HOCHSCHULE
ESSLINGEN

Informatik und
Informationstechnik

IT Innovationen

Band 33
Juni 2024

Grußwort der Fakultät

Liebe Leserinnen und Leser,

Der technische Fortschritt führt zu immer neuen Sensationen bei der KI-Anwendung und KI-generierten Inhalten. Allein in diesem Band der IT-Innovationen sind fast 30 Abschlussarbeiten zu finden, bei denen sich der KI-Bezug bereits aus dem Titel ergibt. KI ist hier. Heute lassen sich beliebige Text-, Audio- und Videoinhalte mit hoher Qualität verblüffend echt erzeugen. Hollywood-Special-Effects-Techniker sind schockiert und bangen um ihren Arbeitsplatz. Und auch im Privaten ist der Einsatz von ChatGPT, Midjourney, Dall-E, Stable Diffusion, Suno, Claude usw. inzwischen alles andere als ungewöhnlich.

So entstehen lustige Videos, Geburtstagsgrüße, peinliche Videos von Stars und Sternchen oder sogar Filmschnipsel mit kontroversen Aussagen von Politikern, die diese niemals getätigt haben. Fake News – ein Begriff, der noch gar nicht so lange existiert.

Gleichzeitig leben wir in einer Welt, in der vielen Menschen nicht mehr klar ist, worin der Unterschied zwischen Meinung, Behauptung und Tatsachen besteht. Sachverhalte werden aus dem Zusammenhang gerissen, „Fakten“ alternativ interpretiert oder gar falsche „Informationen“ geschaffen. Zusammengenommen führen diese beiden Entwicklungen zu einer sehr bedenklichen Entwicklung: Die Wahrheit wird subjektiv und aus einer glasklaren, nüchternen Betrachtung wird ein Gefeilsche um die Interpretation der Tatsachen.

Habe ich Ihnen nun den Tag verdorben? Warum schreibt dieser Dekan solche Downer in einem Vorwort? Was hat das überhaupt mit den IT-Innovationen zu tun? Die Antwort ist einfach: Die wissenschaftliche Methode ist in unserer heutigen Zeit das Bollwerk gegen subjektive Meinung und Behauptung. Worum geht es? Strukturiert und methodisch an eine Fragestellung herangehen. Thesen entwickeln. Wissen, was der Stand der Technik ist. Experimentell oder argumentativ zeigen oder gar mathematisch beweisen, dass die eigene Annahme richtig oder falsch ist. Der eigenen Arbeit bis zum Ende kritisch gegenüberstehen und für begründete Kritik offen sein. Das ist Wissenschaft.

Als Hochschule der angewandten Wissenschaften ist das unser Auftrag und unsere Berufung. Und weil wir dies tun, ist genau das auch ein entscheidender Schritt bei der Ausbildung zukünftiger Ingenieure und Wissenschaftler. Eine Abschlussarbeit umfasst all dies. Alle Absolventinnen und Absolventen, die in diesem Band ihre Arbeiten beschreiben, haben wissenschaftlich gearbeitet. Sie als Jungwissenschaftlerin oder Jungwissenschaftler kennen diese Methode! Wenden Sie diese auch weiterhin an, sie ist seit Jahrhunderten bewährt und hat uns aus der dunklen Zeit der Inquisition des Mittelalters, in der die Welt noch flach war und im Zentrum des Universums stand, in unsere heutige Fortschrittsgesellschaft gebracht! Gehen Sie danach vor! Im Beruf und im Privaten. Hinterfragen Sie Behauptungen! Argumentieren Sie klug. Zeigen Sie, dass Sie ein kritischer, aber diskursfähiger Akademiker oder eine Akademikerin sind. Sie können das. Das haben Sie bewiesen.

Viel Freude mit den vorliegenden Kurzbeschreibungen der Abschlussarbeiten des Abschlussjahrgangs des Sommersemesters 2024 wünscht Ihnen Ihr

Ihr Prof. Dr. Tobias Heer, Dekan

HOCHSCHULE ESSLINGEN

IMPRESSUM

ERSCHEINUNGSORT

73732 Esslingen am Neckar

HERAUSGEBER

Prof. Dr. Tobias Heer
Dekan der Fakultät Informatik und Informationstechnik
der Hochschule Esslingen - University of Applied Sciences

REDAKTIONSANSCHRIFT

Hochschule Esslingen - University of Applied Sciences
Fakultät Informatik und Informationstechnik
Flandernstraße 101
73732 Esslingen am Neckar

Telefon	+49(0)711.397-4210
Telefax	+49(0)711.397-4214
E-Mail	it@hs-esslingen.de
Website	www.hs-esslingen.de/it

REDAKTION, DESIGN, LAYOUT und SATZ

Dipl.-Inform.(FH) Rolf Gassner

Hochschule Esslingen - University of Applied Sciences
Fakultät Informatik und Informationstechnik
Flandernstraße 101
73732 Esslingen am Neckar

ERSCHEINUNGSWEISE

Einmal pro Semester, jeweils Januar und Juni

ISSN 1869-6457

Vergleich von Multicloud-Plattformen: Eine umfassende Analyse der führenden Multicloud-Plattformen hinsichtlich ihrer Funktionen, Leistung, Skalierbarkeit und Integrationsoptionen.

Nikita Adarycev **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Computacenter AG & Co. OHG, Stuttgart

Einleitung

Cloud Computing wird für Unternehmen weltweit immer wichtiger und umfasst Technologien und Geschäftsmodelle, die IT-Ressourcen dynamisch bereitstellen und ihre Nutzung flexibel abrechnen. Anstatt IT-Ressourcen wie Server oder Anwendungen in eigenen Rechenzentren zu betreiben, sind diese bedarfsorientiert über das Internet oder ein Intranet verfügbar. Unternehmen können durch Cloud Computing langfristige Investitionen (CAPEX) reduzieren, da hauptsächlich betriebliche Kosten (OPEX) für die genutzten IT-Ressourcen anfallen. Dies ermöglicht eine flexible und effiziente Nutzung von IT-Dienstleistungen. Cloud Computing bietet somit eine dienstleistungsbasierte Lösung für IT-Bedarf [2].

Cloud-Arten

Wie in Abbildung 1 zu sehen ist, gibt es verschiedene Arten von Clouds. Die **Private-Cloud** lässt sich wie ein Intranet beschreiben. Das bedeutet, dass Applikationen sowie Ressourcen sich in einem firmeneigenen Rechenzentrum befinden [6]. Im Gegensatz zur Private-Cloud, befinden sich bei der **Public-Cloud** die genutzten Daten und Dienste in der Obhut der Cloud-Anbieter [6]. Die Kombination der beiden erwähnten Clouds ist die **Hybrid-Cloud**. Diese Mischform wird häufig genutzt, da unkritische Daten, Applikationen oder IT-Ressourcen in die Public-Cloud ausgelagert werden können und gleichzeitig die geschäftskritischen Ressourcen weiterhin in der Private-Cloud betreut werden [6]. Jedoch wollen immer mehr Unternehmen ihre Cloud nicht nur von einem Cloud-Service-Provider (CSP) beziehen, sondern von verschiedene. Die Nutzung mehrerer Clouds von verschiedenen Providern wird als Multi-Cloud bezeichnet. Es gibt deutliche Vorteile im Vergleich zur Single-Cloud: Unternehmen nutzen verschiedene Innovationen sowie Technologien, die die verschiedenen CSP bereitstellen. Zudem können Unternehmen auf mögliche Angebotsänderungen der CPS flexibel und schnell reagieren. Finanziell spielt die freie Wahl mehrere CSP auch eine Rolle. Die Betriebskosten können deutlich sinken, da die Freiheit besteht, für bestimmte Aufgaben spezielle, kostengünstigere Angebote einzuholen, ohne das gesamte Angebot des CSP in Anspruch nehmen zu müssen. Somit können die unterschiedlichen Kostenmodelle der CSP aufgabengerecht genutzt werden. Die Nutzung verschiedener Anbieter kann ebenfalls höhere Verfügbarkeit oder auch schnellere Antwortzeiten ermöglichen. Bei Notfällen könnten die Unternehmen auch schnell auf andere CSP als Backup zurückgreifen und schlimmeres verhindern [4].

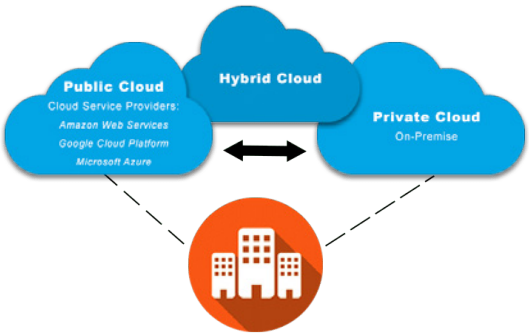

Abb. 1: Die verschiedenen Formen der Cloud. [3]

Herausforderungen

Es gibt Herausforderungen für die Kompatibilität beim Einsatz von verschiedene Cloud-Service-Providern. Die **Interoperabilität** bezieht sich auf die Fähigkeit verschiedener Cloud-Systeme, miteinander zu arbeiten. Technische Hindernisse können in Bereichen wie Programmiersprachen, APIs und Datenbanktechnologien auftreten. Unterschiedliche CSP bevorzugen oft verschiedene Programmiersprachen wie Java, Python oder Ruby, was die Integration erschwert. Auch die APIs der verschiedenen CSP können inkompatibel sein, was die Orchestrierung und Integration von Cloud-Diensten erschwert. Unterschiede in den verwendeten Datenbanktechnologien können dazu führen, dass Daten oder Anwendungen nicht ohne Weiteres zwischen Anbietern migriert werden können [4]. Beim Cloud-Brokerage müssen Daten zwischen verschiedenen Cloud-Diensten und -Anwendungen ausgetauscht werden, was **Sicherheits- und Datenschutzrisiken** birgt. Obwohl die Systeme grundsätzlich voneinander isoliert sind, nutzen sie die gleiche physische Hardware, was Sicherheitslücken begünstigen kann. Auch beim Datenübertragungsweg gibt es Risiken, da trotz verschlüsselter Übertragungswege theoretisch immer die Möglichkeit besteht, dass Sicherheitslücken ausgenutzt werden. Diese Herausforderungen erfordern strenge Sicherheitsmaßnahmen, um die Integrität und Vertraulichkeit der Daten zu gewährleisten [4]. Unterschiedliche Standards von Cloud-Service-Providern können die Einhaltung von Vorschriften erschweren und zusätzliche **Sicherheitsmaßnahmen oder Compliance-Prüfungen erfordern**, um sicherzustellen, dass alle Cloud-Dienste den Unternehmensrichtlinien entsprechen [4].

Führende Cloud-Provider und Multicloud-Plattformen

Die wohl prominentesten Beispiele für Cloud-Provider sind wie in Abbildung 2 zu sehen [5]:

- Microsoft Azure
- Salesforce.com App Cloud
- Google Cloud
- Oracle Cloud

Abb. 2: Top Cloud-Provider. [1]

Zum Bereich der Multicloud-Plattformen gehören unter anderem bspw.:

- Google Anthos
- Microsoft Azure Arc
- IBM Multicloud Manager
- Aviatrix

Im laufe der Arbeit werden einzelne Plattformen welche noch bestimmt werden müssen, verglichen.

Ziel der Arbeit

Die Hauptaufgabe besteht darin, verschiedene Multicloud-Provider wie zum Beispiel Aviatrix mit anderen Multicloud-Provider zu vergleichen und deren Funktionen, Leistung, Skalierbarkeit und Integrationsoptionen zu analysieren und anhand zu bestimmender Kriterien zu kategorisieren.

Literatur und Abbildungen

[1] Mahesh Chand. Top 10 Cloud Service Providers in 2024. https://www.c-sharpcorner.com/article/top-10-cloud-service-providers/, 2024.

[2] Christop Fehling and Frank Leymann. Cloud Computing. https://wirtschaftslexikon.gabler.de/definition/cloud-computing-53360/version-276453, 2018.

[3] Edward Jones. Arten von Cloud Computing – ein umfangreicher Leitfaden zu Cloud-Lösungen und -Technologien im Jahr 2024. https://kinsta.com/de/blog/arten-von-cloud-computing/, 2023.

[4] Claus-Peter Präg and Jochen Günther. Bedeutung und Management von Cloud Computing, Multi-Cloud und Cloud Brokerage in Unternehmen. *HMD Praxis der Wirtschaftsinformatik*, 2023.

[5] Stefan Reinheimer. *Cloud Computing: Infrastruktur der Digitalisierung*. Springer Vieweg, 2018.

[6] Jonas Repschläger, Danny Pannicke, and Rüdiger Zarnekow. Cloud Computing: Definitionen, Geschäftsmodelle und Entwicklungspotenziale. *HMD Praxis der Wirtschaftsinformatik*, 2014.

KI in der Hochschullandschaft

Nils Aichele **Catharina Kriegbaum-Kling**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Künstliche Intelligenz (KI) verändert die Hochschullandschaft immens. Insbesondere die Veröffentlichung des generativen Chatbots ChatGPT Ende 2022 trägt maßgeblich zum Wandel in der Hochschullandschaft bei, da Studierende und Dozenten nun rasch auf detaillierte Informationen zugreifen und Unterstützung in diversen Bereichen erhalten können. Dieser Paradigmenwechsel kann zu einer effizienteren Lernumgebung führen, da komplexe Fragen direkt beantwortet werden und Verständnisprobleme somit gelöst werden können. Des Weiteren können Lehrende KI einsetzen, um die Lernprozesse der Studierenden zu analysieren und didaktisch zu verbessern. Derzeit konzentrieren sich die Anwendungen auf lokale didaktische Maßnahmen, die Lehrende bei Routineaufgaben wie beispielsweise der Erstellung von Lehrmaterialien oder der Prüfungsbewertung entlasten. [4] Gleichzeitig stellt dies die Hochschulen vor neue Herausforderungen hinsichtlich der Integrität und Eigenständigkeit von studentischen Arbeiten.

Zielsetzung

Ziel dieser Arbeit ist es festzustellen, wie KI die Hochschullandschaft verändert. Dies betrifft bestehende KI-Systeme, die an Hochschulen zum Einsatz kommen, sowie die Verbreitung von KI-Tools unter Studierenden und Lehrenden. Daraus ergibt sich die Forschungsfrage: Wie verändert KI die Hochschullandschaft? Dabei wird betrachtet, wie KI in die Hochschullandschaft einfließt und wie sie idealerweise einfließen sollte. Außerdem soll festgestellt werden, wie verbreitet der Einsatz von KI-Tools unter Studierenden und Lehrenden ist und wie diese Tools von den Nutzern eingeschätzt werden. Diesbezüglich werden im empirischen Teil der Bachelorthesis zwei repräsentative Umfragen durchgeführt, eine für Studierende und eine für Lehrende von Hochschulen sowie Universitäten. Diese quantitativen Umfragen zielen darauf ab, die Nutzung von KI-Tools sowie die Einschätzungen zu diesen in der Hochschullandschaft repräsentativ zu erfassen.

Generative KI

Um einen Überblick über die Thematik zu bekommen, wird kurz auf die Grundlagen der Generativen KI eingegangen. Wie in Abbildung 1 zu erkennen, ist Generative KI ein Teilbereich von Deep Learning und verwendet leistungsstarke neuronale Netzwerke um neue Inhalte zu erstellen. Generative KI ist in der Lage neue Bilder, neue Texte in natürlicher Sprache, Musik oder Videos zu erstellen. Dabei werden generative KI-Anwendungen extra auf große Datenmengen trainiert um Muster der Daten zu übernehmen. Chatbots wie ChatGPT basieren auf dieser Technologie. [1]

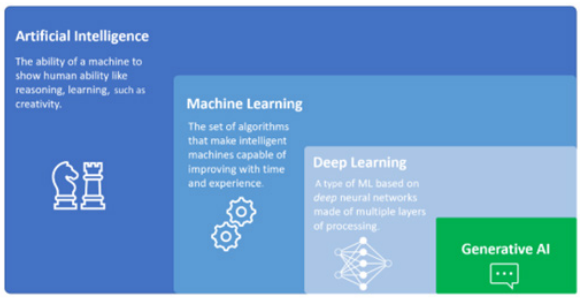

Abb. 1: Einordnung von Generativer KI [1]

KI in der Hochschullehre: Learning Analytics

Ein Trend der von dem EDUCAUSE Horizon Report 2022 präsentiert wurde ist die Kombination von KI und Learning Analytics (LA). [5] [3] Dabei werden Daten wie die Studienleistung von Studierenden, soziodemografische Daten und andere relevante Daten von Studierenden, die zum größten Teil aus Learning Management Systemen (LMS) wie beispielsweise dem LMS Moodle kommen, verwendet, mit dem Ziel, die Lehr- und Lernprozesse an Hochschulen zu modellieren und zu verbessern. Dabei kommen Algorithmen zum Einsatz, mit denen man beispielsweise berechnen kann, wie wahrscheinlich es ist, dass Studierende einen Kurs erfolgreich abschließen, um dadurch ge-

fährdete Studierende erkennen zu können. Darüber hinaus erhalten die Studierenden ein personalisiertes Feedback zu ihrem Lernfortschritt sowie individuelle Unterstützungsempfehlungen. [6]

Umfrage über die Nutzung von KI-Tools

Da für die quantitative Umfrage gezielt nach Lehrenden und Studierenden gesucht wurde, kamen unterschiedliche Vorgehen zur Teilnahmerekrutierung zum Einsatz. Zunächst wurden gezielt, persönliche E-Mails an Dozenten und Lehrende von Hochschulen sowie Universitäten im Raum Deutschland versendet, mit der Bitte, die Umfrage auszufüllen. Der Vorgang der Teilnahmerekrutierung bei der Umfrage für die Studierenden, wurde zum einen durch die Hilfe von Kommilitonen und zum anderen durch die Bewerbung in studentischen Online-Foren umgesetzt. Hier war es ebenfalls wichtig, Studierende aus ganz Deutschland zu erreichen, um eine repräsentative Darstellung der Ergebnisse zu gewährleisten. Wie in Abbildung 2 ersichtlich ist, haben bei den Studierenden 77 Personen und bei den Lehrenden 61 Personen teilgenommen. Auf die Frage, ob KI-Tools im Studium benutzt werden antworteten 92,2% der Studierenden mit ‚Ja'. Nur 7,8% gaben an, keine KI-Tools für studentische Tätigkeiten zu verwenden. Bei den Lehrenden waren die Antworten auf diese Frage noch deutlich ausgeglichener. Hier gaben 44,3% der Befragten an, bereits KI-Tools für die Lehrgestaltung zu verwenden, während hingegen 42,3% keine KI-Tools benutzen. Hinzu kommen die 13,1%, die in Zukunft KI in der Lehre einfließen lassen wollen.

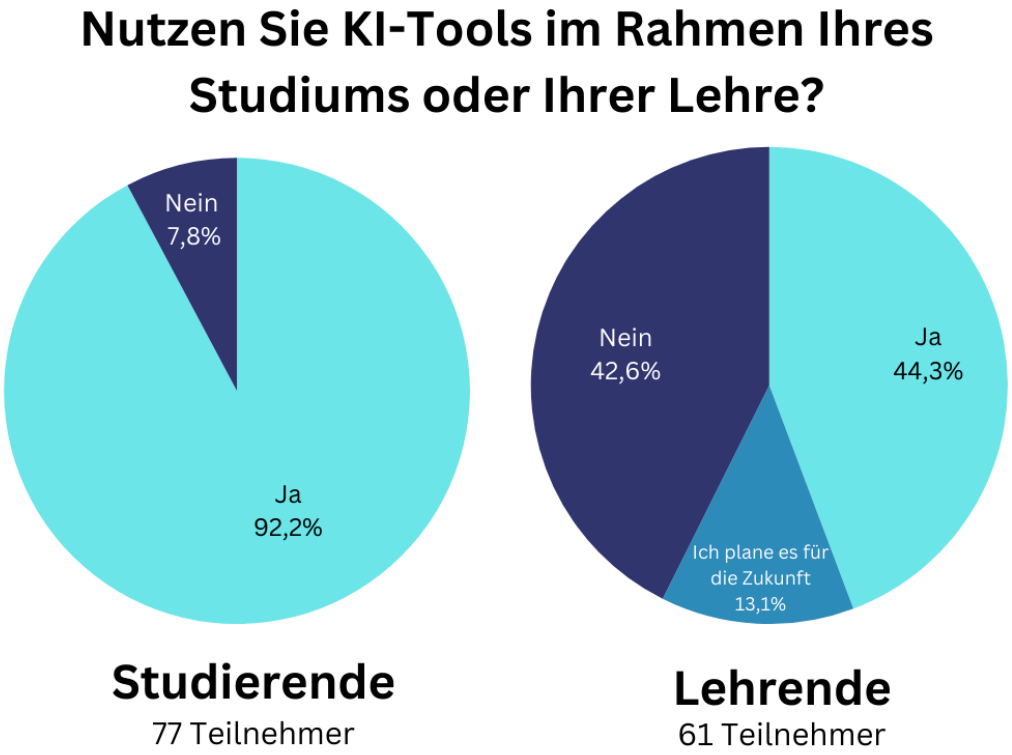

Abb. 2: Nutzung von KI Tools durch Studierende und Lehrende [2]

Ausblick

Durch die Umfrage wurden interessante Einblicke gewonnen, durch die klar wird, dass KI-Tools bereits weit verbreitet in der Hochschullandschaft genutzt werden. Diesbezüglich wird davon ausgegangen, dass die Integration von KI in die Hochschullehre weiterhin schnell voranschreiten und diese zunehmend verändern wird. Ebenfalls wird prognostiziert, dass die Analyse von Lernprozessen durch KI weiter verbessert wird, was zu genaueren Vorhersagen und gezielteren Unterstützungsmaßnahmen für Studierende führen könnte. Gleichzeitig müssen Hochschulen Lösungen finden, um ethische Fragen und Herausforderungen bezüglich der akademischen Integrität zu adressieren. Es wird davon ausgegangen, dass die Entwicklung von KI in der Hochschullehre die Effizienz und Qualität des Bildungswesens nachhaltig steigern wird.

Literatur und Abbildungen

[1] Valentina Alto. *Modern Generative AI with ChatGPT and OpenAI Models*. Packt, 2023.

[2] Eigene Darstellung.

[3] EDUCAUSE Horizon Report. 2022 EDUCAUSE Horizon Report | Teaching and Learning Edition. https://library.educause.edu/resources/2022/4/2022-educause-horizon-report-teaching-and-learning-edition, 04 2022.

[4] Tobias Schmohl and Dennis Schäffer. *Lehrexperimente der Hochschulbildung. Didaktische Innovationen aus den Fachdisziplinen*. wbv media, 2 edition, 2019.

[5] Tobias Schmohl, Alice Watanabe, and Kathrin Schelling. *Künstliche Intelligenz in der Hochschulbildung*, volume 4. Hochschulbildung: Lehre und Forschung, 2023.

[6] Volker Wittpahl. *Künstliche Intelligenz: Technologie | Anwendung | Gesellschaft*. Springer, 2019.

Standards in Business Analytics: Referenzworkflows und Standardprozesse zur Entscheidungsunterstützung

Bengue Akdemir **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Kleine und mittelständische Unternehmen (KMU) suchen nach effektiven Wegen, Business Intelligence (BI) zu integrieren und zu nutzen, stehen jedoch vor der Herausforderung, mit begrenzten Ressourcen und kleineren Datensätzen effiziente BI-Prozesse zu etablieren. Der Schlüssel zum Erfolg liegt in der schnellen Implementierung von standardisierten Prozessen, die sofortige Vorteile, sogenannte "Quick Wins", bieten können. Diese könnten durch die Nutzung vorhandener Infrastrukturen, wie Data Warehouses, unterstützt werden. Entscheidend ist dabei, welche BI-Standardprozesse zuerst umgesetzt werden sollten, wie vorhandene Ressourcen optimal genutzt werden können, welche Rolle die Analyse und Auswahl von Unternehmensdaten in der initialen Phase spielt und wie Handlungsempfehlungen sowie "Best Practices" aus der Literatur KMU unterstützen können. Auf Basis dieser Fragestellungen zielt die Arbeit darauf ab, einen Referenzworkflow für den Einstieg in BI speziell für KMU zu entwickeln. Dies umfasst die systematische Analyse und Priorisierung von BI-Projekten in verschiedenen Unternehmensbereichen wie CRM und Produktionslogistik.
Durch eine fundierte Literaturrecherche und die Anwendung bestehender Industriestandards sollen praxisorientierte Handlungsempfehlungen formuliert werden. Diese Empfehlungen werden dann am Beispiel eines Microsoft Data Warehouses illustriert und bieten praktische Leitlinien für den effektiven Einsatz von BI-Tools in KMU, um sowohl theoretische als auch praktische Aspekte der BI-Implementierung zu erforschen und spezifische Lösungsansätze zu entwickeln.

Data Warehouse

Data Warehouse (DWH) ist ein von operativen Daten getrenntes, zentralisiertes Datenhaltungssystem, das als konsistente Datenbasis für Entscheidungsunterstützungssysteme dient. Wesentliche Merkmale eines DWH, geprägt durch William H. Inmon, umfassen Themenorientierung, Integration, Zeitraumbezug und Nicht-Volatilität. Es wird betont, dass ein DWH entscheidungsrelevante Informationen bereitstellt und eine zentrale Rolle im Bereich BI spielt [2].
Die Architektur eines DWH wird in drei Schichten unterteilt: die Datenschicht, die Bereitstellungsschicht und die Dialog- und Analyseschicht. Die Datenschicht umfasst die originären Datenquellen. Die Bereitstellungsschicht beinhaltet das Core DWH und die Data Marts, und die oberste Schicht bildet die Schnittstelle zu den Endnutzern, einschließlich Berichts- und Analysesystemen. Die nachstehende Abbildung 1 veranschaulicht ein solches Modell. Data Marts bieten dabei datenspezifische Teilmengen, die auf die Bedürfnisse einzelner Abteilungen zugeschnitten sind, und ermöglichen so schnellere Zugriffe und eine vereinfachte Wartung [6].

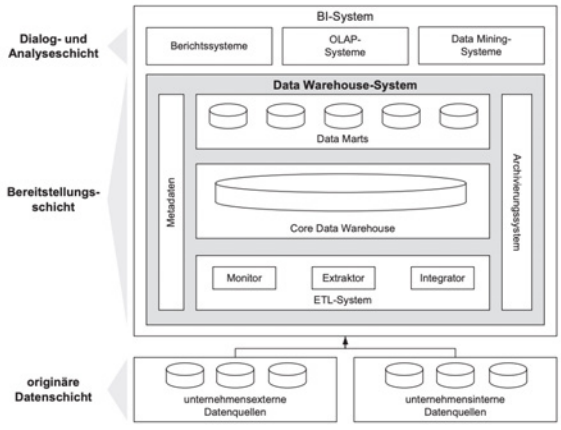

Abb. 1: DWH-Architektur [1]

Business Analytics

Business Analytics befasst sich mit der Anwendung analytischer Methoden und Technologien zur Optimierung geschäftlicher Entscheidungsprozesse. In diesem Kon-

text werden Daten aus diversen Quellen systematisch erfasst, analysiert und in wertvolle Geschäftseinsichten umgewandelt. Die Bedeutung von Business Analytics erstreckt sich über die strategische Planung bis hin zur operativen Steuerung, indem es Unternehmen ermöglicht, auf Basis fundierter Datenanalysen Entscheidungen zu treffen. Dabei kommen verschiedene Techniken und Werkzeuge zum Einsatz, die darauf abzielen, Muster und Trends in großen Datenmengen zu erkennen und nutzbar zu machen. Dies unterstützt Organisationen nicht nur bei der Identifikation von Effizienzsteigerungspotenzialen, sondern auch bei der proaktiven Gestaltung ihrer Geschäftsstrategien [7]. Innerhalb von Business Analytics gibt es vier Hauptarten der Datenanalyse:

- Descriptive Analytics
- Diagnostic Analytics
- Predictive Analytics
- Prescriptive Analytics

Einführung in Standardprozesse

Standardprozesse in Unternehmen spielen eine entscheidende Rolle bei der Optimierung von Arbeitsabläufen und der Steigerung der betrieblichen Effizienz. Diese Prozesse, die sich durch regelmäßige und normgerechte Abläufe auszeichnen, tragen wesentlich zur Vereinheitlichung und Vereinfachung der Geschäftstätigkeiten bei. In Kombination mit BI-Systemen ermöglichen Standardprozesse eine datengestützte Entscheidungsfindung. Durch die systematische Analyse der aus diesen Prozessen gewonnenen Daten können Unternehmen wertvolle Einblicke erlangen, die zur weiteren Optimierung und feineren Anpassung der Prozesse beitragen. Dies führt nicht nur zu einer Reduktion der Kosten, sondern verbessert auch die Gesamtperformance des Unternehmens [3].

Bedeutung von Referenzworkflows

In der Diskussion über Referenzworkflows in BI-Systemen wird hervorgehoben, wie entscheidend diese für die Standardisierung und Optimierung von BI-Prozessen sind. Referenzworkflows stellen bewährte Vorgehensweisen dar, die Organisationen dabei unterstützen, ihre Datenanalyse- und Entscheidungsfindungsprozesse effizient zu gestalten. Sie dienen als vorgefertigte Ablaufpläne, die aufzeigen, wie Daten erfasst, analysiert und interpretiert werden sollten, um konsistente und zuverlässige Ergebnisse zu erzielen.
Diese Workflows sind besonders wertvoll, weil sie eine schnelle und fehlerfreie Implementierung von BI-Systemen ermöglichen, indem sie klare Richtlinien und Schritte bieten. Unternehmen können durch ihre Anwendung sicherstellen, dass ihre BI-Prozesse mit den strategischen Zielen des Unternehmens übereinstimmen und dass die Datenanalyse auf eine Art und Weise durchgeführt wird, die zur Erzielung von maximalen Erkenntnissen und zur Förderung datengestützter Entscheidungsfindung beiträgt. Die Verwendung von Referenzworkflows ermöglicht es auch, die Einarbeitungszeit neuer Mitarbeiter zu verkürzen und die Konsistenz über verschiedene Teams und Abteilungen hinweg zu gewährleisten.
Durch den Einsatz solcher standardisierter Abläufe können Unternehmen nicht nur die Effizienz ihrer BI-Prozesse verbessern, sondern auch ihre Ressourcen optimal nutzen und die Return-on-Investment (ROI) ihrer BI-Investitionen maximieren. Darüber hinaus helfen Referenzworkflows dabei, die Adaptionsrate von BI-Tools innerhalb der Organisation zu erhöhen, indem sie die Benutzerfreundlichkeit und Zugänglichkeit der BI-Systeme verbessern [5].

Optimale BI-Standardprozesse für kleine Unternehmen zur Förderung von kurzfristigen und langfristigen Erfolgen

Die Bedeutung von BI-Standardprozessen für KMU wird umfassend beleuchtet, wobei der Fokus auf den strategischen und operativen Vorteilen liegt, die solche Prozesse diesen Unternehmen bieten können. Besonders hervorgehoben wird, wie maßgeschneiderte BI-Prozesse KMU in die Lage versetzen, ihre Entscheidungsfindung zu verbessern und ihre Wettbewerbsfähigkeit zu steigern.

Abb. 3–7 *Reifegradmodell für die Organisation von BI (vgl. [Dittmar et al. 2013, S. 26])*

Abb. 2: Reifegradmodell [4]

Ein wichtiger Bestandteil dieser Diskussion ist das schrittweise Vorgehen bei der Implementierung von BI, das KMU ermöglicht, sich allmählich mit BI-Technologien vertraut zu machen und deren Anwendung zu optimieren. In diesem Zusammenhang wird ein Reifegradmodell hervorgehoben, das in Abbildung

2 visualisiert ist. Dieses Modell illustriert die verschiedenen Entwicklungsstufen von BI-Kapazitäten in Unternehmen und bietet einen systematischen Rahmen, mit dem KMU ihre BI-Initiativen entwickeln und verfeinern können.
Das Reifegradmodell zeigt auf, wie Unternehmen durch verschiedene Phasen der BI-Nutzung fortschreiten können, von initialen Implementierungen bis hin zu fortgeschrittenen Anwendungen, die tiefergehende Analysen ermöglichen und umfassendere datengesteuerte Entscheidungsprozesse unterstützen. Durch die Einführung von BI-Standardprozessen und das Fortschreiten in den Stufen des Reifegradmodells können KMU nicht nur ihre aktuellen Geschäftsprozesse optimieren, sondern auch ein fundiertes Verständnis für strategische Entscheidungen entwickeln, die ihr langfristiges Wachstum und ihre Innovationsfähigkeit fördern.
Dieses Vorgehen verdeutlicht, dass die gezielte und stufenweise Einführung von BI-Technologien entscheidend für die Effizienzsteigerung und die strategische Ausrichtung von KMU ist. Durch die Anpassung der BI-Prozesse an die spezifischen Bedürfnisse und Kapazitäten kleinerer Unternehmen wird nicht nur die tägliche Datenverarbeitung verbessert, sondern auch eine nachhaltige strategische Entwicklung unterstützt [8].

Ausblick

Der Ausblick dieser Arbeit betont die zunehmende Bedeutung von fortgeschrittenen analytischen Techniken im Rahmen von Business Analytics. Die Entwicklung von branchenspezifischen Standards und individuell angepassten Modellen wird eine Schlüsselrolle spielen, um die Effizienz und Effektivität der Datenanalyse in Unternehmen zu steigern.
Die Implementierung interoperabler Frameworks, die eine umfassende Datenanalyse über verschiedene Plattformen hinweg ermöglichen, wird ebenfalls entscheidend sein. Dies wird nicht nur die Analyseprozesse verbessern, sondern auch helfen, Datensilos abzubauen. Ebenfalls wichtig wird die Integration ethischer Überlegungen und Datenschutzbestimmungen sein, um den Schutz personenbezogener Daten zu gewährleisten und das Vertrauen in Business Analytics-Systeme zu stärken.
Insgesamt eröffnen sich durch die Standardisierung und die Einführung von Referenzworkflows in Business Analytics neue Möglichkeiten für Unternehmen, ihre Entscheidungsprozesse zu optimieren und eine Kultur der datengestützten Entscheidungsfindung zu fördern.

Literatur und Abbildungen

[1] Paul Alpar, Rainer Alt, Frank Bensberg, and Christian Czarnecki. *Anwendungsorientierte Wirtschaftsinformatik*. Springer Vieweg, 2023.

[2] Henning Baars and Hans-Georg Kemper. *Business Intelligence & Analytics – Grundlagen und praktische Anwendungen*. Springer Vieweg, 2021.

[3] Thomas Bergs, Christian Brecher, Robert H. Schmitt, and Günther Schuh. *Internet of Production - Turning Data into Value: Statusberichte aus der Produktionstechnik 2020*. Fraunhofer-Institut für Produktionstechnologie IPT, 2020.

[4] Tom Gansor and Andreas Totok. *Von der Strategie zum Business Intelligence Competency Center (BICC)*. dpunkt.verlag, 2015.

[5] Christoph Mathas. Entwicklung einer Webanwendung mit expliziter Workflow-Steuerung und integrierten BI-Systemausgaben. https://www.degruyter.com/document/doi/10.1515/pik-2015-0020/html?casa_token=wM4Wppp2baoAAAAA:37k_BoA6k6L6xuqe9UOHarzEWIL5ltS5K0zJ6QwTb1lSf490XifTRYRDL 11 2016.

[6] Pascal Schmidt-Volkmar. *Betriebswirtschaftliche Analyse auf operationalen Daten*. Gabler Verlag / GWV Fachverlage GmbH, 2008.

[7] Mischa Seiter. *Business Analytics: Effektive Nutzung fortschrittlicher Algorithmen in der Unternehmenssteuerung*. Verlag Franz Vahlen, 2017.

[8] Engelbert Westkämper and Carina Löffler. *Visionen und strategische Konzepte für das System Produktion*. Springer Vieweg, 2016.

Evaluierung eines Eventsensors Zur Objekterkennung

Ibrahim Al Askar **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Leuze electronic Deutschland GmbH + Co KG, Owen

Einleitung

In den letzten Jahrzehnten hat sich die industrielle Bildverarbeitung zu einer wichtigen Methode für die bildbasierte automatische Inspektion, Prozesssteuerung und Roboterführung in der Industrie entwickelt. In auf maschineller Bildverarbeitung basierenden Systemen ist die Objektklassifizierung ein kritischer Bereich, der die Gesamtleistung des Systems maßgeblich bestimmt. Der in einem Bildverarbeitungssystem eingesetzte Klassifikator sollte eine schnellere Entscheidungsfindung mit erhöhter Klassifizierungsgenauigkeit ermöglichen. Bildverarbeitungsanwendungen verwenden verschiedene Arten von Klassifikatoren wie regelbasierte Klassifikatoren, künstliche neuronale Netze, Naive Bayes-Klassifikatoren und SVM-Klassifikatoren „Support-Vektor-Netzwerke“. Empirische Evidenz in den Bereichen Textcharakterisierung, Gesichtserkennung und Zeichenerkennung zeigt, dass die SVM eine vergleichsweise bessere Datengeneralisierungsleistung bietet.

Events Kamera

Eventkameras sind eine spannende neue Technologie in der Bildverarbeitung. Im Gegensatz zu normalen Kameras, die in regelmäßigen Abständen vollständige Bilder aufnehmen, erfassen Eventkameras nur die Änderungen in der Lichtintensität bei jedem Pixel. Sie reagieren also auf Bewegung und Veränderung mit einem ununterbrochenen Strom von Ereignissen. Diese Idee kommt von der Funktionsweise des menschlichen Auges. Um unsere Umgebung zu sehen, wird Licht von Objekten reflektiert und in unser Auge geleitet. Dieses Licht passiert zuerst die äußeren Schichten des Auges und trifft dann auf die Netzhaut. Dort gibt es spezielle Zellen, die Helligkeit, Schärfe und Farbe erkennen. Die Informationen von diesen Zellen werden über den Sehnerv an das Gehirn gesendet, das dann das Bild zusammensetzt. Ein Event-basierter Vision Sensor (EVS) funktioniert ähnlich. Der Sensor nimmt das einfallende Licht auf und wandelt es in ein elektrisches Signal um. Dieses Signal wird ständig mit einer Referenzspannung verglichen. Wenn die Spannung eine bestimmte Grenze überschreitet, wird dies als Ereignis registriert. So erkennt der Sensor sofort jede Veränderung im Licht und gibt diese als Daten weiter. [4]

Problemstellung:

Die Herausforderung besteht darin, rechteckige Objekte zu erkennen und die Position ihrer vorderen und hinteren Kanten zu bestimmen. Zur Lösung dieses Problems werden verschiedene Arten von Algorithmen experimentell getestet, um eine geeignete Lösung zu finden. Hier werden ein paar Algorithmen dargestellt:

DBSCAN Algorithmus

Der DBSCAN-Algorithmus (Density-Based Spatial Clustering of Applications with Noise) ist ein Clustering-Algorithmus, der verwendet wird, um Gruppen (Cluster) von Punkten in einem Datensatz zu identifizieren und gleichzeitig Ausreißer (Rauschen) zu erkennen. Hier ist eine Erklärung des DBSCAN-Algorithmus:

1. **Kernpunkte**: Der Algorithmus sucht nach sogenannten Kernpunkten. Ein Punkt ist ein Kernpunkt, wenn in seinem Umkreis (definiert durch einen Radius) mindestens eine bestimmte Anzahl von Punkten liegt.

2. **Erweiterung von Clustern**: Wenn ein Kernpunkt gefunden wird, bildet er zusammen mit seinen Nachbarn ein Cluster. Der Algorithmus erweitert dieses Cluster, indem er iterativ die Nachbarn der Nachbarn einbezieht, solange sie Kernpunkte sind.

3. **Rand- und Rauschen-Punkte**: Punkte, die keine Kernpunkte sind, aber in der Nähe eines Clusters liegen, werden als Randpunkte betrachtet und zum nächsten Cluster hinzugefügt. Punkte, die weder Kernpunkte noch Randpunkte sind, werden als Rauschen (Outliers) betrachtet.

DBSCAN ist besonders nützlich, weil es keine Vorabkenntnisse über die Anzahl der Cluster erfordert und Cluster beliebiger Form entdecken kann. [3]

Hough Transform

Die Hough-Transformation ist ein Algorithmus, der verwendet werden kann, um Merkmale einer bestimmten Form innerhalb eines Bildes zu finden. Da die gewünschten Merkmale in einer parametrischen Form angegeben werden müssen, wird die klassische Hough-Transformation am häufigsten für die Erkennung regelmäßiger Kurven wie Linien, Kreise, Ellipsen usw. verwendet. Die Hough-Transformation kann verwendet werden, wenn ein bestimmtes Merkmal (wie eine Form oder ein Objekt) in einem Bild erkannt werden soll, aber dieses Merkmal nicht einfach mit einer mathematischen Formel beschrieben werden kann.
Mit Hilfe der folgenden Abbildung wird verständlich erklärt, wie man durch den Hough Transformations-Algorithmus Linien erkennen kann. Im Bildraum befinden sich vier Punkte. Das Ziel des Hough-Transformations-Algorithmus ist es, die Geradengleichung zu finden, die durch die vier Punkte verläuft. Das funktioniert, indem jeder Punkt im Bildraum eine Gerade im Parameterraum repräsentiert. Zum Beispiel repräsentiert der Punkt P(2,1) im Parameterraum die Gerade 1. Der Schnittpunkt der Geraden im Parameterraum, zum Beispiel hier der Punkt S(-1.5, -2), entspricht im Bildraum der Gerade mit der Gleichung 2, wie in der grünen gestrichelten Linie im Bildraum mit 3. Diese Gerade verläuft durch alle Punkte. [2]

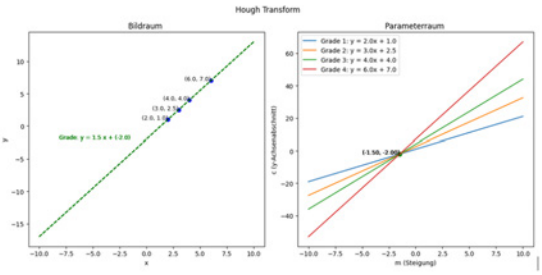

Abb. 1: Hough Transformations Beispiel zur Liniendetektion [1]

$$y = 2x + 1 \tag{1}$$

$$y = (-xs)x + (ys) \tag{2}$$

$$y = 1.5x - 2 \tag{3}$$

Ausblick:

Die Zukunft der Maschine Vision, insbesondere durch die Einführung von Eventkameras, verspricht eine innovative Epoche. Diese Technologie, die darauf spezialisiert ist, Veränderungen in den Lichtverhältnissen zu erkennen und in verwertbare Daten umzuwandeln, könnte eine Revolution in Bereichen wie autonomes Fahren, Robotik und interaktive Medien auslösen. Weiterführende Forschung und Entwicklung könnten darauf abzielen, die Anpassungsfähigkeit und Lernprozesse dieser Systeme zu verfeinern, sodass sie noch präziser auf ihre Umgebung reagieren und komplexe Muster in Echtzeit interpretieren können.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] SIMON PERKINS, ROBERT FISHER, ASHLEY WALKER, and ERIK WOLFART. Hough Transform. https://homepages.inf.ed.ac.uk/rbf/HIPR2/hough.htm, 2000.

[3] ADRIANO R. Fokus auf den DBSCAN Algorithmus. https://datascientest.com/de/machine-learning-clustering, 2023.

[4] Sony Semiconductor Solutions Group. Event-based Vision Sensor EVS Technology. https://www.sony-semicon.com/en/technology/industry/evs.html, 2023.

Portierung einer Legacy PHP Anwendung zur Erfassung behördlicher Bestellvorgängen auf eine effiziente REST-API basierte Fullstack Anwendung

Heba Aladawi **Mirko Sonntag**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einführung

Die Hochschule Esslingen setzt ein internes Bestellsystem ein, um die Bestellungen der IT-Fakultät zu verwalten und automatisch Bestellformulare zu generieren. Die zugrunde liegenden Technologien dieses Systems wurden vor über einem Jahrzehnt veröffentlicht und stehen nun kurz vor dem Ende ihres Lebenszyklus. Eine Aktualisierung auf neuere Versionen würde eine umfassende Neuentwicklung des Systems erfordern, weshalb beschlossen wurde, das System unabhängig von den veralteten Technologien neu zu entwickeln. Vor dieser Bachelorarbeit gab es bereits eine Vorgängerarbeit (Studienprojekt [3]) von Herrn Rico Hofmann, in der er das Technologiekonzept der neuen Software definierte und die Ansichten für Personen, Adressen und Lieferanten entwickelte. Im Rahmen dieser Bachelorarbeit müssen unter anderem die Bestellungsansicht, einschließlich das Anlegen und Editieren von Bestellungen sowie die Generierung von Bestellformularen, implementiert werden. Das Ziel ist, den gesamten Bestellprozess bis zur Generierung von Bestellformularen zu optimieren und langfristige, effiziente Lösungen zu gewährleisten.

Anforderungen

Zu Beginn der Arbeit wurden verschiedene Anforderungen identifiziert und gesammelt, die im Verlauf der Bachelorarbeit schrittweise umgesetzt werden. Diese Anforderungen umfassen:

- Die Einbindung aller bestehenden Datenbanktabellen im Front- und Backend.
- Die Implementierung eines PDF-Exports für Bestellformulare.
- Die Bereitstellung eines Standardtext-Generators für Mail-Korrespondenz. (Optional)
- Die Einrichtung einer Dateiablage für PDF-Rechnungen und Angebote sowie die Gestaltung einer ansprechenden Startseite und Icons für eine verbesserte Benutzerinteraktion. (Optional)
- Die Integration vom SSO-Shibbloleth Login. (Optional)

Bestellprozess

Der Bestellvorgang beinhaltet die Erfassung verschiedener Daten und Einträge aus den verschiedenen Tabellen der Datenbank, darunter Kostenstellen, Personen, Lieferanten und Adressen, sowie das Anlegen von Bestellpositionen, die einer Bestellung zugeordnet werden müssen. In einem übersichtlichen Eingabeformular werden mehrere Abschnitte aufgeführt, um diese unterschiedlichen Daten mithilfe von Auswahlfeldern oder Eingabefeldern zu sammeln. Die Daten müssen konsistent bleiben und stets in der Datenbank- die unverändert vom Legacy PHP System übernommen wurde- gespeichert werden. Der Nutzer navigiert im Bestellsystem zur Bestellansicht und klickt auf die Option "Bestellung anlegen". Dort kann der Nutzer die erforderlichen Daten in den verschiedenen benötigten Eingabefelder eingeben und bei Bedarf neue Informationen über Dialogfelder hinzufügen, um sie später in der Bestellung zu verwenden. Sobald alle relativen Eingabefelder vollständig und korrekt ausgefüllt sind, kann die Bestellung gespeichert werden. Die Herausforderung besteht darin, die Bestellungen und Bestellpositionen gleichzeitig zu speichern, da die Bestellpositionen erst dann in der Datenbank gespeichert werden können, wenn sie einer Bestellung zugeordnet sind (die order_id dient als Fremdschlüssel für die Bestellpositionen-Datensätze). Am Ende kann mit einem Klick das Bestellformular generiert werden. Bestellungen lassen sich zudem nachträglich bearbeitet werden, während die Verwaltung der Bestellungen weiterhin effizient bleibt.

