie zum Beispiel die Filtergröße der Faltungsschicht, Pooling-Größe und Dropout-Rate sind gleich wie die im Klassifikations-CNN-Modell, während die verwendete Dropout-Rate im Regressionsmodell Prozess des Regressions-CNN-Modells beträgt 0,5. Abbildung 10 zeigt die Architektur des entwickeltes Regressionsmodell mit den Dimensionen verschiedener Schichten. Beachten Sie, dass der Dropout und Batch-Normalisierungsebenen werden aus dem Diagramm ausgeschlossen, da sie keinen Einfluss auf die Datendimensionalität. Das Farbschema, das zur Unterscheidung der verschiedenen Schichten in Abbildung 10 verwendet wird , ist übereinstimmend mit denen in Abbildung 7.

Tabelle 4. Informationen zur CNN-Regressionsmodellschicht.

Teil	Schicht-Nr.	Ebenenname	Parameter
Merkmalsextraktion	1	Faltung	64 Filter mit einer Größe von (5 × 5)
	2	Batch-Normalisierung	
	3	Maxpooling	Die Pooling-Größe beträgt (2 × 2)
	4	Ausfallen	Rate = 0,2
	5	Faltung	128 Filter mit einer Größe von (5 × 5)
	6	Batch-Normalisierung	
	7	Maxpooling	Die Pooling-Größe beträgt (2 × 2)
	8	Ausfallen	Rate = 0,2
	9	Faltung	256 Filter mit einer Größe von (5 × 5)
	10	Batch-Normalisierung	
	11	Maxpooling	Die Pooling-Größe beträgt (2 × 2)
	12	Ausfallen	Rate = 0,2
	13	Faltung	512 Filter mit einer Größe von (5 × 5)
	14	Batch-Normalisierung	
	15	Maxpooling	Die Pooling-Größe beträgt (2 × 2)
	16	Ausfallen	Rate = 0,2
Regression	17	Ebene reduzieren	Die Einheitennummer beträgt 51.200
	18	Ausfallen	Rate = 0,5
	19	Vollständig verbundene Schicht	Die Einheitennummer beträgt 64
	20	Vollständig verbundene Schicht	Die Einheitennummer beträgt 128
	21	Vollständig verbundene Schicht	Die Einheitennummer beträgt 256
	22	Vollständig verbundene Schicht	Die Einheitennummer beträgt 512
	23	Ausgabe 1	Die Einheitennummer ist gleich 1
		Ausgabe 2	Die Einheitennummer ist gleich 1
		Ausgabe 3	Die Einheitennummer ist gleich 1
		Ausgabe 4	Die Einheitennummer ist gleich 1

Machine Translated by Google

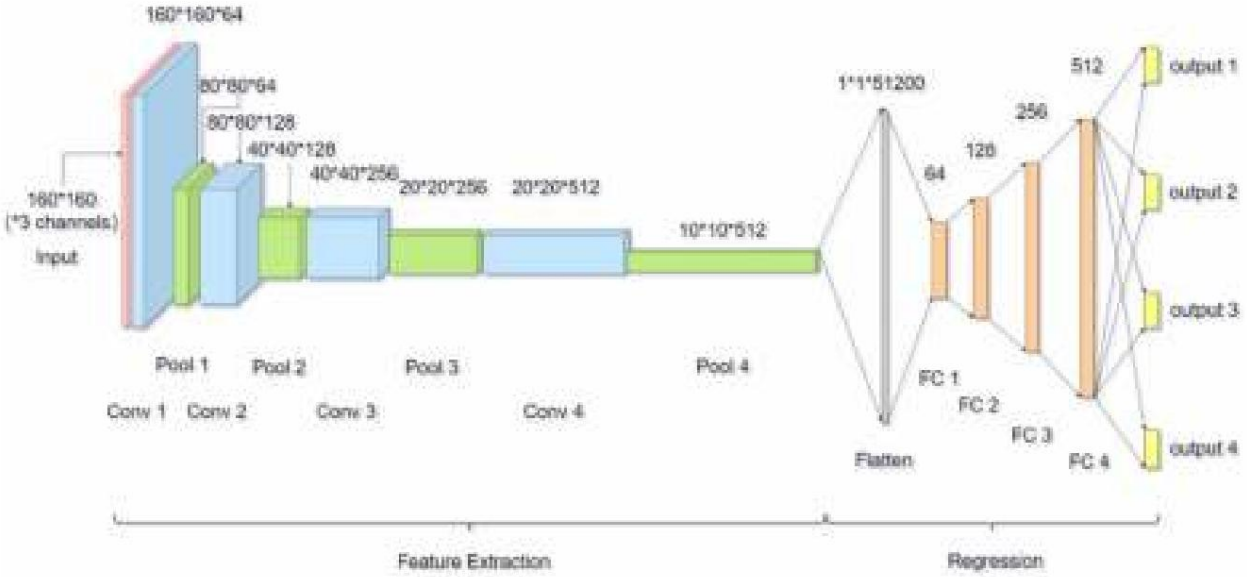


Abbildung 10. Vorgeschlagene Struktur des CNN-Regressionsmodells.

4.2.3. Ergebnisse

Der Determinationskoeffizient (also R **ffi** ^{2}**) wird verwendet, um die Genauigkeit der Vorhersage jedes Parameters (d. h.** rp, **dp/t, h und ÿ) zu messen, die durch die folgende Gleichung definiert ist.**

$$R = 1 - \frac{\sum_{i=1}^{N} (Ywahr,i - Ypred,i)^2}{\sum_{i=1}^{N} (Ywahr,i - \bar{Y}wahr)^2} \quad (4)$$

wobei Ytrue,i **und** Ypred,i **die wahren bzw. vorhergesagten Werte jedes vorhergesagten Parameters für den i-ten** (i **= 1, 2, ... , N) Punkt in einem gegebenen Datensatz mit** N **Datenpunkten** darstellen , **und** Ytrue **den wahren Mittelwert von** N **Datenpunkten bezeichnet, die mit einem vorhergesagten Parameter verknüpft sind. Tabelle 5 fasst das im Regressionsmodell Die Werte für die vier Parameter im Testdatensatz werden enthaltene** R zusammen. Das R **in Abbildung 11. Die wahren Werte der vier Parameter** rp, **dp/t, h und** ÿ **im Vergleich mit den vom CNN geschätzten Werten für die Korrosionen im Testsatz sind in Abbildung 12** dargestellt. Aus Tabelle **5 geht hervor, dass die Vorhersagen des CNN-Modells der Korrosionspositionsparameter (d. h.** h **und** ÿ) stark mit den tatsächlichen Werten übereinstimmen, wie die R- **Werte von 1,0 für alle vier Korrosionskategorien widerspiegeln . In Bezug auf die Größenvorhersagen (d. h.** rp **und** dp/t) ist **das** Modell für innere Korrosionen **(Pin und PICin) genauer als für äußere Korrosionen (Pex und PICex). Diese Ergebnisse zeigen auch, dass die Leistung des Modells in Fällen, die ausschließlich Nadellöcher enthalten, besser ist als in Fällen mit PIC. Dies ist zu erwarten, da PIC zusätzliche Komplexität mit sich bringt, einschließlich des Vorhandenseins allgemeiner Korrosion, die die Genauigkeit des Modells beeinträchtigen kann. Insgesamt ist die** Genauigkeit von R **bei der Vorhersage der Größenparameter hoch, wobei eine höhere Genauigkeit bei Korrosionen auf der Innenfläche und bei Nadellöchern beobachtet wurde. Beachten Sie jedoch, dass die Vorhersagen für** rp **und dp/t** zwar **im Allgemeinen genau sind, Abbildung 12 jedoch relativ große Fehlerbänder sowohl bei** den rp- **als** auch **den dp/t- Vorhersagen aufweist. Beispielsweise liegen die vorhergesagten Werte von** rp **zwischen 1,14 mm und 3,45 mm für** (rp)true **gleich 2 mm. Die großen Fehlerbänder können durch die geringen Abmessungen der Nadellochkorrosion und die begrenzte räumliche Auflösung des Modells erklärt werden. Darüber hinaus kann das Vorhandensein von zusammengesetzten Korrosionsmerkmalen zusätzliche Komplexitäten und Unsicherheiten mit sich bringen, was zu größeren Fehlerbändern in den Vorhersagen führt. Daher sind weitere Untersuchungen erforderlich, um die Faktoren zu ermitteln, die zur Reduzierung der Regressionsfehler bei Nadellochkorrosionen beitragen.**

Machine Translated by Google

Tabelle 5. Ergebnisse des Regressions-CNN-Modells ohne Rauschen.

Kategorie Nr.	Kategorie	Fälle			Test R^2			
		Gesamt	Zug	Prüfen	rp	dp/t	H	ÿ
0	Stift	1600	1280	320	0,91	0,95	1,00	1,00
1	Pex	1598	1277	321	0,88	0,92	1,00	1,00
2	PICin	1599	1280	319	0,89	0,85	1,00	1,00
3	PICex	1598	1279	319	0,86	0,77	1,00	1,00
Gesamt		6395	5116	1279	0,89	0,91	1,00	1,00

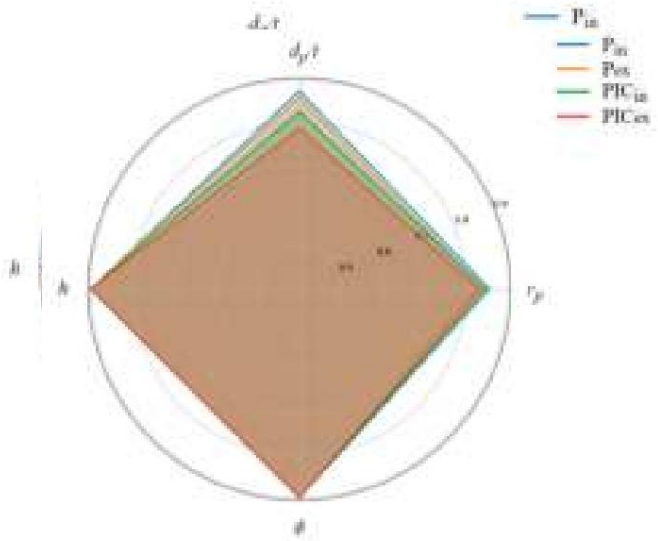


Abbildung 11. Radardiagramm für R^2 **von Metriken** (rp, **dp/t,** *hÿ* **und** ÿ) **für Pin-, Pex-, PICin- und PICex- Korrosionen in der Testdatensatz.**

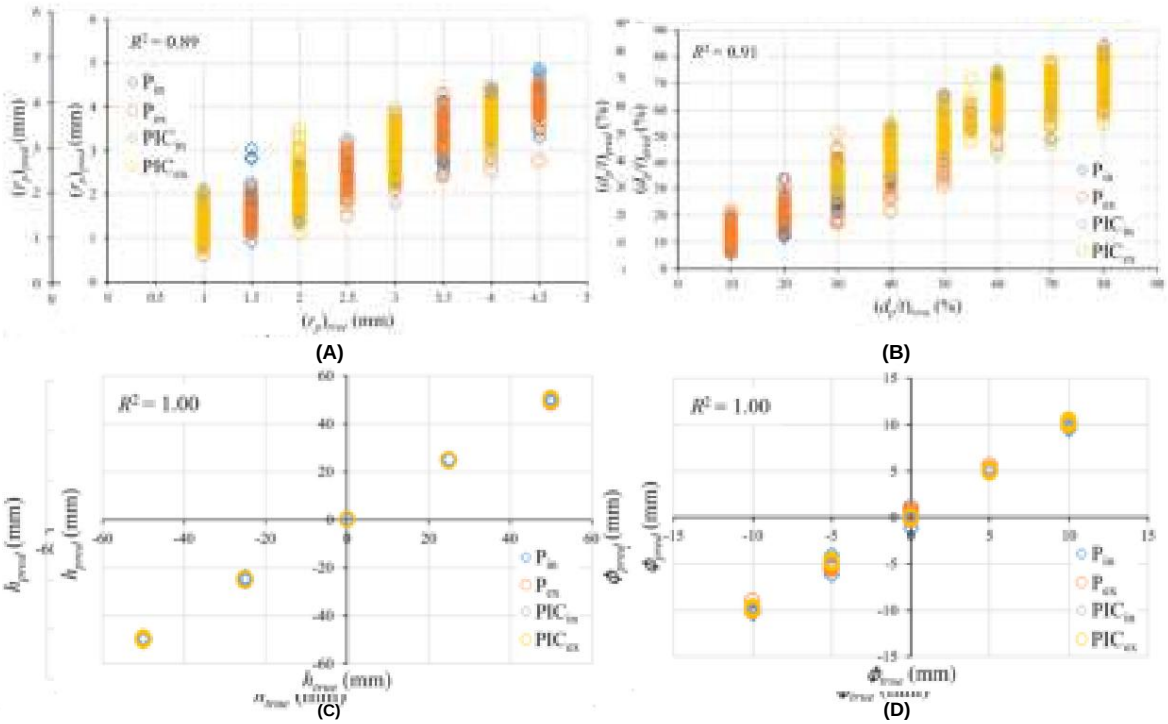


Abbildung 12. Vergleich der wahren und vorhergesagten Werte von rp, **dp/t, h und** *ÿ* **für die Anomalien in der Regressionstestdatensatz ohne Rauschen.**

Machine Translated by Google

4.2.4. Einfluss von Lärm

Abbildung 13 vergleicht die Ergebnisse des Regressions-CNN-Modells für verrauschte Eingaben (SNR = 20) mit denen des rauschfreien Szenarios, das in Abschnitt 4.2.3 beschrieben wird. Bei Vorhandensein von Rauschen nehmen die R- Werte zur Vorhersage von rp und dp/t im Vergleich zu den rauschfreien Werten etwas ab. Dies ist bei Pex- und PICex- Fällen ausgeprägter , was darauf hinweist , dass das Modell bei der Szenario. Der Rückgang von R^2 Vorhersage der Ausmaße externer Korrosion stärker von Rauschen beeinflusst wird.

Trotz der abnehmenden Genauigkeit bei der Vorhersage der Korrosionsgrößen behält das Modell 2 R = 1 für die Lageparameter (h und ÿ) für alle Korrosionsarten außer Pex bei, was auf seine Robustheit bei der Erfassung der räumlichen Informationen selbst bei Vorhandensein von Rauschen hinweist.