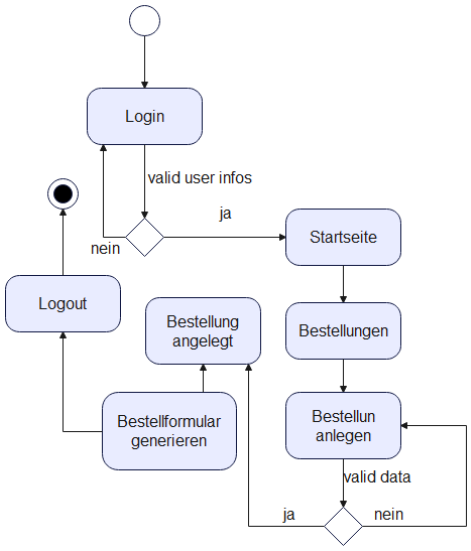

Abb. 1: Aktivitäts-Diagramm eines Bestellprozesses [2]

Architektur

Im Rahmen dieses Projekts kommt im Frontend Angular zum Einsatz, unterstützt durch die Bibliothek „Angular Material UI Components“ für die einfache Erstellung von Material Design Komponenten. Das Backend basiert auf NestJS in Kombination mit TypeORM als Object-Relational Mapping (ORM) Werkzeug. Die Kommunikation zwischen Frontend und Backend erfolgt über HTTP (siehe Abbildung 2), während zur persistenten Datenspeicherung eine MySQL-Datenbank verwendet wird, die auf den bestehenden Schemata der vorherigen Anwendung basiert. Zudem wird TypeScript sowohl im Frontend als auch im Backend verwendet, um eine konsistente und typsichere Codebasis zu gewährleisten. Das Technologienkonzept stammt von Herrn Rico Hofmann und kommt unverändert in dieser Arbeit zum Einsatz. [3]

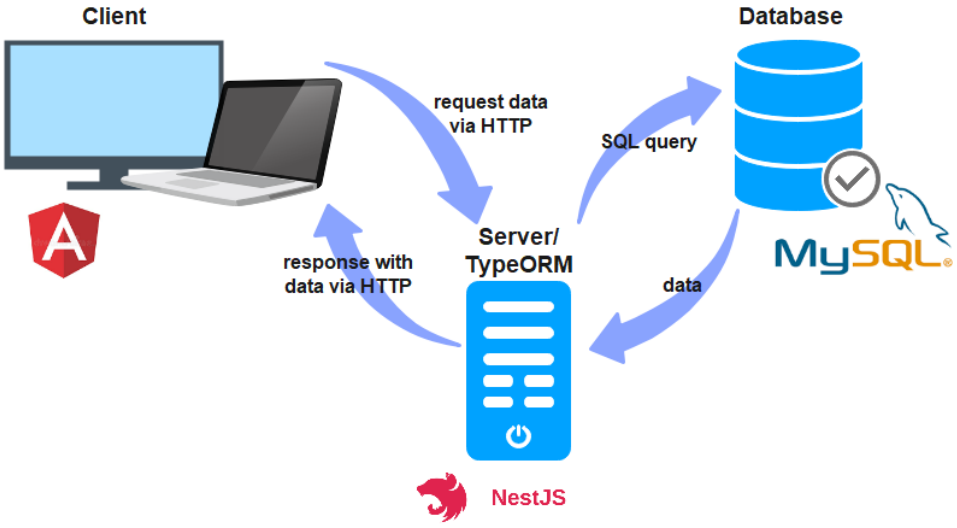

Abb. 2: Technologienkonzept des Bestellsystems [2]

Umsetzung

Zur Umsetzung der Bestellungssoftware wurden Screenshots aus dem alten System bereitgestellt, da Studierenden vom Rechenzentrum kein Zugang auf den fakultätsinternen Server gewährt wird. Diese halfen dabei zu ermitteln, welche Daten angezeigt werden müssen und welche Tabellen aus der Datenbank manipuliert werden können. Schließlich wurden fünf Ansichten definiert: Bestellungen (siehe Abbildung 3), Adressen, Lieferanten, Personen und Inventarisierung, wobei die Implementierung der Inventarisierungsansicht außerhalb des Rahmens dieser Arbeit liegt.

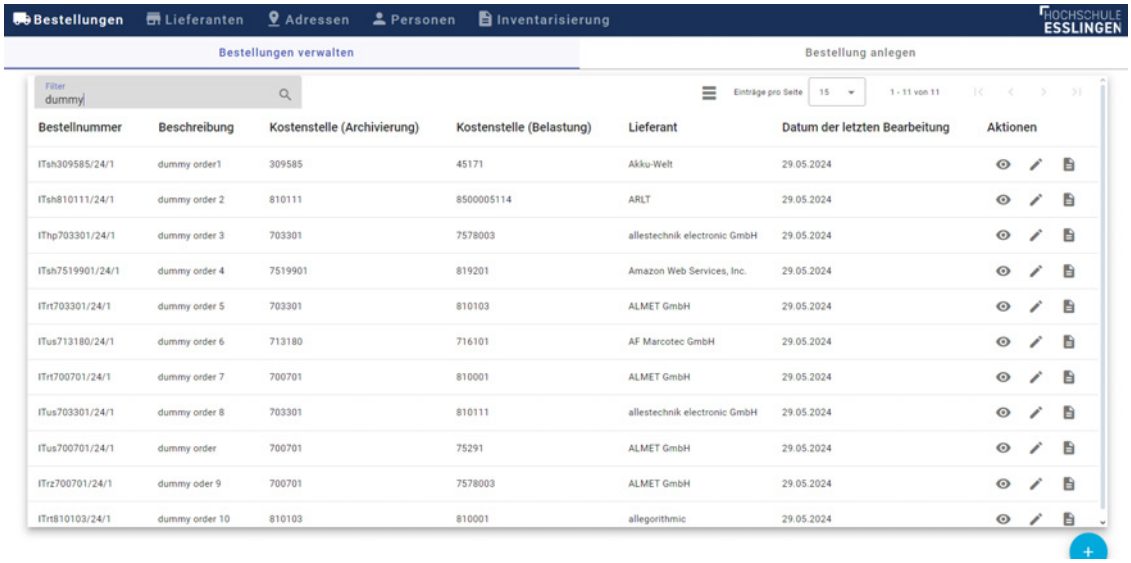

Abb. 3: Bestellungen-Ansicht [2]

Eine Startseite zeigt den Benutzernamen, das Datum sowie die letzten fünf Bestellungen und deren Status an. Die Datenansicht wird mithilfe der Tabellenkomponente der Angular Material Bibliothek realisiert, während weitere Komponenten wie Forms und Dialogs für die Dateneingabe verwendet werden. Im Frontend werden Services verwendet, die HTTP-Anfragen stellen, um die Endpunkte im Backend anzusprechen. Diese Endpunkte werden durch Controller-Klassen repräsentiert, die den jeweiligen Entitäten bzw. Tabellen in der Datenbank zugeordnet sind. Die Controller-Klassen nutzen wiederum Service-Methoden, um CRUD-Operationen (Create, Read, Update, Delete) in der Datenbank durchzuführen. Für das Generieren des ausgefüllten Bestellformulars wurde das Auto-Fill-Tool Apryse [1] verwendet, dass viele wichtige Funktionen bietet, wie:

- Öffnen von Dokumenten,
- Speichern von Dokumenten,
- Zugriff auf Dokumente oder einzelne Seiten eines Dokuments sowie
- Zugriff auf die Inhalte eines Dokuments

Die Finanzabteilung der IT-Fakultät stellt eine Bestellformularvorlage zur Verfügung, die automatisch mit den Bestellinformationen ausgefüllt werden muss. Einen End-point im Backend wird erstellt. Sobald dieser aufgerufen wird, wird mit der Anfrage ein Objekt mit den benötigten Informationen über die ausgewählte Bestellung übermittelt. Ein Backend-Service nutzt diese Informationen, um die .docx-Vorlage zunächst als PDF zu speichern und dann die gespeicherte PDF zu manipulieren, indem die Texte durch die neuen, übermittelten Informationen ersetzt werden. Schließlich wird das Dokument in einem zweiten Tab geöffnet, und die Nutzer können das ausgefüllte Bestellformular herunterladen.

Ausblick

Im weiteren Verlauf sollen die bisher implementierten Anforderungen und Funktionen weiter vervollständigt werden, um eine effiziente und benutzerfreundliche Erfahrung sicherzustellen. Darüber hinaus ist geplant, neue Anforderungen wie die Integration eines SSO-Shibboleth-Logins umzusetzen. Zudem wird die Webanwendung durch verbesserte Styling-Optionen weiter optimiert.

Literatur und Abbildungen

[1] Inc. Apryse Software. PDF and document editing JavaScript library. https://docs.apryse.com/documentation/web/guides/edit/, 2024.

[2] Eigene Darstellung.

[3] Rico Hoffman. Migration einer PHP-basierten Bestellsoftware auf aktuelle REST-API-basierte Technologien, 2024.

Anwendung von Natural Language Processing zur Analyse von Finanzberichten

Jannik Allerdings **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Hintergrund und Motivation

Für Investoren sei es professionell im institutionellen Bereich oder privat, sowie für andere Stakeholder haben von Unternehmen herausgegebene Finanzberichte eine enorme Bedeutung. Die in diesen Berichten enthaltenen quantitativen Kennzahlen und Beschreibungen von Sachverhalten wie z. B. Geschäftsstrategien und Risiken können genutzt werden, um potenzielle Chancen und Risiken zu identifizieren und anhand dieser Informationen wichtige Investitionsentscheidungen zu treffen.

Traditionell war die Analyse und Auswertung von Finanzberichten im institutionellen Bereich bei Investmentunternehmen wie z. B. Hedgefonds Aufgabe von professionellen Analysten. Der Prozess gestaltete sich als sehr zeitaufwendig und ineffizient sowie durch die Gegebenheit der menschlichen Analyse als fehleranfällig.

Die Anwendung von Künstlicher Intelligenz bzw. insbesondere NLP bietet gegenüber manueller Analyse von Finanzberichten einige Vorteile. Große Datenmengen können schnell verarbeitet werden und es ergeben sich neue Möglichkeiten, wie semantische Analyse anhand fester Kennzahlen, welche sonst nur nach der persönlichen Wahrnehmung eines menschlichen Analysten erfolgen würde.

Zielsetzung

Zielsetzung dieser Bachelorarbeit ist es, Finanzberichte mithilfe von NLP-Techniken zu analysieren und die Ergebnisse zu bewerten. Es soll mit dieser Arbeit ein Beitrag bzw. ein Überblick zum Verständnis der Möglichkeiten und Grenzen von NLP zur Anwendung in der Analyse von Finanzberichten geleistet werden.

Natural Language Processing

Bei Natural Language Processing (NLP) handelt es sich um ein interdisziplinäres Forschungsfeld, da Inhalte aus verschiedenen wissenschaftlichen Disziplinen wie Informatik, Linguistik und Künstlicher Intelligenz kombiniert werden. Das Themengebiet umfasst verschiedene Ansätze zur Interaktion zwischen Computern und menschlicher Sprache, wobei Anwendungsfälle die Verarbeitung und Analyse von natürlicher Sprache durch Computer, sowie auch umgekehrt die computerbasierte Generierung menschlicher Sprache sein können. [4]

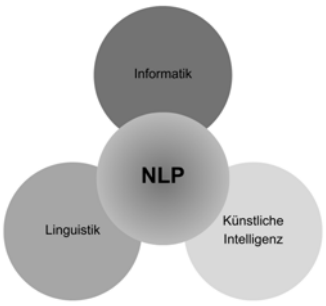

Abb. 1: Forschungsfeld NLP [1]

Neben den in NLP-Anwendungen gängigen Schritten zur Vorverarbeitung von Daten, werden in dieser Arbeit folgende Techniken betrachtet, welche sich zur Analyse von Finanzberichten anbieten:

- Sentiment Analysis: Die Berechnung eines Werts, um die Stimmung eines Textes zu bewerten. Diese kann in unterschiedlich starkem Ausmaß positiv, negativ oder neutral ausgeprägt sein.
- Information Extraction: Das Extrahieren spezifischer Informationen aus Texten. Dazu gehört z. B. auch Named Entity Recognition, welche die Identifikation und Klassifikation von Entitäten wie Personen oder Unternehmen ermöglicht.
- Topic Modeling: Die Identifikation eines Themas, bzw. mehrerer in Texten vorhandener Themen.

- Document Similarity Analysis: Die Bewertung der Ähnlichkeit des Inhalts von verschiedenen Dokumenten bzw. Textabschnitten.
- Text Summarization: Die automatische Erstellung von Zusammenfassungen umfangreicher Texte.

Grundlagen Finanzberichte

Die USA haben mit einem Anteil von 60.5 % am gesamten globalen Aktienmarkt mit Abstand den größten nationalen Aktienmarkt der Welt. [3] Dies verdeutlicht die zentrale Rolle dieses Marktes und das Interesse von internationalen Investoren an US-amerikanischen Unternehmen. Aus diesem Grund sind die Finanzberichte dieser Unternehmen Gegenstand der Analyse der Arbeit.
Berichte US-amerikanischer Unternehmen müssen regelmäßig über verschiedene Formulare bei der US-Börsenaufsichtsbehörde SEC (United States Securities and Exchange Commission) eingereicht werden. Eine der wichtigsten Arten ist der umfangreiche Jahresbericht 10-K, welcher von börsennotierten Unternehmen einmal pro Jahr bei der SEC eingereicht werden muss. Die Gliederung der Inhalte in diesen Berichten erfolgt in vier Parts, welche sogenannte Items mit bestimmten Informationen enthalten. [2]

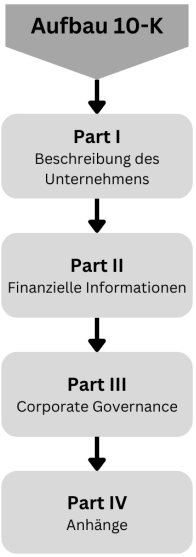

Abb. 2: Aufbau 10-K [1]

Der erste Part enthält wichtige Informationen über das Unternehmen, darunter fallen unter anderem die Items Business, Risk Factors und Legal Proceedings. In Part II sind neben den eigentlichen Finanzaufstellungen auch die Management's Discussion and Analysis of Financial Condition and Results of Operations sowie Quantitative and Qualitative Disclosures about Market Risk enthalten. Die Items in Part III enthalten verschiedene Informationen zu den Führungskräften des Unternehmens. Im vierten Part sind ausschließlich Anhänge enthalten. [2]
Aus den in den Berichten enthaltenen Items sind besonders die folgenden für die Analyse mit NLP-Techniken hervorzuheben:

- Business: Beschreibung der Geschäftstätigkeit des Unternehmens (Hauptprodukte und -dienstleistungen, Märkte, Wettbewerb usw.).
- Risk Factors: Auflistung aller signifikanten Risiken, welche Einfluss auf das Unternehmen haben könnten.
- Legal Proceedings: Aktuell laufende oder potenziell in Zukunft drohende Rechtsstreitigkeiten.
- Management's Discussion and Analysis of Financial Condition and Results of Operations (MD&A): Analyse der Geschäftsergebnisse durch das Management.
- Quantitative and Qualitative Disclosures about Market Risk: Informationen über spezifische Markt- und Finanzrisiken, denen das Unternehmen ausgesetzt ist und wie damit umgegangen werden soll. [2]

Methodik

Die Implementierung von NLP-Techniken mithilfe von Python zur Analyse von Finanzberichten, stellt den allgemeinen Ansatz der Forschung in dieser Arbeit dar. Zur Datenbeschaffung konnte die für die Öffentlichkeit frei zugängliche Datenbank EDGAR (Electronic Data Gathering, Analysis and Retrieval) der SEC genutzt werden. Die Aufbereitung der Daten umfasst klassische NLP-Vorverarbeitung wie Tokenization, Stemming und Lemmatization sowie Stop-Word-Removal.

Ausblick

Im nächsten Schritt der Arbeit werden die ausgewählten NLP-Techniken implementiert. Die Validierung der Ergebnisse erfolgt dabei durch Validierungsmethoden, die für die jeweilige Anwendung passend sind. Für die Sentiment Analysis und Information Extraction werden klassische Metriken wie Precision, Recall und F1-Score genutzt. Beim Topic Modeling wird durch qualitative Analyse manuell überprüft, ob die erzeugten Topics im Hinblick auf die zugrunde liegenden Texte sinnvoll sind. Die Ergebnisse der Document Similarity Analysis werden ebenfalls manuell auf ihre Qualität überprüft. Diesem qualitativen Ansatz entspricht auch die Validierung der Ergebnisse der Text Summarization, welche

mit den in 10-K-Berichten bereits teilweise enthaltenen Zusammenfassungen verglichen werden können. Im Anschluss an Implementierung und Validierung werden die Ergebnisse interpretiert und unter anderem auch im Hinblick auf praktische Anwendbarkeit diskutiert. Diese Herangehensweise soll sicherstellen, dass eine solide Grundlage für die Anwendung von NLP zur Analyse von Finanzberichten entsteht.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] US Securities and Exchange Commission. How to read a 10-K. https://www.sec.gov/investor/pubs/reada10k, 09 2011.

[3] Credit Suisse. Distribution of countries with largest stock markets worldwide as of January 2023, by share of total world equity market value. https://www.statista.com/statistics/710680/global-stock-markets-by-country/, 02 2024.

[4] O. G Yalçın. *Applied Neural Networks with TensorFlow 2: API Oriented Deep Learning with Python*. Apress, 2020.

Adaptierung von Mitarbeiterprofilen an Projektanforderungen mit Hilfe von Large Language Models

Martin Au **Thao Dang**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma adesso SE, Stuttgart

Einleitung

In den letzten Jahren hat sich das Feld der künstlichen Intelligenz (KI), besonders im Bereich des Natural Language Processing (NLP), stark weiterentwickelt. NLP ist ein spezieller Zweig der KI, der sich mit der Interaktion zwischen Computern und menschlicher Sprache befasst. Innerhalb des NLP-Kontextes haben große Sprachmodelle (LLMs) wie ChatGPT-4 von OpenAI eine wichtige Rolle in der Gesellschaft eingenommen. Im Zeitalter der Datenflut gewinnen LLMs, die auf den reichhaltigen Informationsressourcen des Internets trainiert werden, immer mehr an Bedeutung durch ihre beeindruckenden Leistungen. Technologien wie Spracherkennung, maschinelle Übersetzung und automatisierte Textgenerierung erweitern die Interaktionsmöglichkeiten zwischen NLP-Modellen und Menschen und verbessern den Alltag auf vielfältige Weise. Diese Entwicklungen erhöhen das Potenzial von KI in vielen Anwendungsbereichen und schaffen die Grundlage für weitere Forschung in diesem dynamischen und innovativen Bereich.

Problemstellung und Zielsetzung

Das Ziel dieser Arbeit ist es, an spezifische Projektanforderungen angepasste Mitarbeiterprofile mithilfe von LLMs zu generieren. Die Automatisierung dieser Anpassung könnte die Effizienz in der Mitarbeiterverwaltung steigern, da manuelle und zeitaufwändige Tätigkeiten entfallen würden. Sowohl die theoretischen Grundlagen der Struktur und Funktionsweise von Sprachmodellen (LMs) als auch die praktische Anwendung und Implementierung vortrainierter Modelle werden betrachtet. Besonders wird auf die Transformer-Architektur von Modellen wie GPT und BERT eingegangen. Das Modell soll in eine unternehmensinterne Anwendung integriert werden, die es Führungskräften ermöglicht, Projektaufgaben an Mitarbeiter zu delegieren. Diese Anwendung bereits bietet eine Übersicht über die Qualifikationen und Erfahrungen der Mitarbeiter, die in eine PowerPoint-Datei exportiert werden kann. Diese Funktion ist jedoch fehleranfällig und erzeugt oft unvollständige Ergebnisse. Zusätzlich soll das generierte Dokument auch dem Projektkunden übergeben werden. Aufgrund der vertraulichen internen Projekte, auf die der Kunde keinen Zugriff hat, kann die exportierte Kompetenzübersicht nicht direkt weitergeleitet werden. Um eine manuelle Bearbeitung zu vermeiden, soll das Modell in der Lage sein, eine PowerPoint-Datei zu erstellen, die die relevanten Fähigkeiten für das Projekt hervorhebt und vertrauliche Daten herausfiltert.

Aufbau einer Transformer-Architektur

Der aktuelle Aufschwung von LLMs wie ChatGPT und BERT ist der Transformer-Architektur zu verdanken. Diese Architektur setzt einen wichtigen Meilenstein im maschinellen Lernen (ML) und der Verarbeitung natürlicher Sprache (NLP). Im Vergleich zu früheren sequenziellen Ansätzen verarbeitet der Transformer Informationen effizienter mithilfe des Konzepts der Self-Attention, das erstmals in dem Artikel Ättention is All You Need"von Vaswani 2017 eingeführt wurde. [4] Die sequentielle Verarbeitung von Informationen in früheren Modellen wird in der neuen Architektur parallelisiert und durch den Self-Attention-Mechanismus modelliert. Basierend auf der Attention-Variable werden globale Abhängigkeiten zwischen Eingabe und Ausgabe berücksichtigt, wodurch die bisher vorherrschende rekurrente Architektur umgangen wird. Wie Sequence-to-Sequence-Modelle verwendet der Transformer ebenfalls eine Encoder-Decoder-Architektur. (siehe Abbildung 1)

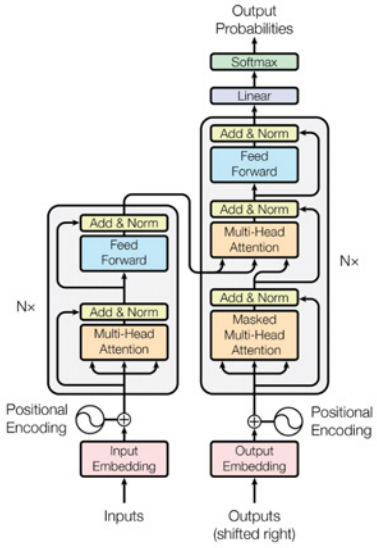

Abb. 1: Aufbau der Transformer Architektur [4]

Konzeption und Umsetzung

In diesem Kapitel werden die Planung und die Auswahl der Methoden erläutert. Zunächst werden die spezifischen Anforderungen an die Anwendung von LLMs in der Personalzuweisung und Projektplanung identifiziert. Diese Anforderungen umfassen die Fähigkeit, Mitarbeiterprofile automatisch zu analysieren und relevante Fähigkeiten und Erfahrungen zu erkennen und diese mit den Anforderungen aktueller Projekte abzugleichen. Zusätzlich soll das Modell lernen, vertrauliche Daten und Projekte aus der Ausgabe herauszufiltern. Für die Implementierung wurden folgender Hauptansatz ausgewählt: ein lokales Modell auf Basis von DistilBERT. DistilBERT wurde aufgrund seiner Effizienz und vergleichbaren Leistungsfähigkeit zu größeren BERT-Varianten gewählt. Es wird auf einem unternehmensinternen Azure-Server in einer virtuellen Maschine (VM) bereitgestellt. Diese Bereitstellung auf einem sicheren und skalierbaren Server bietet die notwendige Flexibilität und Leistung, um große Datenmengen effizient zu verarbeiten und gleichzeitig Datenschutz und Datensicherheit zu gewährleisten. [3] Diese Applikation wird in dem unternehmensinternen Mitarbeiterverwaltungsprogramm Staffinghelper integriert und erhält die nötige Eingabe, um die gewünschte Mitarbeiterübersicht zu generieren. (siehe Abbildung 2)
Die ChatGPT API bietet fortschrittliche generative Fähigkeiten in einer cloudbasierten Umgebung, die für komplexe Textgenerierungsaufgaben nützlich sind. Diese beiden Modelle ermöglichen es, sowohl lokal als auch cloudbasiert verschiedene Szenarien zu vergleichen. Die einfache Implementierung der API benötigt nur einen API-Key, welcher bei OpenAI erworben werden muss. Der Vergleich des lokalen DistilBERT Modelles zur ChatGPT API erfolgt jedoch nur mit anonymisierter Testdaten, aufgrund des Risikos vertrauliche Informationen offenzulegen, und dient lediglich als Orientierungshilfe bei der Evaluation. [1]

Ausblick

Die bisherigen Ergebnisse zeigen, dass die Implementierung von LLMs zur Anpassung von Mitarbeiterprofilen an Projektanforderungen bereits vielversprechende Resultate liefert. Bei Erweiterung des Scopes könnte die Applikation nicht nur das Generieren einer geeigneten Übersicht für die Mitarbeiter übernehmen, sondern auch die Aufgabenverteilung. Um solch eine komplexe Aufteilung vorzunehmen, benötigen LLMs eine große Menge an Datensätzen, welche u.a. empflindliche Daten von Mitarbeitern und Projekten enthalten können. Unter Berücksichtigung des Datenschutzes und einem diskreten Umgang vertraulicher unternehmensinterner Informationen entlasten LLMs die Mitarbeiter bei alltäglichen Aufgaben. Statt eines großen Modells wie ChatGPT zu verwenden, kann man kleiner effizientere Modelle wie DistilBERT auf jeweilige Anwendungsaufgaben trainieren und verfeinern, welche innerhalb des Unternehmensnetzwerkes bereitgestellt werden.

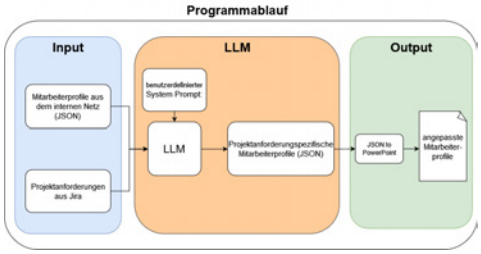

Abb. 2: Programmablauf Konzept [2]

Literatur und Abbildungen

[1] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, et al. Language Models are Few-Shot Learners. In *34th Conference on Neural Information Processing Systems*. NeurIPS, 2020.

[2] Eigene Darstellung.

[3] Victor Sanh, Lysandre Debut, and Thomas Wolf. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. In *5th Workshop on Energy Efficient Machine Learning and Cognitive Computing*. NeurIPS, 2019.

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. In *Advances in Neural Information Processing Systems*. Neural Information Processing Systems Foundation, 2017.

Automatisierung durch Künstliche Intelligenz: Konzeptionierung eines Proof of Concept für eine effizientere Softwarenutzung in der Rohbauplanung

Kerwin Au **Thao Dang**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einführung

Die Rohbauplanung ist ein zentraler Aspekt der Fahrzeugproduktion und hat erheblichen Einfluss auf die Effizienz und Qualität der Endprodukte. Ein wichtiger Bestandteil dieser Planung ist die MTM-Analysen (Methods-Time Measurement), die genaue Zeitvorgaben für verschiedene Arbeitsschritte festlegt. Diese Analysen sind essenziell für die Produktionsplanung und -steuerung, da sie realistische Zeitpläne ermöglichen und die Ressourcenverteilung optimieren. Derzeit werden MTM-Analysen manuell von Planern erstellt.

Problemstellung

Die manuelle Erstellung von der MTM-Analysen ist nicht nur zeitintensiv, sondern auch anfällig für Fehler. Verschiedene Systeme und Methoden führen zu Inkonsistenzen, die die Nachverfolgbarkeit und Übersichtlichkeit der Daten erschweren. Eine KI-basierte Lösung könnte hier Abhilfe schaffen, indem sie historische Planungsdaten nutzt, um MTM-Analysen automatisch zu erstellen. Ziel dieser Arbeit ist es, ein Proof of Concept (PoC) für die Automatisierung dieser Analysen zu entwickeln und damit die Effizienz und Genauigkeit der Rohbauplanung zu verbessern. (Siehe Abbildung 1)

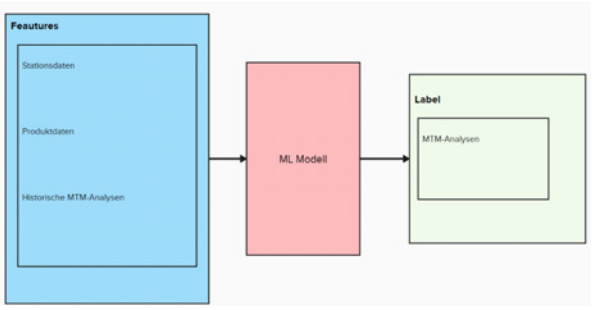

Abb. 1: Modell des PoC's [1]

Definition von Methods-Time Measurement (MTM):

Methods-Time Measurement (MTM) ist ein Verfahren zur Analyse manueller Operationen oder Methoden, indem diese in die grundlegenden Bewegungen zerlegt werden, die zur Ausführung erforderlich sind. Jeder dieser Bewegungen wird ein vorbestimmter Zeitstandard zugewiesen, der durch die Art der Bewegung und die Bedingungen, unter denen sie ausgeführt wird, bestimmt wird. Somit ist MTM im Wesentlichen ein Werkzeug zur Methodenanalyse, das Zeitvorgaben liefert, ohne dass eine Zeitstudie mit einer Stoppuhr notwendig ist. (Siehe Abbildung 2) [3]

Schritt	Grundbewegung	Zeit in m	Beschreibung
1. Greifen des Gegenstands	G1A	0.10	Hand bewegt sich zum Behälter und greift den Gegenstand
2. Transportieren des Gegenstands	M1A	0.20	Hand transportiert den Gegenstand zur Arbeitsstation
3. Positionieren des Widerstands	P1A	0.15	Hand positioniert den Gegenstand auf der Leiterplatte
4. Löten des Widerstands	L1A	0.50	Hand hält den Gegenstand, während er mit dem Lötkolben gelötet wird
5. Visuelle Inspektion	I1A	0.30	Visuelle Inspektion des gelöteten Gegenstands auf der Leiterplatte
Gesamtzeit:		1.25	

Abb. 2: Beispiel einer MTM Analyse [1]

Zusätzlich zur Schrittanalyse werden für jedes Produkt auch Messdaten, wie Gewicht erfasst. Die Messdaten führen dazu, dass jedes Produkt aufgrund seiner individuellen Eigenschaften eine spezifische Zeit-Analyse erhält, was zu unterschiedlichen MTM-Analysen führt. Mithilfe der Zeit-Analysen sollen die Zusammenhänge zwischen diesen Messdaten analysiert und eine Methode entwickelt werden, die MTM-Analysen automatisiert. Durch die Identifikation und Modellierung der Abhängigkeiten zwischen den Produktmerkmalen und den Zeitvorgaben soll ein System geschaffen werden, das die MTM-Analysen effizient und präzise erstellt.

Methodik
Um dieses Ziel zu erreichen, werden relevante Daten lokal gesammelt und bereinigt, darunter fallen Daten wie Mengengerüste, Produktmerkmale und bisherige von verschiedenen Standorten und Werken durchgeführten MTM-Analysen. Diese Daten dienen als Grundlage für das Training eines KI-Modells. [4] Der Prozess umfasst mehrere Schritte: (Siehe Abbildung 3)

1. Datenaufbereitung

- **Datensammlung:** Zunächst werden die relevanten Daten aus verschiedenen Quellen gesammelt. Dazu gehören die Schrittanalysen, Messdaten (Abmessungen und Gewicht) und bisherige MTM-Analysen.
- **Datenbereinigung:** Die gesammelten Daten werden bereinigt, um Ungenauigkeiten, Duplikate und fehlende Werte zu entfernen. Dies gewährleistet eine konsistente und zuverlässige Datenbasis.
- **Datenstrukturierung:** Die Daten werden in einem geeigneten Format strukturiert, das die weiteren Analysen und Modellierungen erleichtert. Dabei werden die Messdaten den entsprechenden MTM-Analysen zugeordnet.

2. Datenexploration

- **Deskriptive Analyse**: Eine erste Analyse der Daten wird durchgeführt, um grundlegende statistische Kennzahlen zu ermitteln und die Verteilung der Daten zu verstehen.
- **Explorative Datenanalyse (EDA)**: Mittels visueller und statistischer Methoden werden Muster und Zusammenhänge in den Daten identifiziert. Hierbei kommen Techniken wie Scatterplots, Korrelationsmatrizen und Boxplots zum Einsatz, um die Beziehungen zwischen den Messdaten und den Zeitvorgaben zu untersuchen.

3. Modellierung

- **Feature Engineering:** Relevante Merkmale (Features) werden aus den Daten extrahiert und transformiert, um die Leistungsfähigkeit der Modelle zu verbessern. Dabei werden auch neue, abgeleitete Merkmale erstellt, die möglicherweise bessere Prädiktoren für die MTM-Analysen darstellen.
- **Modellauswahl und -training:** Verschiedene maschinelle Lernmodelle werden implementiert und trainiert. Dazu gehören lineare Regression, Entscheidungsbäume und neuronale Netzwerke. Die Modelle werden anhand eines Trainingsdatensatzes trainiert, der aus den bereinigten und strukturierten Daten besteht.
- **Modelloptimierung:** Die Hyperparameter der Modelle werden optimiert, um die bestmögliche Leistung zu erzielen. Dies erfolgt durch Techniken wie Grid Search oder Random Search und unter Verwendung von Kreuzvalidierung zur Bewertung der Modellleistung.

4. Evaluation

- **Modellbewertung**: Die Modelle werden anhand eines Testdatensatzes evaluiert, der nicht für das Training verwendet wurde. Relevante Metriken wie Genauigkeit und Präzision werden berechnet, um die Leistung der Modelle zu beurteilen.
- **Vergleich der Modelle**: Die Ergebnisse der verschiedenen Modelle werden verglichen, um das leistungsfähigste Modell auszuwählen. Dabei werden sowohl die Vorhersagegenauigkeit als auch die Interpretierbarkeit der Modelle berücksichtigt.

5. Implementierung und Validierung

- **Prototypentwicklung:** Ein Prototyp des Systems wird entwickelt, der die automatische Generierung von MTM-Analysen ermöglicht. Der Prototyp integriert das ausgewählte Modell und eine Benutzeroberfläche, die es den Planern ermöglicht, die generierten Analysen zu überprüfen und zu validieren.
- **Validierung in der Praxis:** Der Prototyp wird in einer realen Produktionsumgebung getestet, um seine praktische Anwendbarkeit und Leistung zu bewerten. Feedback von den Anwendern wird gesammelt und zur weiteren Verbesserung des Systems genutzt. [2]

Abb. 3: CRISP-DM (Cross Industry Standard Process for Data Mining) [1]

Schlussfolgerung
Die Automatisierung der MTM-Analyse durch Künstliche Intelligenz hat das Potenzial, die Effizienz und Qualität der Rohbauplanung erheblich zu verbessern. Durch die Reduktion manueller Arbeit und die Minimierung von Fehlern können Produktionsprozesse optimiert und Kosten gesenkt werden. Die bisherigen Ergebnisse sind vielversprechend und deuten darauf hin, dass die Integration von KI in die Rohbauplanung nicht nur realisierbar, sondern auch vorteilhaft ist. Mit der weiteren Entwicklung und Validierung des Prototyps wird ein bedeutender Schritt hin zu einer vollständig automatisierten Produktionsplanung unternommen.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] IBM International Business Machines et al. Data Mining [Zugriffsdatum 09.06.2024]. https://www.ibm.com/topics/data-mining, 2023.

[3] Harold Bright Maynard, Gustave James Stegemerten, John L Schwab, et al. *Methods-time measurement [Zugriffsdatum 09.06.2024]*. McGraw-Hill Book Company INC., 1948.

[4] the free encyclopedia Wikipedia et al. Cross-industry standard process for data mining [Zugriffsdatum 09.06.2024]. https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining, 2005.

Plausibility Check of Redundant LiDAR Data for Safety in Autonomous Vehicles

Julian Baisch **Clemens Klöck**

Department of Computer Science and Engineering, Esslingen University

Work carried out at Alfred Kärcher SE & Co. KG, Winnenden

Introduction

As the field of autonomous robotics progresses, ensuring the safety and reliability of these systems becomes increasingly important. Central to this challenge is the use of sophisticated sensors, such as LiDAR (Light Detection and Ranging), which provide crucial data for navigation and obstacle detection [3]. However, the failure or malfunction of these sensors can have serious consequences. This thesis explores the implementation of redundant LiDAR scanners with a real-time plausibility check mechanism to enhance the safety and reliability of autonomous vehicles.
Employing multiple LiDAR sensors and continuously validating their data against each other allows for prompt detection of inconsistencies and potential failures. This redundancy ensures that the autonomous system maintains its functionality even if one sensor fails, while the plausibility check ensures the integrity and accuracy of the data being used. This thesis delves into the methodologies for implementing these checks and evaluates the performance of redundant systems under various conditions. Through this approach, the sensor chain in autonomous vehicles can be fortified, thereby significantly improving their safety and reliability in operation.

Motivation

The motivation behind this research stems from the critical need to enhance the safety protocols of autonomous cleaning robots. As these robots are increasingly used in different environments, the reliability of their sensor systems has a direct impact on operational efficiency and safety. LiDAR sensors are essential for real-time environmental mapping and obstacle detection, but their failure can lead to operational disruptions and potential hazards. By incorporating redundancy and real-time plausibility checks, a more robust sensor network can be created that withstands individual sensor failures.

Additionally, this approach is cost-effective, allowing the use of non-safety certified hardware while still achieving overall safety through the implementation of redundancy and plausibility checks. This method reduces the need for expensive, high-safety-certified components, making the technology more accessible and scalable. This advancement not only improves the immediate safety and efficiency of autonomous cleaning robots, but also paves the way for wider acceptance and integration of autonomous technologies in various cleaning applications.

Research Objectives

In this thesis, the research objectives are aimed at ensuring the safety and reliability of autonomous cleaning robots equipped with LiDAR sensors. The main objectives include:

1. Building and finalizing a simulation framework for rigorous testing of proposed algorithms and redundancy mechanisms.
2. Developing mathematical algorithms to cross-validate laser scans from two independent LiDAR scanners within a fixed reference frame.
3. Implementing developed algorithms and frameworks on real embedded hardware, following best practices like modular programming and unit testing.

Overview

To enable plausibility checks of the two LiDAR sensors, a ground truth must be established by fusing the two sensor inputs into a combined reference system [2]. The LiDAR sensors provide measurements about the distance to a detected reflection, its angle, and the intensity of the laser reflection. This process is continuous but covers a field of view (FoV) of just 270 degrees due to the design of the optical disk of

the LiDAR scanner. This arrangement is also mapped one-to-one in a simulation environment, as shown in 1, where tests can be carried out quickly without relying on the hardware.

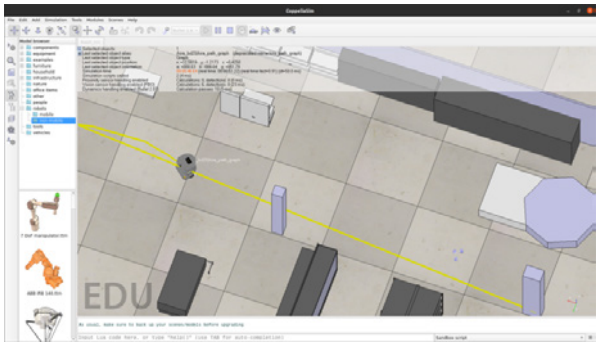

Fig. 1: Simulation environment with Coppeliasim [4]

Grid map and data fusion

Both sensors are positioned at the front of the robot, one on the left and the other on the right side. By rotating each sensor approximately 45 degrees clockwise and counterclockwise, respectively, the combined coverage area only lacks information directly behind the robot, where the body of the robot shades the LiDAR sensors.
A virtual grid map is laid out in front of the robot and sized to meet safety requirements while enabling fast processing with short computation times to avoid overloading the microcontroller, which runs various other tasks simultaneously. In its current state, an area of 3 by 1.5 meters is covered by 450 cells, each measuring 0.1 by 0.1 meters.

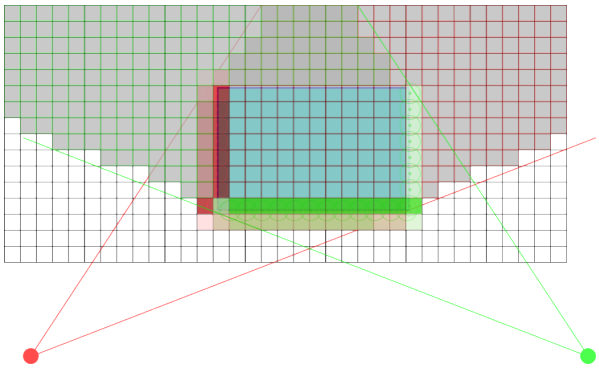

Fig. 2: Virtual grid map with both LiDAR sensors active. Not to scale. [4]

2 displays the sensor layout together with the virtual grid map directly in front of the robot. An obstacle, a rectangular box colored in light blue, is set in the scene together with the robot. Each LiDAR sensor only detects two sides of this box, colored in its respective colors. The side facing the robot, however, is detected by both sensors simultaneously and can therefore be used directly to perform a plausibility check of the two sensors against each other. Each cell of the grid map stores the laser beams passing through it or terminating in it, their angle, and, if applicable, their intensity. The affected cells on the front side of the object then contain information on rays from the left and right LiDAR scanners. Transformations between the object and the robot caused by the movement of the robot, the object, or both are of no concern, as they are tracked and compensated for by the transformation buffer of the robot operating system (ROS).
The recorded hit-and-miss history of each cell and for each LiDAR scanner respectively is then used to determine the plausibility of this detection event in a certain cell.

Outlook

A solution still needs to be found that also takes into account the requirements of important standards, such as IEC61508 [1].
Future improvements will include enhancements for virtual detections which are shaded by an object but cannot be less plausible. In a first draft, this might be solved in a fashion similar to object detection by closing gaps, like the distant side in 2, and matching this against known shapes.
Furthermore, extensive development is being done parallel to the main research to enable visualization of the information, making it more accessible to humans.

References and figures

[1] Alessandro Frigerio. Functional-safety analysis of ASIL decomposition for redundant automotive systems, 2021.

[2] Florian Geissler, Alex Unnervik, and Michael Paulitsch. A Plausibility-based Fault Detection Method for High-level Fusion Perception Systems. *IEEE Open Journal of Intelligent Transportation Systems*, pages 176–186, 2020.

[3] Sharath Patil, Bhanu Singh, Darrell Livezey, Saad Ahmad, and Martin Margala. *Functional Safety of a Lidar Sensor System*. IEEE, 2020.

[4] Own representation.

Abstandserkennung mit einem Convolutional Neural Network in Python

Achim Baumgaertner **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Leuze electronic, Owen

Motivation

Durch die Entwicklung der Technik ist es möglich, zahlreiche Vorgänge zu automatisieren. Ein Hauptreiber dieser Innovation sind Sensoren. Aufgrund der sehr vielen Einsatzgebiete gibt es entsprechend viele verschiedene Sensoren. Ein Sensor ermöglicht z. B. die Erkennung von Mustern und deren Bewegung in Videodaten. Ein anderer Sensor bestimmt z. B. die wechselnde Helligkeit in Bilddaten. Ein weiterer Sensor detektiert die Änderung von Abständen mit Hilfe eines Lasers. Die Sensoren sollten nahezu in Echtzeit ihre Aufgabe ausführen können. Im Fehlerfall können Sensoren im besten Fall lediglich einen Arbeitsstau erzeugen. Im schlimmsten Fall wird die Sicherheit gefährdet. Aus diesem Grund ist auf dem Arbeitsgebiet der Sensoren mit äußerster Sorgfalt zu handeln und in Simulationsmodellen möglichst realitätsnah und nachvollziehbar zu arbeiten.

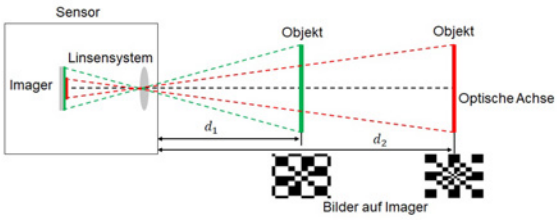

Abb. 1: Erfassung eines graphischen Musters mit Bildsensor [1]

Die Motivation dieser Arbeit ist es, herauszufinden, ob der Abstand zwischen Bildsensor und Objekt mit Hilfe von neuronalen Netzen bestimmt werden kann. Abbildung 1 zeigt eine entsprechende Sensoranordnung. In Abhängigkeit des Abstandes ändert sich der Bildinhalt der Bilddaten auf dem Imager. Im Abstand d_1 erhält man einen vergrößerten Ausschnitt des Bildes, welches im Abstand d_2 aufgenommen wird. Eine hohe Genauigkeit, mit der die Abstände bestimmt werden, ist hierbei von Vorteil. Um Daten auswerten zu können, werden Bilder mit Hilfe von optischen Abbildungsformeln erzeugt. Die Bilder simulieren einen Sensor der Bilddaten in Abhängigkeit vom Abstand zwischen Sensor und Objekt generiert. Um das Netz zu trainieren, werden tausende Paare aus Bilddaten und zugehörigem Abstandswert verwendet. Kann das neuronale Netz die Distanzwerte mit hoher Genauigkeit bestimmen, ist ein Einsatz dieses Systems in Automated Guided Vehicles (AGV) denkbar. [3]

Neuronale Netze

Über die letzten Monate und Jahre hinweg hat sich die künstliche Intelligenz und das Machine Learning sehr stark hinsichtlich Algorithmen und Zuverlässigkeit verbessert. Dadurch hat sich die Anzahl an Einsatzmöglichkeiten stark erhöht. Die verschiedensten Arten der neuronalen Netze sind Grund für die zahlreichen Einsatzgebiete. Abbildung 2 zeigt beispielhaft ein neuronales Netz mit Eingangsschicht, verborgenen Schichten und Ausgangsschicht. Themen die dabei riesiges Interesse wecken sind Autonomes Fahren, Gesundheitswesen, Finanzen und Bildverarbeitung. [2]

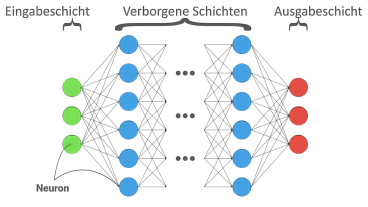

Abb. 2: Struktur Convolutional Neural Network [1]

Das Lernen und erkennen von Mustern ist inzwischen so weit fortgeschritten, dass damit medizinische Fälle gelöst werden, bei denen Faltungsnetze, sogenannte Convolutional Neural Networks (CNN), eine vergleichbar hohe Präzision wie Experten erzielen. Dadurch kommen auch in heutigen Systemen, wie autonomes Fahren, neuronale Netze zum Einsatz.

Durchführung

Da in dieser Arbeit mit Bildsensoren gearbeitet wird, ist ein CNN besonders nützlich. Ein CNN eignet sich speziell für die Bildervearbeitung und somit zum Erkennen von Merkmalen, da diese nahezu unabhängig der Positionen detektiert werden können. Die Datengenerierung für das CNN erfolgt über ein kleines Programm, welches in Python geschrieben ist. Hierbei sollen nicht nur optimale Bilder erzeugt werden, da die Genauigkeit mit realitätsnahen Bilddaten, die Störungen beinhalten, getestet werden soll. Effekte die zur Störung beitragen sind Schmier-, Kontrast-, Licht- und Rauscheffekte. Somit werden vier Störeffekte berücksichtigt. Die Effekte können entweder separat oder gleichzeitig aktiviert werden. Dadurch können pro Distanzwert, $2^4 = 16$ Bilder generiert werden, wobei ein Bild das optimale Bild darstellt.

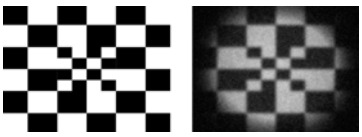

Abb. 3: Erzeugtes Roh- und Störbild [1]

Den Bildern in Abbildung 3 liegt der gleiche Distanzwert zugrunde. Der Unterschied ist, dass das rechte Bild mit Störeffekten hinterlegt ist. Durch die Störeffekte wird die Genauigkeit bei der Abstandserfassung mutmaßlich verringert. Neben der bisher beschriebenen Regressionsaufgabe, um den Distanzwert zu ermitteln, ist auch eine Klassifikationsaufgabe denkbar. Keras bietet Funktionen, um die erzeugten Bilder in Abhängigkeit des Distanzwertes zu klassifizieren. Durch diese Vorgehensweise erhofft man sich eine höhere Genauigkeit. Anders als beim One-Hot Encoding, welches eine Einheitsmatrix widerspiegelt, wird die Klassifikation durch das Erzeugen einer bestimmten Ordnerstruktur ermöglicht. Für das Szenario dieser Arbeit werden z. B. Ordner für Bilder im $0,1mm$ Abstand erzeugt und dadurch die Bilder dementsprechend klassifiziert.

Resultat

Durch die Entwicklung eines Netzes mit angemessener Struktur lässt sich durch das Training die Verlustfunktion einer Regression reduzieren und die Bestimmung der Distanzwerte ist damit möglich. Durch die Verwendung von unterschiedlichen Störeffekten wird die Robustheit des Systems erhöht. Abbildung 4 zeigt die Abweichung der durch das neuronale Netz ermittelten Distanzwerte zu den tatsächlichen Distanzwerten (Soll-Distanzwerten) in Abhängigkeit der Soll-Distanz.

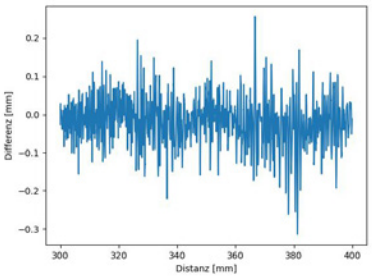

Abb. 4: Abweichung über ermittelte Distanz (in Millimeter) [1]

Die erzielte Genauigkeit der Distanzwerte lässt darauf schließen, dass ein Einsatz dieses Systems in autonomen Systemen möglich ist.

Ausblick

Das System kann in AGVs implementiert werden, um damit das autonome Fahren zu unterstützen. Um die Funktionalität des Systems zu erhöhen, wäre die Detektion von Verdrehungswinkeln, die ein Fahrzeug zusätzlich zum Abstand gegenüber den Merkmalsbildern besitzen kann, wünschenswert. Für diese zusätzliche Erkennung müsste das neuronale Netz neben den Distanzwerten auch die Verdrehungswinkel bestimmen. Alternativ wäre auch eine Lösung mit zwei neuronalen Netzen, Eines zur Detektion der Distanzwerte und eines zur Ermittlung der Verdrehungswinkel, denkbar.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Alex Krizhevsky et al. ImageNet classification with deep convolutional neural networks. https://dl.acm.org/doi/10.1145/3065386, 2017.

[3] Ihab S. Mohamed et al. Detection, localisation and tracking of pallets using machine learning techniques and 2D range data. https://www.researchgate.net/publication/324150758_Detection_localisation_and_tracking_of_pallets_using_machine_learning_techniques_and_2D_range_data, 2018.

Domain-Driven Design of a Metering-Related Back Office System With Remotely Readable Meters via LoRaWAN: Design, Implementation and Evaluation

Philipp Bender **Michael Scharf**

Department of Computer Science and Engineering, Esslingen University

Work carried out at EM Energiemanagement, Weinstadt

Problem

This work concerns an experimental software development of the metering system of a tenant current provider. According to the taxonomy of IEC 61968, the software system to be developed is a "metering-related back office system".
Since 2017, tenant electricity has been subsidized by the Renewable Energy Sources Act (EEG) in Germany to allow tenants to participate financially in the generation of photovoltaic electricity and to increase the installation of photovoltaic systems on apartment buildings [4]. Two measurement concepts have been established to measure the direct consumption of participating tenants and their residual load. The choice depends on the participation of the tenants in the tenant electricity model [4]:

- Physical summation meter, economically suitable for high tenant participation
- Virtual summation meter, economically suitable for low tenant participation

Only if contractual relationships with the tenant and annual billing are efficient and as automated as possible, profitable operation can be achieved. Remotely readable meters are relevant for this purpose.
The problem is summarized:

- Annual billing and contractual relationships with participating tenants must be automated as effectively and efficiently as possible.
- The technological change due to legislation towards the smart meter gateway with external market participant backend is inevitable while the existing meters with LoRaWAN communication modules need to be supported.
- Change of the problem domain by changing existing or introducing new business areas such as submetering of heat meters.

The technical context is depicted in Fig. 1.

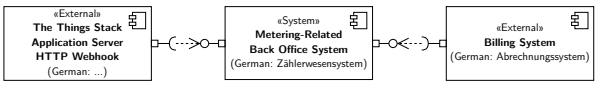

Fig. 1: Technical context diagram of the metering system to be developed [5]

Quality goal	Scenario	Solution approach
Change of the problem domain	All	DDD: Subdomains are identified and mapped one-to-one to problem-induced bounded contexts.
Technological change	All	DDD: One-to-many mapping of a subdomain to solution-induced bounded contexts
Metering-related data model	Development	Common Information Model (CIM)
Evaluate usability	All	Start by implementing domain models directly in the frontend. Attach user interface directly to the domain models. Review and carry out usability tests.
Avoid "Big Ball of Mud"	Development	Modular monolith with hexagonal architecture
Developers have experience with the frontend framework.	Development	TypeScript, Vue.js
Developers have experience with the backend framework.	Development	Java, Spring Boot

Fig. 2: Solution strategy [5]

Methods

The solution strategy is depicted in Fig. 2. The concepts of subdomain and bounded context from Domain-Driven Design (DDD) are used to address the quality goals with regard to changeability. DDD is introduced by Eric Evans in the "blue book" [2]. The implementation of DDD is covered in detail by Vaughn Vernon [6].
The Common Information Model (CIM) is developed by the IEC Technical Committee 57 and the CIM User Group. CIM consists of the three standards IEC 61970, IEC 61968 and IEC 62325. The question arises whether and how CIM can be used for the domain source code. "Domain Prototyping" by Tobias Goeschel [3], which is based on Eric Evans' "Model Exploration Whirlpool" [2], is used as an iterative process model. "Domain Prototyping" aims to develop the user interface and the domain model alternately. A domain model is the domain-related source code of a bounded context and is first developed in TypeScript directly in the frontend. A domain model consists of the classes: entities, value objects and domain services. Over time, further classes are added: application services, repositories (initially only as hash maps), and domain events with observer pattern. Acceptance tests are written against the methods of the domain services. The user interface (as Vue.js components) calls the methods of the domain services directly.
Once the domain models in TypesScript are sufficiently mature, they are implemented in the backend with Java, Spring Boot and Postgres.

Realization

An event storming workshop is carried out. However, the identified subdomains from the event storming workshop are too coarse-grained to create a context map. Together with two domain experts, who are also users, the information and work objects they need to make a decision or perform an action are documented using domain stories in the as-is state.
A context map (Fig. 4) is created and the corresponding source code is implemented in TypeScript in the frontend:

- A decision is made to adopt a domain model for the source system in a bounded context, which is shown as a conformist pattern (CF). The purpose of this solution-induced bounded context is to ensure changeability with regard to multiple source systems. For this purpose, the source code relating to The Things Stack Application Server is encapsulated.
- The bounded context "Parser Management" implements the uplink parser for each LoRaWAN communication module connected to a meter. A manufacturer-independent DLMS/COSEM communication profile for LoRaWAN is described in IEC 62056-8-12:2024-01. However, the standard is still in draft stage.
- The bounded context "Meter Data Management" subscribes to the domain event CimMeterReadings, which is generated by the parser. Internally, "Meter Data Management" uses a domain model that is formed from the simplest possible aggregates of the canonical data model of the CIM. According to the taxonomy developed at Fraunhofer IESE [1], the simplest aggregate is the immutable aggregate. The observable state of an immutable aggregate is updated by deleting the aggregate as a whole and creating it as new with a new identifier. In this case, CimMeterReading (Fig. 5) is selected as the immutable aggregate that is generated from the domain event. This immutable aggregate is extended with derived aggregates.
- With the bounded context "Destination System Integration Billing System", the domain model of the billing system is adopted, whose internal domain model differs from CIM, which is represented by the conformist pattern (CF).
- The other bounded contexts "Device Input and Inventory Data", "Device Management", and "Project Creation", are problem-induced bounded contexts whose domain models are specifically adapted to the company. The product master data is managed by the bounded context "Device Product Master Data".

A low-fidelity prototype is developed and attached to the domain model (Fig. 3).

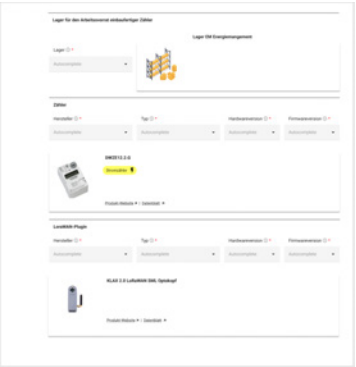

Fig. 3: Example low-fidelity prototype [5]

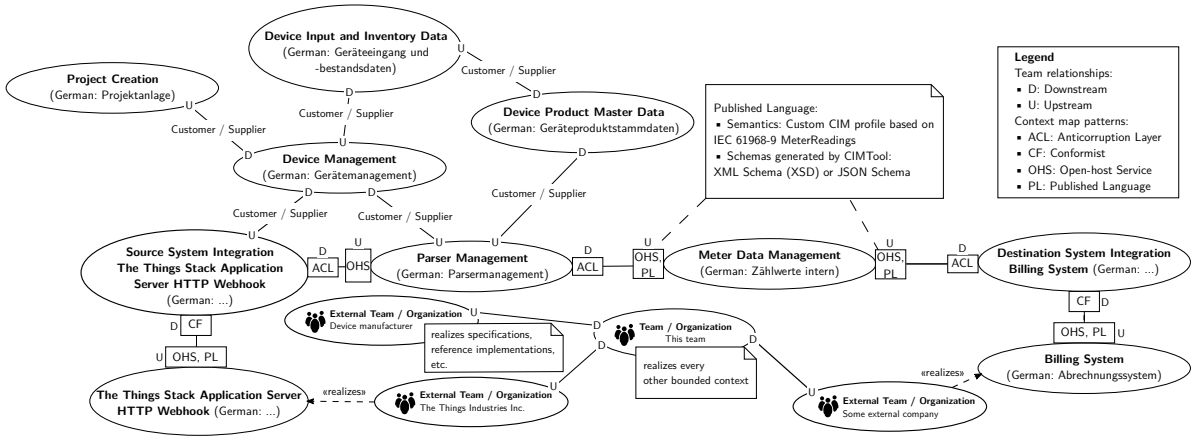

Fig. 4: Context map [5]

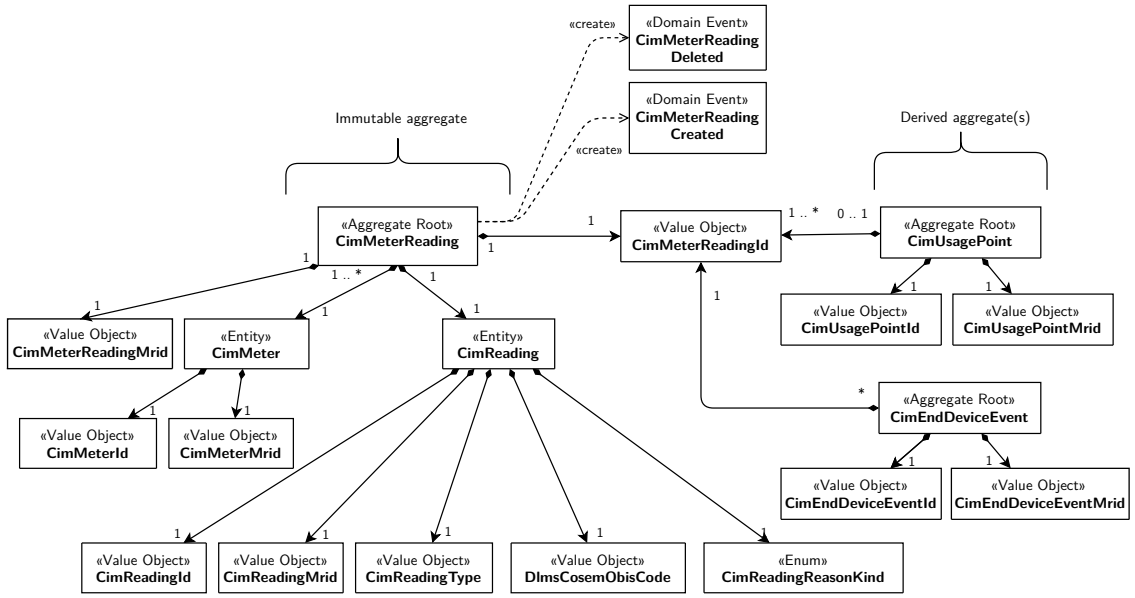

Fig. 5: Derived monotonic state pattern [5]

Evaluation

The results of the evaluation are not yet available. Usability tests are being carried out with the low-fidelity prototypes and static code analysis.

Conclusions and Future Work

The contribution of this work lies in the initial prototypical development of a metering-related back office system. The main innovation compared to the existing low-code implementation lies in the application of methods from domain-driven design and the adoption of CIM, which is currently used in particular by larger transmission and distribution network operators.
Bounded contexts limit the use of specific domain models and differentiate between problem-induced and solution-induced domain models. This is expected to result in high cohesion and low coupling. The appropriate size of the bounded context is relevant, as the messaging between bounded contexts increases the effort. The proposed decomposition of the aggregates enables fine-grained data consistency of the domain model with comparatively little effort.
An evaluation with static code analysis and usability tests has not yet been completed. The results of the evaluation are expected in this bachelor thesis.
Deployment into production will take place after this thesis. So far, it is assumed that the performance of a modular monolith replicated behind an API gateway is sufficient. If a higher load is expected, the Command Query Responsibility Segregation (CQRS) pattern can be implemented with two monoliths. This is less complex than implementing each bounded context in a separate microservice.

References and figures

[1] Susanne Braun and Stefan Deßloch. A Classification of Replicated Data for the Design of Eventually Consistent Domain Models. In *2020 IEEE International Conference on Software Architecture Companion (ICSA-C)*, pages 33–40. IEEE, 2020.

[2] Eric Evans. *Domain-Driven Design: Tackling Complexity in the Heart of Software*. Addison-Wesley, 2004.

[3] Tobias Goeschel. Domain Prototyping or Design Is How It Works. https://youtu.be/gDT5PKIYsT0, 2019.

[4] Michael Knoop, Matthias Littwin, Martin Kesting, and Tobias Ohrdes. Modell zur ökonomischen und ökologischen Bewertung von Gebäudeversorgungsverfahren im Rahmen des Mieterstromgesetzes – Langfassung. https://isfh.de/mieterstrom/, 2018.

[5] Own representation.

[6] Vaughn Vernon. *Implementing Domain-Driven Design*. Addison-Wesley, 2013.

Datengetriebene Berechnung eines teilebasierten Product Carbon Footprints zur Optimierung des Maschinenbetriebs komplexer mechatronischer Systeme

Michael Beyer **Gabriele Gühring**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma TRUMPF SE + Co. KG, Ditzingen

Einleitung

Der Klimawandel stellt eine der größten globalen Herausforderungen dar. Technologische Entwicklungen, insbesondere in der Industrie 4.0, haben zu einer Steigerung der Produktionsmengen und damit auch des Ressourcenverbrauchs und der Emissionen geführt. Um den Treibhausgasausstoß zu reduzieren und nachhaltige Praktiken zu etablieren, müssen Industrie und Gesellschaft ihre Umweltauswirkungen präzise quantifizieren. Insbesondere in der emissionsintensiven Metallverarbeitungsindustrie besteht großes Potenzial zur Verbesserung der Nachhaltigkeit [6]. Der Product Carbon Footprint (PCF) ermöglicht eine detaillierte Erfassung der CO2-Emissionen über den gesamten Lebenszyklus eines Produkts und ist entscheidend für die Bewertung und Reduktion der Umweltauswirkungen.

Motivation

Gesetzliche Maßnahmen wie das Pariser Klimaschutzabkommen und die Corporate Sustainability Reporting Directive fordern die Berichterstattung und Reduktion von Treibhausgasemissionen [5]. Unternehmen wie TRUMPF SE + Co. KG müssen innovative Ansätze entwickeln, um ihre Produktionsprozesse durch präzise Messung und Reduktion von Emissionen nachhaltiger zu gestalten und die gesetzlichen Anforderungen zu erfüllen. Das Greenhouse Gas Protocol (GHG) ist ein weltweit anerkanntes Framework zur Messung und Verwaltung von Treibhausgasemissionen. Es teilt die Emissionen in drei Kategorien (Scopes) ein 1: direkte Emissionen aus eigenen Quellen (Scope 1), indirekte Emissionen aus eingekaufter Energie (Scope 2) und alle anderen indirekten Emissionen entlang der Wertschöpfungskette (Scope 3) [4]. Der PCF umfasst die gesamte Menge an CO2-Emissionen, die während des gesamten Lebenszyklus eines Produkts entstehen, von der Rohstoffgewinnung über die Produktion und Nutzung bis zur Entsorgung. Der PCF basiert auf den Kategorien des GHG-Protokolls, um die Emissionen systematisch zu erfassen und zu quantifizieren. Dadurch ermöglicht der PCF eine detaillierte und standardisierte Bewertung der Umweltauswirkungen eines Produkts und unterstützt Unternehmen dabei, gezielte Maßnahmen zur Reduktion von Treibhausgasemissionen zu ergreifen [3].

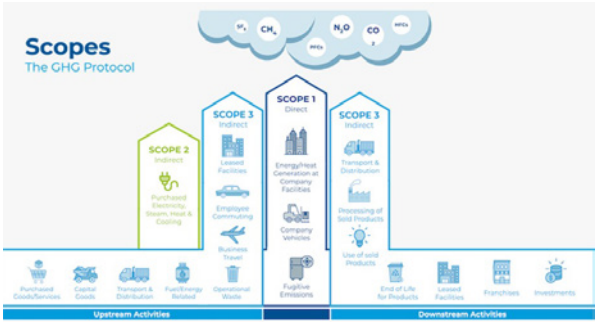

Abb. 1: Drei Scopes der Emissionen nach GHG [4]

Zielsetzung

Diese Arbeit adressiert zwei zentrale Herausforderungen:
Praxisproblem: Unternehmen müssen die Umweltauswirkungen ihrer Produkte präzise quantifizieren und berichten, stehen jedoch vor Herausforderungen aufgrund fehlender standardisierter Methoden und unzureichender Datenqualität. Dies beeinträchtigt die Compliance mit gesetzlichen Vorschriften und die Umsetzung gezielter Emissionsreduktionsmaßnahmen.
Wissenschaftliche Herausforderung: Es bedarf einer methodisch fundierten Grundlage für die präzise und reproduzierbare Berechnung des PCFs auf Teileebene. Die Methode muss zuverlässig CO2-Emissionen quantifizieren, auch bei suboptimaler Datenqualität, und branchenübergreifend anwendbar sein. Das Ziel

dieser Arbeit ist die Entwicklung einer standardisierten Berechnungsmethode für den PCF, die es Unternehmen ermöglicht, ihre Umweltauswirkungen genau zu erfassen und zu verbessern. Hierbei soll der PCF auf Teileebene berechnet werden können, das heißt jedes einzelne produzierte Teil soll einen eigenen PCF bekommen, anstatt wie in den meisten aktuellen Anwendungen, mit Durchschnittwerten und groben Annäherungen zu arbeiten. Ein standardisiertes Vorgehen erleichtert die Vergleichbarkeit und unterstützt den Übergang zu einer nachhaltigeren Produktionsweise.

Vorgehensweise

Diese Arbeit orientiert sich an der Design Research Methodology (DRM) von Blessing und Chakrabarti [1]. Nach einer initialen Literaturrecherche zur Ermittlung des aktuellen Forschungsstandes und der Klärung des Forschungsgegenstandes werden im nächsten Schritt die Anforderungen an eine PCF-Berechnung auf Teileebene identifiziert.

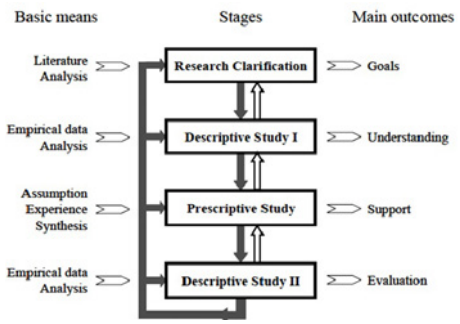

Abb. 2: Phasen der DRM [1]

Anschließend wird eine Methode zur Berechnung des PCF auf Teileebene entwickelt. Diese Methode wird prototypisch implementiert und getestet. Die Methode wird zudem iterativ durch Expertenfeedback und Nutzertests angepasst und optimiert.
Im letzten Schritt wird die entwickelte Methode im Produktentwicklungsprozess validiert. Hierbei wird der PCF als Key Performance Indicator (KPI) verwendet, um die Nachhaltigkeit der von Laserschneidmaschinen hergestellten Teile zu bewerten und zu verbessern. Zur Validierung wird der Ëntwicklungssimulator der Blechkonstruktion"nach Maass et al. genutzt [7]. Dabei wird mittels CAD ein Blechgrill konstruiert Abb. 3. Dieser wird anschließend in einen Produktionsauftrag umgewandelt. Aus diesem Auftrag wird dann der PCF des Teils mittels simulierter Verbrauchsdaten berechnet und in Iterationsschleifen optimiert um das Teil nachhaltiger zu gestalten.

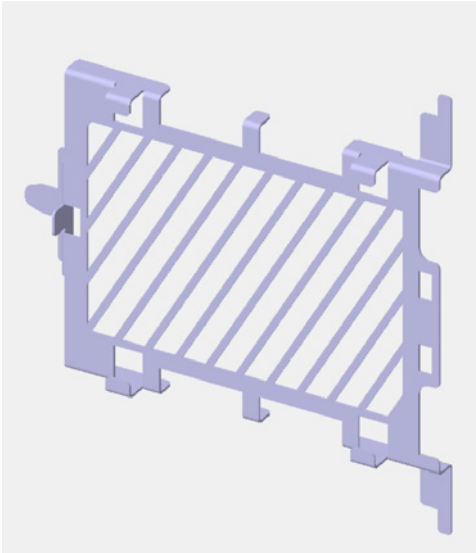

Abb. 3: Beispiel eines während des Entwicklungssimulator konstruierten Teiles [2]

Ergebnisse

Im ersten Durchlauf wurden bereits vielversprechende Ergebnisse erzielt (vgl. Abb. 4), die jedoch in weiteren Durchgängen verifiziert werden müssen. Hierbei konnte der PCF des Blechgitters innerhalb eines Nachmittages und in weniger als 15 Versuchen auf etwa die Hälfte des ursprünglichen Wertes reduziert werden.
Die entwickelte Methode zur Berechnung des PCF auf Teileebene hat gezeigt, dass präzise Emissionsdaten erfasst und analysiert werden können. Die Anwendung des PCF als KPI in der Produktentwicklung hat zu verbesserten nachhaltigen Praktiken geführt. Der Einsatz des Entwicklungssimulators bestätigte die Praxistauglichkeit der Methode und ermöglichte eine realistische Evaluierung unter praxisnahen Bedingungen.

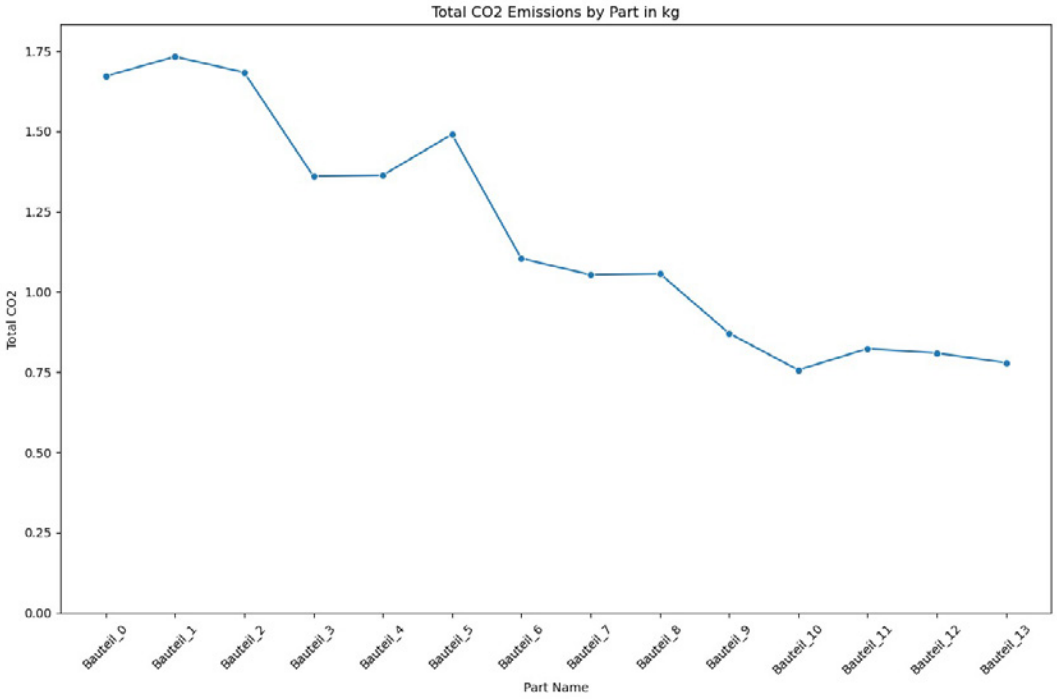

Abb. 4: Ergebnisse Entwicklungssimulator mit PCF als KPI [2]

Schlussfolgerung und Ausblick

Die Arbeit hat gezeigt, dass eine präzise und standardisierte Methode zur Berechnung des PCF nicht nur zur Einhaltung gesetzlicher Vorschriften beiträgt, sondern auch signifikante Verbesserungen in der Nachhaltigkeit von Produktionsprozessen ermöglicht. Zukünftig sollte sich die Forschung auf die Verbesserung der Datenqualität und die Erweiterung der Methode auf andere Industriebereiche konzentrieren. Die vorläufigen Ergebnisse bestätigen, dass eine standardisierte PCF-Berechnungsmethode die Umweltbewertung und -verbesserung in der Metallverarbeitungsindustrie erheblich unterstützen kann. Es wurden jedoch einige Herausforderungen im Zusammenhang mit der Datenqualität und -verfügbarkeit identifiziert, die weiter untersucht werden müssen. Der Entwicklungssimulator hat sich als effektives Werkzeug zur Validierung neuer Methoden erwiesen, welches auch in anderen Kontexten eingesetzt werden kann.

Literatur und Abbildungen

[1] L. T. M. Blessing and A. Chakrabarti. *DRM: A design research methodology.* Springer, 2009.

[2] Eigene Darstellung.

[3] International Organization for Standardization. ISO 14067:2018 Greenhouse gases - Carbon footprint of products - Requirements and guidelines for quantification. *International Standard (Edition 1 - 2018-08)*, 2018.

[4] The World Business Council for Sustainable Development and WOrld Resources Institute. The Greenhouse Gas Protocol. https://ghgprotocol.org/corporate-standard, 2015.

[5] Bundesministerium für Arbeit und Soziales. Corporate Sustainability Reporting Directive (CSRD): Die neue EU-Richtlinie zur Unternehmens-Nachhaltigkeitsberichterstattung im Überblick. https://www.csr-in-deutschland.de/DE/CSR-Allgemein/CSR-Politik/CSR-in-der-EU/Corporate-Sustainability-Reporting-Directive/corporate-sustainability-reporting-directive.html, 2022.

[6] Hoesung Lee et al. *Climate Change 2023: Synthesis Report.* Intergovernmental panel on climate change, 2023.

[7] Ben Maass, Katharina Ritzer, et al. Entwicklungssimulator als Validierungsumgebung für Methoden der Blechkonstruktion. *34. DfX-Symposium 2023*, 2023.

Entwurf eines Zero Trust Implementierungsleitfadens für Großkonzerne

Maxim Bickel **Martin Mink**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einführung & Motivation

Zero Trust ist ein modernes Sicherheitskonzept, das darauf basiert, dass kein Benutzer oder Gerät, unabhängig davon, ob es sich innerhalb oder außerhalb des Unternehmensnetzwerks befindet, automatisch als vertrauenswürdig angesehen wird. Stattdessen wird jeder Zugriff überprüft und authentifiziert, bevor der Zugriff auf Ressourcen gewährt wird. Satya Nadella, CEO von Microsoft, betont, dass Zero Trust nicht nur eine technologische Architektur ist, sondern ein Mindset, das Unternehmen in einer perimeterlosen Welt schützt. Traditionelle Sicherheitskonzepte basieren auf klaren Grenzen, die das Unternehmensnetzwerk schützen, wobei interne Ressourcen als vertrauenswürdig gelten. Dieses Modell war jahrzehntelang Standard, wird aber durch die zunehmende Vernetzung und den Wunsch nach hybriden Arbeitsmodellen überholt.

Die Pandemie von 2020 bis 2021 veränderte die Arbeitswelt erheblich: Vor der Krise arbeiteten lediglich 4% der Beschäftigten von zu Hause, während des Lockdowns stieg diese Zahl auf 27% und stabilisierte sich später bei 24%. [4] Eine Studie von 2023 zeigt, dass 38% der Beschäftigten zumindest teilweise im Homeoffice arbeiten, wobei 65% ein hybrides Arbeitsmodell bevorzugen und 22% dauerhaft im Homeoffice arbeiten möchten. [2] Diese Entwicklung stellt die perimeterbasierten Sicherheitsmodelle vor Herausforderungen, da sie die Netzwerkzugriffe außerhalb der Unternehmensgrenzen nur mit erheblichem Mehraufwand ermöglichen können. Eine Statistik von 2024 zeigt zudem, dass die Cyberkriminalität in Deutschland von 35.000 Fällen im Jahr 2007 auf 135.000 Fälle im Jahr 2023 gestiegen ist, wobei die geschätzten Schäden 200 Milliarden Euro betragen. Angriffe umfassen Datenausspähung, Computerbetrug und Sabotage. [3] Zero Trust, basierend auf dem Prinzip der minimalen Rechte, bietet hier einen modernen Ansatz, der interne und externe Zugriffe absichert und die möglichen Schäden eines Angriffs minimiert.

Zielstellung

Die Zielsetzung dieser Masterthesis ist die Entwicklung eines branchenunabhängigen Leitfadens zur Umsetzung der Zero Trust Strategie für Großkonzerne. Der Leitfaden soll dabei unterstützen, notwendige Schritte, Risiken und potenzielle Probleme vor einer Unternehmensumstrukturierung zu identifizieren. Meilensteine werden definiert und in spezifische Umsetzungsphasen unterteilt. Dies soll gewährleisten, dass zukünftige Umstrukturierungen möglichst erfolgreich stattfinden und Ressourcenaufwände realistisch abschätzen werden können. Hierfür werden branchenunabhängige Anforderungen und Herausforderungen, wie rechtliche, technische oder organisatorische Aspekte, identifiziert und in den Leitfaden integriert. Bereits dokumentierte Umstrukturierungen, wie die Einführung des BeyondCorp Sicherheitsmodells von Google [5], werden analysiert und deren Erkenntnisse einbezogen. Als praktische Umsetzung soll eine Fallstudie, welche die Umstrukturierung des Identity and Access Managements (IAM) anhand einer typischen Großkonzerninfrastruktur als Ziel verfolgt, beispielhaft umgesetzt werden.

Des Weiteren wird ein Überblick der aktuellen Zero Trust Modelle geliefert, um Herausforderungen und Vorteile der jeweiligen Modelle aufzuzeigen. Die Funktionsweise und eine kritische Bewertung der Modelle im Hinblick auf ihre Tauglichkeit für Großkonzerne werden ebenfalls dargestellt. Zudem werden die Grenzen und Limitationen des Zero Trust Ansatzes erläutert. Diese Erkenntnisse sollen Großkonzernen einen realistischen Überblick der aktuellen Marktsituation bieten und ihnen ermöglichen, die bestmögliche Entscheidung für ihr Unternehmen zu treffen. Die Einführung von Zero Trust ist ein langer und ressourcenaufwändiger Prozess, der sorgfältig geplant und strukturiert werden muss. Die Ergebnisse dieser Masterthesis sollen Unternehmen dabei unterstützen, diesen Prozess erfolgreich zu gestalten und langfristig davon zu profitieren.

Integrationsmodell des Bundesamts für Sicherheit in der Informationstechnik

Das Bundesamt für Sicherheit in der Informationstechnik (BSI) hat ein Integrationsmodell entwickelt, das dabei unterstützt, Zero Trust-Prinzipien in bestehende IT-Infrastrukturen zu integrieren. Dieses Modell kann als Grundlage eines möglichen Integrationsleitfadens verstanden werden und ist in fünf themenspezifische Säulen unterteilt. Jede Säule enthält spezifische Funktionen, die in drei Reifegraden (klassisch, fortschrittlich und ideal) umgesetzt werden können.
Die erste Säule, welche als Identität bezeichnet wird, konzentriert sich auf die Authentifizierung und Autorisierung von Nutzern und Geräten. Einfache Authentifizierungsmethoden wie Benutzernamen und Passwörter werden im klassischen Reifegrad verwendet. Der fortschrittliche Entwicklungsstand führt mehrstufige Authentifizierung (MFA) für sensible Ressourcen ein, während der ideale Reifegrad eine durchgängige MFA für alle Zugriffe und kontinuierliche Überprüfung beinhaltet. Die Gerät Säule legt den Fokus auf die Sicherheitsbewertung und das Management der Geräte, die auf die IT-Ressourcen zugreifen. Während im klassischen Reifegrad die Bewertung der Gerätesicherheit manuell erfolgt, werden im fortschrittlichen Reifegrad automatisierte Sicherheitsprüfungen durchgeführt und der Zustand der Geräte kontinuierlich überwacht. Der ideale Reifegrad umfasst dynamische, kontextabhängige Sicherheitsüberprüfungen und sofortige Maßnahmen bei Verstößen. Des Weiteren befasst sich die Netzwerk-Säule mit der Absicherung der Netzwerkinfrastruktur. Im klassischen Ansatz werden statische Sicherheitsrichtlinien und segmentierte Netze verwendet. Dynamische Netzwerksegmentierung und Richtlinien basierend auf aktuellen Bedrohungsinformationen werden im fortschrittlichen Reifegrad implementiert. Und der idealen Reifegrad wird erreicht, wenn eine vollständige Mikrosegmentierung und dynamische Anpassung der Netzwerksicherheit in Echtzeit vorliegt. Die vierte Säule, Anwendung, behandelt die Sicherheit der innerhalb der Organisation genutzten Anwendungen. Im klassischen Entwicklungsstand werden Anwendungen isoliert und grundlegende Sicherheitsrichtlinien angewendet. Der fortschrittliche Reifegrad integriert sicherheitsrelevante Maßnahmen in den gesamten Anwendungslebenszyklus. Der ideale Reifegrad beinhaltet kontinuierliche Überwachung und automatische Anpassung der Sicherheitsmaßnahmen basierend auf dem Anwendungskontext und der aktuellen Bedrohungslage. Abschließend konzentriert sich die Daten-Säule auf den Schutz der Daten. Im klassischen Reifegrad erfolgt der Schutz der Daten durch statische Zugriffsrichtlinien und grundlegende Verschlüsselung. Innerhalb der fortschrittlichen Fortschrittsstufe werden dynamische Zugriffsrichtlinien implementiert, die auf Echtzeit-Bewertungen basieren. Der ideale Entwicklungsstand umfasst eine kontinuierliche Überwachung und Analyse der Datenzugriffe, um sofort auf verdächtige Aktivitäten reagieren zu können. [1]
Zusätzlich zu den fünf themenspezifischen Säulen gibt es zwei Querschnittsfunktionen: Detektion & Reaktion und Anforderungen an Verschlusssachen (VS). Die Funktion Detektion & Reaktion ist in allen Säulen präsent und umfasst die Erkennung von Sicherheitsvorfällen und die entsprechende Reaktion darauf. Im klassischen Ansatz erfolgt die Detektion und Reaktion größtenteils manuell. Der fortschrittliche Reifegrad nutzt automatisierte Detektionsmechanismen und vordefinierte Reaktionspläne. Im idealen Reifegrad werden fortschrittliche Technologien wie maschinelles Lernen eingesetzt, um Bedrohungen in Echtzeit zu erkennen und zu neutralisieren. Die Anforderungen an VS beziehen sich auf die Einhaltung gesetzlicher und regulatorischer Anforderungen sowie interner Sicherheitsrichtlinien. Im klassischen Reifegrad werden diese Anforderungen manuell umgesetzt. Der fortschrittliche Reifegrad integriert automatisierte Compliance-Checks und regelmäßige Überprüfungen. Der ideale Reifegrad umfasst eine vollständige Automatisierung und kontinuierliche Anpassung der Compliance-Maßnahmen an aktuelle Anforderungen. [1]

Aktueller Stand & Ausblick

Der aktuelle Stand der Masterarbeit beschränkt sich auf eine extensive Literaturrecherche, welche bereits einige vielversprechende Erkenntnisse zur Umsetzung eines Implementierungsleitfadens herausarbeiten konnte. Ebenso konnte bereits ein Überblick der unterschiedlichen Zero Trust Modelle herausgearbeitet werden, welcher nun noch aufgearbeitet werden muss. Das technische und regulatorische Umfeld für Großkonzerne sollte nachfolgend analysiert und eingeordnet werden, wodurch die bereits gewonnenen Erkenntnisse strukturiert werden können. Anhand dieser Strukturierung können einzelne Phasen des Implementierungsleitfadens herausgearbeitet und mittels der Analyse von bereits dokumentierten Umstrukturierungen überprüft werden. Auf Basis dieser Erkenntnisse können mögliche Risiken und potenzielle Probleme aufgezeigt und innerhalb des Leitfadens eingebettet werden. Abschließend muss mittels der Fallstudie der Leitfaden validiert werden.

Literatur und Abbildungen

[1] BSI BSI et al. Positionspapier Zero Trust 2023. *Positionspapier Zero Trust 2023*, 2023.

[2] GmbH Statista and GmbH Appinio. Wie würdest du in Zukunft am liebsten arbeiten? https://de.statista.com/statistik/daten/studie/1296962/umfrage/umfrage-arbeitsplatz-der-zukunft/, 2022.

[3] Statista GmbH Statista. Polizeilich erfasste Fälle von Cyberkriminalität in Deutschland von 2007 bis 2023. https://de.statista.com/statistik/daten/studie/295265/umfrage/polizeilich-erfasste-faelle-von-cyberkriminalitaet-im-engeren-sinne-in-deutschland/, 2024.

[4] Hans-Böckler Stiftung. Anteil der im Homeoffice arbeitenden Beschäftigten in Deutschland vor und während der Corona-Pandemie 2020 und 2021. https://de.statista.com/statistik/daten/studie/1204173/umfrage/befragung-zur-homeoffice-nutzung-in-der-corona-pandemie/, 2021.

[5] Rory Ward and Betsy Beyer. BeyondCorp: A New Approach to Enterprise Security. *login:*, 2014.

Evaluierung und Potenzialanalyse eines PIM-Systems als Lösungsansatz zur Erfüllung von Marktanforderungen im Produktdatenmanagement

Fabian Brummer **Thomas Rodach**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Pilz GmbH & Co. KG, Ostfildern

Einleitung

In einer Zeit, die von Digitalisierung geprägt ist, werden Produktdaten zunehmend als Öl des Kaufprozesses bezeichnet [4]. Produktinformationen findet man häufig verstreut in verschiedenen Abteilungen in diversen Versionen, Sprachen und Formaten, was die internen Abläufe verkompliziert und die externe Kommunikation anfällig für Fehler macht. Problematisch kann es insbesondere sein, wenn veraltete oder inkonsistente Daten verwendet werden [2]. Immerhin sind eine aussagekräftige Produktbeschreibung sowie korrekte Produktdaten für 28% der Befragten einer Studie aus Statista das wichtigste Qualitätsmerkmal für einen Online-Shop [5]. Aus diesem Grund greifen jährlich immer mehr Unternehmen zu PIM-Systemen, was sich nun von einer optionalen Ergänzung zu einer unverzichtbaren Notwendigkeit entwickelt und der Marktanteil sich in allen Ländern deutlich vergrößert hat [7].

Zielsetzung der Arbeit

Im Rahmen der Abschlussarbeit wird das Ziel verfolgt, das Potenzial eines Produktinformationsmanagement-Systems (PIM) zur Optimierung des Produktdatenmanagements bei der Pilz GmbH & Co. KG zu evaluieren. Dabei sollen die aktuellen Herausforderungen und Bedürfnisse im Produktdatenmanagement analysiert und spezifische Anforderungen ermittelt werden, die von der aktuellen Systemlandschaft nicht erfüllt werden. Zudem wird eine umfassende Marktanalyse der verfügbaren PIM-Systeme durchgeführt, um geeignete Anbieter zu evaluieren und zu vergleichen. Schließlich soll eine Implementierungsstrategie entwickelt werden, die eine detaillierte Roadmap für die Einführung eines PIM-Systems umfasst. Die Arbeit zielt darauf ab, fundierte Handlungsempfehlungen für eine erfolgreiche Implementierung eines PIM-Systems zu liefern und damit zur Verbesserung der Datenqualität, der Effizienz interner Prozesse und zur Steigerung der Kundenzufriedenheit beizutragen.

Produktinformationsmanagement-System (PIM)

Ein PIM-System ist eine IT-gestützte Lösung, die speziell entwickelt wurde, um sämtliche technischen und vertriebsrelevanten Informationen zu den Produkten und Dienstleistungen eines Unternehmens zentral zu erfassen, zu verwalten und zu pflegen. Darüber hinaus erleichtert ein PIM-System den effizienten internen und externen Austausch dieser Daten in medienneutraler Form, was die Konsistenz und Aktualität der Informationen sicherstellt und die Kommunikation und Zusammenarbeit sowohl innerhalb des Unternehmens als auch mit externen Partnern und Kunden verbessert [2].

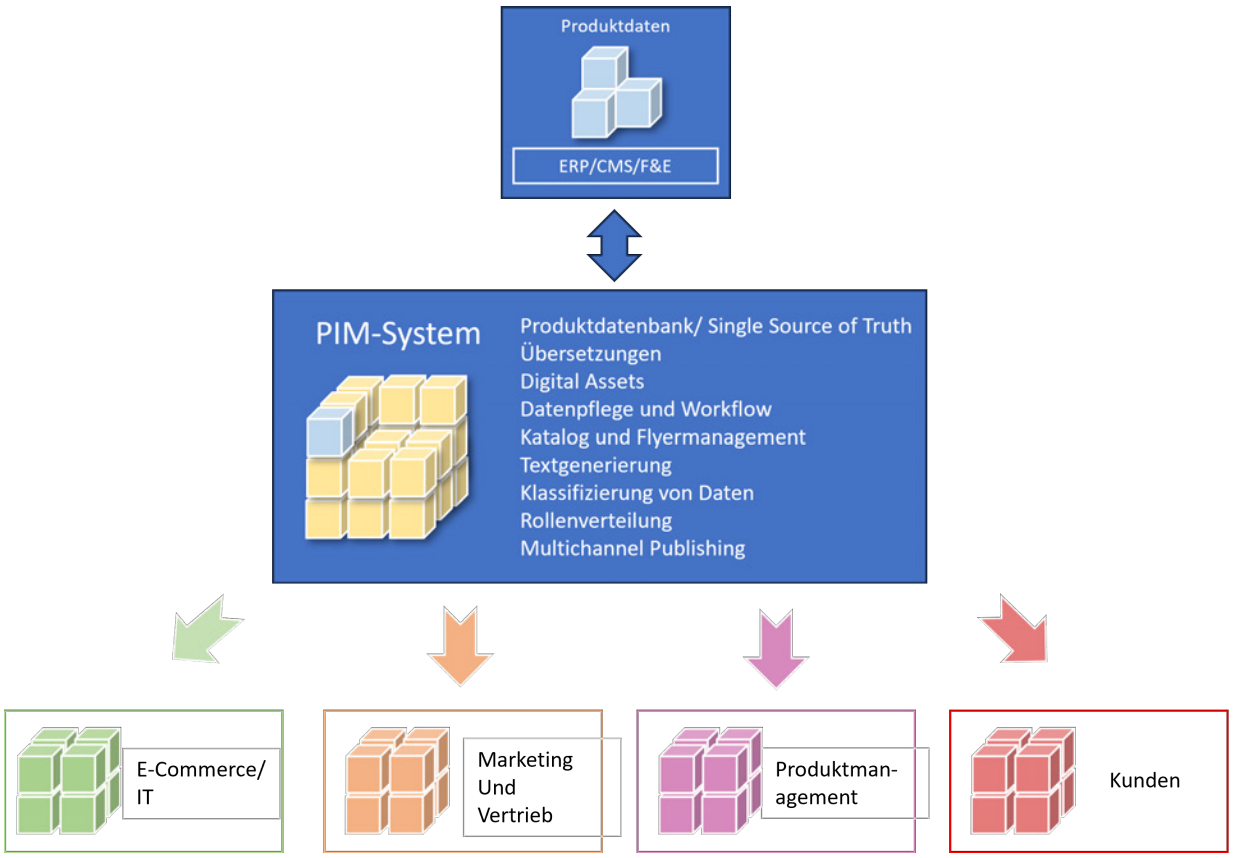

Abb. 1: Funktionsweise eines PIM-Systems [3]

Die Abbildung 1 veranschaulicht die zentrale Rolle eines PIM-Systems innerhalb eines Unternehmens. Oben in der Abbildung sind die Produktdatenquellen dargestellt, die aus verschiedenen Systemen wie Enterprise Resource Planning System (ERP) und Content-Management-System (CMS) stammen. Darüber hinaus liefert die Forschung und Entwicklung (F&E) wichtige Daten, die meist aus dem Produktdatenmanagement (PDM) stammen, an das PIM-System. Dieses System ermöglicht es, CAD- und 2D/3D-Zeichnungen, technische Produktinformationen sowie weitere technische Dokumente, wie Betriebsanleitungen, die während des Produktentwicklungsprozesses erstellt werden, auf elegante und effiziente Weise zu speichern [1]. Während das PDM-System hauptsächlich technische Daten und Dokumente für die interne Produktentwicklung und Produktion verwaltet, konzentriert sich das PIM-System auf die zentrale Verwaltung und Bereitstellung vertriebs- und marketingrelevanter Produktinformationen für externe Kanäle und Kunden [6] [2]. Diese Daten fließen in das PIM-System, das als zentrale Datenbank fungiert und eine *Single Source of Truth* für alle produktbezogenen Informationen bereitstellt. Das PIM-System ermöglicht die Verwaltung und Pflege der Daten durch Funktionen wie Übersetzungen, die Verwaltung von Digital/Media Assets, das Datenpflege- und Workflow-Management, Katalog- und Flyermanagement, Textgenerierung, Klassifizierung von Daten, Rollenverteilung sowie Multichannel Publishing. Diese zentrale Plattform gewährleistet, dass alle produktbezogenen Informationen konsistent und aktuell sind [1] [2]. Unterhalb des PIM-Systems sind die verschiedenen Nutzergruppen dargestellt, die von den bereitgestellten Informationen profitieren. Zu diesen Nutzergruppen gehören die E-Commerce- und IT-Abteilungen, die die Daten für Online-Shops und IT-Anwendungen verwenden. Auch Marketing und Vertrieb nutzen die Informationen intensiv für die Erstellung von Werbematerialien und die Entwicklung von Verkaufsstrategien. Das Produktmanagement greift auf die zentralisierten Daten zu, um sie effektiv zu verwalten und stets auf dem neuesten Stand zu halten. Kunden profitieren von präzisen und aktuellen Produktinformationen, was ihre Zufriedenheit erheblich steigert. Zusammengefasst zeigt die Abbildung, wie ein PIM-System als zentrale Plattform die Produktinformationen aus verschiedenen Quellen integriert und an unterschiedliche Abteilungen und Nutzergruppen im Unternehmen verteilt. Dies sorgt für konsistente und aktuelle Daten, verbessert die Effizienz und unterstützt fundierte Entscheidungsprozesse [1] [2].