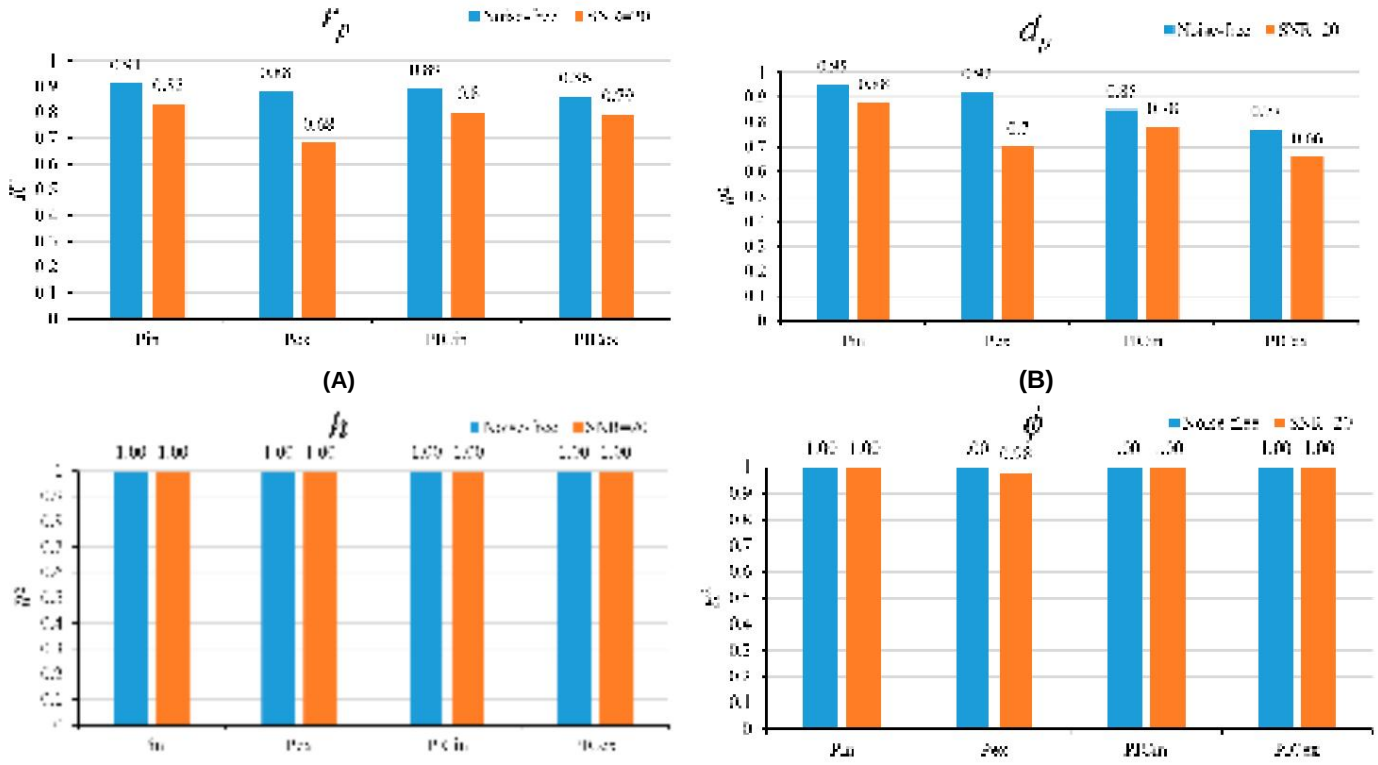


Abbildung 13. Vergleich von R^2 zwischen rauschfreien und SNR = 20-Szenarien für verschiedene Korrosionsarten.

Abschließend muss darauf hingewiesen werden, dass die in dieser Studie entwickelten CNN-Klassifizierungs- und Regressionsmodelle speziell auf bestimmte Rohrleitungs- und MFL-Werkzeugattribute abgestimmt sind, wie etwa Rohrwandstärke, Magneteigenschaften, Lift-off-Werte und Defektformen. Daher sind weitere Untersuchungen erforderlich, um zu bestimmen, wie gut das CNN-Modell bei verschiedenen Rohrwandstärken und unterschiedlichen MFL-Werkzeugparametern funktioniert. Weitere Studien sind auch erforderlich, um die Einbeziehung des MFL-W

5. Schlussfolgerungen

Dieser Artikel berichtet über eine neuartige Studie, die sich auf die Klassifizierung und Größenbestimmung von Nadelstichkorrosionen an Pipelines unter Verwendung von Deep-Learning-Algorithmen konzentriert. Wir schlagen ein CNN-Klassifizierungsmodell zur Klassifizierung von sechs verschiedenen Korrosionsarten (Gin, Gex, Pin, Pex, PICin und PICex) an Pipelines und ein CNN-Regressionsmodell zur Schätzung der Größe und Position der Nadelstichdefekte basierend auf MFL-Signalen vor, die mithilfe der magnetostatischen FEA erzeugt werden. Umfangreiche 3D-parametrische FEA-Fälle mit kastenförmigen allgemeinen Korrosionen und zylinderförmigen Nadelstichen werden verwendet, um die 3D-MFL-Signale zu simulieren, indem die Tiefe, Länge, Breite sowie Längs- und Umfangsposition des Defekts variiert werden. Die CNN-Klassifizierungs- und Regressionsmodelle werden dann mithilfe der simulierten MFL-Signale trainiert und validiert.

Das vorgeschlagene Klassifizierungsmodell weist eine ausgezeichnete Genauigkeit bei der Klassifizierung von sechs Arten von Korrosionsfehlern mit einer Fehlvorhersagerate von 0,34 % (33/9600 = 0,34 %) auf. Das

Machine Translated by Google

Das vorgeschlagene Regressionsmodell ist bei der Vorhersage der Position des Nadellochs sehr genau und erreicht eine gute Genauigkeit bei der Vorhersage der Tiefe und des Durchmessers des Nadellochs. Durch eine vergleichende Analyse des CNN-Regressionsmodells mit rauschfreien und rauschbehafteten Signalen (SNR = 20) als Eingabe wird festgestellt, dass der Einfluss des Rauschens auf die Vorhersagegenauigkeit des Regressionsmodells moderat ist. Diese Studie demonstriert die Anwendung von Deep-Learning -Algorithmen zur Erleichterung des Integritätsmanagements von Pipelines mit komplex geformten Korrosionsdefekten.

Beiträge der Autoren: YS: Untersuchung, Visualisierung, formale Analyse, Methodik, Schreiben – Originalentwurf . WZ: Mittelbeschaffung, Betreuung, Konzeptualisierung, Methodik, Schreiben – Überprüfung und Bearbeitung. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und ihr zugestimmt.

Finanzierung: Die Autoren danken dem Natural Sciences and Engineering Research Council of Canada (NSERC) (Zuschuss-Nr. RGPIN-2019–05160) und der Fakultät für Ingenieurwissenschaften der University of Western Ontario für die finanzielle Unterstützung.

Erklärung zur Datenverfügbarkeit: Daten sind auf Anfrage bei den Autoren erhältlich.

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Verweise

1. **Zhou, Q.; Wu, W.; Liu, D.; Li, K.; Qiao, Q. Schätzung der Korrosionsversagenswahrscheinlichkeit von Öl- und Gaspipelines basierend auf dem Fuzzy-Logik-Ansatz.** Eng. Fail. Anal. **2016,** 70, **48–55.** [CrossRef]
2. **Murphy, JF Albtraumhafte Pipeline-Ausfälle, Fantasieplanung, schwarze Schwäne und Integritätsmanagement – Eine Übersicht.** Process Saf. Prog. **2015,** 34, **207.** [CrossRef]
3. **Shen, Y.; Zhou, W. Ein Vergleich der Unfallstatistiken bei Öl- und Gaspipelines an Land in Kanada und den Vereinigten Staaten.** Int. J. Crit. Infrastruct. Prot. **2024,** 45, **100679.** [CrossRef]
4. **Vanaei, HR; Eslami, A.; Egbewande, A. Eine Übersicht über Pipelinekorrosion, Inline-Inspektion (ILI) und Korrosionswachstumsratenmodelle .** Int. J. Press. Vessels Pip. **2017,** 149, **43–54.** [CrossRef]
5. **Song, H.; Yang, L.; Liu, G.; Tian, G.; Ona, DI; Song, Y.; Li, S. Vergleichende Analyse von Inline-Inspektionsgeräten und Technologien.** IOP Conf. Ser. Mater. Sci. Eng. **2018,** 382, **032021.** [CrossRef]
6. **Pipeline Operators Forum. Spezifikationen und Anforderungen für die Inline-Inspektion von Pipelines. Version 2021. Online verfügbar: https://pipelineoperators.org/documents (abgerufen am 18. Mai 2022).**
7. **Sutherland, J.; Bluck, M.; Pearce, J.; Quick, E. Die Validierung der MFL-Inline-Inspektionstechnologie der neuesten Generation führt zu einer verbesserten Erkennung und Größenbestimmung von Nadellöchern, Lochfraß, axialen Rillen und axialen Schlitzen. In Proceedings der ASME 2010 8. International Pipeline Conference, Calgary, AB, Kanada, 27. September–1. Oktober 2010.**
8. **Peng, X.; Anyaoha, U.; Liu, Z.; Tsukada, K. Analyse von Daten zur magnetischen Flussstreuung (MFL) zur Beurteilung der Korrosion von Rohrleitungen.** IEEE Trans. Magn. **2020,** 56, **1–15.** [CrossRef]
9. **Han, W.; Yang, P.; Xia, F.; Xue, Y. Magnetische Flusslecksignalumkehrung von Korrosionsfehlern basierend auf modifizierter genetischer lokaler Suche Algorithmus.** J. Shanghai Jiaotong Univ. (Sci.) **2009,** 14, **168–172.** [CrossRef]
10. **Priewald, RH; Magele, C.; Ledger, PD; Pearson, NR; Mason, JSD Schnelle Inversion des magnetischen Streuflusssignals zur Rekonstruktion beliebiger Defektprofile in Stahl unter Verwendung finiter Elemente.** IEEE Trans. Magn. **2012,** 49, **506–516.** [CrossRef]
11. **Hwang, K.; Mandayam, S.; Udpa, SS Charakterisierung von Gaspipeline-Inspektionssignalen mit Hilfe der Wavelet-Basisfunktion neuronaler Netzwerke.** NDT E Int. **2000,** 33, **531–545.** [CrossRef]
12. **Ramuhalli, P.; Udpa, L.; Udpa, SS Auf neuronalen Netzwerken basierende Inversionsalgorithmen zur zerstörungsfreien Prüfung von magnetischem Streufluss** 2003 , 93, **82748276.** [CrossRef]
13. **Han, WH; Que, PW 2-D Defektrekonstruktion aus MFL-Signalen basierend auf genetischen Optimierungsalgorithmen. In Proceedings of the IEEE 2005 Internationale Konferenz für Industrietechnologie, Hongkong, China, 14.–17. Dezember 2005.**
14. **Kandroodi, MR; Araabi, BN; Bassiri, MM; Ahmadabadi, MN Schätzung der Tiefe und Länge von Defekten anhand von Messungen des magnetischen Streuflusses: Überprüfung mit Simulationen, Experimenten und Molchdaten.** IEEE Trans. Magn. **2016,** 53, **1–10.** [Querverweis]
15. **Feng, J.; Li, F.; Lu, S.; Liu, J. Schnelle Rekonstruktion von Defektprofilen aus Messungen magnetischer Streuflüsse unter Verwendung einer auf RBFNN basierenden Fehlerkorrekturmethode.** IET Sci. Meas. Technol. **2017,** 11, **262–269.** [CrossRef]
16. **Yao, G.; Lei, T.; Zhong, J. Eine Übersicht über die auf Convolutional-Neural-Networks basierende Aktionserkennung.** Mustererkennung. Lett. **2019,** 118, **14–22.** [CrossRef]
17. **Taye, MM Theoretisches Verständnis von Convolutional Neural Network: Konzepte, Architekturen, Anwendungen, zukünftige Richtungen.** Berechnung **2023,** 11, **52.** [CrossRef]
18. **Lu, S.; Feng, J.; Zhang, H.; Liu, J.; Wu, Z. Eine Methode zur Schätzung der Defektgröße aus einem MFL-Bild mittels visueller Transformation Faltungsneuronales Netzwerk.** IEEE Trans. Ind. Inform. **2018,** 15, **213–224.** [CrossRef]