Umsetzung und Ausblick

Im Rahmen der Forschung zu dieser Bachelorarbeit werden qualitative Experteninterviews mit Fachexperten aus den jeweiligen Abteilungen durchgeführt, um detaillierte Einblicke in die verwendete Software und deren Vor- und Nachteile zu gewinnen. Ziel dieser Interviews ist es, die Herausforderungen zu identifizieren, mit denen die Experten konfrontiert werden, sowie Verbesserungspotenziale aufzuzeigen. Darüber hinaus wird eine ausführliche Roadmap entwickelt siehe Abbildung 2, die als Strategie zur System- und Dienstleisterselektion von PIM-Systemen dient.

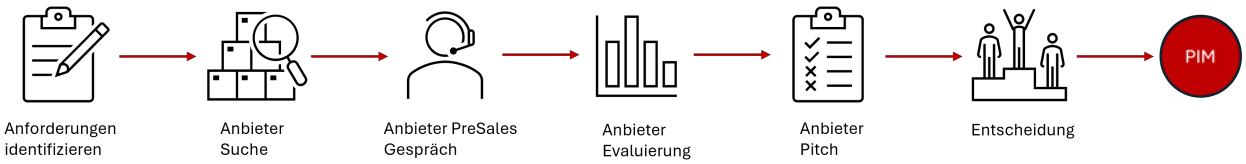

Abb. 2: Roadmap für eine Strategie zur System- und Dienstleisterselektion von PIM-Systemen [3]

Diese Roadmap dient dazu, den Auswahlprozess für das optimale PIM-System zu strukturieren und sicherzustellen, dass die Effizienz und Datenqualität im Unternehmen maximiert werden. Zu Beginn werden die spezifischen Anforderungen und Kriterien für das PIM-System präzise definiert. Daraufhin folgt eine systematische Suche nach potenziellen Anbietern, die diesen Anforderungen entsprechen. In den anschließenden PreSales-Gesprächen wird eine detaillierte Evaluierung der Anbieter durchgeführt. Hierbei kommt eine Nutzwertanalyse zum Einsatz, um die Anbieter objektiv und umfassend anhand festgelegter Kriterien zu bewerten und zu vergleichen. Die vielversprechendsten Anbieter haben dann die Gelegenheit, ihre Lösungen in detaillierten Präsentationen vorzustellen und ihre Leistungsfähigkeit und Eignung zu demonstrieren. Basierend auf diesen methodischen Schritten wird schließlich der geeignetste Anbieter ausgewählt und das PIM-System implementiert.

Literatur und Abbildungen

[1] Jorij Abraham. *Product information management*. Springer International Publishing Switzerland, 2014.

[2] Lars. Binckebanck, R. Elste, A Haas, et al. *Digitalisierung im Vertrieb: Strategien zum Einsatz neuer Technologien in Vertriebsorganisationen*. Springer Fachmedien Wiesbaden, 2023.

[3] Eigene Darstellung.

[4] Marina Eichinger, Marc Knoppe, et al. *Produktinformationen als innovativer Rohstoff der Digitalisierung*. Springer Fachmedien Wiesbaden, 2022.

[5] KPMG KPMG. Online-Shopping: Einkaufsverhalten - wer kauft was, wann und wie? https://de.statista.com/statistik/daten/studie/1233451/umfrage/erfolgskriterien-von-online-shops-aus-konsumentensicht-in-oesterreich/, 04 2021.

[6] Thomas Mechlinski. *Schnittstellen zu anderen Systemen, Datenaustausch*. Springer Berlin Heidelberg, 2021.

[7] Market Research Polaris. Global Product Information Management (PIM) Market Size, Share Analysis Report. https://www.polarismarketresearch.com/industry-analysis/product-information-management-market, 11 2022.

Analyse und Bewertung der Java Frameworks Spring Boot und Quarkus

Fotini Chatzi **Jürgen Koch**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma pep.digital GmbH, Esslingen

Einleitung

In der heutigen schnelllebigen Welt der Softwareentwicklung suchen Unternehmen ständig nach Möglichkeiten, ihre Anwendungen effizienter, skalierbarer und benutzerfreundlicher zu gestalten. Spring Boot und Quarkus gehören zu den führenden Java-Frameworks, die diese Herausforderung angehen. Diese Frameworks ermöglichen es Entwicklern, moderne Anwendungen zu erstellen, die den Anforderungen der heutigen digitalen Wirtschaft entsprechen. Das Ziel dieser Arbeit ist es, die beiden Frameworks Spring Boot und Quarkus im Detail zu analysieren und zu bewerten. Dabei werden die Leistungsfähigkeit, Benutzerfreundlichkeit und Eignung für Cloud-native Anwendungen untersucht. Da diese Faktoren in der heutigen Softwareentwicklung von entscheidender Bedeutung sind, liegt ein besonderes Augenmerk darauf, die Frameworks hinsichtlich ihrer Startzeit, Speichernutzung und Skalierbarkeit miteinander zu vergleichen. Die Benutzerfreundlichkeit und die Entwicklungserfahrung dieser Frameworks sind weitere bedeutende Aspekte in dieser Studie. Durch diese Analyse und Bewertung von Spring Boot und Quarkus soll Entwicklern und Entscheidungsträgern geholfen werden, fundierte Entscheidungen bei der Auswahl des geeigneten Frameworks für ihre Projekte zu treffen.

Einfürung zu Spring Boot

Spring Boot, ein Projekt der Spring-Community, hat sich zu einem gängigen Standard für die Entwicklung von Unternehmensanwendungen und Microservices entwickelt. Es bietet eine vielfältige Auswahl an Werkzeugen und Bibliotheken, die die Komplexität der Anwendungsentwicklung verringern können. Durch Spring Boot können sich Entwickler auf die Geschäftslogik konzentrieren, ohne sich mit technischen Details auseinandersetzen zu müssen, da es viele Konfigurationsaufgaben automatisiert und eine robuste Infrastruktur bereitstellt.

Spring Boot zeichnet sich durch mehrere Hauptmerkmale aus, die es besonders benutzerfreundlich und effizient machen. Die Auto-Konfiguration ist eines dieser Merkmale, bei der Spring Boot die Konfiguration automatisch anhand der im Klassenpfad vorhandenen Bibliotheken vornimmt. Auf diese Weise entfällt der manuelle Konfigurationsufwand.
Der Spring Initializr stellt ein weiteres bedeutendes Werkzeug dar. Mit dem Spring Initializr können Entwickler schnell und unkompliziert neue Spring Boot-Projekte erstellen. Das bevorzugte Build-Tool wie Maven oder Gradle sowie die grundlegenden Projektmetadaten wie Gruppen-ID, Artefakt-ID und Version können mit nur wenigen Klicks festgelegt werden. Darüber hinaus ermöglicht der Spring Initializr das Hinzufügen unterschiedlicher Abhängigkeiten, die für das Projekt benötigt werden, und generiert eine vollständige Projektstruktur, die alle notwendigen Konfigurationen und Bibliotheken enthält.

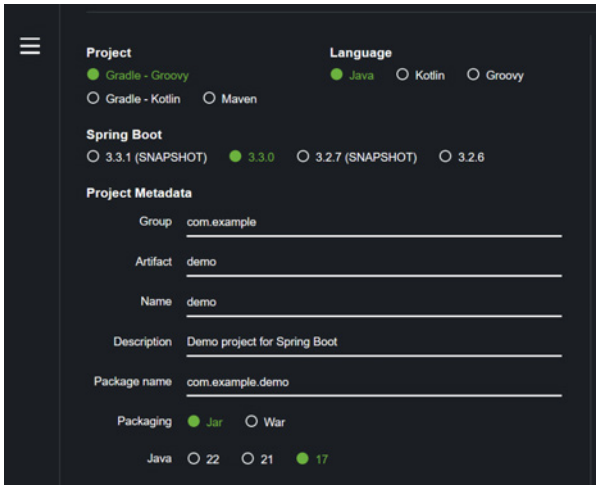

Abb. 1: Spring Initializr [2]

Darüber hinaus verfügt Spring Boot über einen Actuator, der verschiedene Produktionsüberwachungs- und Managementfunktionen bereitstellt. Dazu gehören

Metriken, Überwachungsendpunkte sowie die Möglichkeit, den Zustand der Anwendung in Echtzeit zu überwachen.

Einführung zu Quarkus

Quarkus ist ein Java-Framework, das speziell entwickelt wurde, um die Entwicklung von Cloud-native Anwendungen zu optimieren. Es wurde von Red Hat entwickelt und zielt darauf ab, die Lücke zwischen traditioneller Unternehmensentwicklung und den Anforderungen moderner Cloud-Umgebungen zu schließen. Ein zentrales Merkmal ist seine Fähigkeit, Java für Containerumgebungen zu optimieren. Quarkus kann mithilfe von GraalVM Anwendungen in native ausführbare Dateien umwandeln, was die Startzeit reduzieren und den Speicherbedarf minimieren kann. Des Weiteren bietet Quarkus eine benutzerfreundliche Umgebung, die Funktionen wie Live-Coding beinhaltet. Mit Live-Coding können Entwickler Änderungen am Code in Echtzeit sehen, ohne die Anwendung neu starten zu müssen. Sowohl imperatives als auch reaktives Programmieren werden von Quarkus unterstützt, wodurch Entwickler skalierbare und reaktive Anwendungen entwickeln können. Entwickler können mithilfe des Quarkus-Anwendungsstarters neue Projekte schnell und effizient erstellen. Er bietet eine einfache Möglichkeit, grundlegende Projektmetadaten wie Gruppen-ID, Artefakt-ID und Version festzulegen und ermöglicht die Auswahl des bevorzugten Build-Tools wie Maven oder Gradle. Außerdem können spezifische Erweiterungen für die Anwendung integriert werden, sodass mit wenigen Klicks eine vollständige Projektstruktur erstellt werden kann, die als Basis für die weitere Entwicklung dient.

Abb. 2: Quarkus Starter [3]

Architektur der Anwendungen und Buchungsszenario

Beide Frameworks werden anhand der Architektur von vier Anwendungen sowie eines spezifischen Buchungsszenarios untersucht, um einen praxisnahen Vergleich zwischen Spring Boot und Quarkus durchzuführen. Hierfür eignet sich ein Online-Buchladen ideal, um die Fähigkeiten und Effizienz der Frameworks in realitätsnahen Anwendungsszenarien zu testen. Dieses Szenario umfasst eine Vielzahl von CRUD-Funktionen (Create, Read, Update, Delete) für die Entitäten: Buch, Kunde und Bestellung.

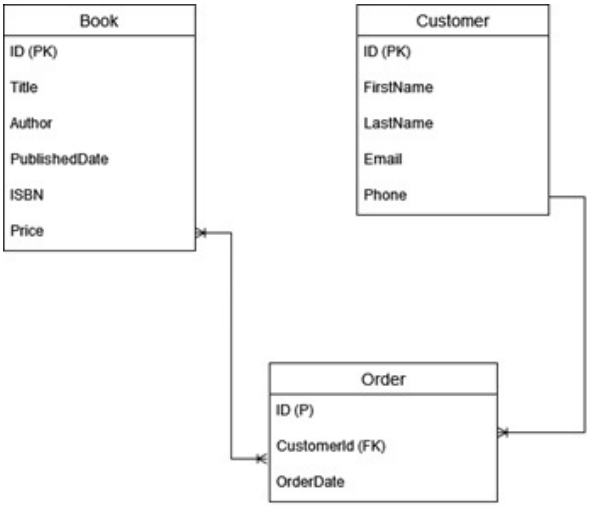

Abb. 3: Entity-Relationship Diagramm [1]

Durch diese komplexe Datenstruktur ist es möglich, die Fähigkeiten des Object-Relational Mapping (ORM) von Spring Boot und Quarkus zu untersuchen. Die Implementierung von RESTful Services ist ein weiterer wesentlicher Bestandteil des Szenarios, um die verschiedenen Endpunkte für Bücher, Kunden und Bestellungen zu verwalten. Ein weiterer wichtiger Aspekt ist die Leistungsbewertung in Szenarien, in denen Non-Blocking I/O und hohe Performance entscheidend sind, beispielsweise bei der Suche nach Büchern oder der Bearbeitung von Aufträgen. Auf diese Weise ist es möglich, die Asynchronität und die Performance-Optimierung von Spring Boot und Quarkus zu bewerten. Das Buchungsszenario ist somit ausreichend komplex, um die Effektivität der Integrationstest-Tools und -Methoden von Spring Boot und Quarkus zu bewerten. Durch die Untersuchung dieser Aspekte wird eine umfassende Bewertung der Vor- und Nachteile von Spring Boot und Quarkus in verschiedenen Bereichen der Anwendungsentwicklung ermöglicht.

Ausblick

In zukünftigen Untersuchungen zu Spring Boot und Quarkus sollten mehrere Bereiche weiter untersucht werden. Hierzu gehören eine detaillierte Analyse der Startzeiten und Speichernutzung unter verschiedenen Lastbedingungen sowie die Leistung der Frameworks bei Datenbankoperationen und der Implementierung von RESTful Services.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Spring Start. Spring Initializr. https://start.spring.io/, 2024.

[3] Quarkus Starter. Starter Quarkus. https://code.quarkus.io/, 2024.

Nutzung von Differenziellen Updates für High-Performance-Computer im Fahrzeug

Cafer Cicek **Rainer Keller**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Vector Informatik GmbH, Stuttgart

Einleitung

Software-Defined Vehicles (SDV) sind mit fortschrittlichen Steuergeräten wie High Performance Computer (HPC) ausgestattet. Diese Steuergeräte enthalten komplexe Software (z. B. ADAS-Algorithmen) und erfordern regelmäßige Updates. Diese Updates können in kurzen Zeitintervallen angefordert werden und können große Update-Pakete umfassen. Je größer die Update-Pakete sind, desto höher sind die zusätzlichen Kosten für den Original Equipment Manufacturer (OEM). Um die Größe der Update-Pakete zu reduzieren, kann die Delta-Update-Methode verwendet werden.

Grundlagen

In der Regel werden Updates für SDV über drahtlose Netzwerke mittels „Over-The-Air“ (OTA) Technologie durchgeführt. Durch diese Technologie können Fahrzeuge jederzeit und überall aktualisiert werden, so dass Sicherheitspakete, Leistungsverbesserungen und neue Funktionen stets aktuell zur Verfügung stehen.

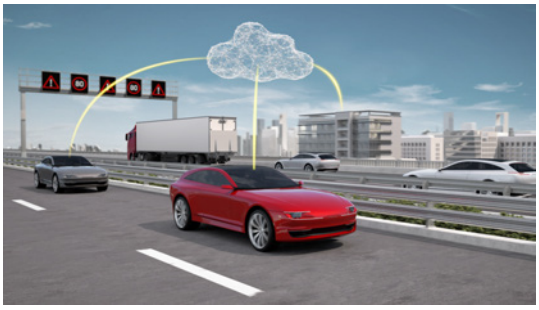

Abb. 1: Automotive Over-The-Air [2]

Die AUTOSAR Adaptive Platform ist ein Standard für moderne Steuergeräte, um die unabhängige Entwicklung zu erleichtern. Der Service „Update Configuration Management“ (UCM) dieser Plattform ermöglicht die standardisierte Ausführung von Installationen, Updates und Deinstallationen. [3]

In dieser Arbeit wird die AUTOSAR Adaptive Platform auf dem POSIX-OS QNX aufgebaut. QNX ist ein Echtzeit-Mikrokernel-Betriebssystem. Im Mikrokernel-Betriebssystem QNX ist jede Anwendung, jeder Treiber und jedes Dateisystem in einem eigenen Adressraum außerhalb des Kernels isoliert. Der Ausfall einer Komponente führt nicht zum Ausfall anderer Komponenten oder des Kernels. [5]

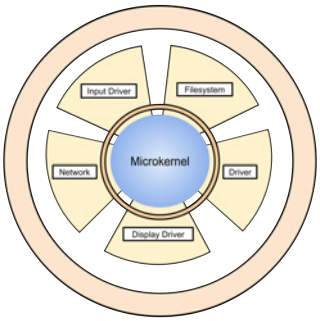

Abb. 2: Microkernel Architektur [1]

Ein weiteres wichtiges Merkmal für Fahrzeugupdates ist die Suspend-Resume-Funktion. Wenn der Update-Prozess unterbrochen wird, muss das System in der Lage sein, den Download oder das Update an der Unterbrechungsstelle fortzusetzen.

Problemdarstellung

Bei Over-The-Air (OTA) spielt die Größe der zu übertragenden Daten eine wichtige Rolle, weil große Daten höhere Kosten für die OEMs bedeuten. Kompression und Delta-Updates sind zwei gängige Methoden zur Reduzierung der Datenmenge.

- Kompression: Die gesamten neuen Daten werden mit entsprechenden Komprimierungsalgorithmen verkleinert und an das Steuergerät übertragen, wobei die alte Version vollständig durch die neue Version ersetzt.

- Delta-Update: Die Unterschiede zwischen der alten und der neuen Version werden erzeugt. Diese Delta-Daten werden an das Steuergerät übertragen. Das Steuergerät wendet dann einen Patch mit den Delta-Daten auf die bestehende Version an, um eine neue Version zu erhalten.

Obwohl Delta-Updates nützlich sind, gibt es spezifische Herausforderungen und Einschränkungen. Es ist möglich, dass die Änderung im Quellcode und das erzeugte Delta nicht immer proportional sind, was dazu führt, dass die Kompression der neuen Version kleiner ist als das erzeugte Delta. Darüber hinaus erzeugen die Algorithmen Delta-Daten in unterschiedlichen Maßstäben und benötigen zum Teil erhebliche Zeit- und Speicherressourcen.

Verfahren für Differenzielle Software-Updates

Delta-Algorithmen wurden ursprünglich in den 1960er Jahren entwickelt, um Textdateien zu vergleichen und Änderungen zwischen verschiedenen Versionen zu identifizieren. Später wurden Delta-Algorithmen weiterentwickelt, um auch komplexere Datentypen wie Binärdateien zu unterstützen. Durch den Vergleich von Blöcken konnten Unterschiede zwischen Text- und Binärdateien effizienter erkannt und Delta-Daten erzeugt werden. [6]
Traditionelle Algorithmen erzeugen oft große Delta-Daten aufgrund von Zeigern in binären Daten. Moderne Algorithmen indizieren die alte Datei und finden exakte Übereinstimmungen in der neuen Datei, die dann zu „approximate matches“ erweitert werden. Finding-Matching-Algorithmen, z.B. Suffix-Bäume, werden verwendet, um identische Blöcke zwischen zwei Versionen zu finden. [4]
Die Vorgehensweise dieses modernen Delta-Algorithmus kann wie folgt erklärt werden:

- Die alte Datei wird in ihre Endungen zerlegt und mit dem Suffix-Array-Algorithmus sortiert. Dies ermöglicht eine effiziente Suche nach den längsten Übereinstimmungen zwischen alten und neuen Datenblöcken.
- Das Suffix-Array wird verwendet, um die längste Übereinstimmung eines Präfixes der neuen Datei in der alten Datei zu finden.
- Der längste gemeinsame Teil zwischen den Dateien wird durch Vergleiche und Verschiebungen gefunden, dann werden die Unterschiede gespeichert.
- Die Unterschiede und zusätzliche Daten wie Header werden komprimiert und in eine Delta-Datei geschrieben.
- Die Blöcke werden mit einem Kompressionsalgorithmus komprimiert, um die Delta-Datei klein zu halten.

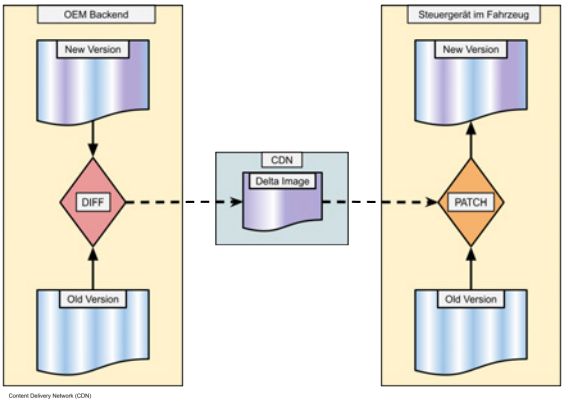

Abb. 3: Verfahren des Delta-Algorithmus [1]

Diese Technik ermöglicht eine effizientere Erstellung von Delta-Dateien, da die Unterschiede zwischen den Versionen genau identifiziert und unnötige Datenübertragungen vermieden werden. Dies führt zu kleineren und effizienteren Software-Updates.

Ausblick

Zukünftige Arbeiten sollten die Effizienz und Geschwindigkeit der Delta-Generierung und der Patch-Anwendung untersuchen. Weitere Evaluierungen an praktischen Beispielen wie MICROSAR Adaptive Images für QNX sowie in Steuergeräten sind notwendig. Optimierungen in diesen Bereichen können zu effizienten und zuverlässigen Software-Updates führen.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Vector Informatik GbmH. Automotive Over-The-Air. https://www.vector.com/de/de/know-how/automotive-ota, 2024.

[3] AUTOSAR Organization. Specification of Update and Configuration Management. https://www.autosar.org/fileadmin/standards/R21-11/AP/AUTOSAR_SWS_UpdateAndConfigurationManagement.pdf, 2021.

[4] Colin Percival. Naive Differences of Executable Code, 2003.

[5] Blackberry QNX. QNX Neutrino Real-Time Operating System (RTOS). https://blackberry.qnx.com/en/products/foundation-software/qnx-rtos, 2024.

[6] Tsviatko Yovtchev. The definitive guide to diff algorithms and patch formats. https://ably.com/blog/practical-guide-to-diff-algorithms, 2023.

Trajectory and Behavior Prediction of Road-Agents using Large Language Models

Martin Dell **Markus Enzweiler**

Department of Computer Science and Engineering, Esslingen University

Work carried out at Robert Bosch GmbH, Stuttgart

Introduction

The development of autonomous vehicles has been a long-standing goal in the field of artificial intelligence, with the aim of creating safe and efficient transportation systems. Because the task of autonomous driving is highly complex, it is often broken down into three subsequent tasks using the divide and conquer principle: perception, prediction, and planning. Perception involves detecting and identifying road-agents and their surroundings, while the prediction task focuses on forecasting their future movements and behaviors. Finally, the planning task uses this information to determine the optimal actions for the autonomous vehicle to take. Notably, human drivers perform these tasks implicitly, relying on their cognitive abilities to navigate complex traffic scenarios [6].
The rapid advancements in natural language processing (NLP) have led to the development of large language models (LLMs). These models have shown their ability in a wide range of NLP tasks, including language translation, text classification, and text generation. Currently, the potential of LLMs in non-traditional applications, such as autonomous driving, is being explored [6], [4]. The key advantage of using LLMs lies in their ability to generalize well across a spectrum of tasks. This is possible because they are pre-trained on large-scale data that allow them to have a base understanding of the world [6], [1].

Problem Formulation

Core objective of this thesis is to analyze the potential of LLMs for the trajectory and behavior prediction task in autonomous driving.
Initially, the possibility of achieving this goal without fine-tuning (zero-shot) the LLM should be examined. This is then compared to the results of a fine-tuned LLM. Besides trajectory prediction, the suitability of using LLMs for ranking or selecting distinct output trajectories of existing state-of-the-art prediction models should also be explored.
The results of this thesis are evaluated on the nuScenes dataset following the guidelines of the nuScenes prediction challenge [2]. This enables a detailed comparison with the current state-of-the-art prediction models. According to the challenge, participants are required to predict up to 25 possible trajectories, called modes, for the next six seconds, given two seconds of history [2].

Large Language Models

LLMs work by breaking down each word in the input into its numerical representation, known as tokens, using a tokenizer. The model's task is to predict the next token in the given input sequence. Since there are multiple potential words or tokens that can follow, the results form a distribution of possible tokens.
Various strategies exist for selecting the next token. Deterministic selection involves choosing the most probable next token, known as greedy selection. Alternatively, beam search can be used to construct multiple token paths and select the most likely path.
In many cases, having the same answer for a sequence is not desired. As a distribution is available, stochastic sampling techniques can be utilized. The simplest method involves randomly selecting the next token based on the probability distribution. The distribution can be adjusted using a temperature parameter to make it more uniform. Unlikely tokens can also be filtered out by initially choosing from a subset of tokens whose combined probability reaches or exceeds a certain threshold p, known as Top-p or nucleus sampling [3].
After selecting the next token, this process is repeated autoregressively until a specified limit is reached or the end-of-sentence token is encountered.

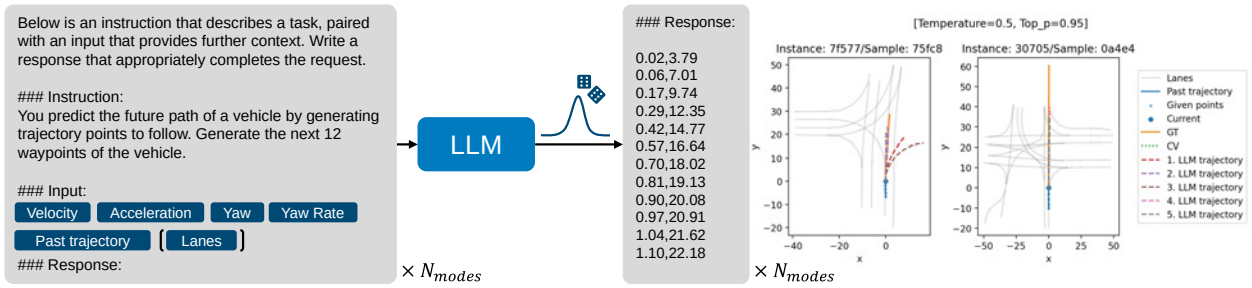

Fig. 1: Flow of predicting trajectories [5]

Trajectory Prediction

The first approach involves a LLM to predict the trajectory of a target vehicle. This is done by providing the LLM with specific instructions. The model is given various contextual information, such as velocity, acceleration, yaw, yaw rate, past trajectory, and lane details, all in text format. The goal of the model is to predict the future trajectory, represented by 12 points relative to the current position of the vehicle. Stochastic sampling is used to generate multiple possible trajectories, as shown in Figure 1.

However, the difficulties lie in trying to represent environmental information, like lanes, as text. There are different ways to describe lanes as text, such as discretizing them into equidistant points, using a fixed number of points, employing cubic Bézier curves, or describing them using semantic language.

The last method is particularly challenging because the nuScenes dataset [2] does not provide such descriptions, and generating them based on lane segments alone is very difficult. One potential solution is to use a vision language model (VLM) and prompt it to provide a map description based on a rasterized image. In this case, it may be more efficient and more correct to skip the step of encoding environmental information as text and instead directly represent it as an image.

Cubic Bézier curves are a compact method for representing lanes, but they may be too abstract for the LLM to understand. On the other hand, dividing the lanes into equidistant points is a straightforward way to describe them, but it is challenging to manage the number of points and, consequently, the number of tokens required. It is crucial to control the number of tokens because of limitations on context size and GPU memory. Therefore, the only viable solution is to encode lanes using a fixed number of points.

Initial results indicate a systematic error in the model, as it does not show improvements when provided with additional context, such as lane information. As a result, alternative ways in which LLMs can be used for trajectory prediction are explored.

Ranking output trajectories

An alternative way to use the LLM is to rank or select distinct trajectories from a set of existing trajectories. When predicting trajectories, multiple possible paths are often identified, but they tend to be very similar to each other. It would be better to have a wider variety of distinct paths. For example, when a vehicle approaches a 4-way intersection, it could turn left, right, or go straight. Ideally, only three distinct trajectories are necessary. However, prediction models often generate multiple trajectories for each direction. This makes it challenging to select the most relevant trajectories for the planning task. A LLM could be helpful in making this selection.

As an initial implementation, the LLM should rank existing trajectories based on other predictors, as depicted in Figure 2. These output trajectories will be referred to as proposal trajectories. For training and evaluation, an oracle is constructed that orders the proposal trajectories based on their average displacement error (ADE) compared to the ground truth trajectory. The LLM should learn to imitate the oracle.

The results indicate that the LLM is unable to accurately mimic the oracle and therefore fails to rank proposal trajectories.

In a slightly different task, the LLM is also asked to predict the best (top-ranked) proposal trajectory. The model is prompted multiple times, and the response is sampled. The frequency of the selected proposal trajectory determines its probability. The trajectory with the highest probability is then evaluated. Compared to the ranking task, similar or slightly inferior results are obtained.

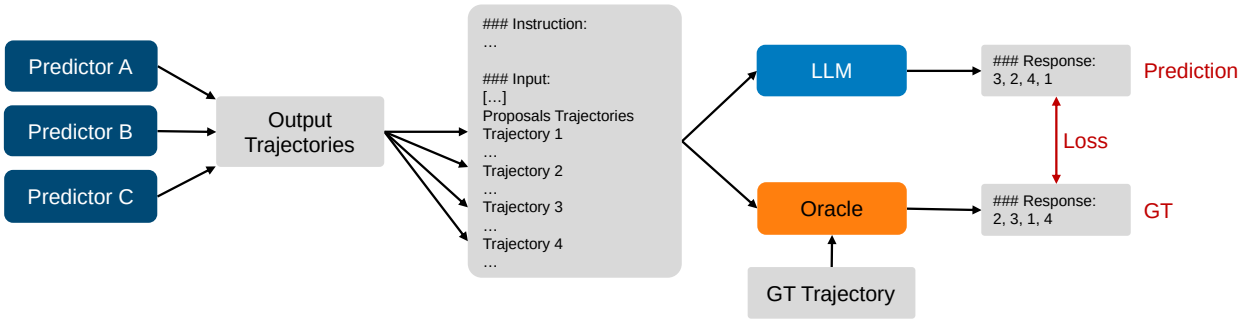

Fig. 2: Flow of ranking proposal trajectories [5]

Outlook

The results from all approaches indicate a systematic error. Additionally, encoding environmental context remains a challenge. Solving this will improve the LLM and let it stay within GPU memory constraints.
One potential solution is to enhance the efficiency and accuracy of the LLM by incorporating a specialized encoder for prediction. This encoder serves as a bottleneck for information between the environmental context and the model, reducing the number of input tokens and subsequently decreasing the required GPU memory. This approach should improve the overall performance of the LLM.

References and figures

[1] Inhwan Bae, Junoh Lee, and Hae-Gon Jeon. Can Language Beat Numerical Regression? Language-Based Multimodal Trajectory Prediction. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2024.

[2] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuScenes: A Multimodal Dataset for Autonomous Driving. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11618–11628. IEEE, 2020.

[3] Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. The Curious Case of Neural Text Degeneration. In *International Conference on Learning Representations (ICLR)*. ICLR, 2020.

[4] Jiageng Mao, Yuxi Qian, Junjie Ye, Hang Zhao, and Yue Wang. GPT-Driver: Learning to Drive with GPT. In *NeurIPS 2023 Foundation Models for Decision Making Workshop*. NeurIPS, 2023.

[5] Own representation.

[6] Chonghao Sima, Katrin Renz, Kashyap Chitta, Li Chen, Hanxue Zhang, Chengen Xie, Ping Luo, Andreas Geiger, and Hongyang Li. DriveLM: Driving with Graph Visual Question Answering. *arXiv preprint arXiv:2312.14150*, 2023.

Implementierung und Vergleich von Zugriffssteuerungen mit NAC-Lösungen in Enterprise-Netzwerken

Yagmur Demiral **Tobias Heer**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma CSNETWORKS GmbH, Göppingen

Motivation und Problemstellung

In der heutigen digital vernetzten Welt sind Unternehmensnetzwerke einer Vielzahl von Sicherheitsbedrohungen ausgesetzt. Diese Bedrohungen reichen von Phishing-Angriffen bis hin zu komplexen, gezielten Angriffen bei denen unautorisierte Geräte ins Netzwerk eingeschleust werden, um Informationen abzufangen. Vor diesem Hintergrund gewinnt die Implementierung von Sicherheitsmechanismen an Bedeutung.
Network Access Control (NAC) Systeme bieten eine gängige Lösung zur Kontrolle und Überwachung des Zugriffs auf Netzwerkressourcen. Ihre Hauptfunktion besteht darin sicherzustellen, dass nur authentifizierte und autorisierte Benutzer oder Geräte auf das Netzwerk zugreifen können. Bei einer Netzwerkverbindung wird beispielsweise mittels Zertifikat die Authentizität des Nutzers oder Geräts geprüft. Zertifikate sind digitale Dokumente, die die Authentizität und Integrität von Benutzern oder Geräten bestätigen, wodurch eine zusätzliche Sicherheitsstufe etabliert wird.
In dieser Arbeit untersuchen wir die Verbesserung des Sicherheitsniveaus in Enterprise-Netzwerken durch den Einsatz von Zertifikaten im Kontext der Absicherung von Netzwerkzugriffskontrollen. Für diese Analyse definieren wir Angriffsszenarien in einer virtuellen Testumgebung und wenden sie auf verschiedene NAC-Systeme und ihre Sicherheitsstufen an. Ziel ist es mittels eines Vergleichs von NAC-Lösungen deren Belastbarkeit im Umgang mit diesen Szenarien anhand geeigneter Evaluationsmetriken zu ermitteln.

Design

In diesem Abschnitt beschreiben wir den Aufbau der virtuellen Testumgebung und die Sicherheitsstufen, anhand derer die Angriffsszenarien durchgeführt werden. Für unsere Untersuchung implementieren und vergleichen wir die drei NAC-Lösungen *Cloudpath Enrollment System von Ruckus*, *Extreme Control von Extreme Networks* und *macmon NAC von der macmon secure GmbH* in einem Referenznetzwerk.

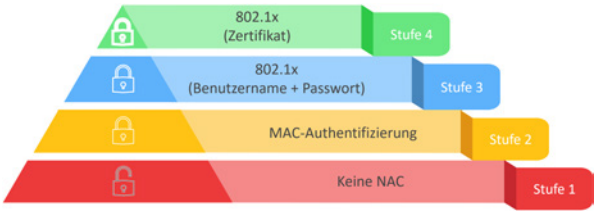

Abb. 1: Sicherheitsstufen der NAC-Lösungen [1]

Unsere Wahl, ein virtuelles Testbed zu nutzen, basiert auf dessen Flexibilität und Skalierbarkeit. Insbesondere fördert die Verwendung einer virtuellen Umgebung die Reproduzierbarkeit der Tests, da Parameter genau dokumentiert und eingestellt werden können.
Wir untersuchen im Referenznetzwerk die verschiedenen, aufeinander aufbauenden Sicherheitsstufen, die in Abb. 1 dargestellt sind:

1. **Keine NAC:** Basisszenario, ohne Netzwerkzugangskontrollmechanismen.
2. **NAC mit MAC-Authentifizierung:** Netzwerkzugriff auf Basis der MAC-Adresse des Endgerätes.
3. **NAC mit 802.1x (Benutzername + Passwort):** Authentifizierung mit Benutzername und Passwort.
4. **NAC mit 802.1x (Zertifikat):** Gültiges Zertifikat erforderlich, um Netzwerkzugriff zu authentifizieren.

Das Basisszenario zeigt Angriffsvektoren und potenzielle Risiken, die durch unbefugten Netzwerkzugriff in einem ungeschützten Umfeld auftreten können. Die MAC-Authentifizierung bietet einen ersten Schritt zur Erhöhung des Sicherheitsniveaus. Die Zugangskontrolle mittels Benutzername und Passwort ist eine weiterführende Absicherung des Netzwerkzugangs, die den Netzwerkzugriff auf autorisierte Benutzer einschränkt. Bei der Untersuchung der höchsten Sicherheitsstufe

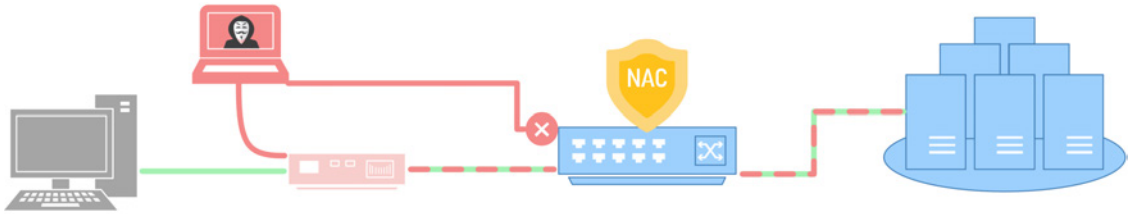

Abb. 2: Man-in-the-middle Attacke mit Hilfe einer Silent Bridge auf der Sicherheitsstufe (3) [1]

ergeben sich Herausforderungen im Umgang mit Endgeräten, die keine Zertifikate unterstützen. Die Herangehensweise der verschiedenen NAC-Systeme bezüglich dieser Problematik beeinflussen ihre Effizienz und Wirksamkeit.

Evaluation

In diesem Abschnitt werden die Methoden zur Bewertung der verschiedenen NAC-Lösungen dargelegt. Die im Folgenden vorgestellten Evaluationskriterien beruhen auf vier zentralen Aspekten, die für die Beurteilung von NAC-Lösungen in unterschiedlichen Umgebungen angewandt werden können.

- **Angriffsszenarien:** Bei der Bewertung der unterschiedlichen Sicherheitsstufen analysieren wir gezielt die Belastbarkeit verschiedener NAC-Lösungen. Auf zweiten Sicherheitsstufe (vgl. Abb. 1) können bereits einfache Angriffe wie das MAC-Spoofing erfolgreich eingesetzt werden. Die Bewertung anhand von Angriffsszenarien stellt sicher, dass wir die Abwehrfähigkeit der verschiedenen NAC-Lösungen gegenüber einer Vielzahl von Bedrohungen auf allen Sicherheitsstufen bewerten können. Wir berücksichtigen dabei Angriffsmethoden verschiedener Schwere, darunter Unauthorized Port Reconnection, Port-Mirroring-basiertes Eavesdropping, Man-in-the-Middle (vgl. Abb. 2), Credential Compromise und Certificate Compromise.

- **Granularität der Funktionen:** Eine weitere Bewertungsgrundlage stellt die Tiefe und Breite der sicherheitsrelevanten Funktionen einer NAC-Lösung dar. Hierbei werden Aspekte wie die Kombination der unterstützten RADIUS-Attribute, Protokolle bei der Kommunikation mit dem Switch und die Verschlüsselungstiefe der Zertifikate betrachtet. Eine Unterstützung dieser Sicherheitsmerkmale zeigt die Flexibilität und Anpassungsfähigkeit der Lösung in verschiedenen Netzwerkumgebungen auf.

- **Skalierbarkeit:** Unter Skalierbarkeit wird nicht nur die technische Ausbaufähigkeit der Lösung betrachtet, sondern auch die Lizenzstruktur und die damit verbundenen Kosten. Besonders für Managed Service Provider (MSPs) sind zudem Multi-Tenancy-Fähigkeiten von Bedeutung, um diverse Kunden in einer gemeinsamen Instanz bedienen zu können, ohne dass diese gegenseitig Einblick in ihre Daten erhalten. Dieser Aspekt beurteilt also, wie gut sich eine Lösung für unterschiedliche Unternehmensgrößen und Geschäftsmodelle eignet und wie wirtschaftlich sie im Verlauf ihrer Nutzung bleibt.

- **Redundanz:** Redundanz gewährleistet die kontinuierliche Verfügbarkeit und Funktionalität der NAC-Lösung in verschiedenen Netzwerkszenarien. Hierbei sind die Implementierung von Load Balancing Mechanismen und das Verhalten im Failover-Fall (wie die Lösung auf Ausfälle reagiert) von zentraler Bedeutung. Ein hoher Redundanzgrad sichert gegen Ausfälle ab und stellt sicher, dass die Sicherheitsrichtlinien konsequent durchgesetzt werden, selbst bei unvorhergesehenen Netzwerkproblemen.

Durch die systematische Bewertung dieser Aspekte können fundierte Schlussfolgerungen über die Eignung und Effektivität von NAC-Lösungen in unterschiedlichen Umgebungsbedingungen gezogen werden. Basierend auf den Ergebnissen der Evaluation werden Best Practices zur Erhöhung der Sicherheit in NAC-Systemen vorgestellt. Diese Best Practices bieten Netzwerkadministratoren gezielte Handlungsempfehlungen zur Verbesserung der Netzwerksicherheit durch geeignete NAC-Lösungen.

Verwandte Arbeiten

In den folgenden Abschnitten stellen wir verwandte Arbeiten vor, die jeweils unterschiedliche Aspekte und Ansätze in Bezug auf den Vergleich und die Bewertung von NAC-Systemen beleuchten.
Matthews [2] untersucht die Implementierung von NAC Systemen und die damit verbundenen Herausforderungen. Dabei schlägt er für seinen Vergleich der NAC-Lösungen von Portnox, Cisco und Bradford Networks, die Anforderungen *Vollständige Sichtbarkeit aller Geräte*, *Zentrales Management*, *Benutzerfreundlichkeit durch Automatisierung*, *Arbeitsintensivität*,

Granulare Regelumsetzung und Kontrolle, *automatische Integration neuer Systeme und Gäste*, *Einhaltung der Unternehmensrichtlinien für eigene Geräte* und *Flexibilität der unterstützten Geräteverwaltung* vor.
Serrao [5] führt eine eingehende Untersuchung von NAC Systemen aus Stakeholder Perspektive durch. In diesem Kontext wurden Lösungen von PacketFence, Cisco und Microsoft analysiert, wobei insgesamt acht Kategorien von Anforderungen identifiziert und zusammengestellt wurden. Diese Kategorien umfassen das *Administrator-Interface*, *Authentifizierung*, *Integrität*, *Sicherheit* sowie *funktionale* und *nicht-funktionale Aspekte*.
Omar und Abdelaziz [3] vergleichen NAC und Software-Defined Perimeter (SDP) Systeme. Dabei werden spezifische Anforderungen für NAC identifiziert, darunter *Authentifizierung*, *Verschlüsselung*, *feingranulierter Zugriff*, *Transparenz*, *Vereinfachung der Netzwerkinfrastruktur* und *Minimierung der Angriffsfläche*. Die beiden unterschiedlichen technologischen Lösungen werden miteinander in Beziehung gesetzt und anhand dieser Anforderungen evaluiert.
Parhi [4] dokumentiert die Schwachstellen von Protokollen, die in verschiedenen NAC Lösungen eingesetzt werden. Dabei werden je nach verwendeten Protokollen verschiedene Arten von Angriffen identifiziert, darunter z.B. *User Name Embezzlement*, *Session-Hijacking* und *Denial-of-Service-Angriffe*. Die Arbeit identifiziert Schwachstellen in den Protokollen und präsentiert einen angriffsbasierten Ansatz zur Evaluierung von NAC-Protokollen.
Zusammenfassend erweitert diese Arbeit die vorhandene Literatur durch einen angriffsbasierten und vergleichenden Ansatz, der auch auf andere NAC Lösungen angewendet werden kann, die nicht in dieser Arbeit betrachtet werden. Unser Fokus auf Zertifikate in Kombination mit einem angriffsbasierten Evaluationsmodell bietet einen differenzierten Ansatz zur bestehenden Forschung.

Ergebnis

Die vorliegende Arbeit untersucht wie der Einsatz von Zertifikaten in der Netzwerkzugriffskontrolle den Raum für potenzielle Angriffsvektoren in Enterprise-Netzwerken reduziert. Hierfür bieten wir Vergleichskriterien, wie die Bewertung von NAC-Lösungen in verschiedenen Angriffsszenarien, um Entscheidungsträgern eine objektive und angriffsbasierte Evaluation zu ermöglichen. Damit leistet diese Arbeit einen Beitrag zur Diskussion zwischen IT-Sicherheit und praktischer Anwendbarkeit.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Joseph Matthews. Challenges to Implementing Network Access Control. In *Challenges to Implementing Network Access Control*. SANS Institute, 2017.

[3] Rami Radwan Omar and Tawfig M. Abdelaziz. A Comparative Study of Network Access Control and Software-Defined Perimeter. In *Proceedings of the 6th International Conference on Engineering & MIS 2020*, volume 22, pages 1–5. Association for Computing Machinery, 2020.

[4] Snehasish Parhi. Attacks Due to Flaw of Protocols Used In Network Access Control (NAC), Their Solutions and Issues: A Survey. In *I. J. Computer Network and Information Security*, volume 3, pages 31–46. MECS Press, 2012.

[5] Gloria J. Serrao. Network access control (NAC): An open source analysis of architectures and requirements. In *44th Annual 2010 IEEE International Carnahan Conference on Security Technology*, volume 44, pages 94–102. Engineers, Institute of Electrical and Electronics, 2010.

Ermittlung und Prototypische Umsetzung von KI-basierten Use Cases - „KI im Unternehmen wertschöpfend einsetzen“

Niko Deuschle **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Festool GmbH, Wendlingen

Einführung und Problemstellung

Trotz der rapiden Entwicklung und Verbreitung von Künstlicher Intelligenz (KI) wird das Potenzial von KI in vielen Unternehmen bislang nur unzureichend genutzt. Dies liegt häufig an Unsicherheiten hinsichtlich der Identifikation relevanter Anwendungsfälle sowie der praktischen Umsetzung von KI-Lösungen. Die fehlende Expertise, unklare Nutzenabschätzung und mangelnde Erfahrung mit KI-Technologien stellen wesentliche Barrieren dar [4].

Zielsetzung

Diese Bachelorarbeit zielt darauf ab, mithilfe eines methodischen Ansatzes KI-Use Cases zu ermitteln und ausgewählte Beispiele umzusetzen. Basierend auf bestehenden und neuen Ansätzen werden relevante Use Cases identifiziert und exemplarisch ein oder zwei Modelle umgesetzt. Durch die Ermittlung dieser Anwendungsfälle wird das umfassende Optimierungs- und Innovationspotenzial von KI-basierten Lösungen aufgezeigt. Zur Validierung dieser Use Cases wird sowohl ein Use Case mit Optimierungspotenzial als auch ein Use Case mit Innovationspotenzial prototypisch umgesetzt, um den Nutzen der jeweiligen Ansätze zu demonstrieren.

Methode

Für die Identifizierung der Use Cases wird ein strukturiertes, iteratives Verfahren angewendet, das in fünf Phasen aufgeteilt ist: Vorbereitung, Ideenfindung, Bewertung, Priorisierung und Umsetzung.

Vorbereitung

In der Phase der Vorbereitung erfolgt zunächst die präzise Abgrenzung des Untersuchungsbereichs sowie die Festlegung der beteiligten Personen, die an diesem Projekt mitwirken. Zusätzlich wird eine Einführung in das Thema Künstliche Intelligenz durchgeführt, um allen beteiligten Personen einen klaren Rahmen zu vermitteln, was mit KI realisierbar ist und welche Aspekte utopisch sind [1]. Zudem ist es von entscheidender Bedeutung, dass fundiertes Fachwissen über die spezifischen Bereiche besteht. Dieses Wissen wird durch Experteninterviews, Job-Shadowing und dokumentarische Analysen erarbeitet.

Ideenfindung

In der Ideenfindungsphase ist es unumgänglich, sich sowohl der Probleme des Unternehmens als auch der Fähigkeiten Künstlicher Intelligenz und der Form und Qualität der verfügbaren Daten bewusst zu sein. Dieses umfassende Bewusstsein ermöglicht eine gezielte Identifikation von Use Cases und die Entwicklung effektiver KI-Lösungen. Durch die Kombination der problemorientierten, technologieorientierten und datenorientierten Ansätze können nicht nur relevante Probleme identifiziert, sondern auch die technischen Machbarkeiten geprüft werden, was im weiteren Verlauf zu einer realistischen und umfassenden Bewertung der Lösungsansätze führt [3].

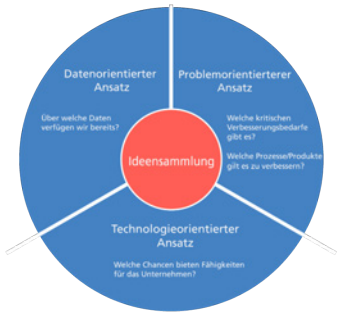

Abb. 1: Kernansätze für die Ideenfindung [3]

Bewertung

Zur Bewertung der Use Cases wird, abhängig von der Art des Use Cases ein spezifisches Verfahren eingesetzt, um den Nutzen bzw. Wert zu bestimmen. Problemorientierte Use Cases werden hinsichtlich Problemlösung und Effizienzsteigerung bewertet, während technologieorientierte Use Cases auf Innovationsgrad und Marktpotenzial untersucht werden.

Priorisierung

Die Priorisierung erfolgt in einem Zusammenspiel zwischen dem erwarteten Wert und der technischen Umsetzbarkeit der identifizierten Use Cases [1]. Basierend auf diesen Indizes wird die Priorität der ermittelten Use Cases festgelegt, wodurch die attraktivsten Fälle priorisiert werden, um sie anschließend prototypisch umzusetzen. Diese sorgfältige Auswahl ermöglicht eine effiziente Nutzung der Ressourcen und eine zielgerichtete Entwicklung von KI-Lösungen, die einen signifikanten Mehrwert bieten können.

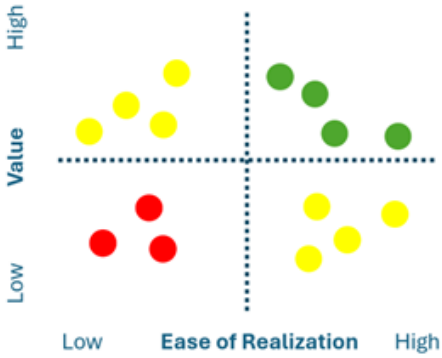

Abb. 2: Use Case Priorisierung [2]

Erste Umsetzung

Beim Testen von Sägemaschinen werden unteranderem die Sägeblätter auf ihrer Abnutzung getestet. Hierbei wird der Verschleiß der einzelnen Sägezähne des Blattes als Abnutzungsgrad festgelegt. Die Ermittlung dieses Verschleißes wird unter einem Mikroskop halbautomatisch durchgeführt. Durch die hohe Anzahl an Sägezähnen der Sägeblätter besteht hier ein sogenannter „Pain-Point“ in Form eines hohen Zeitaufwandes. Ein intelligentes Bildverarbeitungsmodell hätte das Potenzial, den zeitlichen Aufwand des Prozesses zu minimieren und die Fehlerquote zu senken. Im besten Fall müsste der Mitarbeiter lediglich den einzelnen Sägezahn unter das Mikroskop legen, ein Foto aufnehmen und der Verschleiß würde automatisch ermittelt werden.

Abb. 3: High Layer - Use Case Sägezahn [2]

Literatur und Abbildungen

[1] Hendrik Brakemeier et al. How to find and prioritize AI use cases. https://www.appliedai.de/insights/how-to-find-and-prioritize-ai-use-cases, 03 2024.

[2] Eigene Darstellung.

[3] Maximilian Feike. KI-Toolbox für Versorgungsunternehmen: Modul Use Cases. https://www.kodis.iao.fraunhofer.de/de/ki-toolbox-fuer-versorgungsunternehmen/use-cases.html, 05 2024.

[4] Peter Hofmann, Jan Jöhnk Jöhnk, and Nils Urbach. KI-Anwendungsfälle zielgerichtet identifizieren. *Wirtschaftsinformatik & Management*, pages 814–193, 2020.

Analyse der Auswirkungen der Einführung einer Cloud-basierten Customer Data Plattform auf die Erfolgsfaktoren der Kundeninteraktion in einem Unternehmen

Tim Drexler **Thomas Rodach**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Pilz GmbH & Co. KG, Ostfildern

Einleitung

Die Einführung einer Customer Data Plattform (CDP) bietet eine vielversprechende Möglichkeit, Herausforderungen in der Kundeninteraktion zu bewältigen. Die zentrale Fragestellung dieser Bachelorarbeit ist, wie die Implementierung einer CDP die Kundeninteraktionen optimieren kann. Dabei werden Anforderungen, Prozesse sowie potenzielle Herausforderungen und Chancen der CDP-Implementierung untersucht, um die Erfolgsfaktoren der Kundeninteraktion zu verbessern. Diese Arbeit basiert auf einer qualitativen-empirischen Forschung. Es wurden Experteninterviews durchgeführt und ausgewertet, um fundierte und verlässliche Aussagen im Kontext des Unternehmens treffen zu können. Die aus der empirischen Untersuchung gewonnenen Erkenntnisse wurden in die weitere Arbeit integriert, um valide Ansätze zur Implementierung der CDP zu erarbeiten.

Theoretischer Rahmen

Eine CDP ermöglicht es, Kundendaten aus verschiedenen Quellen zu sammeln, zu speichern und zu verarbeiten, um einheitliche Kundenprofile zu erstellen und personalisierte Marketing- und Vertriebsaktivitäten zu unterstützen. Das CDP-Institut beschreibt die Funktionen einer CDP wie folgt [1]:

1. Aufnahme von Daten aus beliebigen Quellen
2. Erfassung aller Details der erfassten Daten
3. Speicherung der aufgenommenen Daten auf unbestimmte Zeit (vorbehaltlich der Einschränkungen des Datenschutzes)
4. einheitliche Profile von identifizierten Personen erstellen
5. Daten mit jedem System teilen, das sie benötigt
6. in Echtzeit auf neue Daten und auf Profilanfragen reagieren
7. Kundendaten in Übereinstimmung mit den lokalen Datenschutz- und Sicherheitsvorschriften verwalten

Damit Kundendaten als Leitfaden zwischen verschiedenen Marketing-, Werbe- und kundenorientierten Anwendungen dienen können, müssen CDPs bestimmte Funktionen bereitstellen. Diese werden anhand des USPA-Frameworks für CDPs deutlich. Dieses Framework beschreibt die Funktionsweise einer CDP anhand der verschiedenen Schritte des Datenzyklus (siehe Abb. 1).

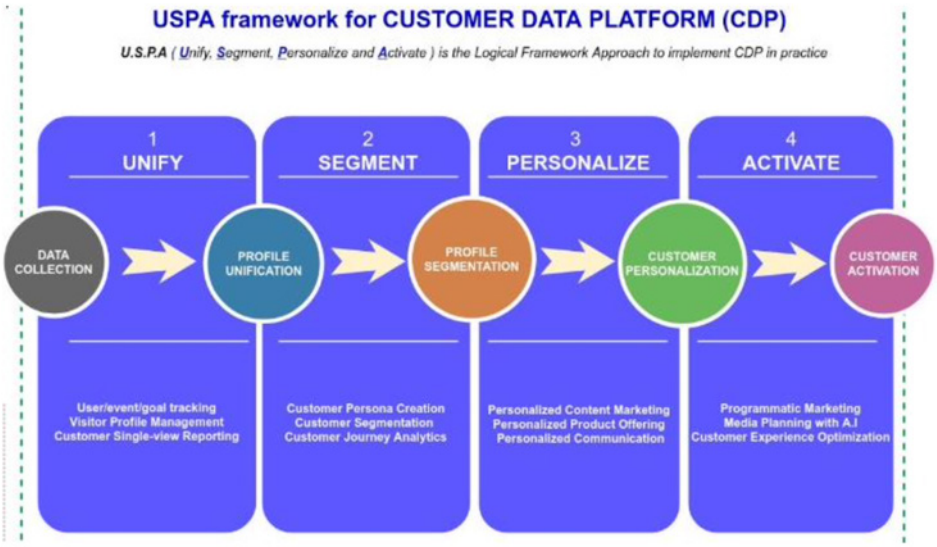

Abb. 1: USPA-Framework für eine Customer Data Plattform [2]

Marktanalyse und Anbieterbewertung

Eine umfassende Marktanalyse und ein Vergleich verschiedener CDP-Anbieter wurden durchgeführt, um die geeignetsten Lösungen für das Unternehmen zu identifizieren. Dabei wurde festgestellt, dass die Anbieter Oracle und SAP die Anforderungen am besten erfüllen. Besondere Beachtung fanden die Integrationsmöglichkeiten und die Funktionsumfänge der CDPs, die die spezifischen Anforderungen des Unternehmens berücksichtigen können.

Implementierungsstrategie

Die Implementierung einer CDP folgt einem strukturierten Ansatz (siehe Abb. 2), der von der grundlegenden Konfiguration der Plattform über die Sammlung und Integration relevanter Kundendaten bis hin zur finalen Aktivierung reicht. Wichtige Schritte beinhalten die Trennung von Entwicklungs- und Produktionsumgebungen, die sorgfältige Planung von Hard- und Soft-IDs sowie die Konfiguration von Kundenschemata [4]. Ein entscheidender Teil des Implementierungsprozesses ist die kontinuierliche Überwachung und Qualitätssicherung der Datenflüsse und Kundeninteraktionsszenarien.

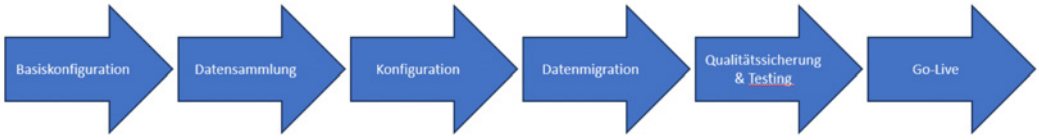

Abb. 2: CDP-Implementierungsansatz in Anlehnung an Rashedi & Maurer, 2023, S. 87-98 [3]

Potenziale und Herausforderungen

Die Einführung einer CDP in Unternehmen birgt erhebliche Potenziale zur Verbesserung der Kundenbindung und Personalisierung von Kundeninteraktionen. Herausforderungen bestehen insbesondere in der Integration bestehender Systeme und der Sicherstellung der Datenqualität. Die Arbeit zeigt, dass eine CDP helfen kann, Datensilos zu überwinden und eine ganzheitliche Sicht auf die Kunden zu ermöglichen.

Schlussbetrachtung und Ausblick

Die Bachelorarbeit zeigt, dass die Implementierung einer CDP in einem Unternehmen erhebliches Potenzial bietet, die Kundeninteraktionen zu verbessern und den Unternehmenserfolg zu steigern. Eine sorgfältige Planung und Umsetzung der CDP-Integration, unter Berücksichtigung der spezifischen Anforderungen und Ressourcen des Unternehmens, ist entscheidend für den Erfolg des Projekts. Zukünftige Forschungen könnten sich auf die langfristigen Auswirkungen der CDP-Implementierung und die kontinuierliche Optimierung der Kundeninteraktionsstrategien konzentrieren.

Literatur und Abbildungen

[1] CDP Institute. What is a RealCDP? https://www.cdpinstitute.org/learning-center/what-is-a-realcdp/, 2024.

[2] Tan Trieu Nguyen. What is USPA framework ? https://www.bigdatavietnam.org/p/about.html, 2019.

[3] Jonas Rashedi and Lena Mauer. *Customer-Data-Plattformen*. Springer Gabler, 1 edition, 2023.

[4] SAP SE. Customer Schema. https://help.sap.com/docs/customer-data-platform/user-guide/customer-schema?locale=en-US, 2024.

Data Mesh Konzeptionen als moderne Datenarchitektur

Merve Duman **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Unternehmen müssen datengesteuert arbeiten, um sich gegenüber der Konkurrenz zu behaupten, da das Datenvolumen kontinuierlich zunimmt. Die International Data Corporation (IDC) prognostiziert, dass sich die Datenmenge zwischen 2022 und 2026 mehr als verdoppeln wird. Derzeitige zentralisierte Datenarchitekturen sind jedoch nicht in der Lage, die steigende Vielfalt und Anzahl der Daten- und Analyseanwendungen zu bewältigen. Da die Datenflüsse nicht in die Unternehmensstruktur integriert sind, weisen Architekturen, die auf zentralisierten Data Warehouses und Data Lakes beruhen, erhebliche Nachteile auf. Dies führt zur Dezentralisierung der Erzeugung und des Verbrauchs analytischer Daten. Die Diskrepanz zwischen Datenproduktion und -konsum führt zu Engpässen in der zentralen IT-Abteilung, unklaren Verantwortlichkeiten und Datensilos im gesamten Unternehmen. Dies verhindert die effiziente Nutzung der Daten für Entscheidungsfindungen sowie die Entwicklung von Produkten und Dienstleistungen. [2]

Ziel der Arbeit

Die Abschlussarbeit gibt eine Einführung in die Grundlagen des Data Mesh und dessen Architekturen. Ziel ist es, zu untersuchen, ob das soziotechnische Datenkonzept Erfolg verspricht. Da Data Mesh ein neues Konzept ist und die Forschung dazu noch begrenzt ist, wird untersucht, welche Nutzungspotenziale Data-Mesh-Architekturen bieten und welche Anforderungen sich für die Implementierung, Datenintegration, Datenqualität und die Zusammenarbeit in Unternehmen ergeben. Abschließend werden konkrete Handlungsempfehlungen für die Implementierung von Data Mesh in mittelständischen Unternehmen gegeben.

Data Mesh

Laut Dehghani, Förderin des Data-Mesh-Konzepts ist Data Mesh eine dezentrale Methode, die technische und organisatorische Methode kombiniert, das darauf abzielt, analytische Daten in umfangreichen und großen Umgebungen zugänglich zu machen, zu verwalten und bereitzustellen. Unabhängig, ob im Rahmen eines Unternehmens oder über mehrere Organisationen hinweg. [3]

Data Mesh-Prinzipien

Die Basis der konzeptionellen Architektur und das Geschäftsmodell wird anhand der folgenden vier Prinzipien beschrieben: *Domain Ownership*, *Data as a Product*, *Self-Serve Data Platform*, *Federated Computational Governance* . Die genannten vier Prinzipien helfen Data Mesh dabei, seine Ziele zu erreichen, wie die Wertschöpfung aus Daten in großen Unternehmen zu generieren, die Agilität eines expandierenden Unternehmens zu bewahren und flexibel auf die Transformation der Wirtschaft zu reagieren.

1. Domain Ownership:
Die Teams, die den Daten am nächsten sind, tragen die Verantwortung für diese. Durch diese Aufteilung der Datenverantwortung können Teams die Datenverwaltung und -nutzung effizienter gestalten. Dies führt zu einer Steigerung der Datenqualität und zur Vermeidung von Engpässen, die durch die zentrale Datenverwaltung entstehen.

2. Data as a Product:
Die Daten gelten als separate Produkte und werden so behandelt. Das heißt, dass alle Datenprodukte über die erforderlichen Metadaten, Policies und Infrastrukturen verfügen müssen, damit sie autonom und nutzbar sind. Für die gesamte Lebensdauer und Qualität der Datenprodukte sind die Teams zuständig, die sie erzeugen. Dazu gehört auch, Daten mit einem festgelegten Servicelevel bereitzustellen, um eine verlässliche Nutzung sicherzustellen.

3. Self-Serve Data Platform:
Eine Self-Serve Data Platform bietet die erforderlichen Ressourcen und Normen, damit Teams Datenpipelines und -produkte erstellen und verwalten können. Querschnittsfunktionen wie Monitoring, Logging und gemeinsame API-Standards werden von dieser Plattform den Teams angeboten. Die Kombination des

„as-a-Service"-Konzepts mit dem „as Code"-Ansatz dient der Sicherstellung von Flexibilität und Effizienz.
4. Federated Computational Governance:
Eine Federated Computational Governance gewährleistet die Einhaltung globaler Standards und Interoperabilitätsanforderungen, während den Domänenautonomien die Möglichkeit zur Umsetzung in ihren eigenen Gebieten gewährt wird. Automatisierte Richtlinien und Standards dienen der Durchsetzung der Governance, um ein Gleichgewicht zwischen wesentlichen Anforderungen und domänenspezifischem Wissen zu gewährleisten. [1]

Zusammenspiel der Prinzipien

Die genannten Prinzipien sind für das Gesamtpaket erforderlich und ausreichend. Alle haben eine Gemeinsamkeit und befassen sich mit Hürden, die sich aus der Umsetzung der anderen ergeben. In Abbildung 1 kann man das Zusammenspiel der vier Prinzipien sehen.

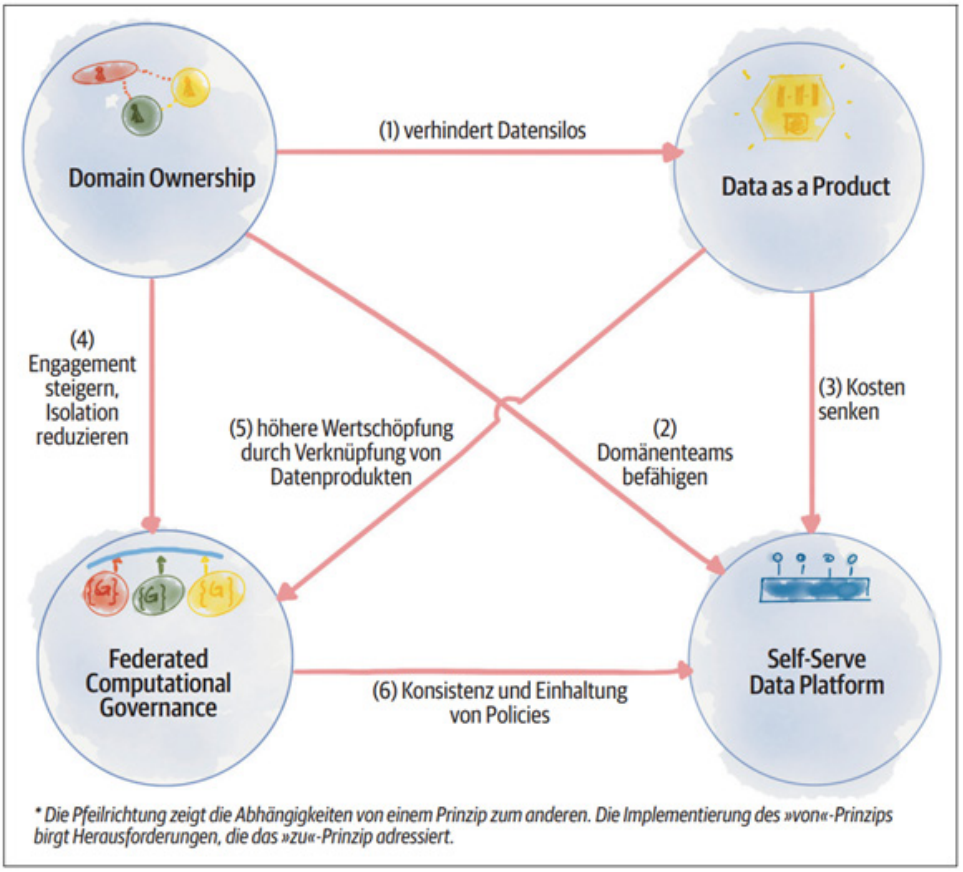

Abb. 1: Zusammenspiel der Data Mesh-Prinzipien [3]

Das Prinzip der Domain Ownership mit verteilter Datenspeicherung kann domäneninterne Datensilos verursachen. Dies kann durch das Prinzip Data as a Product verhindert werden, indem Domänen dazu angehalten werden, ihre Produkte sowohl intern als auch extern zur Verfügung zu stellen. Es ist möglich, dass die Domain Ownership von Datenprodukten zu einer Steigerung der Kosten, des Aufwands und der Leistungsfähigkeit führt. Durch die Self-Serve Data Platform können Domänenteams Datenprodukte effizient anbieten und einsetzen. Dies geschieht durch eine Verringerung des kognitiven Aufwands, Minimierung unnötiger Arbeiten, Steigerung der Produktivität und eine Senkung des Gesamtaufwands. [3]

Ausblick

Im Jahr 2023 führte BARC (Business Application Research Center) eine weltweite Anwenderbefragung unter 345 Führungskräften durch und griff dabei auf die Projekterfahrung der eigenen Analysten zurück. Im Mittelpunkt standen die vier Prinzipien des Data Mesh sowie das Verständnis und die Wertschätzung dieser Prinzipien. [4] Wie aus der zweiten Abbildung ersichtlich, gewinnt Data Mesh an Bedeutung. Da das Konzept jedoch neu ist, sind Unternehmen noch zögerlich und unerfahren in dessen Anwendung und unsicher, ob Data Mesh für sie geeignet ist. Die Vor- und Nachteile sind jedoch noch unklar, und Prognosen sowie Fallstudien, die den positiven Effekt von Data

Mesh unterstreichen, fehlen noch. Dennoch betrachten 35% der Befragten das Konzept als relevant und haben es bereits geplant.

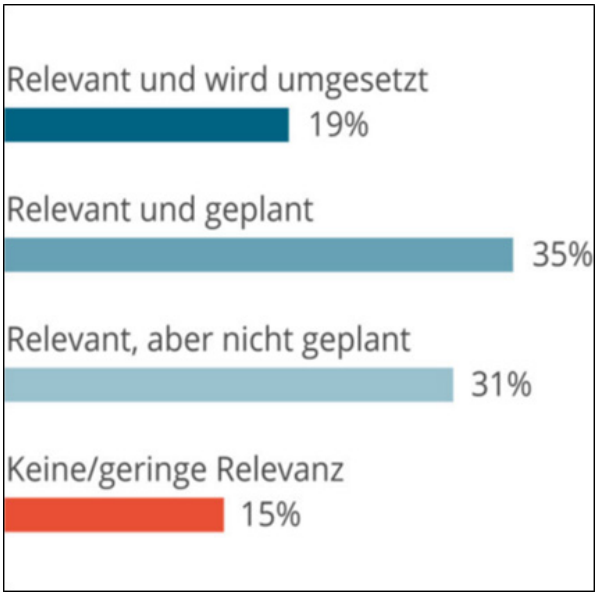

Abb. 2: Wie relevant ist das Konzept Data Mesh für Ihr Unternehmen? [4]

Literatur und Abbildungen

[1] Sebastian Baumann et al. *Data Mesh – Datenpotenziale finden und nutzen*. Bitkom e. V., 2022.

[2] Jan Bode, Niklas Kühl, Dominik Kreuzberger, Sebastian Hirschl, and Carsten Holtmann. Towards Avoiding the Data Mess: Industry Insights from Data Mesh Implementations. In *keins*. IBM, 2023.

[3] Zhamak Dehghani. *Data Mesh: Eine dezentrale Datenarchitektur entwerfen*. dpunkt.verlag GmbH, 1 edition, 2023.

[4] Martin Hensel. Data Mesh gewinnt für Unternehmen an Bedeutung. https://www.bigdata-insider.de/data-mesh-gewinnt-fuer-unternehmen-an-bedeutung-a-950230e80b97d936e720577f92972a19, 2023.

Entwicklung und Implementierung einer GitOps-gesteuerten, Multi-Tenant-fähigen DevSecOps Kubernetes-Plattform für die Cloud-basierte Softwareentwicklung in einer Hybrid-Cloud-Umgebung

Alexander Efremidis **Mirko Sonntag**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma IT-Designers GmbH, Esslingen am Neckar

Motivation und Problemstellung

Die Digitalisierung hat die Methoden und Anforderungen im Bereich der Softwareentwicklung in den letzten Jahrzehnten stark verändert. Während klassische Modelle wie das Wasserfallmodell heute in den meisten Unternehmen von agilen Modellen abgelöst wurden, haben sich auch im Bereich der Infrastruktur große Veränderungen ergeben. In den letzten 20 Jahren wurde mit dem Aufkommen zahlreicher Cloud-Anbieter wie Amazon AWS oder Microsoft Azure eine neue Ära in der Softwareentwicklung eingeläutet [6]. Das Thema Cloud Computing spielt in Unternehmen eine immer wichtigere Rolle, da sich durch die stetig steigende Rechenleistung zahlreiche neue strategische, operative und finanzielle Vorteile durch Themen wie Big Data und Machine Learning ergeben haben, von denen Unternehmen in vielen Bereichen profitieren [6]. Damit gehen aber auch einige technische und finanzielle Hürden einher, die unter anderem durch eine engere Zusammenarbeit zwischen Entwicklungs- und Betriebsteams überwunden werden sollen. Dieser Kulturwandel hat dazu geführt, dass im Zeitalter des Cloud Computing die schnelle Entwicklung und Auslieferung von Software mit agilen Methoden zum De-facto-Standard geworden ist. Die daraus entstandene Cloud-native Softwareentwicklung konnte sich zusammen mit DevOps in den letzten Jahren zunehmend am Markt etablieren [7] und [3].
Bei näherer Betrachtung des Konzepts der Cloud-nativen Softwareentwicklung zeigt sich, dass häufig nicht der gesamte Softwareentwicklungsprozess (SDLC) in der Cloud stattfindet. Ein Großteil der eigentlichen Entwicklung findet lokal in den Entwicklungsteams statt und wird lediglich in Cloud-Umgebungen getestet und ausgeliefert. Dieses klassische Vorgehen führt je nach Größe und Komplexität der Anwendung sowie der Größe des Teams zu einem erheblichen Mehraufwand an Konfiguration, Wartung und Zeit. Gerade bei Teams mit hoher Fluktuation führt dies zu langen Einarbeitungsphasen, minimiert die Effizienz und erfordert viel Dokumentation und technisches Know-how, um neue Teammitglieder in bestehende Prozesse einbinden zu können. Gerade bei Themen rund um Künstliche Intelligenz (KI) stoßen Entwickler immer schneller an die Kapazitätsgrenzen der verfügbaren Hardware, was sie häufig dazu zwingt, auf cloudbasierte Lösungen auszuweichen [3].
Unter anderem dadurch hat sich ein neues Feld der Cloud-nativen Softwareentwicklung eröffnet, das sogenannte Remote Development, bei dem der gesamte Entwicklungsprozess zentral in der Cloud stattfindet. Dies ermöglicht es Entwicklern, von nahezu jedem internetfähigen Endgerät aus zu arbeiten, und bietet neue flexible Lösungen im Bereich DevOps, um Kosten und Verwaltungsaufwand zu reduzieren. Gleichzeitig entstehen jedoch auch Herausforderungen im Ressourcenmanagement und in der Sicherheit, für die es noch keine allgemeingültige Lösung gibt.

Zielsetzung

Ziel dieser Arbeit ist der Entwurf und die Implementierung eines mandantenfähigen DevSecOps Frameworks, das speziell für Remote Development optimiert ist und bestehende Entwicklungsteams effektiv bei der Anwendungsentwicklung unterstützt. Der Fokus liegt dabei auf der Erstellung eines optimalen Konzepts, das versucht, die zahlreichen am Markt existierenden Lösungen zu vergleichen, um anschließend ein anfängerfreundliches Self-Service-Modell zu entwickeln, das es vor allem neuen Entwicklern ermöglicht, ohne umfangreiche Unterstützung durch Betriebsteams und ohne Kenntnisse im Bereich der Cloud-nativen Entwicklung auf die für ihre Arbeit benötigten Ressourcen

zuzugreifen und diese zu verwalten. Dabei geht es vor allem um die generische Erstellung neuer Entwicklungsumgebungen in der Cloud sowie um die Optimierung bestehender Release-Strukturen für die finalen Releases der verschiedenen Anwendungen.

DevOps und die DevOps-Kultur

Um die DevOps-Kultur zu verstehen, muss zunächst die Herkunft und Bedeutung des von DevOps definiert werden. Konkret beschreibt DevOps ein Konzept zur Optimierung der Softwareentwicklung und -bereitstellung, welches durch eine verbesserte Zusammenarbeit von Entwicklungs- und Betriebsteams erreicht werden soll [5]. DevOps unterteilt sich dabei in Continuous Integration, Continuous Delivery und Continuous Deployment. Diese drei Praktiken stellen sicher, dass automatisierte Software-Releases in heutigen Entwicklungsprozessen besser optimiert, häufiger und qualitativ hochwertiger bereitgestellt werden können [4].
Darüber hinaus wird in der Literatur häufig von einer DevOps-Kultur gesprochen [5], [4]. Dies ergibt sich aus den Anforderungen, die sich für Unternehmen ergeben, die DevOps erfolgreich einsetzen wollen. Dazu gehört neben der Anpassung vieler Softwaresysteme auch die Umstrukturierung und Organisation im Unternehmen. Ein hohes Maß an Transparenz, Kommunikation und Zusammenarbeit, aber auch gegenseitiges Vertrauen und Respekt sind notwendig, damit DevOps funktioniert [8]. Dies betrifft vor allem den Austausch zwischen Entwicklungs- und Betriebsteams, denn nur durch Kommunikation können Qualität, Verfügbarkeit und Erfolg in der Praxis sichergestellt werden. Diese Prämissen bilden unter anderem das Fundament von DevOps, welches oft auch als DevOps-Kultur bezeichnet wird, da diese immer individuell angepasst werden muss und jedes Team in der Umsetzung frei ist.

Infrastruktur der Private-Cloud

Für die Bereitstellung der Private Cloud und Basis der Self-Service Plattform wird ein Kubernetes Cluster auf Basis von Talos Linux realisiert. Dabei handelt es sich um ein skalierbares, für Kubernetes optimiertes Betriebssystem, das vollautomatisch über eine CLI und Konfigurationsdateien nach dem GitOps-Ansatz konfiguriert werden kann. Es verfügt ausschließlich über eine API, die alle Kommandos an das Betriebssystem abstrahiert und undokumentierte Änderungen z.B. über SSH verhindert. Da für diesen Cluster High Availability (HA) Anforderungen gelten, müssen Funktion, Skalierbarkeit und Uptime jederzeit gewährleistet sein. Wie in Abbildung 1 zu sehen ist, wird dieser Cluster als "Management Cluster"bezeichnet und verwaltet die verschiedenen Tenant Control Planes, die jeweils einen eigenen Kubernetes Scheduler, Controller und API Server beinhalten. Dies wird durch die Kamaji Software erreicht, die auf dem Management Cluster installiert ist und dafür sorgt, dass neue dedizierte Kubernetes Cluster erzeugt werden, die dann je nach Anwendungsfall in Projekten oder Teams eingesetzt werden können. Diese Cluster sind vollständig voneinander isoliert und können keine Verbindung zum Management Cluster oder zu anderen Tenants aufbauen [1]. Es ist nicht möglich, die Existenz anderer Tenants innerhalb eines Kamaji Tenant Clusters herauszufinden. Damit ist die Bedingung der sog. harten Mandantenfähigkeit erfüllt, die aufgrund von Datenschutzrichtlinien häufig gefordert wird. Konkret bedeutet dies, dass die Tenants in jeder Hinsicht vollständig voneinander isoliert sein müssen. Dies entspricht zwei dedizierten Hardware-Clustern.
Generell bietet Kamaji eine Vielzahl von Vorteilen, die für die Anforderungen von Kubernetes-as-a-Service nützlich sind. Einer der Hauptvorteile ist, dass verwaltete Cluster keine neuen Control Plane Nodes benötigen, was die Skalierbarkeit verbessert und die Hardwarekosten erheblich reduziert, da jeder Tenant Cluster nur dedizierte Worker Nodes benötigt, anstatt mindestens drei Control Plane Nodes plus Worker Nodes.

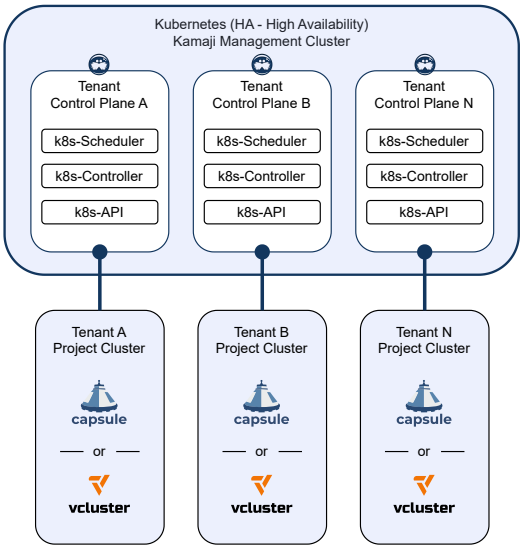

Abb. 1: Übersicht des Kamaji Kubernetes Management Clusters mit virtuellen Tenant Clustern [2]

Aufbau eines Tenant-Clusters

Innerhalb eines Tenant Clusters sind mehrere Schichten zur Verwaltung und Isolation der Entwicklungsumgebungen implementiert. Verschiedene zentrale Dienste wie Flux für das GitOps-Management, Capsule für die Namespace-Aggregation sowie Shared Applications wie ArgoCD für Continuous Delivery und LinSTOR für das persistente Speichermanagement sind im Plattformraum integriert und können je nach Bedarf eines Teams erweitert oder modifiziert werden. Die Verwaltung und Isolation der virtuellen Tenant-Umgebungen, mit denen die einzelnen Teammitglieder später arbeiten, wird durch die Verwendung von Namespaces durch die Software Capsule gewährleistet. Jeder Tenant, z.B. Entwickler Alice oder Bob in Abbildung 2, erhält isolierte Namespaces und eine Entwicklungsumgebung (DevPod). Capsule sorgt dabei für die notwendigen Kubernetes Sicherheits- und Ressourcenrichtlinien, so dass sichergestellt ist, dass alle Ressourcen und Workloads zwischen den verschiedenen Tenants strikt getrennt sind. Alle persistenten Daten von z.B. DevPod oder Datenbanken werden mit Hilfe von LinSTOR in Persistent Volumes (PV) gespeichert. Diese können dann nach Bedarf automatisch mit einem Backup-Plan gesichert werden. Innerhalb der zugewiesenen Namespaces hat jeder Benutzer freien Spielraum für Deployments oder Tests. Der Zugriff wird von Capsule mit gängigen Authentifizierungsmethoden geprüft und kontrolliert. Jeder Tenant hat somit seine eigene virtuelle Umgebung, ohne dass ein neuer dedizierter Cluster angelegt werden muss.

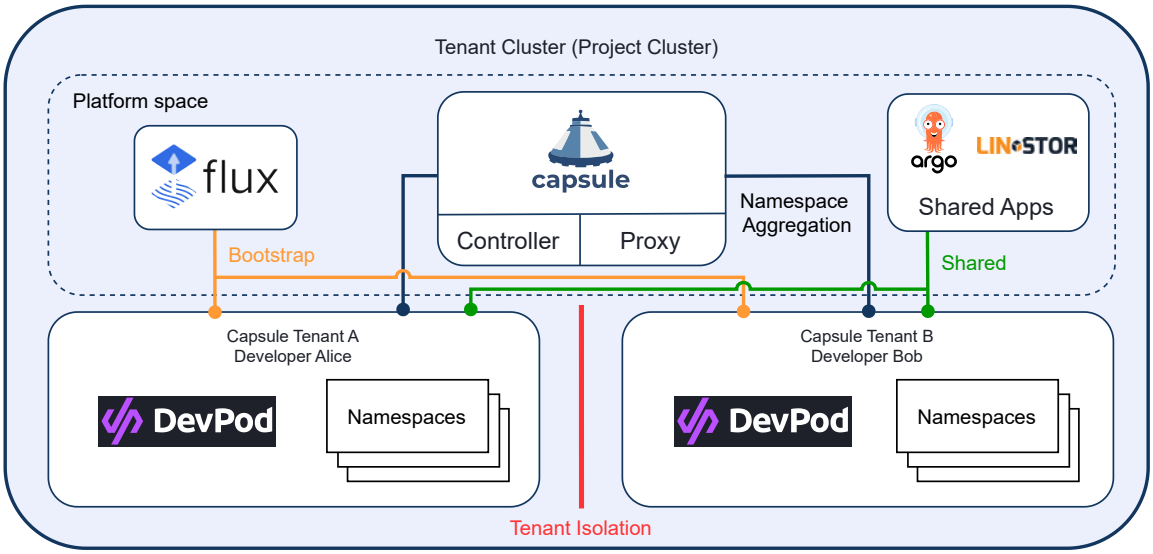

Abb. 2: Übersicht eines Capsule Tenant Clusters mit den zugehörigen virtuellen Entwicklungsclustern [2]

Bisher wurden mehrere Storage-Lösungen, Remote Development Tools, CI/CD Tools sowie mandantenfähige Frameworks miteinander verglichen, um den Use Case bestmöglich abzudecken. Dabei wurde vor allem auf Open Source, Konfigurierbarkeit, Geschwindigkeit und Funktionsumfang geachtet. Die in Abbildung 1 und 2 dargestellten Tools und Frameworks stellen zum jetzigen Zeitpunkt die optimale Lösung dar, wobei versucht wurde, diese optimal miteinander zu verknüpfen, um automatische Bootstrapping-Aufgaben sowie die Konfiguration mit GitOps zu ermöglichen. Die vorgestellte Infrastruktur wurde bereits in eine lauffähige Testumgebung implementiert. Ebenso wurde eine generische GitOps-Repository-Struktur entwickelt, um die einzelnen Tenants, Cluster und Entwicklungsumgebungen auch in großem Umfang verwalten zu können.

Ausblick

Neben dem Aufbau der Infrastruktur ist eine Kommandozeilenschnittstelle (CLI) in der Programmiersprache Go geplant, die das Self-Service-Modell ergänzen soll. Ziel ist es, Anwender bei alltäglichen DevOps-Aufgaben, wie zum Beispiel Deployments in Kubernetes, zu unterstützen. Insbesondere Entwickler sollen in die Lage versetzt werden, Routineaufgaben effizienter und sicherer durchzuführen, ohne auf manuelle Eingriffe oder tiefe Kenntnisse der zugrundeliegenden Infrastruktur angewiesen zu sein. Die CLI soll später mit den Remote-Entwicklungsumgebungen ausgeliefert werden, um mit wenigen Benutzereingaben automatisch einen lauffähigen Docker-Container mit lauffähiger CI/CD-Pipeline und entsprechendem Deployment-Ort zu erzeugen. Dabei greift es auf ein zentrales GitOps-gesteuertes Repository zu, um

relevante Templates, Cluster-Informationen etc. zu erhalten. Diese soll nahtlos mit der entwickelten Infrastrukturplattform harmonieren. Dadurch entstehen keine Abhängigkeiten von bestimmten Tools, wodurch ein hohes Maß an Flexibilität, niedrigen Kosten und Erweiterbarkeit erhalten bleibt. Außerdem werden dadurch Ressourcen bei den Betriebsteams eingespart, die diese Aufgaben sonst manuell erledigen müssten. Darüber hinaus wird eine umfassende qualitative und quantitative Erhebung im Kundenunternehmen durchgeführt, um einen tiefen Einblick in das bereits vorhandene Wissen in den Bereichen DevOps, SecOps, Remote Development und Kubernetes zu erhalten und im weiteren Verlauf die Anforderungen an das zu entwickelnde Tool besser definieren zu können.

Literatur und Abbildungen

[1] Labs Clastix. Kamaji Concepts. https://kamaji.clastix.io/concepts/, 2024.

[2] Eigene Darstellung.

[3] Derek DeBellis and Claire Peters. 2022 Accelerate State of DevOps Report: A deep dive into security. https://cloud.google.com/blog/products/devops-sre/dora-2022-accelerate-state-of-devops-report-now-out?hl=en, 2022.

[4] Tom Hall. Was ist DevOps-Kultur? https://www.atlassian.com/de/devops/what-is-devops/devops-culture, 2020.

[5] Nane Kratzke. *Cloud-native Computing*. Carl Hanser Verlag GmbH & Co. KG, 2 edition, 2023.

[6] S. Reinheimer. *Cloud Computing: Die Infrastruktur der Digitalisierung*. Springer Fachmedien Wiesbaden, 2018.

[7] Lionel Sujay Vailshery. Breakdown of software development methodologies practiced worldwide in 2022. https://www.statista.com/statistics/1233917/software-development-methodologies-practiced/, 05 2022.

[8] M. Walls. *Building a DevOps Culture*. O'Reilly Media, 2013.

Edge System zur Darstellung und Überwachung von Qualitätskennzahlen verbundener Sensoren

Edgar Ehremann **Thao Dang**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Leuze electronic GmbH + Co. KG, Owen

Einleitung

Die fortschreitende Digitalisierung und Automatisierung in der Industrie erfordern zunehmend zuverlässige und präzise Systeme zur Überwachung und Analyse von Sensordaten. Ein wesentlicher Bestandteil dieser industriellen Systeme sind Sensoren, die Daten aus der Umgebung erfassen und zur weiteren Verarbeitung bereitstellen. Sensor Health, zu sehen in Abbildung 1, ist eine von Leuze entwickelte Webanwendung, die in dieser Arbeit eine zentrale Rolle spielt. Sie ermöglicht die Überwachung und Analyse von Sensordaten in Echtzeit und trägt so zur Optimierung industrieller Prozesse bei. Diese Arbeit untersucht die Qualitätsmerkmale von Leuze Sensoren und erweitert das bestehende System, um eine bessere Integration und Datenvisualisierung zu ermöglichen.

Stand der Forschung und Technik

Die bereits bestehende Webanwendung Sensor Health ist ein von Leuze entwickeltes Tool zur Überwachung und Analyse von Sensordaten. Im Backend der Anwendung kommt Java zum Einsatz. Die Sensordaten werden über das Protokoll MQTT (Message Queuing Telemetry Transport) empfangen. Es arbeitet nach dem Publish-Subscribe-Modell, bei dem Geräte (Publisher) Daten an einen zentralen Broker senden, der diese dann an interessierte Clients (Subscriber) weiterleitet [3]. Für die Speicherung der Sensordaten wird die relationale Datenbank Postgres verwendet. Diese Daten werden aufbereitet und im Webbrowser über das Open-Source Python Framework Dash dargestellt. Dash ist ein Framework für den Aufbau von Webanwendungen mit einem interaktiven und ansprechenden Benutzerinterface. Es ist speziell für die Erstellung von analytischen Anwendungen und Dashboards konzipiert [4].
Die gesamte Anwendung wird in Docker-Containern betrieben. Durch die Nutzung von Containern kann sichergestellt werden, dass die Anwendung in isolierten und konsistenten Umgebungen läuft, was die Skalierbarkeit und Wartbarkeit erheblich verbessert [5].

Analyse der Qualitätsmerkmale von Leuze Sensoren

In der Datenbank gespeicherte Qualitätsdaten umfassen beispielsweise den Winkel, in dem ein Barcode erkannt wurde, die Anzahl erfolgreicher Lesungen und die Anzahl der Fehllesungen. Zusätzlich gibt es die Qualität der Lesung, die von den gegebenen Lichtverhältnissen, der Geschwindigkeit der Fließbänder und dem Zustand des Barcodes beeinflusst wird. Die Daten müssen analysiert und für die vom Benutzer ausgewählten Zeiträume berechnet werden, um die Effizienz, Performanz, Erreichbarkeit und Qualität, wie in Abbildung 1 dargestellt, zu ermitteln und zu visualisieren.

Durchführung von Experimenten zur Gewinnung von Testdaten

Um die Qualitätsmerkmale von Leuze Sensoren umfassend zu untersuchen und fundierte Erkenntnisse zu gewinnen, ist es notwendig, Datensätze erfolgreicher Lesungen mit gescheiterten Lesungen zu vergleichen. Zur Gewinnung der Daten von fehlerhafter Lesungen müssen zuerst gezielte Experimente durchgeführt werden. Diese Experimente sollen verschiedene Einflussfaktoren simulieren und deren Auswirkungen auf die Codeerkennung zeigen. Im Folgenden werden mögliche Experimente detailliert beschrieben, die in dieser Arbeit vertieft und miteinander verglichen werden können, um die besten Ansätze zur Datengewinnung von fehlerhaften Lesungen zu identifizieren. Diese Experimente können bspw. mit einem Leuze BCL, wie in Abbildung 2 dargestellt, durchgeführt werden.

- Beleuchtung: Die Beleuchtungsbedingungen haben einen wesentlichen Einfluss auf die Qualität der Codeerkennung. Lesevorgänge werden unter verschiedenen Lichtverhältnissen, wie Tageslicht, künstliches Licht (Neon, LED), Dunkelheit/Nachtbedingungen und Gegenlichtsituationen, durchgeführt.