Machine Translated by Google

19. **Shen, Y.; Zhou, W. Entwicklung eines Convolutional Neural Network-Modells zur Vorhersage der Größe und Lage von Korrosionsfehlern auf Pipelines basierend auf magnetischen Streuflusssignalen.** Int. J. Press. Vessels Pip. **2023,** 207, **105–123.** [CrossRef]
20. **Wang, HA; Chen, G. Methode zur Schätzung der Defektgröße für magnetische Streuflusssignale unter Verwendung von Convolutional Neural Networks.** Insight **2020,** 62, **86–91.** [CrossRef]
21. **EGIG.** 10. Bericht der European Gas Pipeline Incident Data Group (Zeitraum 1970–2016); **EGIG: Groningen, Niederlande, 2018; Dok.-Nr. VA 17.R.0395.**
22. **Zhang, H.; Sha, S.; Willis, C.; Qingshan, F.; Chen, P. Machbarkeitsstudie zur Lochkamerainspektion mittels magnetischem Streufluss und hydrostatische Prüfung in Öl- und Gaspipelines.** IOP Conf. Ser. Mater. Sci. Eng. **2021,** 1043, **022053.**
23. **Feng, Q.; Yan, B.; Chen, P.; Shirazi, SA Fehleranalyse und Simulationsmodell der Nadelstichkorrosion der raffinierten Ölpipeline.** Eng. Fail. Anal. **2019,** 106, **1–28.** [CrossRef]
24. **Subramanian, C. Lokale Lochfraßkorrosion von API 5L Klasse A Rohren, die in industriellen Löschwasserleitungen verwendet werden.** Eng. Fail. Anal. **2018,** 92, **405–417.** [CrossRef]
25. **Askari, M.; Aliofkhazraei, M.; Afroukhteh, S. Eine umfassende Übersicht über innere Korrosion und Rissbildung bei Öl- und Gaspipelines.** J. Nat. Gas Sci. Eng. **2019,** 71, **102971.** [CrossRef]
26. **Kadhim, KN; Al-Rufaye, AHR Die Auswirkungen eines gleichmäßigen transversalen Magnetfelds auf lokale Strömung und Geschwindigkeitsprofil.** Int. J. Civ. Eng. Technol. **2016, 7, 140–151.**
27. **Ji, F.; Wang, C.; Sun, S.; Wang, W. Anwendung von 3-D FEM in der Simulationsanalyse für MFL-Signale.** Insight **2009,** 51, **32–35.** [Querverweis]
28. **Bubenik, T.** Elektromagnetische Methoden zur Erkennung von Korrosion in unterirdischen Rohrleitungen: Magnetischer Streufluss (MFL); **Untergrund Pipeline-Korrosion; Woodhead Publishing: Cambridgeshire, Großbritannien, 2014; S. 215–226.**
29. **Shi, Y.; Zhang, C.; Li, R.; Cai, M.; Jia, G. Theorie und Anwendung der magnetischen Leckflusserkennung in Rohrleitungen.** Sensors **2015,** 15, **31036–31055.** [CrossRef]
30. **Zhang, Y.; Ye, Z.; Wang, C. Eine schnelle Methode zur Rekonstruktion rechteckiger Rissgrößen bei der Prüfung auf magnetischen Streufluss.** NDT E Int. **2009,** 42, **369–375.** [CrossRef]
31. **Walker, J.** In-Line Inspection of Pipelines: Fortschrittliche Technologien für den wirtschaftlichen und sicheren Betrieb von Öl- und Gaspipelines; **Verlag Moderne Industrie: Landsberg am Lech, Deutschland, 2010.**
32. **Liu, Y.; Gao, X.; Wang, Y.; Yang, X. Optimierung des Permanentmagnet-Stützmechanismus durch empfindliche Parameter.** J. Mech. Sci. Technol. **2014,** 28, **2707–2714.** [CrossRef]
33. **Yang, L.; Zhang, G.; Liu, G.; Gao, S. Auswirkungen des Abhebens auf die Inspektion des magnetischen Leckflusses in Rohrleitungen. In Proceedings of the 2008 17th Weltkonferenz für zerstörungsfreie Prüfung, Shanghai, China, 25.–28. Oktober 2008.**
34. **Ireland, RC; Torres, CR Finite-Elemente-Modellierung eines Umfangsmagnetisierers.** Sens. Actuators A Phys. **2006,** 129, **197–202.** [Querverweis]
35. **Azizzadeh, T.; Safizadeh, MS Dreidimensionale Finite-Elemente- und experimentelle Simulation von NDT mit magnetischem Streufluss zur Erkennung von Lochkorrosion. In Proceedings der 4. Iranian International NDT Conference (IRNDT) 2017, Teheran, Iran, 21.–22. Februar 2017.**
36. **Chen, L.; Li, X.; Qin, G.; Lu, Q. Signalverarbeitung der magnetischen Flussstreuung bei Oberflächenfehlern in Rohrleitungsstahl.** Russ. J. Zerstörungsfrei. Test. **2008,** 44, **859–867.** [CrossRef]
37. **Piao, G.; Guo, J.; Hu, T.; Leung, H. Die Wirkung bewegungsinduzierter Wirbelströme auf magnetischen Streufluss bei hoher Geschwindigkeit (MFL) Inspektion für dickwandige Stahlrohre.** Res. Nondestruct. Eval. **2020,** 31, **48–67.** [CrossRef]
38. **Li, F.; Feng, J.; Zhang, H.; Liu, J.; Lu, S.; Ma, D. Schnelle Rekonstruktion beliebiger Rohrleitungsdefektprofile aus MFL-Messungen unter Verwendung eines modifizierten Harmoniesuchalgorithmus.** IEEE Trans. Instrum. Meas. **2018,** 67, **9.** [CrossRef]
39. **Fan, X.; Dai, M.; Liu, C.; Wu, F.; Yan, X.; Feng, Y.; Feng, Y.; Su, B. Einfluss von Bildrauschen auf die Klassifizierung von Hautläsionen mit tiefe Convolutional Neural Networks.** Tsinghua Sci. Technol. **2019,** 25, **425–434.** [CrossRef]
40. **Acharya, UR; Oh, SL; Hagiwara, Y.; Tan, JH; Adam, M.; Gertych, A.; San Tan, R. Ein tiefes Convolutional-Neural-Network-Modell zur Klassifizierung von Herzschlägen.** Comput. Boil. Med. **2017,** 89, **389–396.** [CrossRef] [PubMed]
41. **Tesfai, H.; Saleh, H.; Al-Qutayri, M.; Mohammad, MB; Tekeste, T.; Khandoker, A.; Mohammad, B. Leichtgewichtiges Shufflenet-basiertes CNN zur Arrhythmieklassifizierung.** IEEE Access. **2022,** 10, **111842–111854.** [CrossRef]
42. **West, NE; O'shea, T. Tiefe Architekturen zur Modulationserkennung. In Proceedings des IEEE International Symposium 2017 über Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA, 6.–9. März 2017.**
43. **Baranwal, S.; Khandelwal, S.; Arora, A. Deep Learning Convolutional Neural Network zur Erkennung von Apfelblattkrankheiten. In Proceedings der 2019 International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur, Indien, 26.–28. Februar 2019.**
44. **Virupakshappa, K.; Marino, M.; Oruklu, E. Eine mehrauflösende Convolutional-Neural-Network-Architektur zur Ultraschall-Fehlererkennung . In Proceedings des 2018 IEEE International Ultrasonics Symposium (IUS), Kobe, Japan, 22.–25. Oktober 2018.**
45. **Bilmes, J. Underfitting und Overfitting beim maschinellen Lernen. UW ECE Course Notes. 2020. Online verfügbar: https://people.ece. uw.edu/bilmes/classes/ee511/ee511_spring_2020/overfitting_underfitting.pdf (abgerufen am 20. März 2023).**
46. **Ioffe, S.; Szegedy, C. Batch-Normalisierung: Beschleunigung des Trainings tiefer Netzwerke durch Reduzierung der internen Kovariatenverschiebung. In Proceedings der Internationalen Konferenz zum maschinellen Lernen 2015, Lille, Frankreich, 6.–11. Juli 2015.**

Machine Translated by Google

47. Thakkar, V.; Tewary, S.; Chakraborty, C. Batch-Normalisierung in Convolutional Neural Networks – Eine vergleichende Studie mit CIFAR-10-Daten. In Proceedings der IEEE 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), IIEST Shibpur, Westbengalen, Indien, 12.–13. Januar 2018.

48. Bjorck, N.; Gomes, CP; Selman, B.; Weinberger, KQ Verständnis der Batch-Normalisierung. Adv. Neural Inf. Process. Syst. **2018,** 31, **1–12.**

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere

Machine Translated by Google

Algorithmen

MDPI

Artikel

Das Problem der parallelen Maschinenplanung mit unterschiedlichen Geschwindigkeiten und Freigabezeiten im Erztransport

Luis Tarazona-Torres, Ciro Amaya *, Alvaro Paipilla, Camilo Gomez und David Alvarez-Martinez

Fakultät für Wirtschaftsingenieurwesen, Universität Los Andes, Cra 1 N18A12, Bogota 111711, Kolumbien; 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.)
* Korrespondenz: ca.amaya@uniandes.edu.co

Zusammenfassung: Erztransportvorgänge sind in der Bergbauindustrie von entscheidender Bedeutung, da sie Produktionsanlagen mit wichtigen Mineralien versorgen. Diese Vorgänge werden mit hochentwickelter und teurer Betriebsausrüstung durchgeführt und erfordern eine sorgfältige Planung, um sicherzustellen, dass die Produktionsziele erreicht und gleichzeitig die Geräteauslastung optimiert wird. In dieser Studie stellen wir einen Algorithmus vor, mit dem die Mindestmenge an Transportausrüstung bestimmt werden kann, die erforderlich ist, um das Erztransportziel zu erreichen. Zu diesem Zweck wurde ein mathematisches Modell entwickelt, das es als paralleles Maschinenplanungsproblem mit unterschiedlichen Geschwindigkeiten und Freigabezeiten betrachtet und sich darauf konzentriert, sowohl die Fertigstellungszeit als auch die mit der Gerätenutzung verbundenen Kosten zu minimieren. Darüber hinaus wurde ein weiterer Algorithmus entwickelt, der die taktische Bewertung dieser beiden Variablen ermöglicht. Diese Verfahren und das Modell leisten einen erheblichen Beitrag für Entscheidungsträger, indem sie einen systematischen Ansatz zur Ressourcenzuweisung bieten und sicherstellen, dass Lade- und Transportausrüstung unter Einhaltung der Budgetbeschränkungen und Betriebspläne optimal genutzt werden . Dieser Ansatz optimiert die Ressourcennutzung und verbessert die Betriebseffizienz, wodurch eine kontin

Schlüsselwörter: Problem der parallelen Maschinenplanung; Erztransportausrüstung; mathematisches Modell; Bergbauindustrie

check for updates

Zitat: Tarazona-Torres, L.; Amaya, C.; Paipilla, A.; Gomez, C.; Alvarez-Martinez, D. Das Problem der parallelen Maschinenplanung mit unterschiedlichen Geschwindigkeiten und Freigabezeiten im Erztransportbetrieb. Algorithms 2024, 17, 348. https://doi.org/10.3390/a17080348

Wissenschaftliche Herausgeber: Alexandre Dolgui, David Lemoine, María I. Restrepo und Frank Werner

Empfangen: 2. Mai 2024
Überarbeitet: 24. Juli 2024
Akzeptiert: 30. Juli 2024
Veröffentlicht: 8. August 2024

Copyright: © 2024 bei den Autoren. Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und Bedingungen der Creative Commons

Namensnennungslizenz (CC BY) (https:// creativecommons.org/licenses/by/4.0/).

1. Einleitung

Eine der wichtigsten Aktivitäten im Bergbau ist der Erztransport. Bei diesem Prozess wird zerkleinertes Gestein mithilfe von Lade- und Transportgeräten vom Sprengplatz abtransportiert [1]. Anschließend wird das Erz zum Zerkleinerungsprozess transportiert, um es zu zerkleinern. Anschließend wird es zur Rückgewinnungsanlage geschickt, wo wertvolle Metalle wie Kupfer, Gold, Molybdän und Aluminium gewonnen werden [2–4]. Abbildung 1 zeigt das Schema des beschriebenen Prozesses. Eine schlechte Planung des Erztransports kann die Produktivität verringern, das Erreichen der Produktionsziele innerhalb der festgelegten Zeitpläne beeinträchtigen und sich folglich auf die Betriebskosten auswirken. Für eine effektive Planung des Erztransports müssen mehrere Faktoren berücksichtigt werden, wie die Menge der verfügbaren Lade- und Transportgeräte und deren planmäßige Wartung [5]. Außerdem müssen die festgelegten Produktionsziele, Arbeitszeiten und topografischen Ebenen berücksichtigt werden . Weitere wichtige Aspekte sind die Erzqualität, die Anzahl der verfügbaren Laderampen für Ladevorgänge und die Umweltauswirkungen [6].

Bergbauunternehmen verfügen über eine bestimmte Menge an Lade- und Transportausrüstung, die auf den für einen bestimmten Zeitraum festgelegten Produktionszielen basiert. Es ist von entscheidender Bedeutung, die erforderliche Menge dieser Ausrüstung für den effizienten Transport von Erzgestein zu bestimmen, um die damit verbundenen Kosten zu kontrollieren, die je nach der zur Erreichung dieser Ziele erforderlich

Die Kenntnis der erforderlichen Mindestmenge an Ausrüstung optimiert den Betrieb und erleichtert die taktische Analyse. Diese Analyse ermöglicht eine bessere Entscheidungsfindung bei unvorhergesehenen Ereignissen und stellt sicher, dass die Produktionsziele zu den geringstmöglichen Kosten erreicht werden.

Machine Translated by Google



Abbildung 1. Schematische Darstellung des Erztransportvorgangs.

Wie in Abbildung 1 dargestellt, umfasst der Erztransportprozess die Zuordnung von Transportgeräten zu Ladegeräten. Diese arbeiten nicht unbedingt mit derselben Ladegeschwindigkeit, da einige Geräte unterschiedliche Ladekapazitäten haben. Das Laden erfolgt gleichzeitig und wird als System paralleler Maschinen mit unterschiedlichen Geschwindigkeiten beschrieben. Da die Anzahl der Transportgeräte begrenzt ist, werden sie so oft wie nötig verwendet, um das Produktionsziel zu erreichen. Während des Transports des Erzes zum Zerkleinerungsprozess sind sie jedoch nicht verfügbar und haben eine „Freigabezeit“, bevor sie wieder zum erneuten Laden verfügbar sind. Daher kann dieses Szenario an ein Problem paralleler Maschinen mit unterschiedlichen Geschwindigkeiten und Freigabezeiten angepasst werden.

Das Neue an dieser Arbeit ist, dass unterschiedliche Geschwindigkeiten und Freigabezeiten in das mathematische Modell für die parallele Maschinenplanung im Mineralientransport (PMSPOHO) integriert werden. Diese Funktionen ermöglichen eine realistischere und genauere Darstellung der Betriebsbedingungen im Bergbau. Die Geschwindigkeit der Ladeausrüstung spielt in diesem Modell eine entscheidende Rolle. Schwankungen in der Ladegeschwindigkeit können die Effizienz des Erztransports erheblich beeinträchtigen. Durch die Berücksichtigung unterschiedlicher Ladegeschwindigkeiten kann das Modell Ressourcen präziser zuweisen, um der dynamischen Natur des Bergb Dadurch wird sichergestellt, dass schneller ladende Geräte optimal genutzt werden, während langsameren Geräten Aufgaben zugewiesen werden, die besser zu ihren Fähigkeiten passen. So wird der gesamte Arbeitsablauf op

Durch die Einbeziehung der Geschwindigkeit der Ladeausrüstung in den Planungsalgorithmus wird die Ausrüstungsnutzung maximiert und die Leerlaufzeit der Transportausrüstung minimiert. Dies verbessert die Betriebseffizienz, da Verzögerungen reduziert und ein kontinuierlicher Materialfluss sichergestellt wird. Das entwickelte mathematische Modell berücksichtigt diese kritischen Variablen, optimiert die Zuweisung der Transportausrüstung, maximiert ihre Nutzung und reduziert Leerlaufzeiten. Dieser Ansatz gewährleistet eine höhere Betriebseffizienz, indem die Fähigkeiten der Ausrüstung an die Anforderungen des Bergbauprozesses angepasst werden. Die strategische Zuweisung von Ressourcen, die sich an der Geschwindigkeit und Verfügbarkeit jeder Lademaschine orientiert, trägt zu einem rationalisierteren und kostengünstigeren Betrieb bei und unterstreicht die Bedeutung dieser Variablen für die Erzielung optimaler Leistung.

Das mathematische Modell trägt wesentlich zu mehreren wichtigen Aspekten bei. Es ermöglicht eine effizientere Planung und Terminierung von Transportvorgängen und ermöglicht dynamische Anpassungen auf Grundlage sich ändernder Bedingungen. Darüber hinaus ermöglicht es eine vergleichende Analyse zwischen Kosten und Fertigstellungszeiten und optimiert so wirtschaftliche und produk Durch die Reduzierung von Ausfallzeiten und die Verbesserung der Ressourcennutzung bewältigt das Modell effektiv die logistischen und finanziellen Herausforderungen im Bergbau und passt sich an verschiedene Betriebsszenarien an , wodurch seine Anwendbarkeit und Relevanz in der Branche erhöht wird. Der Rest dieses Dokuments ist wie folgt strukturiert: Abschnitt 2 stellt die damit verbundene Arbeit vor. Abschnitt 3 beschreibt die Methodik im Detail. Computerexperimente und Ergebnisanalysen werden in Abschnitt Abschließend werden in Abschnitt 5 die aus der Studie gezogenen Schlussfolgerungen vorgestellt.

2. Verwandte Werke

Tabelle 1 bietet einen Überblick über verwandte Arbeiten, die sich mit dem Problem befassen. Sie enthält Einzelheiten zu den in unserem Interessengebiet berücksichtigten Merkmalen, wie z. B. Freigabezeiten (ri) und unterschiedliche Geschwindigkeiten (Qm). Die verwendeten Zielfunktionen werden ebenfalls analysiert, wie z. B. Fertigstellungszeit (Cmax) und Kosten (k). Darüber hinaus werden die verwendeten Lösungsmethoden untersucht und als exakt oder annähernd kategorisiert, zusammen mit der A

Machine Translated by Google

die jeweils angewandten Ansätze. Dieser Vergleich ermöglicht eine Visualisierung der verschiedene Methoden und Techniken, die in der Literatur implementiert sind, und hebt ihre Unterschiede und Gemeinsamkeiten hinsichtlich der vorliegenden Analyse.

Die Eigenschaft ri ist insbesondere in der Bergbauforschung noch nicht ausreichend erforscht wo die Verfügbarkeit von Transportausrüstung entscheidend ist, wie [8,9] zeigt. Im Gegensatz dazu Das Studium von Qm hat mehr akademische Aufmerksamkeit auf sich gezogen, obwohl der Schwerpunkt weitgehend auf identische Maschinen. In Bezug auf die Zielfunktion konzentrierten sich die Studien im Allgemeinen auf Minimierung von Variablen wie k und Cmax isoliert. Meza [2] veranschaulicht eine komplexere Integration, die eine Zeit- und Kostenanalyse der Lade- und Transportausrüstung beinhaltet durch ein Simulationsmodell. Hong und Lin [10] konzentrierten sich auf die Minimierung der maximalen Verspätung (Lmax), während Koryagin und Voronov [11] und Oliskevych et al. [12] die Erfüllung von P in Betracht zogen als unverzichtbar für die Aufrechterhaltung der Betriebskontinuität in Produktionsanlagen.