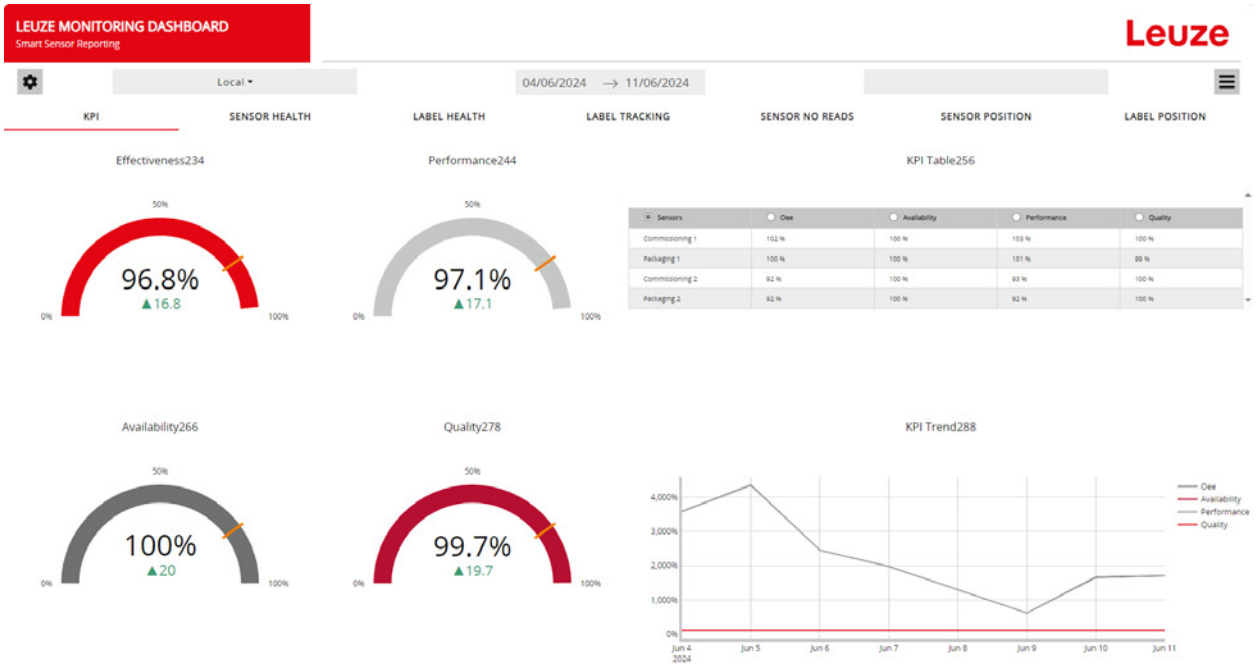

Abb. 1: Die Webansicht der Sensor Health Webanwendung [1]

- Verschmutzung: Sensoren und Barcodes werden mit verschiedenen Arten von Verunreinigungen (Staub, Öl, Wasser) konfrontiert.
- Vibrationen und Erschütterungen: Um die Auswirkungen von Vibrationen und Erschütterungen zu testen, werden die Sensoren in realen Produktionsumgebungen installiert, in denen natürliche Vibrationen und Erschütterungen auftreten.
- Platzierung und Ausrichtung: Die Platzierung und Ausrichtung der Barcodes auf den Objekten werden variiert, um ihre Auswirkungen auf die Erkennungsgenauigkeit zu untersuchen. Tests werden mit Barcodes an unterschiedlichen Positionen und in verschiedenen Ausrichtungen (horizontal, vertikal, schräg) durchgeführt.
- Druckqualität und Papierbeschaffenheit des Barcodes: Die Barcodes werden nicht nur in unterschiedlichen Druckqualitäten (hochauflösend, niedrig auflösend) erstellt, sondern auch auf verschiedenen Arten von Druckpapier (glänzendes, mattes, gestrichenes) gedruckt.
- Beschädigung oder Abnutzung: Barcodes werden bewusst teilweise verdeckt, zerkratzt oder verfärbt, um die Sensorleistung unter solchen Bedingungen zu testen.

Ziel der Arbeit

Im Rahmen der Bachelorarbeit sollen, die von Leuze Sensoren bereits gelieferten Qualitätsmerkmale untersucht werden und ein bestehendes System erweitert werden. Die bisherigen Datenstrukturen werden derzeit in einem isolierten System gespeichert. Diese Vorgehensweise entspricht nicht mehr den aktuellen Anforderungen und erfordert eine Integration mit bestehenden Systemen. Daher muss das aktuelle Projekt in die vorhandenen Systeme eingebunden und die Datenstrukturen entsprechend angepasst werden. Zudem wird das System durch grafische Komponenten ergänzt, welche die ausgelesenen Daten visuell darstellen. Verschiedene Faktoren, die Einfluss auf die Qualitätsdaten der Barcodeerkennungen von Leuze Sensoren haben, werden definiert und untersucht. Auf Basis dieser Untersuchungen sollen Grenzwerte abgeleitet werden, um konkrete Handlungsempfehlungen zu entwickeln und diese an die Benutzenden zu senden. Abschließend wird der entwickelte Prototyp evaluiert.

Abb. 2: BCL308i - Stationärer Barcodeleser [2]

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Leuze electronic GmbH und Co KG. Produkte und Lösungen zur Identifikation. https://files.leuze.com/Volumes/Volume0/opasdata/d100001/medias/docus/273/$v2/PIN_Identifikation_de_-1070-_96dpi.pdf, 01 2021.

[3] MQ Telemetry Transport MQTT. The Standard for IoT Messaging. https://mqtt.org/, 2022.

[4] Hossain Shammamah. Visualization of Bioinformatics Data with Dash Bio. In *Proceedings of the 18th Python in Science Conference (SciPy 2019)*, volume 11, pages 126–133. Chris Calloway and David Lippa and Dillon Niederhut and David Shupe, 2019.

[5] Tuomas Vase. Advantages of Docker, 2015.

Design and Implementation of a Modular Cybersecurity Attack and Defense Platform

Stefan Eisele **Tobias Heer**

Department of Computer Science and Engineering, Esslingen University

Work carried out at Department of Computer Science and Engineering, Esslingen

Motivation and Problem

In our digital world, companies are increasingly vulnerable due to complex threats arising from enhanced connectivity, sophisticated cyberattacks, and a globalized threat landscape. Consequently, the development of robust cybersecurity defenses has become more critical than ever. Therefore, it is necessary to build a workforce that is not only theoretically knowledgeable but also practically skilled in identifying, combating, and reducing risks.

Due to the dynamic of cyberattacks, it is crucial to equip computer scientists with hands-on expertise against evolving threats. Such approaches must transcend conventional classroom learning to immerse students in the cybersecurity challenges they will face in their professional lives.

The creation and implementation of practical training exercises, such as Capture The Flag (CTF) games, present an opportunity to enhance current educational frameworks by providing hands-on experience in a structured environment [4]. While traditional CTF games offer valuable practice in offensive strategies, there is a clear need to enhance defensive training within these exercises. Also, the development of these exercises often encounters challenges due to the lack of an infrastructure that supports reusable building blocks, making the process time-intensive.

This thesis proposes designing and implementing a attack and defense CTF platform. This initiative aims to prepare students for real-world cyber threats by making learning fun and practical. Above all, the development of defensive options should be encouraged, as these are still rare. Through real-life defense in a competitive game environment, students will gain a better understanding of how to act in critical situations. By dividing participants into teams that simultaneously perform offensive and defensive roles, the platform will simulate real-world situations, ensuring students gain comprehensive hands-on experience.

Design

The work builds on an existing infrastructure using proxmox, which is a virtualization platform for operating virtual machines. It exists a web interface with a MySQL database and a back-end to create attacking courses. Figure 1 shows our design that integrates and connects the following components into the infrastructure.

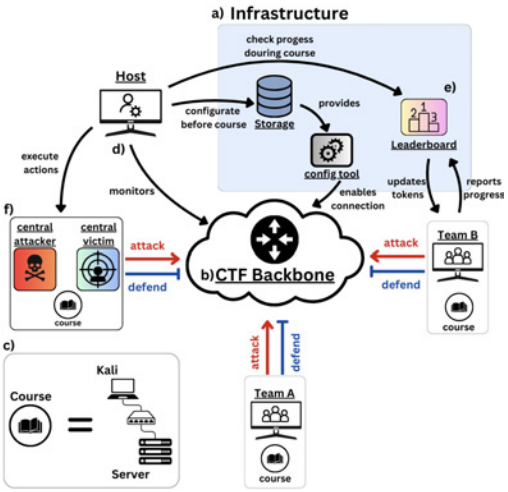

Fig. 1: Components of the CTF Training Platform [5]

The VMs are accessible via jump hosts through the university and can attack teams via the CTF backbone. The challenge is to create a configuration tool, monitor the actions, and connect all the components. The most important features are:

a) Infrastructure: The platform includes a structured and secure environment, which users can reach via VPN. Also, the host is able to manage and replicate the course.

b) CTF Backbone: Connects the teams using IP addresses. It ensures that teams, the central attacker and victim, as well as monitoring tools can all communicate effectively.

c) Course with Kali VM and vulnerable systems: The course component includes Kali Linux virtual machines that have a server connection and a range of tools for exercises.
d) Service Monitoring: The system automatically monitors active services. This design element introduces a feature that allows the session leader to monitor the actions of the teams.
e) Progress Check: To see progress, there will be a leaderboard with the points achieved. To score points, a team collects dynamic tokens that can be found in various vulnerabilities on teams or the central victim. The students can also collect points by closing security gaps to learn defensive actions.
f) Dynamic Actions: The instructor can set dynamic actions to guide the session. These actions include attacks and setting up a victim that teams can attack. Each feature aligns with the educational goals, focusing on scalability, accessibility, real-time feedback, adaptability, and modularity of the platform.

Evaluation

The evaluation of the training platform includes the systematic measurement of its effectiveness and impact on cybersecurity training. The assessment is based on the following specific requirements that measure its educational effectiveness:
Engagement and Motivation: Maintaining student interest and motivation through gamification with interactive and competitive elements. To measure commitment and motivation, we plan to conduct test runs of the platform within the IT security team and hacking AG, which will be completed with surveys.
Scalability and Accessibility: Ensuring the platform can support a growing number of users and is accessible to students with varying levels of expertise and resources. This involves creating a system that can adapt to different class sizes and central control of exercise complexity. Accessibility assessments also include user experience studies.
Central Progress Monitoring: The development of a mechanism for the immediate observation of participant actions within the game facilitates real-time learning from errors and achievements. The effectiveness of central progress monitoring is verified through feedback from test-run hosts.
Modularity and Reproducibility: Focusing on the design of a modular framework that allows for the easy addition, removal, or modification of content and scenarios. This ensures the platform can evolve with emerging cybersecurity trends and educational needs. To evaluate modularity and reproducibility, we will document the process and feedback from multiple educators who attempt to adapt the platform to different curricular requirements.
The evaluation aims not only to confirm the effectiveness of the platform in providing a dynamic learning environment but also to highlight areas that require continuous improvement to ensure that the platform covers broad areas of cybersecurity education.

Related Work

Cybersecurity education has seen various approaches aimed at enhancing practical skills and theoretical understanding, from traditional classroom practices to platforms designed to simulate real-world scenarios. Chattopadhyay et al. [2] show the importance of aligning cybersecurity educational games with academic and industry benchmarks. This study serves as a critical backdrop for our platform, suggesting the necessity for game designs that not only engage but also educate.
Yamin et al. [6] highlight the crucial role of real-world games in simulating complex cybersecurity scenarios, providing a dynamic platform for participants to engage mostly in attack but also defense strategies. Our platform aims to expand defensive capabilities to the same extent as offensive capabilities.
Kuo et al. [3] demonstrate the use of Emulab, a controllable platform that supports both attack and defense cybersecurity exercises that transcends traditional classroom settings. Similarly, our platform employs scenario-based training and expands upon this concept by offering a modular environment where scenarios can be customized to meet diverse learning objectives and skill levels.
The King of the Hill (KotH) game by Kevin Bock et al. [1] explores the benefits of cybersecurity competitions designed to give students hacking experience. The developed KotH competition involves a network topology requiring students to attack targets and defend themselves against other teams. Although this approach shares similarities with the goals of our thesis, our platform offers additional levels of modularity and adaptability.

Result

The result of this thesis is a well documented modular attack and defense training platform. The resulting platform consists of multiple key components, i.e., linking topologies, dynamic actions, and monitoring the progress. We evaluate these elements based on the criteria set out in our analysis. Consequently, the platform will improve the quality of cybersecurity education in our infrastructure, with a particular focus on strengthening defense mechanisms.

References and figures

[1] Kevin Bock, George Hughey, and Dave Levin. King of the Hill: A Novel Cybersecurity Competition for Teaching Penetration Testing, 2018.

[2] A. Chattopadhyay, C. Maschinot, and L. Nestor. Mirror Mirror On The Wall - What Are Cybersecurity Educational Games Offering Overall: A Research Study and Gap Analysis. In *IEEE Frontiers in Education Conference*. FIE, 2021.

[3] Cheng-Chung Kuo, Kai Chain, and Chu-Sing Yang. Cyber Attack and Defense Training: Using Emulab as a Platform. International Journal of Innovative Computing, Information and Control. *International Journal of Innovative Computing, Information and Control*, 14, 2018.

[4] L. McDaniel, E. Talvi, and B. Hay. Capture the Flag as Cyber Security Introduction. In *49th Hawaii International Conference on System Sciences*. HICSS, 2016.

[5] Own representation.

[6] Muhammad Mudassar Yamin, Basel Katt, and Mariusz Nowostawski. Games as a Tool to Model Attack and Defense Scenarios for Cyber-Security Exercises. *Computers & Security*, 110, 2021.

Vergleich von Clustering-Ansätzen für Positionsdaten einer free-floating Flotte im Shared-Micromobility Kontext

Mehmet Sinan Eris **Mirko Sonntag**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Mikromobilität beinhaltet kompakte und leichte Fortbewegungsmittel, welche für den individuellen Gebrauch genutzt werden [3]. In einer geteilten Mikromobilität teilen Nutzer Fahrzeuge mit anderen Nutzern. Ein Anbieter stellt eine Menge an Fahrrädern, E-Bikes, Scootern usw. zur Verfügung, auf die die Nutzer gegen eine Gebühr Zugang haben. Diese kann man entweder an einer Station oder in einem erlaubten Bereich abstellen. Hier gibt es die Unterscheidung zwischen Free-Floating und stationsbasierten Systemen. Free-Floating -Systeme erlauben es dem Nutzer, ein Fortbewegungsmittel überall im Geschäftsgebiet abzustellen, wobei die Stationsbasierten diese nur bei festgelegten Stationen erlauben. In Großstädten gewinnt diese Art der Fortbewegung zunehmend an Bedeutung als Alternative zu öffentlichen Verkehrsmitteln. Diese Systeme speichern eine große Anzahl an Buchungsdaten. Die Daten sollten effektiv genutzt werden, um Prognosen über die Nachfrage zu erstellen und das System effizient auszunutzen. Durch die Daten ist es möglich, das Verhalten einer Stadt bezüglich der Fortbewegung zu analysieren. Zur welcher Uhrzeit gibt es die meisten Fahrten? Wie lange fährt ein Nutzer durchschnittlich? Doch um Positionen zu analysieren, Hotspots zu finden oder um Nachbarschaften herauszufinden, müssen diese Daten in eine andere Darstellung transformiert werden, welche die Positionen besser darstellen.
Es stellt sich die Frage, wie man Positionsdaten clustern kann, um diese in eine Form zu bringen, die es erlaubt, weitere Analysen durchzuführen, um bekannte Probleme bei Mikromobilitätsflotten zu lösen.

Use Case

CHANGE ist ein Projekt des Fraunhofer IAO | Anwendungszentrum KEIM, das sich mit Free-Floating Mikromobilitätsangeboten beschäftigt und sich auf die Ladeinfrastruktur und die Verteilung der Fahrzeuge konzentriert. Es soll mit einem dynamischen Anreizsystem die Nutzer selbstständig dazu bringen, die Fahrzeuge an den optimalen Standort abzustellen, um eine bessere Verteilung der Fahrzeuge zu gewährleisten. Seit 2016 liegen von Stella, einem Mikromobilitätsanbieter, Buchungsdaten vor, welche für das Anreizsystem genutzt werden sollen. Der Geschäftsbereich soll nun in Zonen unterteilt werden, in denen die Daten aggregiert werden können. Diese Zonen ermöglichen es, spezifische Bereiche zu identifizieren, in denen Fahrzeuge abgestellt werden sollen, um die Verteilung zu optimieren. So können Nutzer gezielt dazu angeregt werden, Fahrzeuge in diesen Zonen abzustellen, was zu einer verbesserten Verteilung und effizienteren Nutzung der Mikromobilitätsangebote führt. Die Frage ist, wie der Geschäftsbereich aufgeteilt werden sollte.

Problemstellung

Wie lässt sich ein Gebiet in verschiedene Bereiche unterteilen? Es gibt verschiedene Möglichkeiten, die in der aktuellen Literatur verwendet werden. Die populärsten Unterteilungen sind Quadrat-Gitter und Hexagon-Gitter, welche über einen Bereich mit festgelegtem Abstand gelegt werden. Diese Gitterarten gehören zu den regulären Gittern 1. Diese Gitter sind statisch und ändern sich nicht.

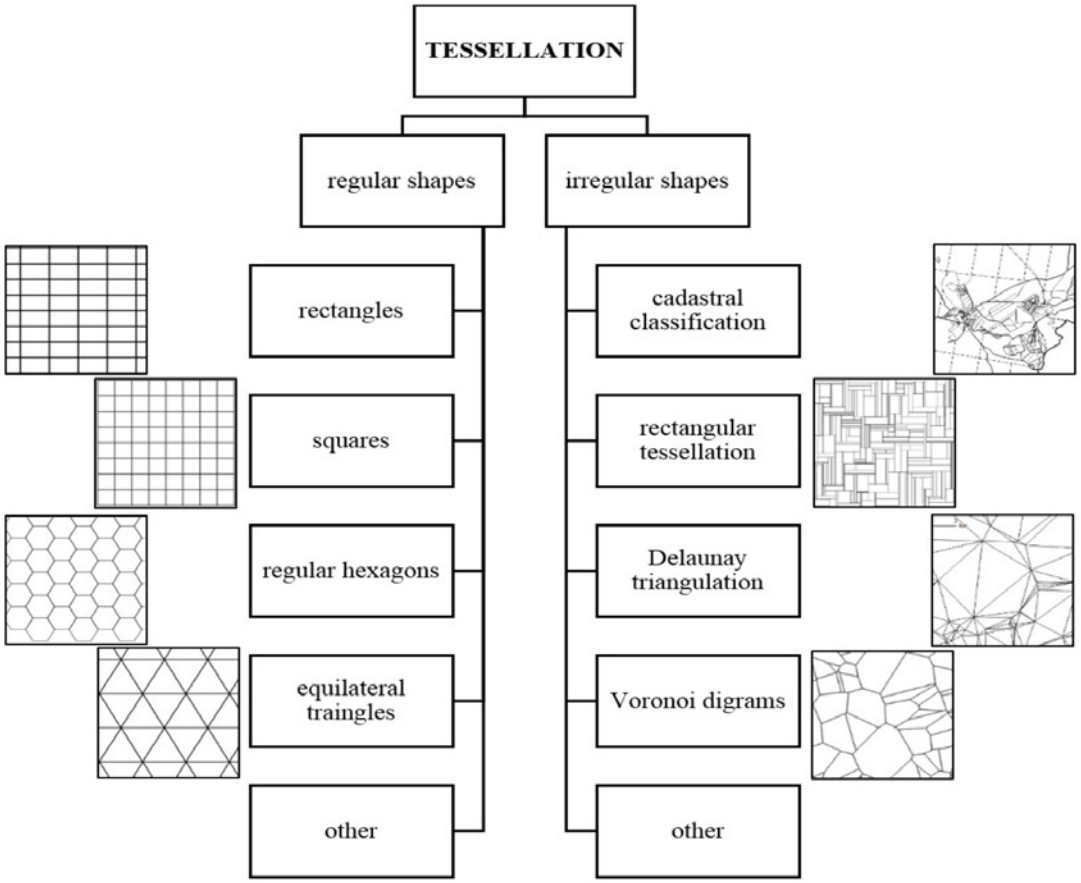

Abb. 1: Unterschiedliche Tesselierungsarten [1]

Diese Art der Gitter berücksichtigt jedoch zwei Aspekte nicht. Die vorhandenen Daten und das Straßennetz. Demzufolge wird in dieser Arbeit die Erstellung eines Gitters mithilfe von Voronoi-Zellen untersucht, die es ermöglichen, dynamische Zellen zu generieren. Voronoi-Zellen zählen zu den irregulären Gittern 1. Es gibt jetzt mehrere Möglichkeiten, ein Voronoi-Diagramm zu erstellen. Deswegen werden hier verschiedene Ansätze betrachtet. Danach sollen die Positionsdaten in den generierten Zellen aggregiert werden. Dadurch können neue Erkenntnisse gewonnen und Nachbarschaften entdeckt werden.

Zielsetzung

Am Ende soll ein Voronoi-Diagramm das Geschäftsgebiet unterteilen. Es sollen verschiedene Methoden zur Generierung der Voronoi-Zellen betrachtet werden. Diese Zellen sollen dann die Buchungsdaten aggregieren und visualisieren 2.

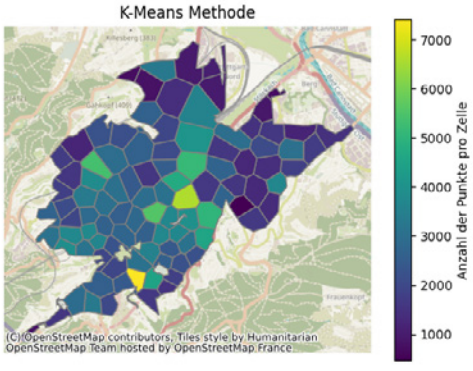

Abb. 2: Voronoi Zellen im Geschäftsgebiet [2]

Forschungsfragen sind hier:

1. Wie kann man ein Gitter bewerten?
2. Wie unterscheiden sich die Voronoi-Zellen bei verschiedenen Generierungsarten?

Vorgehen

1. Datenanalyse: Es liegen Buchungsdaten von Stella, einem Mikromobilitätsunternehmen, vor. Diese Daten werden auf die Eigenschaften der vorhandenen Attribute, ihre räumliche Verteilung und zeitliche Betrachtung untersucht. Damit die Parameter besser für die folgenden Schritte ausgewählt werden können.

2. Datenbereinigung: Die Daten sollen hinsichtlich Ausreißern bereinigt werden. Hier wird der modifizierte Z-Score verwendet, um Buchungsdaten anhand ihrer Position zu entfernen. Zudem werden die Datensätze entfernt, welche unsinnige Attribute haben.

3. Generatorpunkte bestimmen: Damit Voronoi-Zellen generiert werden können, reicht es nicht aus, alle Positionsdaten zur Generierung zu wählen, weil dies eine hohe Rechenleistung fordert und es nicht ermöglicht, Daten per Zelle zu aggregieren. Mithilfe von verschiedenen Verfahren werden die Generatorpunkte für die Voronoi-Zellen bestimmt. Ein Generatorpunkt in einem Voronoi-Diagramm ist ein Punkt, von dem aus die Voronoi-Zelle definiert wird. Jede Voronoi-Zelle umfasst alle Punkte, die näher an ihrem zugehörigen Generatorpunkt liegen als an jedem anderen Generatorpunkt. Die Buchungsdaten werden mithilfe von unterschiedlichen Algorithmen geclustert, um daraus die Generatorpunkte zu gewinnen.

4. Generierung der Voronoi-Zellen: Die Generatorpunkte aus dem letzten Schritt werden zur Generierung der Voronoi-Zelle verwendet. Das daraus resultierende Voronoi-Diagramm wird bearbeitet. Die No-Parking-Zonen von Stelle werden aus dem Voronoi-Diagramm entfernt, da sie niemals Fahrzeuge enthalten werden. Das Voronoi-Diagramm wird zudem durch das Geschäftsgebiet begrenzt, damit diese nicht über diesem liegt.

5. Visualisierung der Voronoi-Zellen: Die Voronoi-Zellen werden auf Stuttgart visualisiert und zeigen die verschiedenen Zonen mit einer unterschiedlichen Anzahl an Buchungsdaten pro Zelle. Hier kann man Hotspots erkennen und verschiedene Methoden vergleichen.

6. Analyse der Voronoi-Diagramme: Diagramme werden nach verschiedenen Kriterien analysiert. Die geometrischen Eigenschaften der Zellen, wie die Größe und Anzahl der Zellen. Die Beeinflussbarkeit der Zelleneigenschaften, wie die Größe. Auch bezüglich des Straßennetzes werden die Zellen beurteilt.

Ausblick

Im weiteren Verlauf werden die optimalen Parameter für die Cluster-Algorithmen gesucht. Zudem wird eine Methode verwendet, die das Straßennetz als Generator verwendet. Damit die verschiedenen Methoden verglichen werden können. Die Ergebnisse müssen im Anschluss evaluiert werden.

Literatur und Abbildungen

[1] Mirosław Bełej and Marta Figurska. 3D Modeling of Discontinuity in the Spatial Distribution of Apartment Prices Using Voronoi Diagrams. *Remote Sensing*, page 229, 2020.

[2] Eigene Darstellung.

[3] Wenke-Thiem Sybille. Mikromobilität Begriffe aus der kommunalen Szene, einfach erklärt. *Difu-Berichte*, page 16, 2021.

Kundenbindung und Produktpräferenzen bei einem Hersteller für hochwertige Elektrowerkzeuge: Eine explorative Analyse der Endkunden auf Basis von Garantieregistrierungen

Alex Erler **Thomas Rodach**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Für Unternehmen ist es essenziell, die Bedürfnisse ihrer Kunden zu verstehen. Es ist nicht nur ein entscheidender Faktor für den langfristigen Erfolg eines Unternehmens, sondern auch ein Wettbewerbsvorteil gegenüber Mitbewerbern. Dies ist in einer digitalen Welt umso wichtiger, da unterschiedlichste Daten über Endkunden erfasst und ausgewertet werden können. Ein führender Hersteller hochwertiger Elektrowerkzeuge nutzt Produktbündelungen, um den Absatz zu steigern und den Kunden gleichzeitig attraktive Angebote zu bieten. Diese Arbeit untersucht die Effektivität von Bündelungsstrategien und Preisnachlässen und analysiert das Kaufverhalten der Endkunden auf Basis von Garantieregistrierungen. Ziel ist es, durch das explorative Analysieren dieser Daten Produktpräferenzen zu entdecken. Anhand dieser Erkenntnisse sollen Handlungsempfehlungen ausgesprochen werden, um optimale Bündelungsstrategien und Rabattstrategien zu entwickeln.

Motivation und Ziele

Die Analyse des Kaufverhaltens der Kunden gibt Aufschluss darüber, welche Produkte und Produktkombinationen besonders beliebt sind. Durch die Anwendung von Market Basket Analysen und dem Ableiten von Assoziationsregeln sollen Muster im Kaufverhalten identifiziert werden und Empfehlungen zur Optimierung der Produktbündelungen [6] gegeben werden. Zusätzlich wird untersucht, welche Geräte als typische Einstiegsprodukte dienen und ob bestimmte Akkugeräte Neukunden schneller anziehen und somit zu weiteren Käufen innerhalb des Akkusystems anregen.

Hypothesen:

- Es existieren spezifische Produkte oder Produktkategorien, die signifikant häufiger als erste Kaufentscheidung von Neukunden gewählt werden.
- Es existieren Akkugeräte, die maßgeblich zur Neukundengewinnung beitragen und zum Kauf weiterer Akkugeräte führen.
- Kunden, die Produkte eines bestimmten Systems besitzen, kaufen mehr komplementäre Produkte, da sie die Systemlösung verstehen und nutzen.

Methodik

Die Analysen basieren auf den Transaktionsdaten der Kunden, die ihre Produkte direkt beim Hersteller registriert haben. Nach einer sorgfältigen Aufbereitung der Daten werden Mustererkennungs-Algorithmen wie Apriori und FP-Growth eingesetzt, welche häufig zusammen gekaufte Produkte identifizieren und daraus Assoziationsregeln ableiten. Des Weiteren wird die Wirkung von Akkugeräten auf die Neukundengewinnung und Kundenbindung untersucht. Ergänzend wird das Kaufverhalten von Kunden analysiert, die Produkte eines bestimmten Systems besitzen, um zu prüfen, ob diese Kunden mehr komplementäre Produkte kaufen und die Systemlösungen nutzen.

Theoretischer Hintergrund

Die Warenkorbanalyse (Market Basket Analysis, MBA) ist eine Data-Mining-Technik (siehe Abbildung 1), die Muster im Kaufverhalten aufdeckt, indem analysiert wird, welche Produkte häufig gemeinsam gekauft werden. Diese Technik liefert wichtige Erkenntnisse für das Verständnis von Kundenpräferenzen und zur

Optimierung von Cross-Selling-Strategien. Durch das Identifizieren häufig zusammen auftretender Artikelsets (Frequent Itemsets), ermöglicht die MBA tiefe Einblicke in die Produktwahl der Kunden [4].

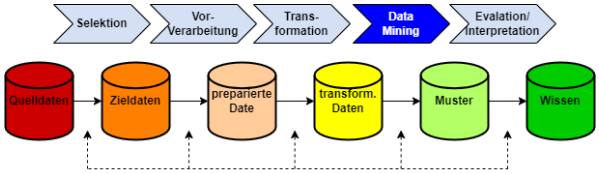

Abb. 1: Data-Mining Prozess [2]

Ein zentrales Element der Warenkorbanalyse ist das Ableiten von Assoziationsregeln. Diese Regeln helfen zu verstehen, welche Produkte typischerweise zusammen erworben werden und sind entscheidend für die Entwicklung effektiver Marketing- und Verkaufsstrategien. Eine typische Assoziationsregel wie „Wenn ein Kunde Produkt A und Produkt B kauft, kauft er wahrscheinlich auch Produkt C“, verdeutlicht nicht nur die Beziehung zwischen den Produkten [1], sondern bietet auch wertvolle Einsichten, die für die Gestaltung von Produktbündeln und die Förderung von Zusatzverkäufen genutzt werden können.
Die Algorithmen Apriori und FP-Growth, die zur Identifikation dieser Regelsets verwendet werden, operieren beide mit den Metriken
Support: $\text{Support}(A \to B) = \frac{|T(A \cup B)|}{|T|}$
Confidence: $\text{Confidence}(A \to B) = \frac{|T(A \cup B)|}{|T(A)|}$
Der Support gibt an, wie häufig ein Artikelset innerhalb aller Transaktionen vorkommt, während die Confidence die Wahrscheinlichkeit beschreibt, dass beim Kauf eines bestimmten Produkts auch ein anderes gekauft wird. Diese Metriken bilden die Grundlage für das Ableiten robuster Assoziationsregeln in der Analyse [3].
Der Apriori-Algorithmus generiert Kandidatenmengen schrittweise und verwendet einen iterativen Ansatz, der die Häufigkeit jeder Item-Kombination überprüft [1], wodurch er bei großen Datenmengen ineffizient sein kann. Im Gegensatz dazu verwendet der FP-Growth-Algorithmus eine kompakte Datenstruktur namens FP-Baum (siehe Abbildung 2), die es ermöglicht, häufige Itemsets effizienter und ohne Kandidatengenerierung zu finden, was ihn besonders bei umfangreichen Datensätzen leistungsfähiger macht [5].

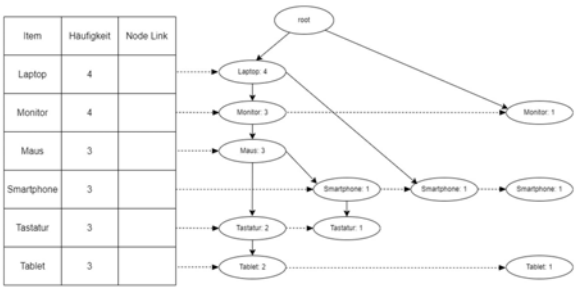

Abb. 2: FP-Baum Traversiert [2]

Bisherige Ansätze und Herausforderungen

Bisherige Ansätze zu Produktündelungen basieren auf älteren Daten und qualitativen Auswertungen, welche möglicherweise nicht mehr zum aktuellen Marktgeschehen und sich weiterentwickelnden Produktportfolio passen. Dies wirft die Frage auf, ob die bestehenden Bündelungsstrategien durch aktuelle Analysen optimiert werden können. Ein unzureichendes Verständnis könnte dazu führen, dass das Potenzial von Marketingstrategien nicht vollständig ausgeschöpft wird.
Es fehlt noch an Erkenntnissen darüber, ob es spezifische Akkugeräte gibt, die Neukunden in das Akkusystem ziehen und zum Kauf weiterer Akkugeräte anreizen. Die Identifikation solcher Produkte könnte maßgeblich dazu beitragen, gezielte Marketingmaßnahmen zu entwickeln und die Neukundengewinnung zu fördern. Auch ein Blick auf das Kaufverhalten von Kunden, die Produkte eines bestimmten Systems besitzen, lohnt sich. Der Hersteller entwickelt seine Produkte so, dass sie miteinander kombiniert werden können, wie zum Beispiel Absauggeräte und Feinstaubproduzierende Werkzeuge. Es wird untersucht, ob es einen Zusammenhang gibt, dass wenn der Kunde das System versteht und nutzt, er zusätzliche Käufe tätigt.

Ausblick

Erste Ergebnisse zeigen, dass auf einen Kauf von Akkugeräten signifikant mehr Akkugeräte folgen, was darauf schließen lässt, dass die Kunden das Akkusystem annehmen. Außerdem scheint sich die Annahme zu bestätigen, dass die Kunden, die ein zentrales Gerät der Systemlösung besitzen, ihr Sortiment um weitere Geräte ergänzen. Aus der Assoziationsanalyse konnten einige starke Regeln abgeleitet werden, welche eine gute Grundlage für das Erstellen von Produktbündeln bieten. Durch gezieltes Nutzen dieser Erkenntnisse und einer umfassenden Interpretation der Assoziationsregeln können wertvolle Maßnahmen abgeleitet werden, die dem Hersteller helfen, seine Marketing- und Produktentwicklungsstrategien weiter zu verbessern.

Literatur und Abbildungen

[1] Rakesh Agrawal and Srikant Ramakrishnan. Fast Algorithms for Mining Association Rules in Large Databases. *Proceedings of the 20th International Conference on Very Largge Databases*, pages 487–499, 1994.

[2] Eigene Darstellung.

[3] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques*. Elsevier Ltd., 2012.

[4] Manpreet Kaur and Shivani Kang. Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining. *Procedia Computer Science*, 2016.

[5] M.S. Mythili and Mohamed Shanavas. Performance Evaluation of Apriori and FP-Growth Algorithms. *International Journal of Computer Applications*, 70, 2013.

[6] Gerorge J. Stigler. Competition Between Firms that Bundle Information Goods. *Proceedings of the 27th Annual Telecommunications Policy Research Conference*, 1999.

Specification, Implementation and Integration of Neural Networks in Real Time Control Environments for Optical Character Recognition on Metallic Surfaces

Marc Glaser **Dirk Hesse**

Department of Computer Science and Engineering, Esslingen University

Work carried out at Schuler Group GmbH, Göppingen

Introduction

The recognition of identification codes on metallic surfaces describes a complex task for modern industrial systems. A trade-off between speed and accuracy degrades the performance of state-of-the-art (SOTA) software implementations. This work proposes the use of Neural Networks (NNs) with Optical Character Recognition (OCR) for fast and reliable code identification on metallic surfaces. Moreover, we deliver a concept for seamless integration of NN-based software in real-time control environments.

Motivation and Problem

Producing high-quality metal workpieces, especially for the automotive industry, relies on transparency of process values during the whole production workflow. Each production step seamlessly assigns the process values and tracks the quality parameters for each workpiece. The application of a unique code, named ID code in the following, ensures identification for each workpiece. A camera at the beginning of the production plant takes images of the ID code on the workpieces. A typical OCR task describes the processing of the images and the detection of characters. Traditional OCR solutions mainly work with noise-reducing filters and huge libraries with predefined fonts. Reacting on fast-alternating material surfaces on images with different backgrounds, like metallic surfaces, proposes a different task [1], [2].
Especially, detecting the Region of Interest (ROI) like in Figure 1 is challenging. The position of the ID code on the image taken by the camera depends on various factors. Triggered by the Front-of-Line (FoL) Programmable Logic Controller (PLC), the implemented line camera takes pictures of the workpiece with the ID code. For each product type, the operator trains the system, to configure the time slot, when the workpiece with the ID code is visible for the camera. Additionally, the position of the ID code itself on the workpiece varies for different product types.
The current software implementation optimizes the parameters of the algorithms for aluminum surfaces. In the case of steel workpieces, the performance of the algorithms decreases rapidly, which leads to many unrecognized ID codes. Steel engravings suffer from lower engraving deepness due to high corrosion probabilities caused by deeper scratches on the surface. Additionally, the rough and porous surface of steel leads to less contrast between the background and ID code in comparison to the smooth and hard material surface of aluminum. The currently used software libraries mainly contain thresholding and image operations with predefined values. It requires time as well as high expertise and knowledge, to adapt these parameters to fulfill the optimized settings for different metallic surfaces.

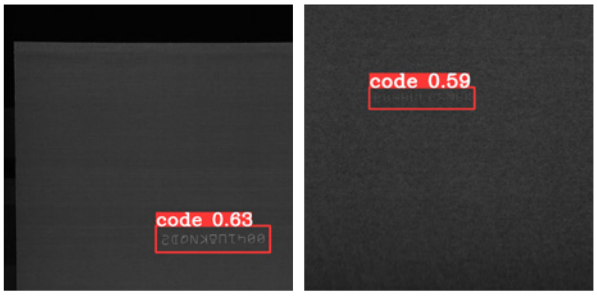

Fig. 1: ROI detection with YOLOv6 model for aluminum (left) and steel (right) plates [3]

For error code detection, the current implementation uses Reed-Solomon-Codes. The characteristics of Reed-Solomon error correction provides valid ID codes, although not all characters are recognized correctly. Reed-Solomon influences the OCR task, whereby the level of support is unclear at the moment.

Research Questions

In the course of this work, we give answers to the following research questions:

1. Do SOTA object detection NNs effectively and robustly detect the ROI on metallic surfaces?
2. Can proven filter and noise reduction strategies improve image quality with fast operations for further image processing on complex backgrounds?
3. Do lightweight but specially designed NNs deliver the same or better results than current OCR libraries?
4. What is the influence of Reed-Solomon-Codes in this application?
5. How performant is the implementation of NNs in the control environment via the Open Neural Network Exchange (ONNX) format?

Design

By the methodology of CRISP-ML(Q) [4], see Figure 2, the design adapts concepts of SOTA models and proposes new approaches related to the work with metallic surfaces. The goal of this work is a brownfield solution realized by NNs. We aim for a software-based adaption with the currently used hardware, existing of camera, lighting, and camera controller with a connection to the FoL PLC.

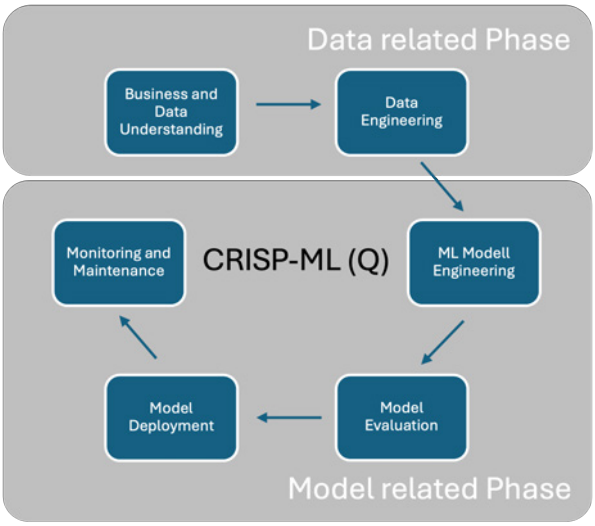

Fig. 2: CRISP-ML(Q) process model, adaption of [4]

Region of Interest

For detecting the ROI, we choose to work with NNs and their implementations, like the latest YOLO (You Only Look Once) models or similar implementations, see Figure 1. In terms of object detection, these models are widely used in the computer vision environment. Speed, accuracy, and the ability to detect thousands of categories in the real-world environment are the main benefits. In the context of the ID code detection on a dark, noisy background with reflections and illuminations, we expect that YOLO & Friends need some adaptions. An additional task is the improvement of the original input image quality in a good binarization of ID code and background. Under consideration of time efficiency, we examine whether image preprocessing methods, like median filter or encoder-decoder architecture offer benefits for the following OCR.

Optical Character Recognition

For the OCR task, we implement and rate SOTA libraries in contrast to specifically trained NNs to detect alpha-numeric characters. In this work, we assume a determined font size and style for the use case. Even for production systems in the industrial environment, it is common to commit to standard parameters. This leads to the question, of whether narrow NN architecture, like CNN or CRNN, performs as well or even better than predefined OCR libraries, like Tesseract.

Control Environment Integration

Implementing the above-described NNs in the industrial production environment marks the final contribution of this work. With the support of an ONNX interface, the first control manufacturers open their framework for standardized integration of NNs in the control environment.

Outlook

With the results of this work, a specified and integrated proof-of-concept for OCR on metallic surfaces by using NNs in the control environment will be presented. The accuracy for correct detection of workpiece IDs should reach 98 %. Moreover, the processing time should not exceed 300 ms due to the near-real-time request of the task.

References and figures

[1] Jakob Grönlund and Angelina Johansson. Defect Detection and OCR on Steel, 2019.

[2] Jing Li, Tao Huang, Yilei Xyang, and Qi Xu. Detection and Recognition of Characters on the Surface of Metal Workpieces with Complex Background. In *2020 IEEE 4th Infirmation Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. IEEE, 2020.

[3] Own representation.

[4] Stefan Studer, Thanh Binh Bui, Christian Drescher, Alexander Hanuschkin, Ludwig Winkler, Steven Peters, and Klaus-Robert Müller. Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. *Machine Learning and Knowledge Extraction*, pages 392–413, 2021.

Wirtschaftlichkeitsbetrachtung von Investitionen - Konzeption eines Tools im Bereich Automatisierungslösungen für die maschinelle Blechbearbeitung

Marco Goerlach **Anke Bez**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma TRUMPF SE + Co. KG, Ditzingen

Einleitung

Automatisierungslösungen spielen heutzutage in der maschinellen Blechbearbeitung eine immer wichtigere Rolle. Dies gilt insbesondere im Hinblick auf die zunehmende Verbreitung von sogenannten Smart Factories, welche sich durch einen hohen Automatisierungsgrad und eine vollvernetzte Fertigung auszeichnen. [2] Doch auch in kleineren Unternehmen im Bereich der Lohnfertigung kommen Automatisierungslösungen immer häufiger zum Einsatz, etwa um aktuellen Herausforderungen wie dem Fachkräftemangel zu begegnen, die Produktivität der Fertigung zu erhöhen oder Kosten zu sparen. [1] Hierbei ist zu beachten, dass es sich bei Automatisierungslösungen in diesem Fall nicht um elektrische Komponenten wie Sensorik, Aktorik oder Steuerungen handelt, sondern vielmehr um periphere Maschinen, die an eine bestehende Werkzeugmaschine angebunden werden können und einen Materialfluss in der Fabrik ermöglichen.

Das Maschinenbauunternehmen TRUMPF bietet seinen Kunden dabei ein breites Spektrum verschiedener Automatisierungslösungen für Werkzeugmaschinen. Im Bereich der 2D-Laserschneidmaschinen reicht dieses von einfachen Beladelösungen über Be- und Entladeautomatisierungen bis hin zu Sortiersystemen zum automatischen Heraustrennen, Sortieren und Stapeln von geschnittenen Blechteilen. [4] Außerdem stehen verschiedene automatisierte Lagersysteme zur Auswahl, mit denen im Zusammenspiel mit der Automatisierungslösung eine Laserschneidmaschine zu einer automatisierten Fertigungszelle ausgebaut werden kann. Auf diese Weise kann eine mannlose Fertigung, etwa in Form einer mannlosen Nachtschicht oder der Fertigung am Wochenende, realisiert werden. [3] Abbildung 1 zeigt eine solche automatisierte Fertigungszelle.

Abb. 1: Automatisierte Fertigungszelle mit Be- und Entladeautomatisierung und Kompaktlager [3]

Im Zusammenhang mit Investitionsentscheidungen stellt sich häufig die Frage, wie wirtschaftlich eine Investition in eine solche Automatisierungslösung ist. Dieser Frage soll im Rahmen dieser Abschlussarbeit nachgegangen werden.

Zielsetzung der Arbeit

Da diese Frage vor allem im Vertrieb bei der Beratung von potentiellen Kunden häufig auftritt und diskutiert wird, soll im Rahmen dieser Arbeit ein Tool erstellt werden, mit dem die Wirtschaftlichkeit von Investitionen in die Automatisierungslösungen berechnet werden kann. Die Arbeit wird dabei in der Abteilung *Produktmanagement Automation* bei TRUMPF durchgeführt. Das erstellte Tool soll anschließend dem Vertrieb vom Produktmanagement zur Verfügung gestellt werden. Um sowohl beim Vertrieb als auch bei den Kunden eine hohe Akzeptanz des Tools zu gewährleisten, sollen bei der Erstellung des Tools die üblichen Vorgehensweisen der Kunden bei Investitionsentscheidungen sowie die Anforderungen des Vertriebs berücksichtigt werden.

Verfahren der Investitionsrechnung

Zur Ermittlung der Wirtschaftlichkeit von Investitionen können verschiedene Verfahren der Investitionsrechnung eingesetzt werden. In diesem Zuge findet vor allem eine quantitative Berechnung der Wirtschaftlichkeit statt. Qualitative Aspekte werden bei diesen Berechnungen nicht berücksichtigt. Im Zusammenhang mit den Investitionsrechnungen spielen neben der voraussichtlichen Nutzungsdauer des Investitionsobjekts auch die jährlichen Rückflüsse als Differenz von Einzahlungen und Auszahlungen eine wichtige Rolle. [6] Die jährlichen Auszahlungen ergeben sich dabei hauptsächlich aus den vom Investitionsobjekt verursachten Kosten, während die Einzahlungen auf zusätzlichen Umsatzerlösen oder auch auf eingesparten Kosten basieren. [6]
Die Verfahren der Investitionsrechnung lassen sich nun allgemein in zwei Gruppen unterteilen.

- Bei den **statischen Verfahren** wird der Zeitbezug der Zahlungen vernachlässigt und für alle Perioden die gleichen Durchschnittswerte angenommen. [5]
- Im Gegensatz dazu berücksichtigen die **dynamischen Verfahren** den Zeitbezug der über die Nutzungsdauer des Investitionsobjekts hinweg anfallenden Ein- und Auszahlungen und machen diese durch eine Abzinsung auf den Investitionszeitpunkt vergleichbar. [5]

Abbildung 2 gibt eine Übersicht über die verschiedenen statischen und dynamischen Verfahren der Investitionsrechnung.

Abb. 2: Übersicht über Verfahren der Investitionsrechnung [5]

Vorgehen und Ausblick

Im Rahmen der Forschung zu dieser Arbeit werden qualitative Experteninterviews mit Kunden aus verschiedenen Kundensegmenten durchgeführt, um ein möglichst umfassendes Bild über das Vorgehen der Kunden bei ihren Investitionsentscheidungen zu erhalten. Bei diesen Interviews geht es unter anderem auch darum, welche Verfahren der Investitionsrechnung die Kunden vorzugsweise einsetzen und wie sie die Ein- und Auszahlungen, die durch eine Automatisierungslösung entstehen, ermitteln. Des Weiteren wird auch ein Experteninterview mit einem Mitarbeiter des Smart Factory Consultings bei TRUMPF geführt, um dessen Erfahrungen aus der Praxis in die Erstellung des Tools mit einzubeziehen. Um die Sichtweisen aller Stakeholder des Projekts zu berücksichtigen, wird außerdem ein Interview mit einem erfahrenen Mitarbeiter des Vertriebs durchgeführt, um so die Anforderungen des Vertriebs an das Tool zu ermitteln. Basierend auf den Auswertungen der Experteninterviews werden die Anforderungen an das zu erstellende Tool abgeleitet. Unter Berücksichtigung dieser Anforderungen wird dann ein Tool erstellt werden, welches später vom Vertrieb bei der Kundenberatung zur Berechnung der Wirtschaftlichkeit der Investition in eine Automatisierungslösung im individuellen Fall eingesetzt werden kann. Abschließend soll das Tool noch anhand realer Maschinendaten aus der Praxis sowie durch das Feedback ausgewählter Kunden validiert werden. Im Falle einer positiven Validierung kann das Tool dann in Zukunft bei der täglichen Arbeit des Vertriebs und des Produktmanagements eingesetzt werden. Auch eine zukünftige Erweiterung oder Anpassung des Tools ist denkbar, sodass neben Automatisierungslösungen für 2D-Laserschneidmaschinen auch Lösungen für andere Arten von Werkzeugmaschinen betrachtet werden können.

Literatur und Abbildungen

[1] HPW Hagelberg Präzisionswerkzeuge GmbH. Die Zukunft der CNC-Lohnfertigung: Warum sich die Investition in automatisierte Fertigung lohnt. https://www.werkzeug-einsatz-optimierung.de/blog/die-zukunft-der-cnc-lohnfertigung-warum-sich-die-investition-in-automatisierte-fertigung-lohnt-hpw-hagelberg-sonderwerkzeuge-strategieberatung, 2023.

[2] Roover GmbH. Smart Factory. https://roover.eu/smart-factory-industrie-4-0/, 2023.

[3] TRUMPF SE u Co KG. Automatisierung: LiftMaster Compact. https://www.trumpf.com/de_DE/produkte/maschinen-systeme/automatisierung/automatisierung-fuer-2d-laserschneidmaschinen/liftmaster-compact/, 2024.

[4] TRUMPF SE u Co KG. Automatisierungskomponenten für 2D-Laserschneidmaschinen. https://www.trumpf.com/de_DE/produkte/maschinen-systeme/automatisierung/automatisierung-fuer-2d-laserschneidmaschinen/, 2024.

[5] Jean-Paul Thommen et al. *Allgemeine Betriebswirtschaftslehre: Umfassende Einführung aus managementorientierter Sicht*. Springer Gabler Wiesbaden, 10 edition, 2023.

[6] Dietmar Vahs and Jan Schäfer-Kunz. *Einführung in die Betriebswirtschaftslehre*. Schäffer-Poeschel Verlag Stuttgart, 7 edition, 2015.

Design und Implementierung eines Echtzeit-Datenbanksystems zur Verwaltung von Positionsdaten anderer Verkehrsteilnehmer im Vehicle-to-everything (V2X) Umfeld

Finn Guist **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Vector Informatik GmbH, Weilimdorf

Einführung V2X

V2X (Vehicle-to-Everything) ist eine fortschrittliche Kommunikations-Technologie, die darauf abzielt, Fahrzeuge mit ihrer Umgebung zu verbinden und somit die Verkehrssicherheit zu erhöhen, den Verkehrsfluss zu optimieren und neue Mobilitätsdienste zu ermöglichen. Diese Technologie umfasst verschiedene Kommunikationsformen wie in Abbildung 1 zu sehen, darunter V2V (Vehicle-to-Vehicle), V2I (Vehicle-to-Infrastructure), V2P (Vehicle-to-Pedestrian) und V2N (Vehicle-to-Network). V2X ermöglicht es Fahrzeugen, miteinander sowie mit der Infrastruktur und anderen Verkehrsteilnehmern zu kommunizieren, um Informationen in Echtzeit auszutauschen. Dadurch können Unfälle vermieden und Gefahren früher erkannt werden indem Fahrzeuge beispielsweise vor abrupten Bremsmanövern anderer Autos oder vor gefährlichen Straßenabschnitten gewarnt werden. Auch Ampeln und Verkehrsmanagementsysteme können effizienter gestaltet werden, um somit den Verkehrsfluss zu verbessern. Darüber hinaus ermöglicht V2X neue Mobilitätsdienste wie automatisiertes Fahren, Car-to-Car-Services und verbesserte Navigationssysteme, die auf Echtzeitdaten basieren.
Dennoch stehen der Verbreitung von V2X einige Herausforderungen gegenüber. Die Standardisierung der V2X-Kommunikation ist von großer Bedeutung, um Interoperabilität zwischen verschiedenen Fahrzeugen und Infrastrukturen zu gewährleisten. Zudem sind robuste Sicherheits- und Datenschutzmaßnahmen notwendig, da sensible Daten übertragen werden. Die größte Schwierigkeit besteht allerdings in der Verbreitung der Technologie, da die Kommunikation nur erfolgen kann wenn möglichst viele Fahrzeughersteller diese verbauen.

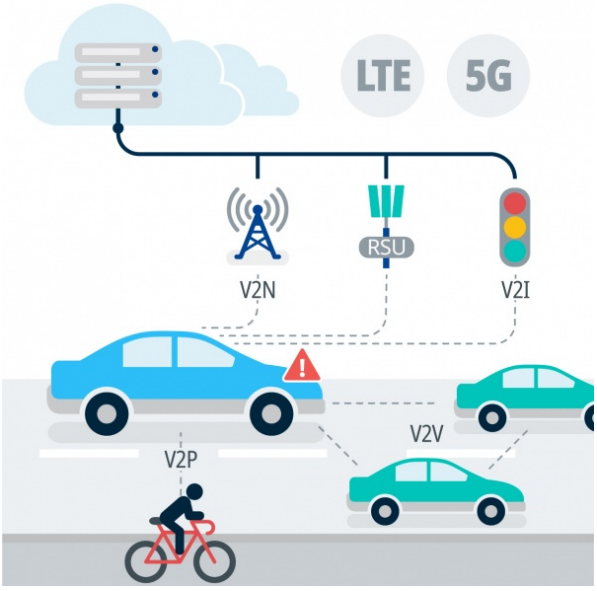

Abb. 1: V2X Kommunikation [1]

Die Standardisierung erfolgt nicht zentral, wodurch es unterschiedliche regionale Standards gibt. Hauptsächlich unterscheiden sich diese in der Verwendung verschiedener Protokolle und teilweise auch diverse Übertragungstechnologien, wodurch sie nicht mehr kompatibel sind. Die bekanntesten Standards sind der europäische, der amerikanische und der chinesische, wobei in dieser Arbeit und insbesondere diesem Artikel vom europäischen Standard ausgegangen wird.

Motivation

Die Kommunikation im europäischen Standard verläuft über verschiedene Nachrichtentypen. Die bisher am häufigsten verwendeten sind hierbei die Cooperation

Awareness Message (CAM), sowie die Decentralized Environmental Notification Message (DENM) und sollen zunächst zur Erklärung dienen. Die CAM wird als Broadcast durch einen Verkehrsteilnehmer zyklisch versendet und teilt allen umliegenden Fahrzeugen mit einer vorgeschriebenen Frequenz den aktuellen Standort mit. Die DENM wird dagegen situationsbedingt versendet, um Informationen über gefährliche Verkehrssituationen, wie Unfälle oder Hindernisse, zwischen Fahrzeugen und der Infrastruktur in Echtzeit zu teilen. Die Inhalte dieser Nachrichten, wie Standort, Zeitpunkt, Geschwindigkeit und Bewegungsrichtung sollen nun V2X Applikationen zur Verfügung gestellt werden. Hierfür müssen sie zwischengespeichert und anhand von bestimmten Kriterien abgefragt werden können. Thema dieser Arbeit ist die Konzeptionierung, Planung und Implementierung einer V2X Database, welche die aufgezählten Anforderungen übernimmt.

Anforderungsanalyse

Grundlage der V2X Database ist die vom Europäische Institut für Telekommunikationsnormen (ETSI) standardisierte Local Dynamic Map (LDM). [3] Basierend auf den Schnittstellen werden die Anforderungen für Kompatibilität definiert. V2X Services und Applikationen können sich als Data Provider registrieren, um Daten an die Datenbank zu liefern. Diese können dann wiederum von Data Consumern abgefragt werden. Hierfür sollen zwei Mechanismen zur Verfügung stehen, zum einen eine Query Funktion, mit der Daten nach verschiedenen Filtern abgefragt werden können und zum anderen eine Publish/Subscribe Funktion, mit der sich Consumer auf bestimmte Datenupdates subscriben können und diese erhalten, sobald sie verfügbar sind.

Einbinden in V2X Stack

Der durch die ETSI standardisierte europäische Softwarestack orientiert sich nach dem OSI-Modell und wird von Unten nach Oben durch eine Access Layer für den Datenempfang, der Networking & Transport Layer mit Netzwerkprotokollen, sowie der Facility Layer für die Nachrichtenservices, welche beispielsweise für CAM und DENM verantwortlich sind zusammengesetzt. [2] Aufsetzend darauf besteht die Application Layer, welche die Anwendungen für die Use-Cases von V2X enthält.

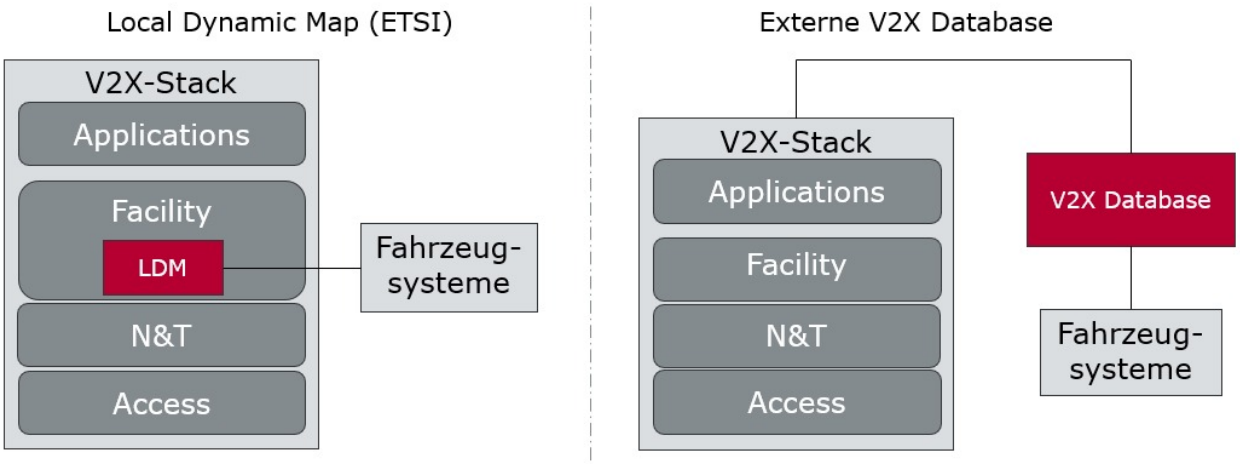

Abb. 2: Gegenüberstellung LDM und V2X Database [1]

Wie in Abbildung 2 zu sehen bestehen zwei Ansätze die Database in den Stack einzubinden, wobei der linke Teil des Schaubilds den Ansatz der ETSI mit der LDM darstellt. Der in der Arbeit verfolgte Ansatz besteht allerdings darin die Database extern des Stacks zu platzieren. Dies führt dazu, dass andere Fahrzeugsysteme wie ADAS auch auf die gespeicherten Objekte zugreifen kann, ohne Zugriff auf die Facility Layer des Stacks haben zu müssen. Ein weiterer Vorteil ist die unabhängige Ressourcenzuweisung sowie Skalierbarkeit, um auf unterschiedliche Mengen an benötigtem Speicher oder Rechenleistung reagieren zu können.

Ausblick

Zusammen mit einem Softwarearchitekturentwurf und Überlegungen über die Struktur eingehender und ausgehender, als auch der intern gehaltenen Daten, kann die Database implementiert werden. Daraufhin müssen Funktionalität und Zusammenspiel mit den Komponenten des Stacks getestet werden. Als abschließender Schritt werden Benchmarks durchgeführt, um quantitative Werte wie Latenz und Bandbreite überprüfen zu können.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] European Telecommunications Standards Institute. Intelligent Transport Systems (ITS); Communications Architecture. https://www.etsi.org/deliver/etsi_en/302600_302699/302665/01.01.01_60/en_302665v010101p.pdf, 09 2010.

[3] European Telecommunications Standards Institute. Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Local Dynamic Map (LDM). https://www.etsi.org/deliver/etsi_en/302800_302899/302895/01.01.01_60/en_302895v010101p.pdf, 09 2014.

Motion Forecasting on German Traffic Data

Jakob Haeringer **Thao Dang**

Department of Computer Science and Engineering, Esslingen University

Work carried out at Robert Bosch GmbH, Stuttgart

Introduction

Promoting trust in autonomous driving algorithms and minimizing the hazards of driving are critical in the realm of autonomous vehicles. The objective of this study aligns with the principles of Vision Zero, a multinational road traffic safety initiative launched in Sweden in 1997 [3] aiming for a roadway system free of fatalities or serious injuries. In the context of autonomous driving, motion forecasting is critical to enable vehicles to anticipate the paths of other road users, allowing autonomous vehicles to choose the safest routes and avoid potential collisions.
To make a meaningful contribution to this field, my thesis focuses on creating a dataset from a self-driving fleet operating in Stuttgart, Germany, and Sunnyvale, California. The aim is to investigate features that improve multi-agent trajectory prediction. Furthermore, this study undertakes the validation of an open-source transformer-based neural network using a publicly available dataset and compares its performance with a novel dataset curated from different geographical locations.

Dataset Creation

The raw CSV data from the self-driving fleet was converted into the Argoverse 2 (AV2) format. This format includes basic information about the agents, such as their positions (x, y), velocities (x, y), directions, and object types. It also provides high-resolution (HD) map data, including features like pedestrian crossings and lane segments, offering a comprehensive view of the scene. AV2 is an open-source dataset specifically designed for autonomous vehicle research, containing autonomous driving data and HD maps from six U.S. cities: Austin, Detroit, Miami, Pittsburgh, Palo Alto, and Washington, D.C. While AV2 offers various datasets, this study focuses on the Argoverse 2 Motion Forecasting Dataset. [4]
Converting the data into this open-source format allows for a meaningful comparison between neural networks trained on our dataset and those trained on existing datasets. This comparison can provide crucial insights into model performance and identify areas for improvement. Moreover, the limitations of the AV2 dataset format, in terms of the number of features used, present an opportunity to experiment with additional features and adapt neural networks accordingly.

	Waymo	Argoverse 2	Ours
# Scenarios	104k	250k	90k
Average Track Length	7.04s	5.16s	3.47s
Total Time	574 h	763h	274 h
Scenario Duration	9.1s	11s	11s
Test Forecast Horizon	8s	6s	6s
Sampling Rate	10 Hz	10 Hz	10 Hz
# Cities	6	6	2
Unique Roadways	1750 km	2220 km	-
Avg. # Tracks Per Scenario	-	73	47
# Evaluated Object Categories	3	5	5
Mined For Interestingness	✓	✓	✗
Download Size	1.4 TB	58 GB	19 GB
Lidar Data Incl.	✓	✗	✗

Fig. 1: Dataset comparison [2]

As shown in Fig. 1, our created dataset can be compared with the original AV2 dataset, but it is roughly a third of the size, containing 90,000 scenarios compared to AV2's 250,000. This smaller size still offers significant data for meaningful analysis. Another key aspect worth mentioning is that AV2 and Waymo [1] scenarios are specifically mined for interestingness, which adds a layer of curated complexity that could be insightful when comparing results.
Furthermore, the Waymo dataset, another popular dataset, is included in the comparison. This highlights that researchers use different approaches regarding scenario duration and forecast horizons. For instance, our dataset and AV2 have longer scenario durations compared to Waymo. Additionally, Waymo includes Lidar data, which is not yet incorporated in AV2 or our dataset, pointing to potential avenues for incorporating more detailed sensory inputs in future iterations.
Both Waymo and AV2 focus on data within the U.S., while our dataset includes scenarios from Stuttgart, Germany, adding geographic complexity. This can contribute to understanding how models trained on

U.S. data adapt to German traffic conditions.

QCNet

QCNet [5] has shown notable performance in trajectory prediction, particularly within the Argoverse Motion Forecasting Benchmarks, securing top positions such as 1st place in both Single-Agent and Multi-Agent Motion Forecasting Benchmarks. It also won the Argoverse 2 Multi-Agent Motion Forecasting Challenge at the CVPR 2023 Workshop on Autonomous Driving. [6]
A significant factor in QCNet's performance is its architecture. The model features a scene encoder with roto-translation invariance in space, which aids in accurate multi-agent prediction. Additionally, a scene encoder with translation invariance in time supports continuous processing. QCNet utilizes a two-stage DETR-like (DEtection TRansformer) trajectory decoder to predict multi-modal and long-term trajectories.
The selection of QCNet for this study is based on its documented efficacy in multi-agent prediction and its open-source availability on GitHub [6]. This open-source nature facilitates replicability and transparency in research. Comparing neural networks trained on this new dataset with those trained on existing datasets is expected to provide insights into performance variations and potential improvements.
The design of QCNet focuses on improving trajectory prediction for autonomous vehicles. Key aspects of its operation include:
Scene Encoding with Query-Centric Paradigm: Traditional trajectory prediction models often grapple with the re-normalization and re-encoding of scene elements, impeding online prediction efficiency. QCNet circumvents this challenge by embracing a querycentric paradigm for scene encoding. By processing all scene elements in their local spacetime reference frames, redundant computations are minimized, resulting in lower inference latency.
Multi-Modal Trajectory Decoding: Addressing the inherent multimodality of agents' future behaviors, QCNet uses anchor-free queries for recurrent trajectory proposal generation. This adaptive approach allows the model to tailor trajectory proposals based on diverse scene contexts and prediction horizons. Subsequent refinement through anchor-based queries enhances prediction quality, ensuring adaptability and robustness. [5]

Initial Training Results

The initial training on our dataset used the same hyperparameter settings as those used for QCNet on the AV2 dataset.

Dataset	minFDE ↓	minADE ↓	MR ↓	brier-minFDE ↓
AV2	1.25	0.72	0.16	1.87
Ours	0.86	0.62	0.09	1.49

Fig. 2: Initial training results [2]

Fig. 2 shows that training on our dataset leads to lower values in all metrics compared to the original AV2 dataset. [6] However, it is important to note that AV2 was mined for interestingness, adding complexity that our dataset does not currently incorporate. The equal distribution in object categories for example, including vehicles, pedestrians, buses, motorcyclists, and cyclists, as seen in AV2 compared to our predominantly vehicle-centric dataset, impact key metrics such as Minimum Final Displacement Error (minFDE), Minimum Average Displacement Error (minADE), and Miss Rate (MR) due to the increased complexity and diversity in motion patterns. Additionally, the training process now requires less hardware. Initially, it needed 8 Nvidia RTX 3090 GPUs (24 GB each), but with our dataset, it can be accomplished using 8 Nvidia V100 GPUs (16 GB each).

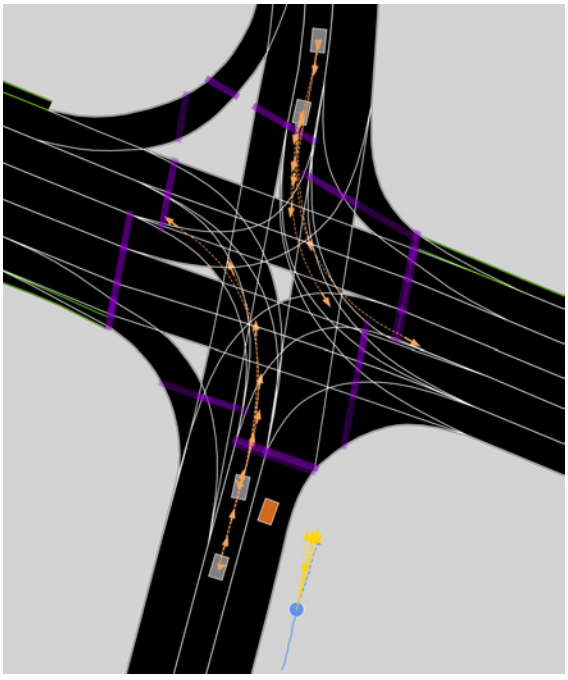

Fig. 3: Qualitative results of predicted agent trajectories and HD map data trained with our dataset. The lanes are marked with different colors: vehicle lanes (white) bike lanes (green), and pedestrian crossings (purple). The ego-vehicle is shown in orange. [2]

Fig. 3 showcases predicted trajectories of the QCNet, with the focal agent (agent with the main focus in the scene) represented in blue, where the solid blue line

denotes its history (5 seconds) and the blue dashed line signifies the ground truth of the future trajectory (6 seconds). The six yellow lines depict the predicted trajectories of the focal agent, specifically a pedestrian in this instance. Additionally, the gray rectangles illustrate other scored agents (agents seen for each timestep in the 11-second scenario) in the scene, with solid and dashed lines indicating their history and future trajectories, respectively.
One can see that the prediction of the focal agent in the scene was very accurate. Some of the predictions for the scored agents correctly predict that the vehicles did not move (orange arrow inside the gray agents), but some also predict a crossing of the intersection (visible by the orange dashed lines). This example illustrates how adding more features, such as traffic light states, could improve the accuracy of the network to prevent incorrect predictions of vehicles crossing the intersection, and provides a good opportunity to be used as a qualitative example for later iterations of the dataset and neural network.

Outlook

The initial success in training QCNet on our dataset, even without additional features, is a promising start. This initial progress ignites our anticipation for future advancements leveraging more comprehensive datasets and customized neural network architectures. The integration of additional features, such as lane-specific speed limits, nuanced kinematic data detailing agent poses within lanes, and traffic light states at intersections presents an opportunity to significantly improve the model's predictive capabilities, particularly in complex scenarios like intersections. These additional insights could prove critical in fine-tuning trajectory predictions and improving overall performance. Furthermore, adjusting the hyperparameters and using the pre-trained checkpoint of the original AV2 dataset during training on our dataset offer further opportunities for refinement and improvement.

References and figures

[1] S. Ettinger et al. Large Scale Interactive Motion Forecasting for Autonomous Driving : The Waymo Open Motion Dataset. *arXiv*, 2021.

[2] Own representation.

[3] Trafikverket. Vision Zero - no fatalities or serious injuries through road accidents. https://www.roadsafetysweden.com/about-the-conference/vision-zero—no-fatalities-or-serious-injuries-through-road-accidents/, 2019.

[4] B Wilson et al. Argoverse 2: Next Generation Datasets for Self-Driving Perception and Forecasting. *arXiv*, 2023.

[5] Z. Zhou, J. Wang, Y. Li, and Y. Huang. Query-Centric Trajectory Prediction. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada*, 2023.

[6] Zikang Zhou. QCNet: An Elegant, Performant, And Scalable Framework For Marginal/Joint Multi-Agent Trajectory Prediction. https://github.com/ZikangZhou/QCNet, 2023.

Ein Modell zur Kostenabschätzung der Datenübertragung bei Cloud-basierter Sammlung von Fahrzeugdaten

Lara Heidenwag **Rainer Keller**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Vector Informatik GmbH, Stuttgart-Weilimdorf

Motivation und Problemstellung

Moderne Fahrzeuge sind mit einer Vielzahl von Sensoren ausgestattet, die Daten über Motorleistung, Fahrprofil und die Umgebung des Fahrzeugs generieren. Diese Datenflut eröffnet neue Möglichkeiten für die Automobilindustrie, birgt aber auch Herausforderungen. Laut [2] sind vernetzte Fahrzeuge mit mehr als 200 verschiedenen physikalischen Sensoren ausgestattet, die Signale und Messwerte an elektronische Steuergeräte (ECUs) übertragen, welche von hier weiter verarbeitet und verteilt werden. Diese Datenströme sind nicht nur innerhalb des Fahrzeug essentiell sondern stellen auch eine wertvolle Informationsquelle dar, die für die Analyse und die Optimierung von Fahrzeugfunktionen, Diagnosezwecken oder die Bereitstellung innovativer Konnektivitätsdienste genutzt werden kann.
Die Verarbeitung dieser Daten, die täglich bis zu mehrere Terrabyte akkumulieren ist jedoch mit erheblichen Kosten verbunden. Dabei gilt es, sicherzustellen, dass möglichst nur relevante Daten übertragen werden, um einerseits die Kosten zu minimieren, andererseits aber die Qualität und den Nutzen der Daten sicherzustellen. Die Herausforderung besteht in der effizienten und kosteneffektiven Entwicklung von Lösungen, bei denen die Vollständigkeit der übertragenen Daten und die damit verbundenen Kosten in einem ausgewogenen Verhältnis zueinander stehen. Um dies zu beurteilen ist es hilfreich Kennzahlen der jeweiligen Datenübertragung zu ermitteln.
Das Ziel dieser Arbeit ist daher, ein Kostenmodell zu entwickeln, welches anhand verschiedener Faktoren eine Kostenabschätzung vornimmt und auf diese Weise Anhaltspunkte für eine Optimierung der Übertragung liefert. Die Berücksichtigung der verschiedenen Datenquellen, Protokolle, Datenformate und Serialisierungsmethoden ist dabei ein entscheidender Faktor. Die von hoher Heterogenität und Dynamik geprägte Datenübertragung im Fahrzeug soll möglichst präzise modelliert werden. Primär bezieht sich das Modell auf die Vorhersage der Daten- und Informationsmenge des Netzwerkverkehrs zwischen Fahrzeug und Cloud. Es werden aber auch andere Kostenpunkte miteinbezogen.

Verwandte Arbeiten

Probleme dieser Art sind bereits aus dem Umfeld des Internet of Things (IoT) bekannt, wo ebenfalls große Datenmengen in hohen Frequenzen effektiv übertragen, verarbeitet und gespeichert werden müssen. Die mit dem Begriff "Big Data"verbundenen Herausforderungen werden durch verschiedene Lösungsansätze adressiert, wie zum Beispiel durch Datenfilterung, Aggregation oder eine dynamische Anpassung der Übertragungsfrequenz [3]. Es gibt jedoch nur wenige Ansätze, die darauf abzielen, die anfallende Datenmenge initial abzuschätzen, um ein fundiertes Verständnis für die Datenmengen und die damit verbundenen Kosten zu entwickeln.

Definition von Kosten

Bei der Datenübertragung können die Kosten in direkte und indirekte Kosten unterteilt werden. Direkte Kosten beziehen sich dabei auf Gebühren, die abhängig von der übertragenen Datenmenge anfallen (beispielsweise pro Kilobyte Daten). Der Prozess der Datenübertragung bindet außerdem Ressourcen wie Bandbreite, Rechenleistung und Speicherplatz. Dies kann zu Leistungseinbußen, Verzögerungen (Latenz) und Beeinträchtigungen der Zuverlässigkeit führen, die in diser Arbeit als indirekte Kosten berücksichtigt werden.

Herausforderungen

Die Datenübertragung von der Quelle im Fahrzeug bis zur Cloud zeichnet sich durch eine hohe Heterogenität und Dynamik aus, was zu einer Reihe von Herausforderungen bei der Modellierung führt.
Die Daten werden aus vielen verschiedenen Quellen gesammelt, die unterschiedliche Kommunikationspro-

tokolle wie CAN, HTTP oder SOME/IP verwenden. Diese Protokolle können entweder ereignis- oder abfragebasiert sein, was zu unterschiedlichen Datenübertragungsmustern führt. Bei ereignisbasierten Protokollen ist die Häufigkeit der zu erwartenden Ereignisse oft nicht bekannt. Dies erfordert Annahmen oder zusätzliche Informationen bei der Modellierung, die die Genauigkeit der Vorhersage beeinträchtigen.
Ein weiteres Problem ist die Vorhersage der Datenmenge bei dynamischen Datenstrukturen. Bei der Abfrage eines Reifendrucks wird immer dieselbe Datenstruktur und Datengröße zurückgeliefert, was die Vorhersage vereinfacht. Bei anderen Datenstrukturen, wie beispielsweise einem Fehlerspeicher, ist dies jedoch nicht der Fall. Um in diesem Kontext Vorhersagen zu treffen, muss bei der Modellierung beispielsweise mit konfigurierbaren Durchschnittswerten oder Vorhersagemodellen gearbeitet werden.
Die übertragenen Daten können außerdem je nach Datentyp und Serialisierungsmethode unterschiedliche Strukturen und Größen aufweisen. Beispielsweise beansprucht eine Fließkommazahl im Binärformat in gängigen Programmiersprachen und Prozessorarchitekturen 4 Bytes, während die Speichermenge in JSON je nach Dezimalstellenanzahl variiert. Diese Variabilität erschwert die genaue Abschätzung der Datenmenge und die einheitliche Modellierung.

Vectors Vehicle Data Collector

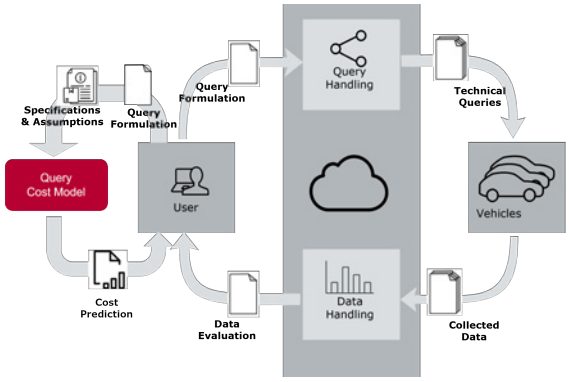

Abb. 1: Gesamtkonzept des Vehicle Data Collector und Positionierung des Kostenmodells (eigene Darstellung basierend auf Vorlage von [4]) [1]

Der Vehicle Data Collector (VDC) (vgl. Abbildung 1) ist ein Produkt der Vector Informatik GmbH, welches dem Kunden die einfache Implementierung von Datensammelkampagnen im Fahrzeug ermöglicht [4]. Ein anschauliches Anwendungsbeispiel ist die Erfassung der Raddrehzahl jedes Rads in Verbindung mit dem Reifendruck über einen Zeitraum von zwei Wochen. Diese Abfrage umfasst 8 Datenpunkte, für jedes der vier Räder je den Reifendruck und die Raddrehzahl. Die Abfragen (*Queries*) werden in einer Cloud verwaltet und in Form von technischen Queries an das betreffende Fahrzeug gesendet. Die gesammelten Daten werden schließlich in die Cloud zurückgesendet (vgl. Abbildung 1). Letztere können für Analyse- oder Optimierungszwecke verwendet werden. Verschiedene Datenquellen werden mit sogenannten DCAs (Data Class Adapters) angebunden. Die DCAs nutzen Übertragungsprotokolle wie CAN, SOVD oder SOME/IP.

Modellierung

Eine Query bezeichnet die Definition und Konfiguration eines Datensammelszenarios über einen gewissen Zeitraum. Dieses kann mehrere Datenpunkte umfassen, welche in regelmäßigen oder unregelmäßigen Abständen Daten liefern bzw. abgefragt werden. Wie in Abbildung 1 ersichtlich, erfolgt eine parallele Positionierung des Kostenmodells in Relation zum VDC. Dabei werden dem Modell sowohl die Query-Definition als auch eine zusätzliche, optionale Konfigurationsdatei (die *Assumptions*) als Eingabe übermittelt.
Je Query sind folgende Parameter für die Kostenabschätzung notwendig:
- Dauer der Datenkollektion
- Übertragungskonfiguration
- Datenpunkte

Die Übertragungskonfiguration umfasst Informationen bezüglich der maximalen Buffergröße je Query, des Transmission Overheads sowie der minimalen Transmission Distance Time (hier definiert als der minimale Abstand, in dem Pakete versendet werden).
Zusätzlich werden je Datenpunkt folgende Parameter übergeben:
- Datentyp mit Serialisierungsgröße
- erwartetes Datenintervall*
- dynamische Datengröße*
- Overhead Funktion*

Der jeweilige Datentyp wird je nach Serialisierungsart in eine Datentypgröße umgewandelt. Alle mit Stern (*) markierten Felder sind nicht notwendigerweise in der Query-Definition enthalten und können vom Nutzer in den Assumptions übergeben werden. Andernfalls wird von festgelegten Standardwerten ausgegangen. Bei abfragebasierten Übertragungsprotokollen wird das Datenintervall in der Query spezifiziert, andernfalls muss dieses in den Assumptions angegeben werden. Für verschiedene Serialisierungsarten kann außerdem eine spezifische Overhead-Funktion definiert werden, die dem Modell als Annahme übergeben wird. Die Verwendung von Overhead-Funktionen ermöglicht eine präzisere Modellierung des Overheads im Vergleich zur Verwendung von festen Overhead-Konstanten. Da-

durch wird eine flexible Modellierung unterschiedlicher Serialisierungsformate möglich.
Zur präzisen Abschätzung der anfallenden Datenmenge erfolgt eine Differenzierung in:
1. die **Informationsmenge** - die Speichermenge, die benötigt wird um die Daten im Fahrzeug darzustellen
2. die **serialisierte Informationsmenge** - die Datenmenge, welche benötigt wird um diese Informationen je nach Serialisierungsart zu übertragen und
3. den zusätzlichen **Serialisierungs- und Protokolloverhead**.
Dies erlaubt einen effektiven Vergleich diverser Serialisierungsarten. Ebenso ist es möglich, verschiedene Datenkollektionsszenarien hinsichtlich ihrer Realisierbarkeit zu analysieren und die Auswirkung einzelner Datenpunkte auf die Gesamtkosten zu untersuchen.

Berechnung

Die Berechnung der anfallenden Datenmenge erfolgt durch die Simulation der Ankunft der Daten der einzelnen Datenpunkte und deren anschließende Verpackung in verschiedene Pakete in Abhängigkeit von der definierten Übertragungskonfiguration. Zunächst wird das kleinste gemeinsame Vielfaches (LCM) der jeweiligen erwarteten Datenintervalle berechnet. Innerhalb dieser Periode wird das Ankommen der Daten simuliert. Die Zusammenfassung der einzelnen Datenpunkte in Pakete erfolgt in Abhängigkeit vom definierten minimalen Übertragungsabstand bzw. der maximalen Upload-Größe. Die jeweiligen Datengrößen, darunter fallen u.a. die Informationsmenge oder der Overhead, werden bei der Hochrechnung kontinuierlich in verschiedenen Variablen mitgezählt. Nach der Simulation einer Periode (LCM) wird die anfallende Datenmenge pro Periode auf die anfallende Datenmenge pro Stunde bzw. die Dauer der Abfrage hochgerechnet. Des Weiteren wird die zu erwartende Zeitdauer, innerhalb derer ein Buffer Overflow zu erwarten ist, berechnet. Bei verschiedenen Fehlern, wie beispielsweise einer zu klein gewählten maximalen Upload-Size, wird eine Warnung ausgegeben. Auf diese Weise können grobe Fehler beim Entwurf der Datensammelkampagne bereits vor der Ausführunng identifiziert werden.

Ausblick

Erste Evaluierungen versprechen sehr gute Ergebnisse des Kostenmodells. Bei kleineren Szenarien liefert dieses Ergebnisse mit 95 bis 100% Genauigkeit. Diese hängt jedoch stark von den abgefragten Datentypen bzw. den getroffenen Annahmen ab. Eine genauere Evaluierung der Korrelation steht dabei noch aus. Ein weiteres interessantes Ergebnis ist die Bedeutung der möglichen Einsparung von Serialisierungsoverhead je nach gewähltem minimalen Abstand zwischen den Paketen. Je mehr Datenpunkte in einem Paket zusammengefasst werden, desto geringer ist der Serialisierungsoverhead. Für JSON kann die Datenmenge so um bis zu 50% reduziert werden.
Die erste Iteration des Modells fokussiert sich hauptsächlich auf die Abschätzung anfallender Datenmengen, was einen wichtigen ersten Schritt darstellt, um die Umsetzbarkeit und Kosten einer Datenkollektionskampagne zu bewerten. Zukünftige Iterationen des Modells könnten verschiedene Optimierungsschritte einbeziehen, wie beispielsweise das dynamische Anpassen der Datensammelfrequenz oder die Implementierung von Filterlogiken, bei denen Daten nur übertragen werden, wenn sie sich geändert haben. So könnten potentielle Optimierungsschritte zur Kosteneinsparungen identifiziert und bewertet werden. Dies ist jedoch mit einer signifikanten Erhöhung der Komplexität des Kostenmodells verbunden.
Zusammenfassend lässt sich sagen, dass das hier vorgestellte Kostenmodell einen wichtigen Beitrag zur Bewertung und Optimierung von Datenerhebungsprozessen im Fahrzeug leistet. Die bisherigen Ergebnisse zeigen, dass das Modell das Potenzial eines zuverlässigen und wertvollen Instruments zur Planung und Kostenkontrolle darstellt und eine solide Grundlage für weitere Forschungen und Entwicklungen bietet.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Chris Dickert. Network Overload? Adding Up the Data Produced By Connected Cars. https://www.visualcapitalist.com/network-overload/, 2023.

[3] Jelena Čulić Gambiroža and Toni Mastelić. Big Data Challenges and Trade-offs in Energy Efficient Internet of Things systems. In *2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, 2018.

[4] Stuttgart-Weilimdorf Vector Informatik GmbH. Data Collector - Die Basis zur flexiblen Analyse von Fahrzeugdaten. https://www.vector.com/de/de/products/products-a-z/embedded-components/microsar-connect/vehicle-data-collector/, 2024.

Kamera basierte Anomalien Erkennung in einem Industirellen Umfeld unter Verwendung von Künstlichen Neuronalen Netzen

Luka Henig **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Motivation

Unter einer Anomalieerkennung versteht man das Problem, Muster in Daten zu finden, welche nicht einer genau definierten Vorstellung von normalen Verhalten entsprechen. Anomalien werden häufig auch als Ausreißer oder Verunreinigungen bezeichnet und können in unterschiedlichsten und meist schwer vorhersehbaren Fällen auftreten [2]. Durch den Fortschritt der Forschung im Bereich der Maschine Vision werden immer mehr Ansätze zur Anomalieerkennung anhand von Bilddaten untersucht. Gerade im Bereich der Qualitätskontrolle können hier immense Vorteile für die Industrie geschaffen werden. Diese Arbeit zielt darauf ab einen Qualitätskontroll-Prozess im Wareneingang zu automatisieren. Hierfür werden verschiedene aktuelle Ansätze aus der Forschung einem in dieser Arbeit vorgestellten Modell gegenübergestellt. Die Auswertung findet auf Basis des öffentlichen MVTec AD Datensatzes sowie diversen selbst erstellten Datensätzen mit Bezug auf den reale Prozess statt.

Grundlagen

Das Forschungsfeld der Maschine Vision beschäftigt sich hauptsächlich mit der Bildklassifizierung und der Objekterkennung, hier haben sich deep-learning basierte Algorithmen mit robusten und präzisen Ansätzen bewehrt. Diese Algorithmen basieren weitest gehen auf künstlichen Neuronalen Netzen. Durch die schlechte Vorhersehbarkeit der Anomalien wird in den meisten Fällen ein unsupervised Lernansatz gewählt. Grundlegend sind neuronale Netze an den Aufbau eines Gehirns angelehnt und setzen sich aus Knoten, welche die Neuronen repräsentieren und aus Kanten, welche die Synapsen darstellen, zusammen. Abbildung 1 veranschaulicht dieses Konstrukt.

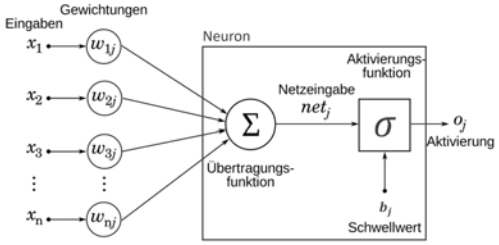

Abb. 1: Ein einfaches Neuron j mit den Eingabewerten (x_1, x_2, x_3,...,x_n), ihren dazugehörigen Gewichten (w_1j, w_2j, w_3j,...,w_nj), einem Bias b_j, der verwendeten Aktivierungsfunktion sigma und der Ausgabe o_j [3]

Die Neuronen zielen darauf ab, mehrere Eingangssignale zu gewichten, mit einem Bias-Wert aufzusummieren und mittels einer Aktivierungsfunktion an das nächste Neuron weiter zu geben. Durch diesen Aufbau ist es dem Netz möglich, über die Gewichte sowie die Bias-Werte lernbare Parameter anzupassen und Informationen zu lernen. Der Bias-Wert dient als Schwellwert, wann eine Ausgabe über die Aktivierungsfunktion stattfinden soll. Das künstliche Neuronale Netzt kann aus einer Vielzahl dieser Knoten in verschiedenen Ebenen zusammengesetzt werden. Abbildung 2 veranschaulicht diese Struktur [12].

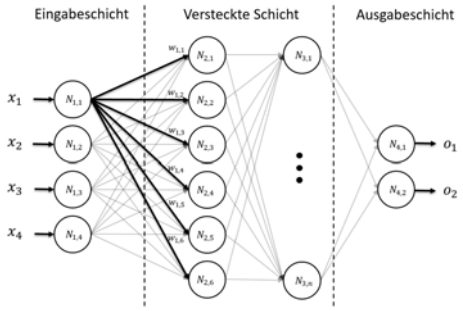

Abb. 2: Ein einfaches Neuronales Netz bestehend aus einer Eingabeschicht, zwei versteckten Schichten und einer Ausgabeschicht. Die Neuronen N_s,j sind mit ihrer Schichtebene s und einem Index j bezeichnet. [3]

Da diese Netze mit einer vektoriellen Form der Daten arbeitet, würde in der Bildverarbeitung ein extrem hoher Rechenaufwand entstehen. Um dem entgegenzuwirken, existieren Convolutional Neuronal Networks (CNN). Diese arbeiten mittels Matrizen oder Tensoren anstelle von Vektoren. Durch diese Dimensionserweiterung ist es möglich, mittels einer Faltung von speziellen rezeptiven Feldern (Filter) und Pooling Operationen die Informationen der Daten wesentlich kompakter zu verarbeiten. Durch die Mehrdimensionalität ist es möglich, mehrere Filter auf eine spezielle Position im Bild anzuwenden und gegebenenfalls mehre Merkmale gleichzeitig zu analysieren. Ein essenzieller Vorteil dieser Netze ist die Beibehaltung der räumlichen Information der Bilder. Abbildung 3 veranschaulicht eine solche Verarbeitung von einem Farbbild und mehren Filtern [5].

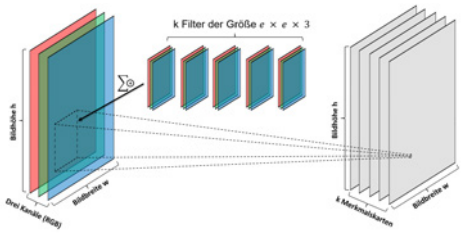

Abb. 3: Veranschaulichung einer Faltungsschicht mit dreidimensionalem Eingabebild und mehreren Filtern der Größe e [3]

Stand der Technik

Die aktuelle Forschung im Bereich der visuellen unüberwachten Anomalie Erkennung beruft sich auf ein breites Spektrum an Algorithmen. Es werden Algorithmen, welche Cluster-basiert [4] [7] [10], Normalizing-Flow-basiert [9], Language-basiert [6] [8] und Reconstruction-basiert [13] [11] [1]arbeiten analysiert. Diese verwenden diverse Methoden aus dem Bereich des maschinellen Lernens, darunter fällt unter anderem die Wissensdestilation mittels Schüler Lehrer Modelle, die Autoencoder Sturktur und Transformer. Auf die einzelne Funktionsweisen der vergleich Algorithmen kann an dieser Stelle nicht eingegangen werden. Dennoch soll diese Aufzählung der betrachteten Modelle eine Übersicht der Diversität bieten.

1. PaDIM [4]
2. DRAEM [13]
3. CPR [10]
4. ReConPatch [7]
5. EfficientAD [1]
6. DDAD [11]
7. PyramidFlow [9]
8. AnomalyGPT [6]
9. WinCLIP [8]

Realer Prozess

Der Prozess aus der realen Anwendung beschäftigt sich mit der Anlieferung von Waren für die Produktion. Die Kontrolle basiert vorerst auf den Transportbehältern selbst. Auf eine genauere Beschreibung des Prozesses, die generierten Daten, sowie die eigen entwickelte Methodik zur Anomalie Erkennung kann aus Geheimhaltungsgründen an dieser Stelle nicht weiter beschrieben werden.

Zielsetzung

Das Ziel dieser Arbeit ist es, eine Methodik zu entwickeln, welche die Qualitätskontrolle voll automatisiert. Hierfür werden in dieser Arbeit diverse Algorithmen und Methodiken aus der aktuellen Forschung untersucht und vergleichen. Im Anschluss wird ein eigenes Modell, welches auf einer Kombination aus den besten Methodiken der aktuellen Forschung basiert, für die betrachtete Anwendung implementiert. Zusätzlich werden verschiedene Bilderfassungsmöglichkeiten in Form von verschiedenen Perspektiven analysiert. Hierbei soll im Anschluss ein Fazit für einen realen Aufbau sowie eine optimierte Datengenerierung für das Training getroffen werden können.

Literatur und Abbildungen

[1] Kilian Batzner, Lars Heckler, and Rebecca König. EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies. -, 2024.

[2] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly Detection: A Survey. *ACM Comput. Surv.*, 2009.

[3] Eigene Darstellung.

[4] Thomas Defard et al. PaDiM: a Patch Distribution Modeling Framework for Anomaly Detection and Localization. -, 2020.

[5] Ian Goodfellow et al. *Deep Learning*. MIT Press, 2016.

[6] Zhaopeng Gu, Bingke Zhu, Guibo Zhu, Yingying Chen, Ming Tang, and Jinqiao Wang. AnomalyGPT: Detecting industrial anomalies using large Vision-Language Models. -, 2023.

[7] Jeeho Hyun, Sangyun Kim, Giyoung Jeon, Seung Hwan Kim, Kyunghoon Bae, and Byung Jun Kang. ReConPatch : Contrastive patch representation learning for industrial anomaly detection. -, 2024.

[8] Jongheon Jeong et al. WinCLIP: Zero-/Few-Shot Anomaly Classification and Segmentation. -, 2023.

[9] Jiarui Lei, Xiaobo Hu, Yue Wang, and Dong Liu. PyramidFlow: High-resolution defect contrastive localization using pyramid normalizing flow. *IEEE*, 2023.

[10] Hanxi Li, Jianfei Hu, Bo Li, Hao Chen, Yongbin Zheng, and Chunhua Shen. Target before shooting: Accurate Anomaly Detection and localization under one millisecond via Cascade Patch Retrieval. -, 2023.

[11] Arian Mousakhan, Thomas Brox, and Jawad Tayyub. Anomaly detection with conditioned denoising diffusion models. -, 2023.

[12] Vivienne Sze, Yu-Hsin Chen, Tien-Ju Yang, and Joel Emer. Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proc. IEEE Inst. Electr. Electron. Eng.*, 2017.