Im Vergleich dazu erweitert unsere Arbeit das Analyseniveau erheblich, indem sie sowohl Qm und ri , wodurch wir die Produktionsumgebungen realistischer modellieren können, insbesondere bei Erztransporten. Dies ist insbesondere dann relevant, wenn die Anzahl von Maschinen übersteigt fünf, ein Szenario, in dem viele Studien Näherungsmethoden bevorzugen, zu ihrer Rechenleistung. Darüber hinaus beinhaltet unser Ansatz eine iterative Studie, die ermöglicht eine Analyse auf taktischer Ebene, um die optimale Nutzung der Ausrüstung zu definieren, ergänzt durch robuste analytische Unterstützung bei der Entscheidungsfindung zur Maximierung der betrieblichen Effizienz. Unser Ansatz beinhaltet auch eine Komplexitätsanalyse, die noch erforscht werden muss [10,13,14]. Unsere Arbeit schließt also eine bedeutende Lücke in der Literatur und liefert wichtige Erkenntnisse, die Erleichtern Sie strategische Entscheidungen und verbessern Sie die Betriebseffizienz in der Bergbauindustrie.

Tabelle 1. Verwandte Arbeiten zum Problem der parallelen Maschinenplanung.

	Eigenschaften Ziel Funktion						Methoden		Anzahl Maschinen			
Referenz	Qm	ri	Cmax	k	Andere	Genau	Ungefähr	m ÿ 5	m ÿ 6			Komplexität Analyse	Ansätze
[2]	X		X	X			X		X				FIFO-Simulationsmodell
[15]			X			X		X					Filial- und Preisalgorithmus mit einzelner Server
[16]			X			X	X	X	X				MIP-Formulierung und Hybrid heuristischer Algorithmus mit einzelner Server
[17]			X			X	X	X					MIP-Formulierung und Tabu-Suche mit Einzelserver
[18]			X			X		X					MIP-Formulierung mit nicht-identischen Auftragsgrößen
[19]	X			X		X		X					MIP-Formulierung
[20]	X			X		X		X					MILP-Formulierung
[21]	X			X		X	X	X	X				MILP-Formulierung und Einfügung Heuristik bei zeitabhängigen Tarifen
[22]	X			X		X		X					MILP-Formulierung unter Zeitabhängige Tarife
[8]		X	X			X	X	X	X				MILP-Formulierung und GA-Algorithmus
[11]	X				P		X	X					Heuristischer Algorithmus
[23]			X			X		X					MIP-Formulierung mit Rüstzeit und Einzelserver
[10]			X		Lmax		X	X				X	Dynamische konstruktive Algorithmen
[24]			X			X	X						Constraint-Programmiermodell unter Nichtverfügbarkeitsbeschränkungen und LPT ändern
[13]			X			X	X	X				X	MIP-Formulierung und Tabu-Suche

Machine Translated by Google

Tabelle 1. Fortsetzung

	Eigenschaften Ziel Funktion					Methoden		Anzahl Maschinen			
Referenz Qm ri	Cmax	k	Andere	Genau	Ungefähr	m ÿ 5	m ÿ 6			Komplexität Analyse	Ansätze
[25]				X		X			X		MIP-Formulierung
[14]			X			X		X	X	X	MIP-Formulierung mit unabhängige Arbeitsplätze
[26]	X		X			X		X			MIP-Formulierung mit Funktionsbeschränkung
[12]	X				P	X		X			MIP-Formulierung
[27]				X			X	X			Monte-Carlo-Simulationsmodell
[9]		X	X				X	X			VNS- und ACO-Algorithmen
[28]	X		X			X	X		X		MIP- und TSP-Algorithmus
[29]			X			X		X			Constraint-Programmiermodell
[30]				X		X	X	X			MIP-Formulierung und MOBSO-Algorithmus
[31]	X			X		X	X	X			MILP-Formulierung mit Time-of-Use und SPT, MDPC, MDEC
[32]	X		X			X	X		X		MIP-Formulierung und GA–GWO Hybridalgorithmus
Unser Papier	X	X	X	X		X		X	X	X	Iterative MIP-Formulierung

Unter Berücksichtigung verwandter Studien zur Optimierung des Erztransportbetriebs war es Meza [2] entwickelte ein Modell für die Anwendung von Simulationskonzepten einer Be- und Entladesystem für Tagebaubetriebe unter Berücksichtigung der Beladungsart und Transportmittel, Ankunftszeit, Wartezeit, Positionierungszeit, Ladezeit, Abfahrtszeit Zeit, Rüstzeit, Entladezeit und Rückfahrzeit, zusätzlich zur Anzahl der Fahrten als Produktionsziel und die Berücksichtigung eines FIFO-Systems für das Be- und Entladen. Auf der anderen Andererseits führten Vasquez et al. [3] ein ganzzahliges Programmiermodell durch, um die Gesamtmenge zu minimieren Arbeitszeit einer Flotte von Niederquerschnittsladern in einem Untertagebergwerk, zusätzlich zur Entwicklung eines optimalen Algorithmus in polynomischer Zeit, der in den Entscheidungsprozess int Ergebnisse, die der optimalen Gesamtzeit sehr nahe kommen. Eivazy und Askari-Nasab [33] entwickelten ein Multi-Destination Mixed Integer Linear Programming (MILP)-Modell zur Minimi Betriebskosten im Tagebau; im Modell wurden neben horizontalen Ladesystemen und Rampenentscheidungen auch Lagerhalden und Mischhaufen berücksichtigt. Anjomshoa et al. [34] entwickelten ein Simulations- und gemischt-ganzzahliges Programmiermodell (MIP) zur Optimierung von Überholbuchten, Dimensionierung und Einsatzplanung des Fuhrparks in Untertagebergbau.

Darüber hinaus entwickelte Gligoric [35] eine Untersuchung über die Überholstellen für den Gütertransport Ausrüstung im Untertagebergbau, bei der versucht wurde, die notwendige Menge zu optimieren von Buchten und der optimalen Lage, um Wartezeiten zu minimieren. Gonen et al. [36] führten eine Studie für den Untertagebergbau, in der sie die jährlichen Produktionskapazitäten mit verschiedene Transportsysteme unter Berücksichtigung der Transportkosten je Einheit und der Minentiefe. Eivazy und Askari-Nasab [37] präsentierte eine Methodik, die auf Produktionsplanungshierarchien basiert für den Tagebau mittelfristig die Entwicklung eines MILP zur Minimierung der Betriebskosten; Die Planung umfasste Lagerstätten, Verarbeitungsanlagen und Deponien sowie eine Auswahl von Routen und Rampen. Tom-Socarras [38] entwickelte ein Werkzeug für das Kostenmanagement auf der Grundlage einer Entscheidungsbaum für den Materialtransport im Tagebau. Uribe [39] entwickelte eine Studie die Funktionalität und Leistung des halbautomatischen Betriebs beim Laden analysierte und Transportausrüstung im Untertagebergbau. Wi ÿecek [40] schlug eine geeignete Verfahren zur Steuerung der Beladung und des Transports von Erz in einem Bergwerk mit Pfählen und

Machine Translated by Google

Kammersystem. Wiest [41] entwickelte ein Planungsmodell für den Tagebau, bei dem die Kostenreduzierung im Vordergrund stand.

Solomon [42] modellierte Erzverlade- und -transportsysteme im Tagebau und entwickelte ein Modell, das die Abbaufront, die Routen, die Ausrüstung und die Erzziele berücksichtigte. Gaspar und Jhasmin [43] analysierten Nutzungs- und Verfügbarkeitsindikatoren zur Kostenoptimierung beim Erztransport. Li et al. [44] entwickelten im Rahmen ihrer Forschungen zum Untertagebergbau einen Bienenschwarmalgorithmus zur Maximierung des Gesamtertrags während eines Planungszeitraums; außerdem erstellten sie einen genetischen Algorithmus zur Planung der Abfertigung von Transportgeräten, um die Wartezeiten dieser Geräte zu minimieren. Elijah [45] entwickelte ein mathematisches Modell zur Berechnung der Ankunfts- und Servicerate unterschiedlicher Mengen von Transportgeräten, die beim Materialtransport eingesetzt werden. Huisa [46] verbesserte die Nutzung von Lade- und Transportgeräten, um die Kosten dieser Vorgänge zu senk

Choi et al. [47] entwickelten ein intelligentes, nicht überwachungsbasiertes System zur Vorhersage der Leistung des Transportausrüstungssystems im Tagebau unter Verwendung einer Kombination verschiedener Optimierungsmodelle. Li et al. [48] konzentrierten sich auf ein mathematisches Modell zur Umplanung von Vorgängen unter Berücksichtigung der Anforderungen an Produktionsräume, Betriebsumgebung und Verschleiß der Produktionsausrüstung, während sie gleichzeitig versuchten, die maximale Planungsabschlussrate und die geringste Schwankung der Erzqualität zu erreichen. Abolghasemian et al. [49] stellten eine Mehrzieloptimierung im Tagebausystem vor, bei der die maximale Fördermenge und die Minimierung der Transportzeit unter Berücksichtigung der Lagerkapazität, der Transportausrüstung und des Budgets angestrebt wurden. Shamsi et al. [50] führten eine IP mit Betriebsbeschränkungen durch, um das Transportsystem im Tagebau zu planen, wobei sie den besten Nettogegenwartswert als Ziel betrachteten. Parichehreh et al. [51] befassten sich mit dem Problem der energieeffizienten, unabhängigen parallelen Maschinenplanung, wobei sie die Arbeitsverschlechterung und die Lerneffekte der Bediener einbezogen und wertvolle Einblicke in die Mehrzieloptimierung für komplexe Produ

Unsere Studie stellt einen neuartigen und bedeutenden Ansatz zur Optimierung von Erzlade- und -transportvorgängen vor, indem ein taktisches Problem mithilfe eines Planungsmodells modelliert wird, das typischerweise für betriebliche Entscheidungen verwendet wird. Dieses Modell berücksichtigt parallele Maschinen mit unterschiedlichen Geschwindigkeiten und sorgt so für mehr Realismus. Insbesondere spiegelt sich die Art des Problems in der dynamischen Verfügbarkeit von Ladegeräten wider, die durch dynamische Freigabezeiten dargestellt wird, wobei Aufträge nicht im Voraus bekannt sind, sondern während ihrer Ausführung bekannt werden. Dieser Ansatz ermöglicht tiefere und detailliertere Analysen, die besser an die Komplexität von Bergbauproduktionsumgebungen angepasst sind. Die Implementierung dieses Modells kann in taktische und betriebliche Strategien integriert werden, um diese effizienter und effektiver zu gestalten und die Entscheidungsfindung in Bergbauunternehmen zu unterstützen, indem die betriebliche Effizienz deutlich verbessert, Kosten gesenkt und die Produktivität gesteigert wird. Darüber hinaus wird diese Studie sowohl zur theoretischen Entwicklung als auch zur praktischen Anwendung im Bergbauingenieurwesen beitragen und w

3. Methodik In

diesem Abschnitt wird die Methodik beschrieben, die in dieser Studie verwendet wurde. Sie ist wie folgt strukturiert : Abschnitt 3.1 definiert das Problem und legt den Kontext für die Forschung fest. Abschnitt 3.2 beschreibt die Modellannahmen im Detail und präsentiert die mathematische Formulierung des Problems, einschließlich der Zielfunktionen und Einschränkungen. Abschnitt 3.3 beschreibt die Algorithmen, die zur Bestimmung und Bewertung der optimalen Menge an Erztransportausrüstung verwendet werden. Dies umfasst das Verfahren zur Bestimmung der Mindestmenge an Transportausrüstung und die Bewertung der Kosten-Zeit-Kompromisse. Die folgenden Unterabschnitte behandeln die Forschungsziele systemat

3.1. Problemdefinition

Aufgrund der entscheidenden Bedeutung des Erztransports für die Produktion zielt diese Forschung darauf ab, die Mindestmenge an Transportausrüstung zu bestimmen, die erforderlich ist, um sowohl die mit der Gerätenutzung verbundenen Kosten als auch die Fertigstellungszeit des Produkti

Machine Translated by Google

innerhalb eines definierten Zeithorizonts. Erztransportvorgänge sind für die kontinuierliche Versorgung von Produktionsanlagen mit wichtigen Mineralien von entscheidender Bedeutung. Technische Entscheidungen bezüglich der erforderlichen Ausrüstung werden jedoch häufig auf der Grundlage von Erfahrungen und unter Zeitdruck getroffen. Diese Entscheidungen können zwar schnell getroffen werden, sind jedoch möglicherweise nur manchmal optimal und können zu Mehrkosten oder Verzögerungen führen, die sich auf die Anlagenproduktion auswirken.

Diese Forschung befasst sich mit der Ineffizienz und potenziellen Suboptimalität traditioneller Entscheidungsprozesse bei der Bestimmung der Mindestmenge an Transportausrüstung, die zur Erreichung des Produktionsziels erforderlich ist. Dies kann dazu führen, dass die Transportausrüstung entweder über- oder unterschätzt wird, was zu unnötigen Ausgaben oder Engpässen beim Erztransport führt. Durch die Entwicklung eines mathematischen Modells, das das Problem der parallelen Maschinenplanung mit unterschiedlichen Geschwindigkeiten und Freigabezeiten berücksichtigt , möchten wir einen systematischen und datengesteuerten Ansatz zur Optimierung der Ausrüstungszuweisung bieten .

Eine detaillierte Analyse des Kosten-/Fertigstellungszeitverhältnisses kann wertvolle Szenarien aufzeigen und im Falle von Unwägbarkeiten im Transportbetrieb fundiertere und angemessenere Entscheidungen ermöglichen. Mit diesem Modell können Entscheidungsträger die Kompromisse zwischen Gerätekosten und Betriebseffizienz abwägen und sicherstellen, dass die richtige Menge an Geräten eingesetzt wird, um die Produktionsziele zu erreichen, ohne unnötige Kosten zu verursachen.

Das Problem kann als paralleles Maschinenplanungsproblem mit unterschiedlichen Geschwindigkeiten und Freigabezeiten konzeptualisiert werden. Während der Hauptfokus auf einer taktischen Entscheidung zur Bestimmung der Mindestmenge an Transportausrüstung und zur Analyse der Kosten im Zusammenhang mit der Fertigstellungszeit liegt, wird eine Anpassung an dieses Planungsproblem unter Verwendung eines ganzzahligen Prog Nach der von Graham [52] entwickelten Notation wird ein Problem der Betriebsplanung durch drei Felder (ÿ|ÿ|ÿ) definiert: ÿ stellt die Systemkonfiguration dar, ÿ die Beschränkungen und Eigenschaften des Systems und ÿ die Zielfunktion. Unter Verwendung dieses Rahmens und der von Pinedo und Hadavi [53] beschriebenen Nomenklatur entspricht die Systemkonfiguration parallelen Maschinen mit unterschiedlichen Geschwindigkeiten (Qm), wobei diese Maschinen die Ladeeinrichtungen mit unterschiedlichen Verarbeitungszeiten sind. Eine wichtige Systemeigenschaft sind die Freigabezeiten (ri), die in der Transportzeit enthalten sind. Die Zielfunktion zielt darauf ab, sowohl die Kosten als auch die Fertigstellungszeit zu minimieren. Abbildung 2 zeigt den Zeitplan des Erztransportvorgangs, der als das zuvor erwähnte Problem konzeptualisiert wurde.

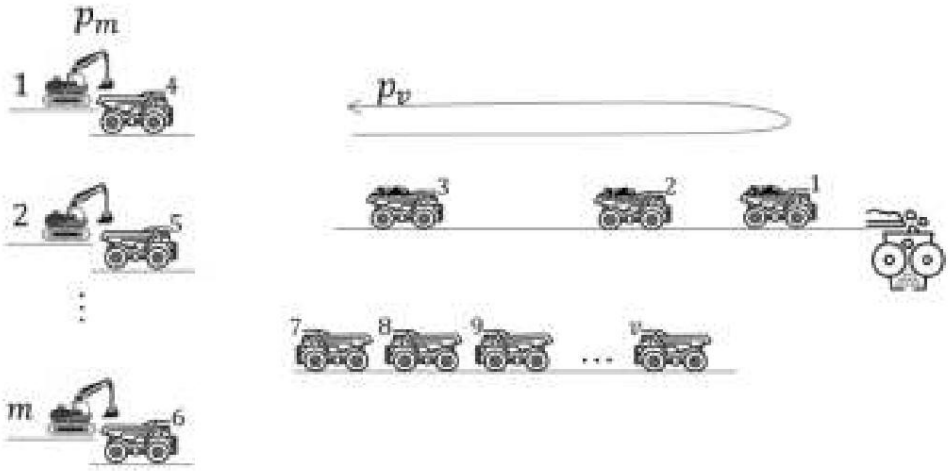


Abbildung 2. Zeitplan unter Berücksichtigung des Erztransportbetriebs.

3.2. Modellierung

Modellannahmen. Die Annahmen des Modells, basierend auf der von Meza [2] vorgestellten Arbeit, lauten wie folgt: • Es wurde nur ein

Erz-Entladepunkt berücksichtigt. • Die Betankungszeit ist vernachlässigbar. • Die Lade- und Transportausrüstung ist jederzeit verfügbar. • Die Wartezeit für das Erz-Entladen ist unbedeutend.

Machine Translated by Google

Notation. Tabelle 2 enthält die Notation für das vorgeschlagene Modell und zeigt den Typ und Definition.

Tabelle 2. Notation für das vorgeschlagene Modell.

Var	Typ	Definition
M	Satz	Satz von Ladegeräten M = {1, 2, . . . , m} .
V	Satz	Satz von Transportgeräten V = {1, 2, . . . , v}.
T	Satz	Satz von Zeiteinheiten: T = {1, 2, . . . , h}.
P_j	Parameter	Ladezeit für Ladegeräte j.
P_i	Parameter	Transportzeit für Transportgeräte i.
P	Parameter	Das Produktionsziel ist die Gesamtzahl der durchzuführenden Fahrten.
	Parameter	Zeithorizont zur Erreichung des Produktionsziels.
hk_j	Parameter	Kosten im Zusammenhang mit der Nutzung von Ladegeräten.
$k_i h$	Parameter	Kosten im Zusammenhang mit der Nutzung von Transportausrüstung.
x_{ijt}	Variable	Hauptentscheidungsvariable, nimmt den Wert 1 an, wenn das Transportgerät i verwendet zum Zeitpunkt t das Ladegerät j, andernfalls 0.
y_{jt}	Variable	Die Hilfsentscheidungsvariable nimmt den Wert 1 an, wenn das Ladegut j wird zum Zeitpunkt t, 0 verwendet, andernfalls.
z_{ickel}	Variable	Die Hilfsentscheidungsvariable nimmt den Wert 1 an, wenn das Transportgerät i wird zum Zeitpunkt t, 0 verwendet, andernfalls.
Cmax	Variable	Fertigstellungszeitpunkt des Produktionsziels
k	Variable	Kosten im Zusammenhang mit der Gerätenutzung

Zielfunktion. Die erste Zielfunktion (Gleichung (1)) versucht, k zu minimieren verbunden mit der Nutzung von Geräten. Die zweite Zielfunktion (Gleichung (2)) versucht, Cmax aller Fahrten.

$$\text{Mindestk} \tag{1}$$

$$\min . \text{Cmax} \tag{2}$$

Einschränkungen. In Bezug auf Einschränkungen stellt der Satz von Einschränkungen in Gleichung (3) sicher, dass die Zuordnung nur eines Transportfahrzeugs i zu einem Ladefahrzeug j zum Zeitpunkt t. Die Menge der Einschränkungen in Gleichung (4) sorgen dafür, dass nur ein Ladegerät j pro Transportgerät i vorhanden ist zum Zeitpunkt t.

$$\sum_{i \in V} x_{ijt} \le 1 \quad j \in M, t \in T \tag{3}$$

$$\sum_{j \in M} x_{ijt} \le 1 \quad i \in V, t \in T \tag{4}$$

Die in den Gleichungen (5) und (6) enthaltenen Beschränkungen stellen sicher, dass das Laden und Transportieren Die Ausrüstung wird bei Zuweisung für die erforderliche Dauer verwendet.

$$\sum_{t=1}^{t+p_j} y_{jt} \le (p_j - 1) \cdot x_{ijt} \quad i \in V, j \in M, t \in T \setminus (t \in h - p_j) \tag{5}$$

$$\sum_{t=1}^{t+p_j+p_i} z_{it} \le (p_i + p_j - 1) \cdot x_{ijt} \quad i \in V, j \in M, t \in T \setminus (t \in h - p_i - p_j) \tag{6}$$

Die Beschränkungen in den Gleichungen (7) und (8) stellen sicher, dass das Laden und Transportieren Geräte werden nicht doppelt belegt und nur bei Verfügbarkeit eingesetzt.

$$x_{ijt} \le 1 - y_{jt} \quad i \in V, j \in M, t \in T \tag{7}$$

$$x_{ijt} \le 1 - z_{it} \quad i \in V, j \in M, t \in T \tag{8}$$

Machine Translated by Google

Die in Gleichung (9) enthaltenen Beschränkungen stellen Cmax sicher .

$$\text{Cmax} \ddot{y} t \cdot xijt + \ddot{p} + p_j \quad \ddot{y}i \ddot{y} V, j \ddot{y} M, t \ddot{y} T \tag{9}$$

Der Satz von Beschränkungen in Gleichung (10) stellt sicher, dass P erfüllt wird.

$$\ddot{y} \ddot{y} \ddot{y} xijt \ddot{y} P \quad \ddot{y}i \ddot{y} V, j \ddot{y} M, t \ddot{y} T \tag{10}$$

Gleichung (11) berechnet k.

$$k = \ddot{y} \ddot{y} h_{k \cdot zit + i} \ddot{y} \ddot{y} l_{kj} \cdot yjt \ddot{y}i \ddot{y} V, j \ddot{y} M, t \ddot{y} T \tag{11}$$

Der Satz von Einschränkungen in den Gleichungen (12) und (13) stellt die Natur der Variablen sicher.

$$xijt, yjt, zit \ddot{y} \{0, 1\} \ddot{y}i \ddot{y} V, j \ddot{y} M, t \ddot{y} T \tag{12}$$

$$\text{Cmax}, k \ddot{y} 0 \tag{13}$$

Das vorgeschlagene Modell kann in taktischen und operativen Entscheidungsprozessen verwendet werden, da es beide Entscheidungen genau spezifiziert. Die Komplexität dieses gemischt-ganzzahligen linearen Programmiermodells in Bezug auf die Anzahl der ganzzahligen Variablen beträgt jedoch i·j·t + j·t + i·j, und es hat zwei weitere kontinuierliche Variablen. Wenn wir die Anzahl der Einschränkungen analysieren, finden wir insgesamt 7 ·i·j·t + j·t + i·j. Die Tatsache, dass das Modell auf binären Vektoren basiert (um die Startzeiten der LKWs und die Aktivitätszeiträume anzuzeigen), erhöht die Komplexität von Natur aus, eröffnet aber auch die Möglichkeit , (i) gültige Ungleichungen einzubeziehen, um Lösungen zu verwerfen, die nicht zur Struktur des Problems passen, oder (ii) Zerlegungstechniken wie die Verwendung kombinatorischer Bender-Schnitte oder die Einführung eines Rolling-Horizon-Frameworks [54] zu verwenden, um Strategien nach dem Prinzip „Teile und herrsche“ zu nutzen; beides ist Teil der laufenden Forschung. Man könnte eine Reduzierung der Zeiträume in Betracht ziehen, aber dies würde die Verwendung dieser Modelle in Steuerungssystemen wie denen verhindern, die von einem digitalen Zwilling bereitg

3.3. Algorithmen zur Bestimmung und Bewertung der optimalen Menge an Erztransportausrüstung Um

die beiden Ziele des Modells, die Minimierung von Fertigstellungszeit und -kosten, zu berücksichtigen , wurde ein iterativer Ansatz verwendet, bei dem Modelle mit nur einer Zielfunktion erstellt werden . Wie in den Algorithmen 1 und 2 gezeigt, ermöglicht der iterative Prozess die Untersuchung des Kompromisses zwischen Fertigstellungszeit und -kosten und bietet eine Reihe von Lösungen, die beide Ziele ausbalancieren.

Unter Berücksichtigung der Analyse auf taktischer Ebene wurde das folgende Verfahren vorgeschlagen, um die optimale Menge an Transportausrüstung zu bestimmen, wie in Algorithmus 1 beschrieben. Der Algorithmus hl beginnt mit der Definition der Eingabedaten, zu denen M, V, T , p i , p j , h und P. Anschließend werden die Ausgabedaten definiert, die die Mindestanzahl an Transportgeräten (vopt) und die Zuordnung der Transportgeräte zu Ladegeräten (Schedule) einschließlich Cmax umfassen.

Weisen Sie dann v der Größe von V zu, der Gesamtzahl der verfügbaren Transportgeräte (Zeile 1). Die Variable Continue wird auf true gesetzt, um die Schleife zu starten (Zeile 2). Solange sie wahr bleibt, wird ein Modell erstellt (unter Berücksichtigung der Gleichungen (2)–(13) des PMSPOHO-Modells, der Eingabedaten und v) und ein Ergebnis erhalten , das Cmax minimiert (Zeilen 3–5). Es wird eine Machbarkeitsbewertung durchgeführt. Wenn das Ergebnis durchführbar ist, wird v um eine Einheit reduziert (Versuch, das Problem mit weniger Transportgeräten zu lösen) (Zeilen 6–8). Wenn es nicht durchführbar ist, wird vopt um eine Einheit erhöht und die Schleife beendet (Zeilen 9–13). Dieser Algorithmus gibt vopt und Schedule zurück (Zeile 14).

Machine Translated by Google

Algorithmus 1: Bestimmer für optimale Transportausrüstung.

Eingabe: M, V, T , p $^{H}_{ich}$ **,** $p^{=}_{j}$ **, h, P;**

Ausgabe: vopt: Mindesttransportausrüstung, Zeitplan: Transportausrüstungszuweisung bis hin zur Beladungsausrüstung einschließlich des Cmax.

```
1: v ÿ |V|
2: Weiter ÿ true
3: während Weiter machen
4:     Modell ÿ ErstelleModell(Gleichungen(2)–(13), Eingabe, v)
5:     Ergebnis ÿ Lösen(Modell, Minimieren)
6:     wenn IsFeasible(Ergebnis) dann
7:         v ÿ v ÿ 1
8:         Ergebnisse, Zeitplan
9:     anders
10:        vopt ÿ v + 1
11:        Weiter ÿ false
12: Ende, wenn
13: Ende während
14: Rückkehr vopt, Zeitplan
```

Für die Analyse der Beziehung zwischen **k wurde folgendes Verfahren vorgeschlagen und Cmax, detailliert in Algorithmus 2. Dieser Pseudocode ermöglicht die Beobachtung des Verhaltens von k wenn Cmax erhöht wird. Der Algorithmus beginnt mit der Definition der Eingabedaten, die Folgendes umfassen: die gleichen Faktoren wie in Algorithmus 1;** k^{H}_{ich} **,** $k^{=}_{J}$ **, und v. Die Definition der Ausgabedaten umfasst dann eine Liste der Kosten und eine Liste der Beendigungszeiten (Cmax). Definieren Sie v (Zeile 1). Das Modellb wird erstellt (unter Berücksichtigung der Gleichungen (2)–(13) des PMSPOHO-Modells und der Eingangsdaten) (Zeile 2) und ein Ergebnislb aus der Minimierung von Cmaxlb erhält man (Zeile 3), und extrahiert dieses Ergebnis (Zeile 4). Ein modelub wird erstellt (unter Berücksichtigung der Gleichungen (1), (3)–(13) des PMSPOHO-Modells und der Eingangsdaten) (Zeile 5), und ein Ergebnis aus der Minimierung von k wird erhalten (Zeile 6), wobei Cmaxub für diese Ergebnis (Zeile 7).**

Algorithmus 2 Progressiver Kosten-Zeit-Evaluator

Eingabe: M, V, T , p $^{H}_{ich}$ **,** $p^{=}_{j}$ **,** $k^{h}_{ich,}$ k^{l}_{j} **, h, P und v;**

Ausgabe: ListPair(Cmax, k): Liste der Kosten in Bezug auf Cmax.

```
1: Definieren Sie v
2: modellb ÿ ErstelleModell(Gleichungen(2)–(13), Eingabe)
3: Ergebnislb ÿ Lösen(modellb, Minimieren)
4: Cmaxlb ÿ ExtrahiereCmax(Ergebnislb)
5: modelub ÿ Createmodel(Gleichungen (1),(3)–(13), Eingabe)
6: Resultub ÿ Lösen(modelub, Minimieren)
7: Cmaxub ÿ Extrahiere Cmax(Ergebnisub)
8: für Cmaxre_F in Cmaxlb bis Cmaxub tun
9:     modelre f ÿ ModellErstellen(Gleichungen (1),(3)–(8),(10)–(13), Eingabe)
10:    Ergebnis fÿ Solve(modelre f , Minimieren)
11:    Kosten ÿ Kosten extrahieren(Ergebnis f)
12:    Listenpaar(Cmax, k) ÿ ZuListe hinzufügen(Cmaxre_F, Kosten)
13: Ende für
14: Rückgabewert Cmax, k
```

Es wird eine progressive Auswertung durchgeführt, die Cmaxre iteriert $_F$ **um 1 von Cmaxlb auf Cmaxub (Zeile 8). Für jeden Wert wird ein Modellre f erstellt (unter Berücksichtigung der Gleichungen (1), (3)–(8), und (10)–(13) des PMSPOHO-Modells und der Eingabedaten) (Zeile 9). Das Ergebnis f wird durch Minimierung von k für jede Iteration erhalten** (Zeile 10); k wird extrahiert und einer Liste (Kosten) für jedes **entsprechende Cmaxre ,k) hinzugefügt (Zeilen 14),** $_F$ **(Zeilen 11–12). Dieser Algorithmus gibt die Liste der Paare zurück (Cmaxre** $_F$

Machine Translated by Google

Bereitstellung einer detaillierten Bewertung, wie sich k bei Änderungen von $_{F}$. Im Algorithmus Cmaxre ändert . Cmaxre dient als Zwischenreferenzwert, um den Kompromiss zwischen der Minimierung von Cmax und k iterativ zu bewerten. Diese iterative Bewertung bietet Einblicke in die Kos Kompromisse im Modell.