[13] Vitjan Zavrtanik, Matej Kristan, and Danijel Skocaj. DRAEM - A discriminatively trained reconstruction embedding for surface anomaly detection. *CoRR*, 2021.

Konzeption und Implementierung einer Datenpipeline zur automatisierten Auswertung von Marketing Automation Daten

Dennis Herzog **Jürgen Koch**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Problemstellung und Ziel der Arbeit

Im Zeitalter von Big Data, dem Internet of Things und der Digitalisierung werden zunehmend Daten generiert. Diese Daten enthalten Informationen zu Produkten, Kunden und Verkaufszahlen. Die Nutzung von Daten bietet Unternehmen die Möglichkeit zur Optimierung von Prozessen, Entscheidungen und Strategien. Somit kann eine kontinuierliche Datenauswertung den Unternehmenserfolg langfristig untermauern. Die Umsetzung einer kontinuierlichen Datenauswertung wird beispielsweise durch die Implementierung von automatisierten Datenpipelines ermöglicht. Pipelines können Daten aus einem Quellsystem extrahieren, in ein gewünschtes Zielformat transformieren und anschließend zur Nutzung bereitstellen oder ein automatisiertes Training von Machine-Learning-Modellen starten [5].
Ziel der Arbeit ist die Erstellung einer Datenpipeline, welche Daten aus einem Marketing-Automation-System extrahiert, veredelt und bereitstellt. Im Anschluss an die Bereitstellung der Daten soll zudem ein Report entwickelt werden, welcher die wichtigsten Kennzahlen zu existierenden Marketingkampagnen visualisiert. Die Veredelung der Daten umfasst den Entwurf eines flexiblen und verständlichen Datenmodells, welches mit weiteren Datenquellen verknüpfbar sein soll sowie die dafür notwendigen Transformationen. Die Pipeline soll täglich automatisiert ausgeführt werden, um einen aktuellen Datenbestand für den Report bereitzustellen. Ein weiteres Ziel der Arbeit ist, die notwendigen Komponenten unter Nutzung der im Unternehmen vorhandenen Architektur und Methoden zu implementieren.
Die Arbeit umfasst die Fachgebiete Data Engineering und Data Analytics. Diese Themen werden im Folgenden genauer beschrieben.

Data Engineering

Data Engineering beschreibt einen Satz von Operationen, die auf Daten angewendet werden, um diese für spätere Analysen durch Data Scientisten, Data Analysten oder andere Spezialisten vorzubereiten und bereitzustellen [6]. Abbildung 1 zeigt den Data Engineering Lifecycle von Reis und Housley [7].

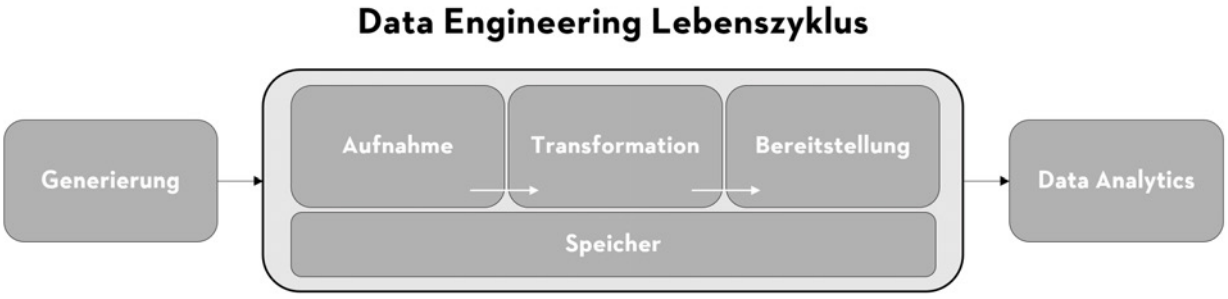

Abb. 1: Der Data Engineering Lebenszyklus in Anlehnung an [7]

Der Data Engineering Lifecycle beginnt mit der Extraktion von Daten, welche in einem beliebigen Quellsystem, wie einem IoT-Gerät oder einem Customer-Relationship-Management-System, erzeugt werden. Die Extraktion von Daten aus dem Quellsystem sowie die anschließende Aufnahme im Zielsystem können sowohl durch eine vom Quellsystem durchgeführte Push-Operation erfolgen als auch durch eine Pull-Operation, welche vom Zielsystem initiiert wird. Die anschließenden Transformationen sind notwendig, um die Daten in einem Datenmodell bereitzustellen, welches die Umsetzung aller Anforderungen der Stakeholder ermöglicht [7]. Gängige Datenmodelle beziehungsweise Datenmodellierungstechniken sind die Normalisierung von Daten, das Data Vault, das Stern-Schema und das Schneeflocken-Schema [1]. Im Anschluss an die Veredelung der Daten werden diese über eine Schnittstelle bereitgestellt, welche die Nutzung der Daten ermöglicht. Hierbei gilt es vor allem die Anforderungen an die Performanz des Systems zu beachten. Die Nutzung der Daten ist dem Fachgebiet Data Analytics zuzuordnen [7].

Data Analytics

Data Analytics beschreibt die Nutzung von Daten, um einen Mehrwert zu erzielen. Dabei gibt es verschiedene Techniken, welche sich in Kategorien gliedern lassen. Diese Kategorien sind in Abbildung 1 zu finden.

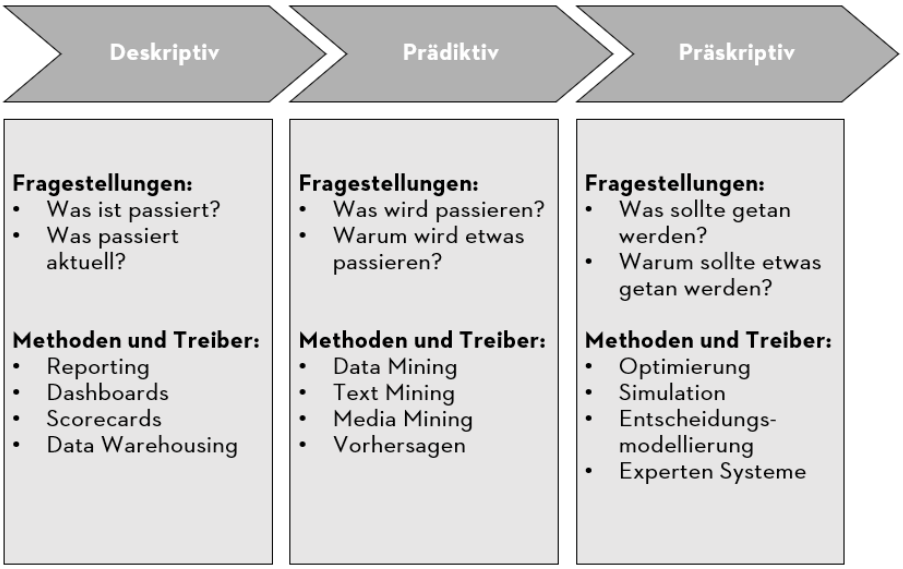

Abb. 2: Data Analytics Kategorien in Anlehnung an [4]

Deskriptive Analysen haben das Ziel, die Vergangenheit und die Gegenwart durch Daten zu beschreiben. Hierzu werden in der Regel statistische Berechnungen wie Durchschnitte oder Varianzen sowie einfache Transformationen und Aggregationen wie Summen angewendet. Anschließend werden die Ergebnisse durch Tabellen oder Visualisierung vorgestellt und an entsprechende Stakeholder verteilt. Dies wird häufig durch Reports oder Dashboards umgesetzt. Prädiktive Analysen zielen darauf ab, zukünftige Möglichkeiten und Events vorherzusagen. Hierbei werden in der Regel Methoden des Data Minings und Vorhersagealgorithmen eingesetzt. Beispielsweise können Regressionsalgorithmen genutzt werden, um zukünftige Verkaufszahlen zu prädizieren. Basierend auf diesen Vorhersagen können die Planung und die Entscheidungsfindung im Unternehmen optimiert werden. Präskriptive Analysen sind am komplexesten und haben das Ziel, die Frage, wie gehandelt werden sollte, zu beantworten. Beispielsweise werden bei einer präskriptiven Analyse verschiedene Szenarien simuliert. Durch eine systematische Optimierung wird das Beste der simulierten Szenarien ermittelt. Neben den bereits genannten Kategorien wird häufig auch die diagnostische Analyse als Kategorie genannt. Diagnostische Analysen sind zwischen den deskriptiven und prädiktiven Analysen einzuordnen und haben das Ziel die Fragestellung warum etwas passiert ist zu beantworten. Diese Analysen werden häufig im Falle von Problemen oder Fehlern eingesetzt, mit dem Ziel, die Ursache für die Problematik beziehungsweise den Fehler zu identifizieren. Hierfür werden diagnostische Algorithmen wie Korrelationsanalysen oder statistische Signifikanztests angewendet [3].

Rollen und User Stories

Im Unternehmenskontext lassen sich verschiedene Rollen identifizieren, welche Interesse an einer automatisierten Auswertung von Daten haben. Ein Product Owner hat Interesse an den Daten einer Quelle und möchte diese Daten für die Beantwortung seiner Fragestellungen nutzen. Daher beauftragt der Product Owner einen Data Analyst oder einen Data Scientist, die Daten auszuwerten, einen Bericht zu erstellen oder ein Machine-Learning-Modell zu trainieren. Anschließend wird geprüft, ob die Daten bereits im notwendigen Format vorliegen, um die Fragestellungen des Product Owners zu beantworten. Sofern dies nicht der Fall ist, wird ein Data Engineer mit der Bereitstellung der Daten im gewünschten Format beauftragt. Aufkommende Fragen bezüglich der Anbindung der Daten klärt der Data Engineer mit dem Data Owner. Hierzu zählen zum Beispiel die Art der Authentifizierung, eine Abschätzung bezüglich der Datenmenge, um eine effiziente Ressourcenplanung zu ermöglichen, sowie bevorzugte Zeitfenster zur Extraktion von Daten aus dem Quellsystem. Im Anschluss an die Klärung offener Fragen führt der Data Engineer die Anbindung der Datenquelle durch. Abbildung 3 zeigt jeweils eine Anforderung der genannten Rollen.

Rolle	User Story
Product Owner	Als Product Owner möchte ich einen Report, welcher täglich aktualisierte Kennzahlen zu Marketing Kampagnen zeigt, um optimierte Marketing Kampagnen zu gewährleisten.
Data Analyst	Als Data Analyst möchte ich ein einfach zu verstehendes und gut dokumentiertes Datenmodell, um die Richtigkeit meiner Analysen zu gewährleisten.
Data Engineer	Als Data Engineer fordere ich eine möglichst effiziente Pipeline, um die Betriebskosten gering zu halten.
Data Owner	Als Data Owner setze ich voraus, dass der Datenabzug keine Auswirkung auf die Leistungsfähigkeit des Quellsystems hat, sodass das System seinen Zweck uneingeschränkt erfüllen kann.

Abb. 3: Rollen und User Stories [2]

Literatur und Abbildungen

[1] Soham Bhatt and Deepak Sekar. Data Warehousing Modeling Techniques and Their Implementation on the Databricks Lakehouse Platform: Using Data Vaults and Star Schemas on the Lakehouse. https://www.databricks.com/blog/2022/06/24/data-warehousing-modeling-techniques-and-their-implementation-on-the-databricks-lakehouse-platform.html, 2022.

[2] Eigene Darstellung.

[3] Andrea De Mauro. *Data Analytics Made Easy*. Packt Publishing, 2021.

[4] Dursun Delen and Haluk Demirkan. Data, information and analytics as services. *Decision Support Systems*, 55:359–363, 2013.

[5] James Densmore. *Data Pipelines Pocket Reference*. O'Reilly Media, Inc., 2021.

[6] AltexSoft Incorporated. Data Engineering Concepts, Processes, and Tools. https://www.altexsoft.com/blog/what-is-data-engineering-explaining-data-pipeline-data-warehouse-and-data-engineer-role/, 2023.

[7] Joe Reis and Matt Housley. *Fundamentals of Data Engineering*. O'Reilly Media, Inc., 2022.

Analyse und Einsatzmöglichkeiten leichtgewichtiger Webframeworks in professionellen Entwicklungsprojekten

Rico Hofmann **Harald Melcher**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma IT-Designers GmbH, Esslingen Zell

Einleitung

Heutzutage nutzen mehr als 97% aller Webseiten und Webanwendungen JavaScript. Allerdings hat sich die Art und Weise, wie JavaScript verwendet wird, im Laufe der Zeit gewandelt. Während vor zehn Jahren noch 60% der Webseiten ausschließlich JavaScript nutzten, ohne Hilfsmittel, kommen aktuell nur noch 20% ohne den Einsatz von Frameworks oder Bibliotheken aus [2]. Abbildung 1 stellt die Daten dieser Aussage dar. In Anbetracht der vorangehend dargelegten Entwicklung ist es nicht verwunderlich, dass in den letzten Jahren eine Zunahme an Vielfalt und Entwicklungstätigkeiten im Bereich von Webframeworks und Bibliotheken zu verzeichnen ist.

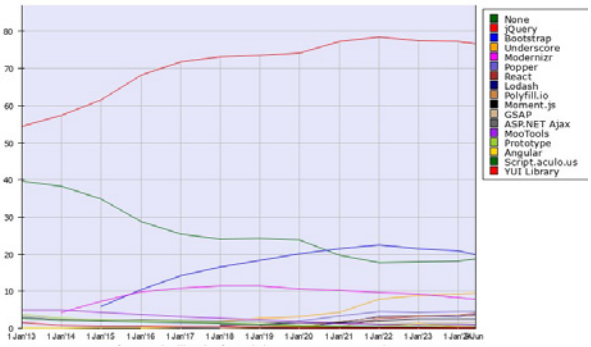

Abb. 1: Webseiten Technologienutzung [3]

Im Rahmen dieser Arbeit erfolgt eine Fokussierung auf Frontend-Webframeworks. Mittels eines Vergleichs wird ermittelt, welche Frameworks zukünftig für das Unternehmen IT-Designers von Relevanz sein können. Dazu wird eine Gegenüberstellung der jeweiligen Vor- und Nachteile vorgenommen, um abschließend eine Aussage darüber zu treffen, für welche Projektarten diese besonders geeignet sind.

Framework-Vergleiche

Zur Durchführung des Vergleichs wurden im voraus Kriterien ermittelt, die als ausschlaggebend für die Wahl eines Frameworks betrachtet werden. Die Kriterien wurden aus mehreren öffentlichen Frameworkvergleichen sowie einer internen Befragung der Mitarbeiter zusammengetragen, in welcher sie verschiedene Kriterien gewichten und ergänzen konnten.
Letztlich wurden die Kriterien in vier Hauptkategorien zusammengefasst.

- **Performance** - Im Rahmen des Performance-Vergleichs wird die Performance einer Anwendung, die in den jeweiligen Frameworks implementiert ist, evaluiert. Dabei werden insbesondere die Geschwindigkeit der Anwendungsbereitstellung sowie die Geschwindigkeit der Reaktion auf dynamische Änderungen untersucht. Zudem wird der Speicherverbrauch, der beim Nutzer anfällt, sowie die Größe der Abhängigkeiten, die das Framework benötigt, gemessen.
- **Dokumentation** - Hierbei wird eruiert, in welchem Umfang das Framework Materialien zur Verfügung stellt, die es ermöglichen, das Framework und seine Funktionen zu erlernen. Dazu zählen beispielsweise aktuelle Dokumentationen, Migrationsleitfäden oder auch Anfängerleitfäden bzw. Tutorials.
- **Popularität** - Der zentrale Aspekt der Popularität ist die Kenntnis und Nutzung des Frameworks durch eine Vielzahl von Personen. Zudem ist die Meinung dieser Personen zum Framework ebenfalls wichtig. In der Regel ist mit der Popularität auch die Anzahl der Inhalte Dritter zum Framework verbunden, weshalb diese ebenfalls von Relevanz ist. Solche Inhalte können beispielsweise Bibliotheken, Codebases oder Tutorials sein.

- **Usability** - Dies umfasst alle Angebote, die das Framework anbietet, die eine Optimierung der Entwicklung zum Ziel haben. Dazu zählen beispielsweise eine ansprechende Syntax, eine bereits vorinstallierte Linter-Konfiguration oder auch ein effektives Command Line Interface.

Ergebniserarbeitung

Zur Ableitung und Messung von Ergebnissen in den Kategorien wurden Demo-Applikationen für jedes Framework erstellt, die ein einheitliches Design und eine einheitliche Funktionalität aufweisen. Um Aussagen über die verschiedenen Projekttypen treffen zu können, wurden zwei Demos für jedes Framework genutzt. Die erste Demo stellt eine kleine Anwendung dar, während die zweite eine mittelgroße Anwendung ist.
Um nicht nur das subjektive Empfinden und die Messungen an den eigenen Demos zu werten, wurden auch öffentliche Erarbeitungen und Umfragen berücksichtigt. Ein Beispiel hierfür ist Abbildung 2, welche die Popularität von Frameworks nach der Umfrage „State Of JavaScript 2022" darstellt. An dieser Umfrage haben nahezu 40.000 Entwickler weltweit teilgenommen. [1].

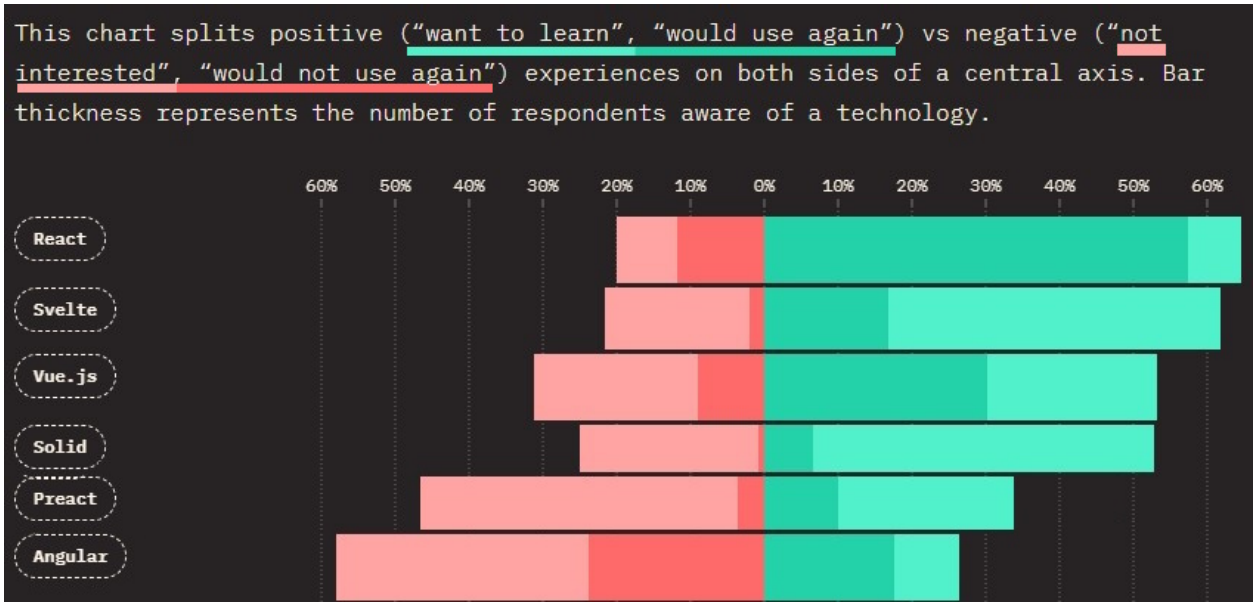

Abb. 2: Frameworknutzung und Einstellung der Entwickler zu diesen [1]

Die Ergebnisse werden schließlich zusammengetragen und ausgewertet, um Erkenntnisse daraus zu ziehen. Diese Erkenntnisse werden in Form von Empfehlungen zu einem Framework unter den jeweiligen Voraussetzungen formuliert.

Ausblick

Die vorliegende Arbeit bietet einen breiten, bewerteten Überblick über eine Vielzahl von Frontend Web-Frameworks. Weiterführende Forschungsarbeiten in diesem Bereich sind möglich, indem detaillierter auf die einzelnen Kategorien und Unterpunkte eingegangen wird. Zudem können spezifischere Demo-Anwendungen genutzt werden, die auf spezielle Teilgebiete zugeschnitten sind.
Denn auch zukünftig werden Webanwendungen oder Webseiten durchaus sehr relevant sein, da es mit modernen Browsern kaum eine andere Anwendungsplattform gibt, die so allgegenwärtig ist und von einer solch großen Anzahl an Endgeräten genutzt wird.

Literatur und Abbildungen

[1] Sacha Greif. State of JS 2022. https://2022.stateofjs.com/en-US/libraries/front-end-frameworks/, 01 2023.

[2] Risto Ollila, Niko Mäkitalo, and Tommi Mikkonen. Modern Web Frameworks: A Comparison of Rendering Performance. *Journal of Web Engineering*, 21:789–813, 2022.

[3] World Wide Web Technology Surveys. Historical yearly trends in the usage statistics of javascript libraries for websites. https://w3techs.com/technologies/history_overview/javascript_library/all/y, 01 2024.

Localization Using High-Resolution Radar Images

Frank Holzmueller **Markus Enzweiler**

Department of Computer Science and Engineering, Esslingen University

Work carried out at Robert Bosch GmbH, Renningen

Introduction

Navigation of mobile robots in unknown environments remains a non-trivial problem. Despite a large variety of existing algorithms for tackling this task, different environments and sensors require different approaches for finding a satisfying solution.
For visual approaches such as camera images, the specification of the camera as well as changing weather and lighting conditions pose great challenges [1], [4]. A more robust solution was introduced with Light Detection and Ranging (LiDAR). LiDAR can offer dense point clouds with high resolution and good robustness against changing weather conditions. However, this technology also fundamentally operates in the near-visible electromagnetic spectra, making it susceptible to floating particles such as fog or rain [1], [4]. A third approach uses radar devices operating in the much lower 76 GHz to 77 GHz range to obtain information from the environment. Even though radar offers a point cloud that is much sparser than LiDAR, these sensors generate reliable range measurements at a greatly improved resistance to environmental conditions such as rain, fog or even snow [1], [4].
All mentioned sensors are prominently featured in simultaneous localization and mapping (SLAM) algorithms to solve the localization and mapping problem. SLAM algorithms usually rely on an inertial measurement unit (IMU) for initial estimation of their current relative pose and combine this estimate with information extracted from the primary sensors [1], [4]. IMU measurements suffer from time-varying biases. IMU based odometry systems face drift and decreasing localization accuracy as the robot moves through the environment. One promising approach for compensating this drift error is the introduction of loop closing.
Within the last years, many algorithms for solving the SLAM problem using visual, LiDAR or radar approaches have been introduced [1], [4]. However, the previously elaborated issues remain a challenge for the algorithms. This thesis focuses on the approach of utilizing high-resolution radar images generated using synthetic aperture radar (SAR) for SLAM in indoor environments. By additionally introducing loop closing to our calculated results, we are looking to minimize the localization drift for the radar SLAM on a custom set of scenarios.

Synthetic Aperture Radar

When recording radar data, the aperture size of the utilized sensor determines its angular resolution. A larger radar aperture results in higher resolution radar images, hence larger apertures are always desirable. Synthetic aperture radar describes the idea of synthetically enlarging the aperture of a radar sensor by using digital signal processing techniques. For this processing step, SAR requires movement of the radar.
SAR is mostly used for airborne or even spaceborne remote sensing of the earth. Figure 1 describes the general setup of SAR imaging and shows the SAR imaging geometry for a spaceborne application. However, the technique and terminology may also be applied to ground-based radar imaging.

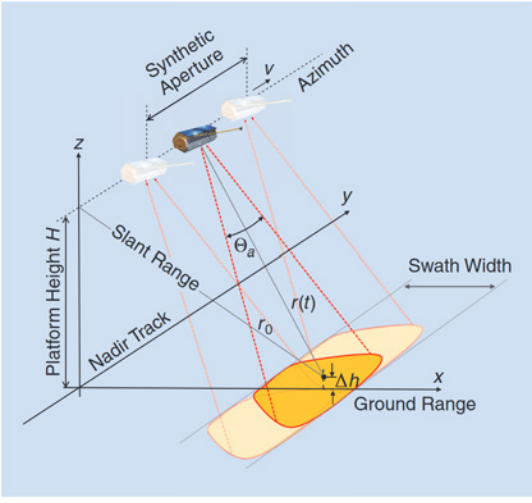

Fig. 1: SAR imaging geometry. © 2013, IEEE [6]

During generation of a SAR image, the information from multiple radar measurements is merged over a

certain time period. This combines the information of multiple measurements into a single image, leading to a potentially improved signal-to-noise ratio and a higher resolution image. For the generation of SAR images (showcased in Figure 2), the SAR process requires movement data that includes information from the IMU and radar data combined by a Kalman filter. This radar-inertial odometry provides the initial pose estimates for the SLAM pipeline. More information on radar image generation using SAR is available in [8].

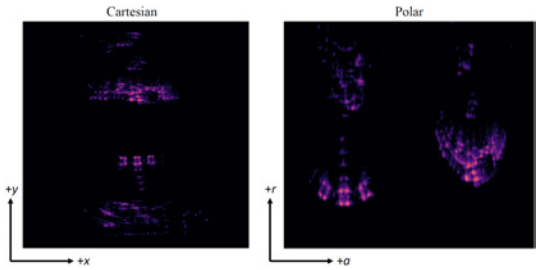

Fig. 2: SAR images in cartesian (left) and polar (right) representation [7]

Dataset

Our data was recorded using an unmanned ground vehicle (UGV) equipped with an IMU and four chirp-sequence frequency modulated continuous wave (FMCW) multiple-input multiple-output (MIMO) radar sensors operating between 76 GHz and 81 GHz, two facing in each direction orthogonal to the vehicle's direction of travel [8]. The data recording platform is visualized in Figure 3. This setup produces the raw baseband time signal for each radar sensor together with the required IMU signal.

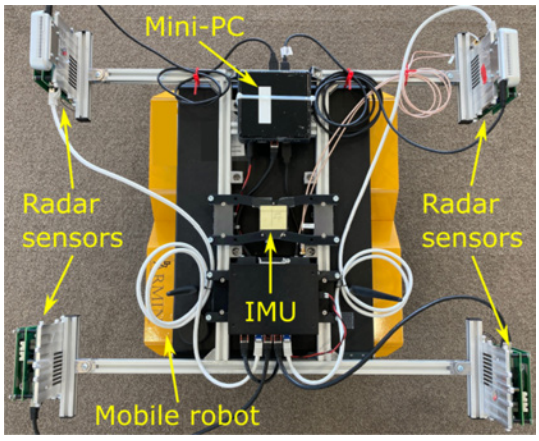

Fig. 3: UGV and sensors used for data recording [8]

In addition to the recorded data, an optical motion capture system was used during the recording of all runs for obtaining precise reference data of the vehicle movement in a controlled laboratory environment.

SLAM Pipeline

In order to extract localization information from the available data, most available algorithms follow a similar pipeline structure. These pipelines often vary only in the underlying implementation steps. One exemplary implementation for radar SLAM can be seen in Figure 4.
The first step for extracting information from an image usually lies in finding keypoints that form a point cloud of the distinctive features in a specific image such that it can be recognized at a later point in time, even with changes in the viewpoint. These keypoints can be obtained from well-known algorithms such as ORB [9] and SIFT [5] or machine learning algorithms such as RoMA [3]. Domain specific key point extractors have also been proposed in the literature [2].

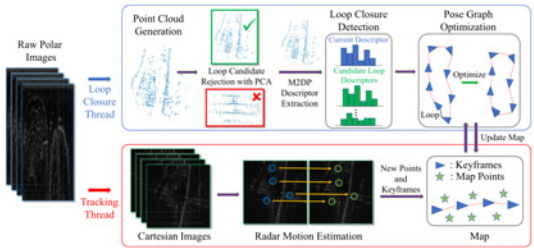

Fig. 4: Exemplary radar SLAM pipeline [4]

After extracting keypoints, the relative pose of the vehicle between two images is estimated. Using the motion and tracking information obtained by running the SLAM pipeline on previous images from the same sequence, the poses are connected in a pose graph. This pose graph introduces new constraints for each calculated relative pose, which allows for later optimization using techniques such as pose graph optimization (PGO). These techniques enable the utilization of constraints from loop closing.
During relative pose estimation, the detected map points are aligned to the map points of previously calculated poses, resulting in a transformation for the currently observed pose relative to the previously calculated poses. Once the poses are recovered, key points from different images may be represented in a common map coordinate system for visual representation of the detected environment.

Objective

The focus of this thesis is implementing a SAR image-based radar SLAM pipeline in Python. This pipeline needs to handle SAR images together with Radar-

Inertial Odometry data. Our work will focus on leveraging the improved resolution of the generated SAR images for indoor radar scenarios. The main objective is to reliably detect true loop closures while rejecting false positives and therefore building a solid foundation for robust localization and mapping.

Method

For enabling this functionality, we will evaluate different approaches to finding and aligning keypoints using multiple different algorithms of feature extraction and matching ranging from purely visual to domain specific methods. Furthermore, we will analyze the impact of loop closing on the resulting accuracy of pose graph optimization in various driving scenarios of the recorded dataset.

In order to decide when a good loop closure candidate is available, we will evaluate approaches to matching sets of extracted keypoints while including prior knowledge of the robot movement.

Evaluation of loop closing will be done by comparing the transformations resulting from the constraints introduced by loop closing to the available reference data.

In a final evaluation step, the whole trajectory generated by our SLAM pipeline will be evaluated against the reference trajectory using commonly utilized metrics such as the root mean squared error (RMSE) of the relative pose error (RPE) or absolute pose error (APE) [1], [4].

References and figures

[1] D. Adolfsson, M. Magnusson, A. Alhashimi, A.J. Lilienthal, and H. Andreasson. Lidar-level localization with radar? The CFEAR approach to accurate, fast and robust large-scale radar odometry in diverse environments. *IEEE Transactions on Robotics*, 39:1476–1495, 2023.

[2] S. H. Cen and P. Newman. Radar-only ego-motion estimation in difficult settings via graph matching. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 298–304. IEEE, 2019.

[3] J. Edstedt, Q. Sun, G. Bökman, M. Wadenbäck, and M. Felsberg. RoMA: Robust Dense Feature Matching. *IEEE Conference on Computer Vision and Pattern Recognition*, 2024.

[4] Z. Hong, Y. Petillot, A. Wallace, and S. Wang. RadarSLAM: A robust simultaneous localization and mapping system for all weather conditions. *The International Journal of Robotics Research*, 41:519–542, 2022.

[5] D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[6] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hejnsek, and K. P. Papathanassiou. A Tutorial on Synthetic Aperture Radar. *IEEE Geoscience and Remote Sensing Magazine*, 1:6–43, 2013.

[7] Own representation.

[8] Y. E. Ritterbusch, J. Fink, and C. Waldschmidt. Indoor Synthetic Aperture Radar Measurements of Point-Like Targets Using a Wheeled Mobile Robot. In *15th European SAR Conference*. VDE, 2024.

[9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: An efficient alternative to SIFT or SURF. *2011 International Conference on Computer Vision*, pages 2564–2571, 2011.

KI im Prozessmanagement stark regulierter Sektoren: Potentialanalyse bezogen auf den gesamten Prozessmanagement Lebenszyklus

Jannis Joos **Thomas Rodach**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Deloitte Consulting GmbH, Stuttgart

Einleitung

Künstliche Intelligenz (KI) hat sich in kürzester Zeit von einer Vision der Zukunft zu einem unverzichtbaren Werkzeug entwickelt. Im aktuellen Bericht von Deloitte zur Lage der generativen KI in Unternehmen, für den über 2.000 Führungskräfte befragt wurden, zeigt sich, dass Unternehmen zunehmend Werte aus ihren KI-Initiativen generieren und die Technologie nicht mehr nur als Experiment betrachten. In einer Welt, die sich rasant verändert, wird KI nicht nur als vorübergehender Hype betrachtet, sondern als Schlüssel-Technologie zum zukünftigen Erfolg, wie die erzielten Vorteile in Abbildung 1 zeigen [5].

Abb. 1: Q: Welche Vorteile erwarten Sie und inwieweit erreichen Sie diese Vorteile bisher? [5]

In stark regulierten Sektoren, wie dem Banken- und Versicherungswesen, sowie dem öffentlichen Sektor, spielt das Prozessmanagement eine zentrale Rolle. Die Einhaltung von strengen regulatorischen Anforderungen und die Notwendigkeit transparenter und gleichzeitig effizienter Prozesse stellen besondere Herausforderungen für die Organisationen dar. Durch den Einsatz von KI im Prozessmanagement können hier entscheidende Vorteile erzielt werden.

Geschäftsprozessmanagement und Künstliche Intelligenz

Geschäftsprozessmanagement (GPM) oder kurz Prozessmanagement ist ein integriertes System aus Führung, Organisation und Controlling, das auf eine zielgerichtete Steuerung und Optimierung von Geschäftsprozessen abzielt, um die Effektivität und Effizienz der betrieblichen Abläufe zu erhöhen und somit den Unternehmenserfolg langfristig zu sichern. Integriert schließt somit ein, dass Aufgaben, Teilsysteme, Ressourcen, Methoden, Tools und IT-Unterstützung in ihrer Gesamtheit betrachtet und aufeinander abgestimmt, koordiniert und zielgerichtet gesteuert werden [6]. Durch die starke Präsenz der KI versuchen auch Hersteller ihre Tools als „intelligente“ GPM-Systeme zu vermarkten. KI hat die Fähigkeit, gezielt Muster zu erkennen, Prognosen aufzustellen, Sprachen zu verarbeiten, Bilder oder Abläufe zu analysieren oder, wie wir es von Chatbots kennen, ausgefeilte Konversationen mit Menschen zu führen. Durch den Einsatz von maschinellen Lernverfahren kann die Qualität der Prognosen weiter verbessert werden, was auch bei großen Datenmengen der Fall ist. KI kann somit in verschiedenen Phasen der Prozessmodellierung unterstützen, indem sie bei Entscheidungen, (Thema „Decision-Management”) hilft, sowie mittels RPA („Robotic-Process-Automation“) oder („Cognitive RPA“) die Automatisierung von Abläufen oder Workflows steuert. Zusätzlich kann diese auch eine beratende Rolle spielen, oder Entscheidungsalternativen aufzeigen [1].

Ziel der Arbeit

Ziel der Arbeit ist es, die gestellte Forschungsfrage: „**Welchen Einfluss hat der Einsatz von Künstlicher Intelligenz auf das Geschäftsprozessmanagement von Unternehmen stark regulierter Sektoren und wie lassen sich diese Auswirkungen im Prozessmanagement Lebenszyklus einordnen?**“ zu diskutieren und somit Potentiale und Einsatzmöglichkeiten aufzuzeigen, in denen KI im Prozessmanagement Lebenszyklus eingeordnet werden kann.
Zusätzlich werden die Anforderungen an Organisationen mit starker Regulatorik mit einbezogen. Da der Einsatz und Kenntnisstand von Künstlicher Intelligenz noch relativ neu ist, wurden Experten mit umfangreichem Fachwissen in den Bereichen Prozessmanagement und Künstlicher Intelligenz befragt, um zusätzlich auf vorhandenen Studien eine umfangreiche Einordnung der Potentiale und Vorteile durch den Einsatz von KI im Prozessmanagement zu erhalten.

Anforderungen an Versicherungen und Banken

Verschiedene Gesetze und Vorschriften erfordern besondere Maßnahmen, um Transparenz zu schaffen und die Einhaltung sicherzustellen. Solvency II bringt strenge Anforderungen an das Risikomanagement und die Eigenkapitalausstattung von Versicherungsunternehmen mit sich, wodurch Versicherer verpflichtet sind, regelmäßig Berichte über ihre Finanzlage, Risikomanagementstrategien und interne Prozesse an die Aufsichtsbehörden zu übermitteln. Diese Transparenz soll das Vertrauen der Marktteilnehmer stärken und eine frühzeitige Identifizierung von Risiken ermöglichen [3].
Die Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin) stellt ebenfalls strenge Anforderungen an Banken, um die Stabilität und Integrität des Finanzsystems zu gewährleisten. Zu den wichtigsten Aspekten gehören: die Sicherstellung der Deckung kurzfristiger Verbindlichkeiten durch entsprechende Liquiditätsanforderungen, ein effektives Risikomanagement und interne Kontrollen zur Bewertung und Steuerung von Risiken sowie regelmäßige Berichtspflichten, welche die finanzielle Lage der Bank transparent machen [4].

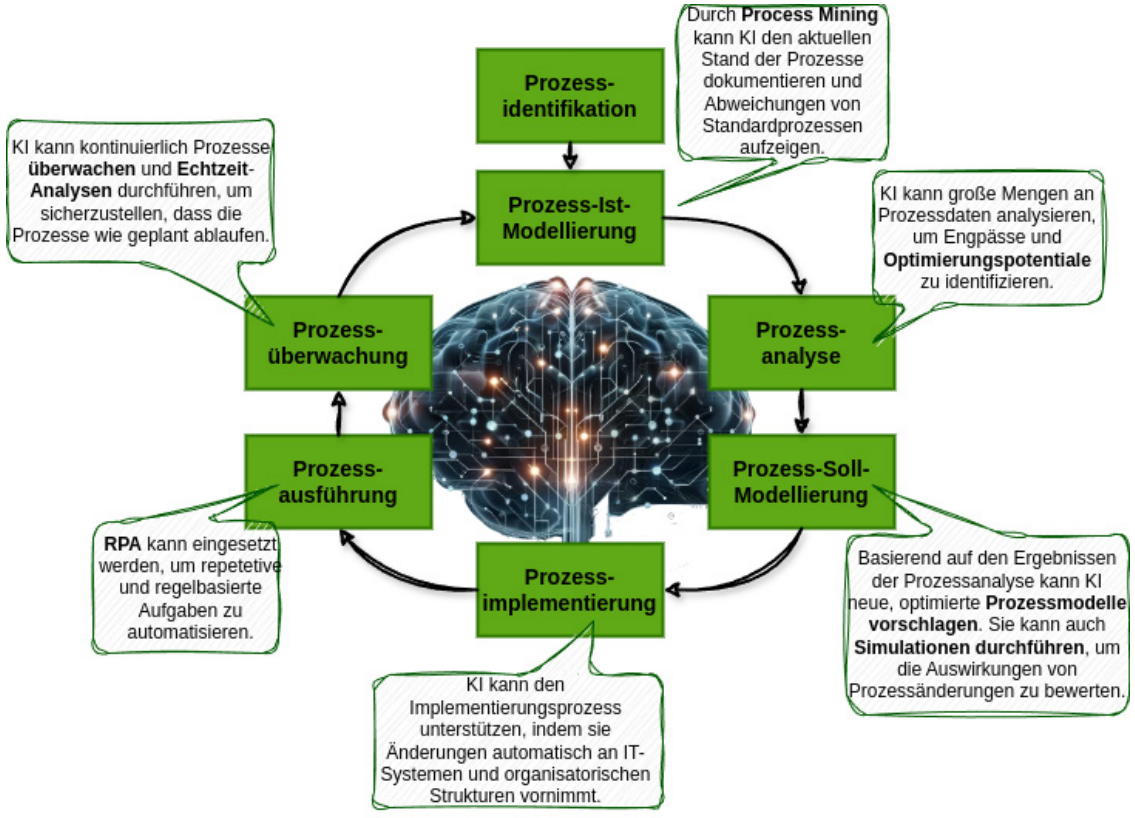

Abb. 2: Einordnung von KI in den IT-Prozessmanagement-Lebenszyklus (Quelle: Angelehnt an Schmelzer & Sesselmann, 2020, S.452), Ilustration erstellt mit Hilfe des Image Generator von ChatGPT) [2]

Ergebnisse der Experteninterviews

KI kann in fast allen Phasen des Prozessmanagement-Lebenszyklus (PMLC) eingesetzt werden. Insbesondere in den Phasen der Prozessanalyse und Prozessüberwachung erweist sich KI als besonders effektiv. In den Phasen werden Daten analysiert, um beispielsweise die Ermittlung von Kennzahlen oder die Erfassung von Durchlaufzeiten zu ermöglichen. Banken können automatische Datenanalysen zur Nutzung von Geldwäsche im Rahmen von Know Your Customer (KYC) Prozessen nutzen. Ein weiteres Beispiel wäre die Nutzung von KI in der Versicherungsbranche, wo sie in Prozessen wie der Schadensbearbeitung eingesetzt wird, um eingehende Schadenmeldungen effizient zu prüfen und zu bearbeiten. Abbildung 2 zeigt weitere Eisatzmöglichkeiten im PMLC.

Zukünftige Entwicklung

Der Einsatz von generativer KI wird weiter zunehmen, um nicht nur Prozesse zu optimieren, sondern auch neue Prozesse zu entwerfen und innovative Lösungen zu entwickeln. Auch mit der Weiterentwicklung von RPA und kognitiver RPA werden immer mehr Prozesse vollständig automatisiert, was zu einer weiteren Effizienzsteigerung führen wird. Der größte Vorteil wird aber in der Analyse von großen Datenbeständen liegen, in denen KI komplexe Datenmuster erkennen und tiefere Einblicke in die Prozessleistung geben kann. Auch die ethischen und regulatorischen Rahmenbedingungen werden sich weiterentwickeln, um den sicheren und verantwortungsvollen Einsatz von KI im Prozessmanagement zu gewährleisten.

Literatur und Abbildungen

[1] Thomas Allweyer. *Technologien für Geschäftsprozesse*. BoD - Books on Demand, Norderstedt, 1 edition, 2023.

[2] Eigene Darstellung.

[3] Bundesanstalt für Finanzdienstleistungsaufsicht. Solvency II - Eigenmittel und Eigenmittelanforderungen. https://www.bafin.de/DE/Aufsicht/VersichererPensionsfonds/Eigenmittelanforderungen/SolvencyII/solvencyII_node.html, 2020.

[4] Bundesanstalt für Finanzdienstleistungsaufsicht. Bankaufsichtliche Anforderungen an die IT (BAIT). https://www.bafin.de/SharedDocs/Downloads/DE/Rundschreiben/dl_rs_1710_ba_BAIT.pdf%3F__blob%3DpublicationFile%26v%3D9, 2021.

[5] Nitin Mittal, Costi Perricos, Kate Schmidt, Brenna Sniderman, and David Jarvis. KI-Studie Q2: Now decides Next: Getting real about Generative AI. https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Innovation/us-state-of-gen-ai-report-q2.pdf, 2024.

[6] Hermann J. Schmelzer and Wolfgang Sesselmann. *Geschäftsprozessmanagement in der Praxis*. Hanser Verlag GmbH & Co.KG, München, 9 edition, 2020.

Qualitative Evaluation KI generierter Texte in einer Beispielanwendung mit ChatGPT und Astro

Nicolas Kahle **Harald Melcher**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma pep.digital GmbH, Esslingen

Zielsetzung

Die vorliegende Masterarbeit untersucht die Qualität von KI-generierten Texten in einer beispielhaften Anwendung unter Verwendung von ChatGPT und Astro. Ziel ist es, ein System zu entwickeln, das aktuelle Nachrichten aus dem Internet automatisch extrahiert, filtert und, mithilfe von KI, Zusammenfassungen davon generiert. Die Qualität der generierten Texte soll durch Anwendung geeigneter Metriken sowie durch Nutzerfeedback bewertet werden.

Astro

Astro ist ein moderner Static Site Generator, der die sogenannte Island Architecture verwendet. Diese Architektur ermöglicht es, Webseiten aus verschiedenen UI-Komponenten zusammenzustellen und im Build-Prozess in statischen HTML-Code umzuwandeln. In Fällen, in denen clientseitige JavaScript-Logik erforderlich ist, lädt Astro nur die absolut notwendigen Komponenten. Dies minimiert die Ladezeit im Browser erheblich, da unnötiger JavaScript-Code entfernt wird. [5], [7]

ChatGPT

ChatGPT, basierend auf der Transformer-Architektur, wird genutzt, um menschenähnliche Texte zu generieren. Das Modell wird zunächst auf einer großen Menge von Textdaten vortrainiert und dann durch spezifische Aufgaben weiter optimiert (fine-tuned). [3], [4], [8], [9]
In dieser Arbeit wird ChatGPT eingesetzt, um Nachrichteninhalte zu filtern und zusammenzufassen, basierend auf den individuellen Präferenzen der Nutzer.

Tagesschau API

Die Tagesschau API wird verwendet, um aktuelle Nachrichteninhalte automatisch aus dem Internet zu extrahieren. Diese API stellt Nachrichteninhalte im JSON-Format bereit, die dann von ChatGPT weiterverarbeitet werden können.

Metriken

Die durch die Anwendung generierten Texte sollen qualitativ Ausgewertet werden. Dafür werden zum einen Nutzerbefragungen durchgeführt, zum anderen sollen auch gängige Metriken angewendet werden.
In dieser Arbeit werden die Metriken BLEU (Bilingual Evaluation Understudy) und ROUGE (Recall-Oriented Understudy for Gisting Evaluation) verwendet. Beide basieren auf dem Vergleich des zu bewertenden Textes und einem Referenztext. Der Referenztext stellt eine optimale Formulierung für eine bestimmte Situation dar. Beide Metriken bewerten grundsätzlich die Übereinstimmung von Wortfolgen zwischen dem zu bewertenden Text und dem Referenztext. Eine solche Wortfolge wird für eine Länge von n Wörtern auch als n-Gram bezeichnet (zum Beispiel ist eine Folge von 2 Wörtern ein 2-Gram oder Bigram). [6], [1], [2]
ROUGE:
Die ROUGE Metrik liefert mehrere wichtige Kennzahlen, darunter die sogenannte Präzision und den Recall. Präzision beschreibt die Anzahl übereinstimmender n-Grame $i\,(n)$ geteilt durch die Gesamtzahl an n-Gramen $|C|$, die im zu bewertenden Text C vorkommen. Man kann sich also zur Veranschaulichung vorstellen, dass dies in etwa der Anteil des generierten Textes ist, der den Inhalt richtig wiedergibt.

$$ROUGE_p = \frac{i\,(n)}{|C|} \tag{1}$$

[2]
Recall beschreibt hingegen die Anzahl übereinstimmender n-Grame $i\,(n)$ geteilt durch die Gesamtzahl an n-Gramen $|R|$, die im Refenztext R vorkommen. Man kann sich vorstellen, dass der Recall in etwa den Anteil des Referenztextes angibt, die das Sprachmodell wiedergibt.

$$ROUGE_r = \frac{i\,(n)}{|R|} \tag{2}$$

[2]

BLEU:
Es gibt mehrere Formulierungen, die ein und denselben Sachverhalt inhaltlich korrekt wiedergeben. Um darauf einzugehen, werden bei BLEU mehrere Referenztexte R_j bereitgestellt, die unterschiedlich formuliert sind, aber inhaltlich identisch sein müssen. Das Ziel dahinter ist, synonyme Formulierungen genauer zu bewerten. Es wird gezählt, wie viele n-Grame im generierten Text C enthalten sind, die auch in einem der Referenztexte vorkommen. Die Anzahl übereinstimmender n-Grame wird durch die Anzahl aller n-Grame in C geteilt, also ein Präzisionswert ermittelt. [