4. Computergestützte Experimente und Ergebnisanalyse

In diesem Abschnitt werden die Computerexperimente zur Bewertung der Leistung von das vorgeschlagene PMSPOHO-Modell und die Algorithmen. Alle Experimente wurden auf einem Computer mit einem Windows 11-Betriebssystem, Intel(R) Core(TM) i5-12450H der 12. Generation, 2,50 GHz Prozessor und 8 GB RAM. Die Programmiersprache Python 3.11 wurde verwendet für die Entwicklung. Der kommerzielle Optimierer Gurobi 10.0.1 und der kostenlose Optimierer CBC wurden wird mit einem Suchzeitlimit von 3600 s und einem GAP von 0,01 verwendet.

Basierend auf tatsächlichen Daten aus den Studien von Meza [2], Vásquez [3] und Anchiraico und Rojas [4], die die Anzahl der Lade- und Transportgeräte, die Lade- Zeiten, Transportzeiten, Zeithorizonte und Produktionsziele wurden 34 experimentelle Fälle generiert. Diese Instanzen sind so konzipiert, dass sie reale Szenarien simulieren und eine robuste Evaluierung unserer vorgeschlagenen Methoden unter verschiedenen Bedingungen. Dieser Ansatz gewährleistet dass unser Versuchsaufbau auf realistischen Betriebsparametern basiert, was die Anwendbarkeit und Relevanz unserer Ergebnisse. Es ist wichtig zu beachten, dass ein direkter Vergleich mit den oben genannten Studien der Autoren wurde nicht durchgeführt, da ihre Forschungsmethoden unterschiedlich sind erheblich von unserem vorgeschlagenen Ansatz ab.

4.1. Anpassen der Werte für die Optimiererparameter

Gurobi Optimizer Version 10.0.1 verfügt über eine umfangreiche Palette von Parametern, die um die Leistung des Optimierers zu optimieren. Angesichts der Vielzahl möglicher Optionen Konfigurationen, Identifizierung der optimalen Parameterkombination, die das Gesamtergebnis verbessern Die Leistung des Modells stellt eine große Herausforderung dar. Die TuningAPI von Gurobi bietet eine wertvolles Tool zum Erkunden verschiedener Parametereinstellungen, die die Modellleistung verbessern können in bestimmten Fällen [56].

Unter Berücksichtigung des von Cuellar-Usaquén et al. [57] angegebenen Verfahrens wird eine Stichprobe von sieben Instanzen wurden erstellt, um die geeigneten Parametereinstellungen für unser Modell zu ermitteln. TuningAPI. Dieses Beispiel umfasst die Fälle, in denen die von das PMSPOHO-Modell mit den Standardparametern (PMSPOHO-D) übertrifft inzwischen die seiner jeweiligen Klasse um zwei Standardabweichungen.

Tabelle 3 zeigt die Ergebnisse der Gurobi-Parameteroptimierung für sieben verschiedene Problemfälle . Jede Tabellenzeile stellt einen Fall dar, die Spalten geben die spezifischen Gurobi- Parameter, die für diese Instanz angepasst wurden. Die Parameter umfassen Symmetry, DegenMove, Heuristik, GomoryPasses, PreDepRow, MIPFocus, Methode, CutPasses und BranchDir.

Tabelle 3. Für das mathematische Modell gefundene Parameterkonfigurationen.

	Zusammenfassung			Wertparameter										
Inst	vm	P	Symmetrie	DegenMoves	Heuristik	GomoryPässe	PreDepRow	MIPFocus	Methode	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 %

Der dritte Fall fällt deutlich auf, da er die höchste Variation aufweist (20,92%) zwischen der Standard-Rechenzeit (Tc-D) und der abgestimmten Rechenzeit Zeit (Tc-T), was auf eine deutliche Leistungsverbesserung mit den angepassten Parametern hinweist. In diesem Fall sind die Parameter Heuristics, GomoryPasses und PreDepRow wurden angepasst, was zu einer deutlichen Reduzierung der Rechenzeit von 69,59 auf 55,03 Zeiteinheiten. Diese Anpassung reduzierte die Gesamtberechnungszeit und optimierte die Nutzung von Rechenressourcen, wodurch der Prozess effizienter und effektiver wird. Die

Machine Translated by Google

Der erhebliche Unterschied zwischen Tc-D und Tc-T unterstreicht die Bedeutung der Anpassung und Feinabstimmung der Modellparameter, um eine maximale Effizienz bei der Lösung komplexer mathematischer Probleme zu erreichen.

In diesem Fall erweist sich der angepasste Parameter Heuristik als Schlüsselfaktor bei der Beschleunigung des Lösungsprozesses. Durch die verstärkte Verwendung heuristischer Methoden, also Lösungsstrategien, die einfache und schnelle Regeln verwenden, um schnell zufriedenstellende Lösungen zu generieren, konnte das Modell potenzielle Lösungen effizienter untersuchen und bewerten. Dieser Ansatz reduzierte die zum Erreichen einer praktikablen Lösung erforderliche Zeit erheblich und stellte seine Praktikabilität und Wirksamkeit unter Beweis. Die Anwendung von Heuristiken ist ein wertvolles Werkzeug, insbesondere in Szenarien, in denen Geschwindigkeit entscheidend ist und ein Teil der Präzision zugun

GomoryPasses, ein weiterer anpassbarer Parameter, steuert die Anzahl der Gomory-Schnittpässe. Diese Durchläufe sind Optimierungstechniken, die die lineare Relaxation des Modells stärken und so zu einer schnelleren Konvergenz führen. Die Anpassung der Anzahl dieser Durchläufe kann die Effizienz der Lösung erheblich beeinflussen und die Notwendigkeit zur Verbesserung der Lösung mit der für jeden Schnitt erforderlichen zusätzlichen Zeit in Einklang bringen. PreDepRow, das die Vorverarbeitung abhängiger Zeilen übernimmt, hat die Effizienz ebenfalls grundlegend verbessert, indem es die Problemgröße reduziert und die numerische Stabilität verbessert. Eine effektive Vorverarbeitung kann das Modell vereinfachen, indem Redundanzen eliminiert und die Gesamtkomplexität reduziert wird. Im dritten Fall reduzierte die Anpassung von PreDepRow die Rechenzeit erheblich, was die Bedeutung einer ordnungsgemäßen Modellverwaltung unterstreicht. Diese Anpassungen unterstreichen die Bedeutung der Feinabstimmung von Parametern zur Verbesserung des Effizienzlösers und ermöglichen schnellere und genau

4.2. Benchmarking-Ergebnisse für Optimierer

Tabelle 4 zeigt die Benchmarking-Ergebnisse von Optimierern in 34 Instanzen. Die Instanzen wurden unter Berücksichtigung des Produktionsziels organisiert, da mit zunehmendem Ziel eine Erhöhung der Rechenzeit (Tc) festgestellt wurde. Die Ergebnisse wurden zwischen dem Standardmodell (PMSPOHO-D), dem abgestimmten Modell (PMSPOHO-T) und dem Modell mit dem kostenlosen Optimierer (PMSPOHO-CBC) verglichen. Die kritischen Parameter für den Vergleich waren Tc und GAP, die das Komplexitätsmodell und den Effizienzlöser messen. In dieser Studie wird der GAP-Wert direkt vom Gurobi-Optimierer als Teil s

Das optimierte Modell (PMSPOHO-T) kommt durchweg mit kürzeren Rechenzeiten zurecht als die anderen Modelle. Dies deutet darauf hin, dass die Parameteroptimierung die Effizienz des Lösers deutlich verbessert hat. Der GAP, also die Differenz zwischen der besten bekannten Lösung und der besten vom Löser gefundenen Untergrenze, beträgt sowohl für PMSPOHO-D als auch für PMSPOHO-T 0,00 %, was darauf hinweist, dass die gefundenen Lösungen optimal sind. PMSPOHO-CBC weist jedoch in mehreren Fällen einen höheren GAP auf, was auf eine geringere Präzision hinweist und die Einschränkungen eines kommerziellen Optimierers verdeutlicht.

PMSPOHO-T sticht mit den besten Ergebnissen in Bezug auf Tc hervor und löst Instanzen schneller, während die Lösungspräzision erhalten bleibt. Dies steht in starkem Kontrast zum Modell mit dem kostenlosen Optimierer (PMSPOHO-CBC), das in vielen Fällen einen höheren GAP aufweist und damit seine Einschränkungen verdeutlicht. Es zeigt auch, dass die Komplexität des Problems NP-schwer ist, was bedeutet, dass mit zunehmenden Instanzen optimale Ergebnisse nach deutlich längerer Zeit erzielt werden könnten, was einen kostenlosen Optimierer zu einer ineffizienten Option für Feldan

Abbildung 3 vergleicht die Rechenzeit (Tc) und Cmax für die Optimierer PMSPOHO-D, PMSPOHO-T und PMSPOHO-CBC für Instanz 34. PMSPOHO-D und PMSPOHO-T erreichen denselben Cmax von 183, aber PMSPOHO-T ist in Bezug auf die Rechenzeit effizienter, mit 671,8 im Vergleich zu 798,8 für PMSPOHO-D. Andererseits zeigt PMSPOHO-CBC mit einem Cmax von 229 und einer Rechenzeit von 3600 die schlechteste Leistung. Dies zeigt , dass PMSPOHO-T der effizienteste der drei bewerteten Optimierer ist und ein gutes Gleichgewicht zwischen Rechenzeit und Cmax erre

Machine Translated by Google

Tabelle 4. Benchmarking-Ergebnisse für Optimierer pro Instanz.

Beispiel	P vm Cmax	PMSPOHO-D			PMSPOHO-T			PMSPOHO-CBC		
			Tc	Lücke Cmax		Tc	Lücke Cmax			Tc-Lücke
	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 %	31	3600	5,21 %
3	2 81 20 4 81 16 4	29	0,00 %	1,41 0,00	29	1,58	0,00 % 1,03	30	3600	4,82 %
4	81 10 4 81 8 4 81	31	% 1,13	0,00 %	31	0,00 %	1,01 0,00	32	3600	3,77 %
5	6 4 81 4 4 81 20 6	34	0,9	0,00 % 28,25	34	% 1,82	0,00 %	35	3600	2,36 %
6	81 16 6 1 10 6 81	40	0,00 %	18,34 0,00	40	26,74	0,00 %	41	3600	2,56 %
7	8 6 81 6 6 90 5 3	51	% 49,23	% 69,59	51	18,61	0,00 %	51	1846	0,00 %
8	90 4 3 90 3 3 200	55	0,00 %	66,54 0,00	55	46,08.	0,0 0 %	60	3600	8,52 %
9	19 4 200 18 4 200	55	% 76,29	0,00 %	55	50,86	0,00 %	60	3600	8,20 %
10	17 4 200 16 4 200	64	42,12	0,00 %	64	63,01	0,00 % 59,78	68	3600	6,45 %
11	15 4 200 14 4 200	78	318,67	0,00 % 94,3	78	0,00 %	42,26 0,00	82	3600	5,50 %
12	13 4 200 12 4 200	100	0,00 %	100,46 0,00	100	% 130,11	0,00 %	106	3600	5,72 %
13	11 4 200 10 4 200	147	% 145,77	0,00 %	147	76,82	0,00 %	152	3600	3,27 %
14	9 4 200 8 4	38	42,98	0,00 % 83,77	38	100,13	0,00 %	42	3600	10,25 %
15		42	0,00 %	44,52 0,00	42	90,32	0,00 % 26,41	46	3600	9,86 %
16		63	% 1784,21	0,00 %	63	0,00 %	40,36 0,00	70	3600	10,52 %
17		77	1693,64	0,00 %	77	% 30,78	0,00 %	85	3600	10,73 %
18		98	1634,63	0,00 %	98	1627,21	0,00 %	110	3600	12,72 %
19		130	1640,24	0,00 %	130	1437,64	0,00 %	163	3600	25,32 %
20		161	1584,04	0,00 %	161	1463,63	0,00 %	202	3600	25,47 %
21		212	1554,43	0,00 %	212	1400,24	0,00 %	259	3600	22,23 %
22		125	1532,23	0,00 %	125	1437,04	0,00 %	161	3600	29,03 %
23		125	1524,64.	0,0 0 %	125	1291,43	0,00 %	161	3600	28,53 %
24		125	1279,45	0,00 %	125	1349,23	0,00 %	161	3600	28,93 %
25		125	1017,6	0,00 % 865,6	125	1372,64.	0,0 0 %	161	3600	28,73 %
26 27		125	0,00 %	798,8 0,00 %	125	1019,45	0,00 %	164	3600	31,45 %
28		125			125	746,6	0,00 % 615,6	159	3600	27,32 %
29		125			125	0,00 %	671,8 0,00	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 %

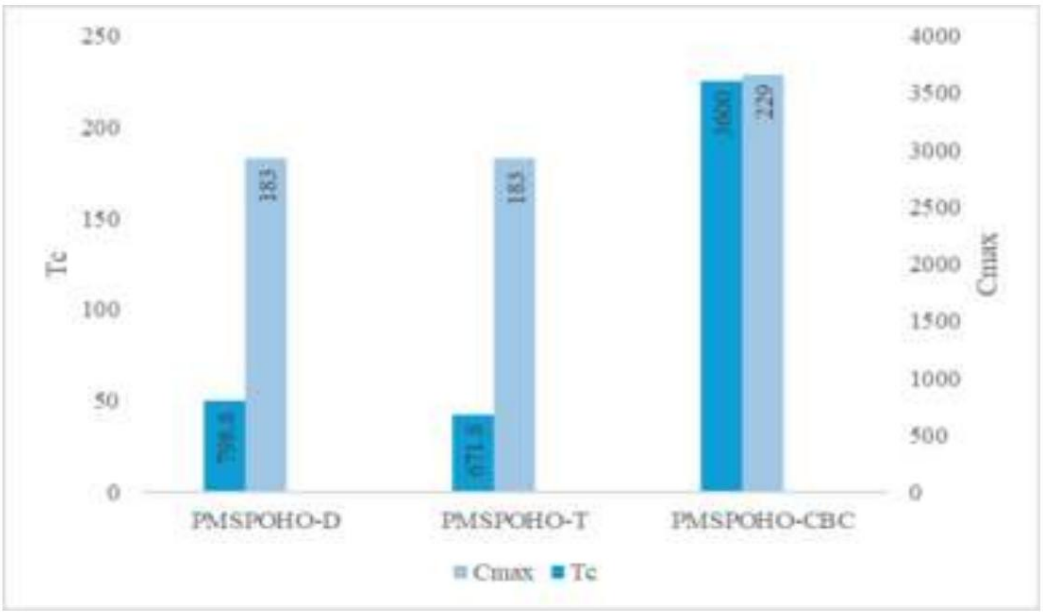


Abbildung 3. Vergleich von Rechenzeit und Cmax für Instanz 34.

4.3. Einblicke in den taktischen Ansatz

Tabelle 5 zeigt die Effizienz der verwendeten Algorithmen. Algorithmus 1 berechnet die minimale Anzahl an Transportgeräten, was zu acht Geräten führt. Auf der anderen Andererseits führt Algorithmus 2 eine k-Cmax- Auswertung durch, die eine angemessenere Analyse ermöglicht zwischen diesen beiden Variablen. Dies verbessert die taktische Entscheidungsfindung, indem es zeigt, wie Variationen von Cmax beeinflussen k und umgekehrt. Die Minimierung von Cmax führt zu deutlich reduzierten Werte (zwischen 33 und 49), obwohl eine Erhöhung von k es Planern ermöglicht, zeitliche Effizienz und Betriebskosten. Algorithmus 2 bietet eine umfassendere und ausgewogenere

Machine Translated by Google

Perspektive für taktische Entscheidungen und wird durch Algorithmus 1 unterstützt, der definiert die Anzahl der Ressourcen, die zum Erreichen des Produktionsziels erforderlich sind.

Tabelle 5. Ergebnisse der Optimierung der vorgeschlagenen Algorithmen.

			Minimieren Sie k		Cmax minimieren		
—	M	P	Cmaxub k	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	21.40
10	4	200	51	2350 103,30	40	3848	32,76
9	4	200	50	2350 44,62	44	3498	19.40
8	4	200	51	2350 76,94	49	3148	15,93

Mit der Verkürzung der Fertigstellungszeiten steigen die Kosten überproportional an, was auf einen Kompromiss zwischen Zeit- und Kosteneffizienz bei Erztransporten hindeutet. Die steileren Kurven bei nie Die Fertigstellungszeiten bedeuten erhebliche Kostensteigerungen für schnellere Operationen, während die flacheren Steigungen bei höheren Zeiten deuten auf eine Mindestkostenschwelle hin. Die breitere Spanne der Fertigstellungszeiten mit mehr Transportausrüstung bedeutet, dass, obwohl zusätzliche Ausrüstung zwar die Abläufe beschleunigen, aber auch höhere Kosten verursachen, was die Bedeutung eines Ausgleichs unterstreicht Budgetbeschränkungen mit betrieblicher Dringlichkeit bei der Planung von Bergbauarbeiten.

Abbildung 4 zeigt die Beziehung zwischen Kosten und Fertigstellungszeit für verschiedene Anzahl der Transportgeräte. Es gibt eine deutliche Kostensenkung bei nur geringfügiger Erhöhung in der Fertigstellungszeit. Ein taktischer Ansatz zur Bestimmung der Mindestmenge an Transport Bei der Auswahl der erforderlichen Ausrüstung zur Erreichung der Produktionsziele sollten nicht nur die Kosten und die Zeit, sondern auch die Steigung jeder Linie, was die Anwendung der Algorithmen unterstützt in diesem Papier vorgeschlagen. Beispielsweise würde eine Umstellung von neun auf acht Transporteinheiten erhöht die Fertigstellungszeit und senkt die Kosten erheblich. Dies deutet darauf hin, dass der Betrieb mit acht Transporteinheiten könnte eine effektive taktische Entscheidung sein, wenn die zusätzliche Zeit andere Prioritäten des Erztransportbetriebs dürfen nicht beeinträchtigt werden.

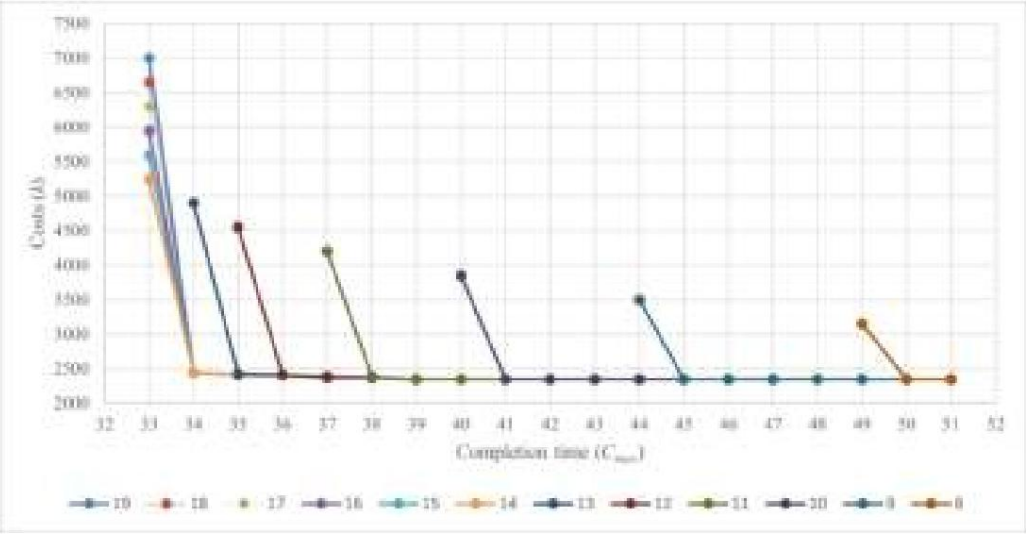


Abbildung 4. Kosten k vs. Fertigstellungszeit Cmax für die in Tabelle **5 vorgeschlagenen Instanzen .**

Machine Translated by Google

Entscheidungsträger müssen die Kostensteigerungen, die mit dem Einsatz zusätzlicher Ausrüstung verbunden sind, gegen die Verkürzung der Transportzeit abwägen. Ab einem bestimmten Punkt verringert sich durch den Einsatz zusätzlicher Transporteinheiten zwar der Vorteil einer Verkürzung der Fertigstellungszeit, es entstehen jedo **Um den Betrieb zu optimieren und die Ressourcen effizient zu verwalten, muss der Zeitpunkt erkannt werden, an dem die Mehrkosten die Verkürzung der Transportzeit nicht mehr rechtfertigen .**

4.4. Erkenntnisse zum operativen Ansatz

Abbildung 5 zeigt den Zeitplan, z. B. 34, der acht Transport- und vier Ladegeräte berücksichtigt. Dies veranschaulicht das Erzladeprogramm für einen Bergbaubetrieb während der ersten 50 Fahrten des Produktionsziels. Es ist interessant zu sehen, wie die Zuteilung organisiert ist, um sicherzustellen, dass die minimal erforderliche Menge an Transportgeräten ständig in Bewegung gehalten wird, was auf eine sorgfältige Bewertung der Ladezeiten und Intervalle zwischen den Aufgaben hinweist. Diese Strategie stellt sicher, dass kein Ladegerät ungenutzt bleibt und dass die Transportgeräte effektiv verteilt werden, um das

Durch die Minimierung der Menge an Transportausrüstung kann der Betrieb kosteneffizienter werden und die Logistik durch die Reduzierung von Staus und potenziellen Engpässen beim Erztransport verbessert werden. Dieser ausgewogene Ansatz zwischen Effizienz und Effektivität bei der Planung ist entscheidend für einen erfolgreichen Betrieb, der sich an die dynamischen und oft unvorhersehbaren Anforderungen des Bergbaus anpasst.

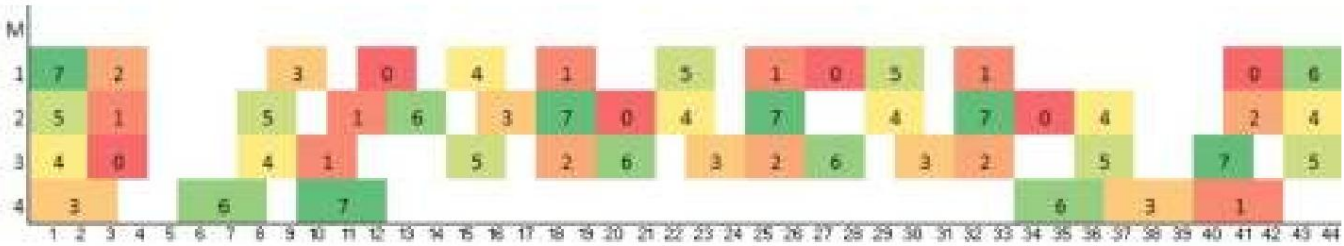


Abbildung 5. Zeitplan für Instanz 34 unter Berücksichtigung der minimalen Fertigstellungszeit Cmax.

Tabelle 6 bietet eine detaillierte Verteilung der Auslastung von Lade- und Transportgeräten in einem Bergbaubetrieb und verdeutlicht die Zuweisung und Nutzung dieser Ressourcen in der Produktion . Es ist zu beobachten, dass die Ladegeräte zwei und drei mit Auslastungsprozentsätzen von 32 % bzw. 30 % am stärksten ausgelastet sind, was auf eine hohe Nachfrage und häufige Nutzung hinweist. Im Gegensatz dazu ist das Ladefahrzeug vier mit nur 12
Dies lässt darauf schließen, dass die Ressourcen nicht vollständig optimiert werden, insbesondere wenn man berücksichtigt, dass der Betrieb des Ladefahrzeugs Nr. 4 am kostspieligsten sein könnte, da weniger häufig genutzte Geräte in der Regel regelmäßig gewartet werden müssen und wertvollen Platz beanspruchen, der besser genutzt werden könnte.

Die Unterauslastung der teuren Verladeausrüstung wirft erhebliche Fragen zum Kostenmanagement und zur betrieblichen Effizienz auf. Die hohen Kosten dieser Ausrüstung können auf spezielle Funktionen zurückzuführen sein, die eine anspruchsvollere Wartung oder eine höhere Anfangsinvestition erfordern. Ihre suboptimale Auslastung bedeutet jedoch, dass diese Investition die erwarteten Erträge bringen muss, was bei Betrieben, die auf maximale Effizienz und Rentabilität abzielen, von entscheidender Bedeutung ist. Diese Situation unterstreicht die dringende Notwendigkeit, die Aufgabenverteilung und Ressourcenzuweisung neu zu bewerten, um sicherzustellen, dass alle Verladeausrüstungen auf eine Weise genutzt werden, die ihre Kosten rechtfertigt, wodurch der gesamte Betrieb optimiert und eine potenzielle Ressourcenverschwendung reduziert wird.

Die Ergebnisse zeigen, dass das vorgeschlagene Modell hinsichtlich der Rechenzeit effizient ist und dass die vorgeschlagenen Algorithmen nützliche taktische Werkzeuge für eine effektivere Entscheidungsfindung hinsichtlich der Zuteilung der Transportausrüstung sind, die zum Erreichen der gesetzten Ziele im Erztransportbetrieb erforderlich ist. Es muss jedoch anerkannt werden, dass die Analyse bedeutenderer Fälle als mathematisches Modell die zum Erzielen von Ergebnissen erforderliche Zeit verlängern kann , selbst wenn die Optimierungsparameter genau abgestimmt sind. Daher wird es wichtig sein, die Zeiträume für die Lösungssuche zu verlängern und Näherungsmethoden zu entwickeln, um diese Zeiten zu verkürzen und praktische und umsetzbare Lösungen zu erreichen.

Machine Translated by Google

Tabelle 6. Auslastung der Lade- und Transportgeräte.

	Ladeausrüstung					
Transportausrüstung	1	2	3	4	Gesamt	Verwendung
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 %
Gesamt	13	16	15	6	50	
Verwendung	26 %	32 %	30 %	12 %		

5. Schlussfolgerungen

Das vorgeschlagene Modell weist eine erhebliche Effizienz in Bezug auf die rechnerische Zeit, wodurch Kosten und Fertigstellungszeit bei Erztransportvorgängen effektiv ausgeglichen werden. Die Anwendung Die beiden Algorithmen lieferten taktische Werkzeuge, die die Entscheidungsfindung in Bezug auf die Zuteilung von Transportmitteln. Die Effizienz des Modells zeigt sich in der rechnerischen Die Ergebnisse zeigen, dass die Parameteroptimierung die Leistung des Lösers deutlich verbessern kann. Das optimierte Modell erreichte durchweg geringere Rechenzeiten als die Standardeinstellungen. was die Wichtigkeit sorgfältiger Parameteranpassungen unterstreicht. Darüber hinaus dynamische Freigabezeiten und unterschiedliche Geschwindigkeiten der Transportausrüstung im Modell fügen eine Ebene hinzu Realismus, der sicherstellt, dass die Planung die tatsächlichen Betriebsbedingungen widerspiegelt. Dieser Aspekt optimiert den Zuordnungsprozess weiter und macht das Modell rechnerisch robust.

Trotz der Ergebnisse weist das Modell Einschränkungen auf. Eine Haupteinschränkung ist seine Skalierbarkeit. Mit zunehmender Größe der Instanzen kann die Zeit, die zum Erhalten von Ergebnissen benötigt wird, prohibitiv, selbst bei optimierten Parametern. Dieses Problem ist mathematischen Modellen inhärent Umgang mit komplexen, NP-schweren Problemen. Darüber hinaus ist die Abhängigkeit des Modells von exakten Optimierungsmethoden sind für Echtzeit-Entscheidungen nur manchmal praktikabel in dynamische Umgebungen, in denen schnelle Reaktionen unerlässlich sind.

Zukünftige Arbeiten sollten sich auf die Verlängerung der Lösungssuchperioden und die Entwicklung von Näherungsmethoden zur Behebung dieser Einschränkungen konzentrieren. Diese Methoden, einschließlich heuristis metaheuristische Ansätze können die Rechenzeiten verkürzen und nahezu optimale praktische und ausführbare Lösungen in realen Szenarien liefern. Darüber hinaus kann die Integration von Maschi Lerntechniken zur Vorhersage und dynamischen Anpassung von Parametern könnten die Leistungsfähigkeit und Anpassungsfähigkeit des Modells [58].

Beiträge der Autoren: Konzeptualisierung, LT-T., CA und DA-M.; Methodik, LT-T. und CA; Software, LT-T. und CG; formale Analyse, LT-T. und AP; Untersuchung, LT-T. und AP; Daten Kuratierung, LT-T.; Schreiben – Vorbereitung des Originalentwurfs, LT-T.; Schreiben – Überprüfung und Bearbeitung, LT-T., CA und DA-M.; Visualisierung, LT-T.; Supervision, CA und DA-M.; Projektadministration, LT-T. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu.

Finanzierung: Das APC wurde von der Universität Los Andes finanziert.

Erklärung des Institutional Review Board: Nicht zutreffend.

Einverständniserklärung: Nicht zutreffend.

Datenverfügbarkeitserklärung: Die in dieser Studie präsentierten Daten sind verfügbar unter: https://github.com/LuisTarazonaTorres/Parallel-Machine-Scheduling-problem---Hualing-Equipment.git (abgerufen am 29. Juli 2024).

Danksagung: Die Autoren danken allen Teilnehmern, die zum Durchführung dieser Studie.

Machine Translated by Google

Interessenkonflikte: Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Verweise

1. Rout, M.; Parida, CK Optimierung von Sprengparametern im Tagebau. Dissertation, National Institute of Technology Rourkela, Odisha, Indien, 2007.
2. Meza Castro, JE. Entwurf eines Modells für die Simulationsanwendung eines Fracht- und Bergbausystems sowie eines Minenbetriebs am Tajo Abierto. Bachelorarbeit, Päpstliche Katholische Universität von Peru, San Miguel, Peru, 2011.
3. Vasquez, OC; Sepúlveda, JM; Córdova, F. Modellierung und Optimierung des Fahrzeugbetriebs im unterirdischen Kupferbergbau. Rem Rev. Esc. De Minas 2011, 64, 365–371. [CrossRef]
4. Anchiraico, A.; Rojas, K. Optimierung des Acarreo- und Transportsystems und Vorbereitungsarbeiten für Tiefenzonen, Vermittlung der Six Sigma-Methodik durch ECM Zicsa und die Minera Inmaculada. Bachelorarbeit , Universidad Peruana de Ciencias Aplicadas, Santiago de Surco, Peru, 2023.
5. Soofastaei, A.; Karimpour, E.; Knights, P.; Kizil, M. Energieeffiziente Lade- und Transportvorgänge; Springer: Cham, Schweiz, 2018; S. 121–146. [CrossRef]
6. Choi, Y.; Baek, J.; Park, S. Überblick über GIS-basierte Anwendungen für den Bergbau: Planung, Betrieb und Umweltmanagement. Appl. Sci. 2020, 10, 2266. [CrossRef]
7. Owolabi, A. Auswahl der Lade- und Transportausrüstung für eine optimale Produktion in einem Granitsteinbruch. 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 limite. Masterarbeit, Universitat Politècnica de València, Valencia, Spanien, 2017.
9. Salazar-Hornig, E.; Soto-Gavilán, G. Minimierung der Bearbeitungszeit für parallele Maschinenprobleme, keine Beziehung zu den Setup-Zeiten, abhängig von der Sequenz, die einen hybriden VNS/ACO-Algorithmus vermittelt. Rev. Ing. Univ. De Medellín 2021, 20, 171–184. [CrossRef]
10. Hong, HC; Lin, BM Parallele dedizierte Maschinenplanung mit Konfliktgraphen. Comput. Ind. Eng. 2018, 124, 316–321. [Querverweis]
11. Koryagin, M.; Voronov, A. Verbesserung der Organisation der Schaufelrad-LKW-Systeme in Kohletagebauen. Transp. Probl. 2017, 12, 113–122. [CrossRef]
12. Oliskevych, M.; Kovalyshyn, S.; Magats, M.; Shevchuk, V.; Sukach, O. Die Optimierung der Einsatzplanung von LKW-Flotten im Hinblick auf ihre Interaktion und die Einschränkungen der europäischen Vereinbarung über die Arbeit der Besatzungen. Transp. Probl. 2020, 15, 157–170. [CrossRef]
13. Wu, L.; Wang, S. Exakte und heuristische Methoden zur Lösung des Problems der parallelen Maschinenplanung bei Multiprozessor-Tasks. Int. J. Prod. Econ. 2018, 201, 26–40. [CrossRef]
14. Elidrissi, A.; Benmansour, R.; Benbrahim, M.; Duvivier, D. Mathematische Formulierungen für die parallele Maschinenplanung Problem mit einem einzelnen Server. Int. J. Prod. Res. 2021, 59, 6166–6184. [CrossRef]
15. Gan, HS; Wirth, A.; Abdekhodaee, A. Ein Branch-and-Price-Algorithmus für den allgemeinen Fall der Planung paralleler Maschinen mit einem Einzelserver. Comput. Oper. Res. 2012, 39, 2242–2247. [CrossRef]
16. Kim, MY; Lee, YH MIP-Modelle und Hybridalgorithmus zur Minimierung der Durchlaufzeit paralleler Maschinen Planungsproblem mit einem einzigen Server. Comput. Oper. Res. 2012, 39, 2457–2468. [CrossRef]
17. Hasani, K.; Kravchenko, SA; Werner, F. Minimierung von Interferenzen bei der Planung zweier paralleler Maschinen mit einem einzigen Server. Int. J. Prod. Res. 2014, 52, 7148–7158. [CrossRef]
18. Kosch, S.; Beck, JC Ein neues MIP-Modell für parallele Batch-Planung mit nicht identischen Auftragsgrößen. In Proceedings of the Integration of AI and OR Techniques in Constraint Programming: 11. Internationale Konferenz (CPAIOR 2014), Cork, Irland, 19.–23. Mai 2014; Proceedings 11; Springer: Berlin/Heidelberg, Deutschland, 2014; S. 55–70.
19. Sharma, A.; Zhao, F.; Sutherland, JW Ökonomische Planung eines produzierenden Unternehmens, das nach einem zeitabhängigen Stromtarif arbeitet. J. Clean. Prod. 2015, 108, 256–270. [CrossRef]
20. Bajany, D.; Xia, X.; Zhang, L. Ein MILP-Modell für die LKW-Schaufel-Planung zur Minimierung des Kraftstoffverbrauchs. Energy Procedia 2017, 105, 2739–2745. [CrossRef]
21. Che, A.; Zhang, S.; Wu, X. Energiebewusste, unabhängige parallele Maschinenplanung unter zeitabhängigen Stromtarifen. J. Clean. Prod. 2017, 156, 688–697. [CrossRef]
22. Cheng, J.; Chu, F.; Zhou, M. Ein verbessertes Modell für die parallele Maschinenplanung unter zeitabhängigen Strompreisen. IEEE Trans. Autom. Wissenschaft. Ing. 2017, 15, 896–899. [CrossRef]
23. Elidrissi, A.; Benmansour, R.; Benbrahim, M.; Duvivier, D. MIP-Formulierungen für identische parallele Maschinenplanungsprobleme mit einem einzigen Server. In Proceedings der 4. Internationalen Konferenz für Optimierung und Anwendungen (ICOA) 2018, Mohammedia, Marokko, 26.–27. April 2018; IEEE: Piscataway, NJ, USA, 2018; S. 1–6.
24. Kaabi, J.; Harrath, Y. Scheduling auf uniformen Parallelmaschinen mit periodischen Nichtverfügbarkeitsbeschränkungen. 57, 216–227. [CrossRef]
25. Quiroz Silva, EE Programmación por Objetivos para el Dimensionamiento y la Asignación de Una Flota de Camiones en una Empresa Minera. Schnittstellen 2021, 87–112. [CrossRef]
26. Hsu, CL; Liao, JR Zwei Parallelmaschinen-Planungsprobleme mit Funktionsbeschränkung. Diskret. Dyn. Nat. Soc. 2020, 2020, 1–6. [Querverweis]

Machine Translated by Google

27. Tapia, E.; Salazar Araya, A.; Saavedra, N.; Nehring, M.; Mora, J. Eine Analyse der Strategien „Voll-LKW“ im Vergleich zu „Voll-Schaufel“ bei Lade- und Transportvorgängen im Tagebau. Int. J. Min. Reclam. Environ. 2021, 35, 1–11. [CrossRef]
28. Tigreros Cardenas, ID Propuesta Para la Programación de Producción en Máquinas Paralelas con Tiempos de Alistamiento en una Empresa del Sector Agroindustrial. 2021. Online verfügbar: https://repositorio.uceva.edu.co/bitstream/handle/20.500.129 93/2744/T00031983.pdf?sequence=1&isAllowed=y (abgerufen am 29. Juli 2024).
29. Yunusoglu, P.; Topaloglu Yildiz, S. Constraint-Programmierungsansatz für das durch mehrere Ressourcen eingeschränkte, unabhängige parallele Maschinenplanungsproblem mit sequenzabhängigen Rüstzeiten. Int. J. Prod. Res. 2022, 60, 2212–2229. [CrossRef]
30. Ma, X.; Fu, Y.; Gao, K.; Zhu, L.; Sadollah, A. Ein multikriterielles Planungs- und Routingproblem für häusliche Gesundheitspflegedienste mittels Brainstorming-Optimierung. Komplexes Syst. Modell. Simul. 2023, 3, 32–46. [CrossRef]
31. Feng, L.; Chen, G.; Zhou, S.; Zhou, X.; Jin, M. Ein energieeffizientes, unabhängiges paralleles Maschinenplanungsproblem mit Batchverarbeitung und zeitabhängigen Strompreisen. Mathematics 2024, 12, 376. [CrossRef]
32. Hu, B.; Xiong, Z.; Sun, A.; Yuan, Y. Einsatzplanung von Containertransportfahrzeugen im Tagebau auf Basis des GA–GWO- Hybridalgorithmus. Appl. Sci. 2024, 14, 3986. [CrossRef]
33. Eivazy, H.; Askari-Nasab, H. Ein gemischt-ganzzahliges lineares Programmiermodell für die kurzfristige Produktionsplanung im Tagebau. Min. Technol. 2012, 121, 97–108. [CrossRef]
34. Anjomshoa, H.; Albrecht, A.; Lee, D.; Pudney, P. Effizienter Fahrzeugtransport im Untertagebergbau. Min. Technol. 2012, 121, 83–90. [Querverweis]
35. Gligoric, Z. Optimierung des Transportsystems in einem unterirdischen Kleinbergwerk unter Verwendung von Fuzzy-Mengen. Systems Design. S. 63. Online verfügbar: https://www.researchgate.net/publication/287397565_Optimization_of_haulage_system_in_an_underground_ Kleinbergbau mit Fuzzy-Sets (abgerufen am 29. Juli 2024).
36. Gonen, A.; Malli, T.; Kose, H. Auswahl eines Erztransportsystems für einen metallhaltigen Untertagebergbau/Dobór systemu transportu rud w kopalni podziemnej rud metalu. Bogen. Min. Wissenschaft. 2012, 57, 779–785. [CrossRef]
37. Eivazy, H.; Askari-Nasab, H. Ein hierarchisches Optimierungsmodell für die Produktionsplanung im Tagebau. Int. J. Min. Miner. Eng. 2012, 4, 89–115. [CrossRef]
38. Tom Socarras, JN Árbol de Decisión para la Gestión y Control de Costo de Transporte Interior Mina. Bachelorarbeit, Universität von Chile, Santiago, Chile, 2014.
39. Uribe Neira, PA, Analyse der Funktion und Ausführung der halbautomatischen Operation bei Frachtausrüstung; Acarreo y Descarga en Minería Subterránea Load Haul-Dump Vehicle (LHD). Bachelorarbeit, Universität Chile, Santiago, Chile, 2014.
40. Wi ÿecek, D.; Burduk, A.; Kuric, I. Der Einsatz neuronaler Netzwerke zur Verbesserung der Effizienz und Gewährleistung der Stabilität des Kupfererzbergbaus Prozess. Acta Montan. Slovaca 2019, 24, 1–14.
41. Wiest Goyeneche, JR Modelo de Programación de Operaciones en a Mina a Cielo Abierto: Aplicación en Organización Corona. Masterarbeit, Universidad de La Sabana, Chía, Kolumbien, 2019.
42. Solomon, L.; Ortiz, A. Konzeptionelles Modell von Erzlade- und Transportsystemen im Tagebau. Univ. Cienc. Y Tecnol. 2020, 24, 41–50.
43. Gaspar, S.; Jhasmin, C. Höhere Produktivität bei Ackerbau-, Mineraltransport- und Desmonte-Geräten in Veta Gavia–Nivel 100, Unidad Minera Huarón. Bachelorarbeit, Universidad Continental, Los Olivos, Peru, 2020.
44. Li, N.; Feng, S.; Ye, H.; Wang, Q.; Jia, M.; Wang, L.; Zhao, S.; Chen, D. Versandoptimierungsmodell für Transportausrüstung zwischen Abbaustellen basierend auf der kurzfristigen Ressourcenplanung der Mine. Metals 2021, 11, 1848. [CrossRef]
45. Elijah, K.; Joseph, G.; Samuel, M.; Mauti, D. Optimierung eines Schaufelrad-Transportsystems in einem Tagebau mithilfe eines Warteschlangenverfahrens. Arab. J. Geosci. 2021, 14, 973. [CrossRef]
46. Huisa Supho, A. Optimización de Equipos de Carguío y Acarreo en el Tramo Botaderos-Trituradora Thyssen Krupp, para el Incremento de Producción en una Empresa Cementera. Bachelorarbeit Universidad Tecnológica Del Perú, Lima, Peru, 2021.
47. Choi, Y.; Nguyen, H.; Bui, XN; Nguyen-Thoi, T. Optimierung der Leistung von Lastwagensystemen für die Erzproduktion im Tagebau unter Verwendung von Big Data und maschinell lernenden Methoden. Resour. Policy 2022, 75, 102522. [CrossRef]
48. Li, N.; Feng, S.; Lei, T.; Ye, H.; Wang, Q.; Wang, L.; Jia, M. Umplanungsplanoptimierung von unterirdischen Minentransportgeräten basierend auf der Simulation zufälliger Ausfälle. Sustainability 2022, 14, 3448. [CrossRef]
49. Abolghasemian, M.; Kanafi, AG; Daneshmand-Mehr, M. Simulationsbasierte multiobjektive Optimierung des Tagebau-Fördersystems : Eine modifizierte NBI-Methode und ein Metamodellierungsansatz. Complexity 2022, 2022, 3540736. [CrossRef]
50. Shamsi, M.; Pourrahimian, Y.; Rahmanpour, M. Optimierung der Produktionsplanung im Tagebau unter Berücksichtigung eines optimalen Transportsystems zwischen Lkw-Transport und halbmobiler Zerkleinerung und Förderung im Tagebau. Int. J. Min. Reclam. Environ. 2022, 36, 142–158. [CrossRef]
51. Parichehreh, M.; Gholizadeh, H.; Fathollahi-Fard, A.; Wong, K. Ein energieeffizientes, unabhängiges paralleles Maschinenplanungsproblem mit Lerneffekt der Bediener und Verschlechterung der Arbeitsplätze. Int. J. Environ. Sci. Technol. 2024, 1–26. [CrossRef]
52. Graham, RL; Lawler, EL; Lenstra, JK; Kan, AR Optimierung und Approximation bei deterministischer Sequenzierung und Planung: eine Übersicht. In Annals of Discrete Mathematics; Elsevier: Amsterdam, Niederlande, 1979; Band 5, S. 287–326.
53. Pinedo, M.; Hadavi, K. Scheduling: Theorie, Algorithmen und Systementwicklung. In Proceedings of the Operations Research Proceedings 1991: Beiträge der 20. Jahrestagung/Vorträge der 20. Jahrestagung; Springer: Berlin/Heidelberg, Deutschland, 1992; S. 35–42.
54. Yu, K.; Yang, J. MILP-Modell und ein Rolling-Horizon-Algorithmus für die Kranplanung in einem hybriden Lagercontainerterminal. Mathe. Probl. Ing. 2019, 2019, 4739376. [CrossRef]

Machine Translated by Google

55. Tarazona-Torres, LE; Amaya-Guio, CA; Álvarez-Martínez, D. Asynchrones Team für das Flow-Shop-Planungsproblem: Eine Fallstudie . In Proceedings der 4. südamerikanischen internationalen Konferenz für Wirtschaftsingenieurwesen und Betriebsmanagement, Lima, Peru, 9.–11. Mai 2023.
56. Gurobi Optimization, LLC. Gurobi Optimizer-Parameter Version 10.0.2; Gurobi Optimization, LLC.: Beaverton, OR, USA, 2024.
57. Cuellar-Usaquén, D.; Palacio, A.; Ospina, E.; Botero, M.; Álvarez-Martínez, D. Modellierung und Lösung des Endpunktschneideproblems . Int. Trans. Oper. Res. 2023, 30, 800–830. [CrossRef]
58. Hou, Y.; Wang, H.; Huang, X. Ein Q-Learning-basierter multi-objektiver evolutionärer Algorithmus für integrierte grüne Produktion und Probleme der Verteilungsplanung. Eng. Appl. Artif. Intell. 2024, 127, 107434. [CrossRef]

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultiere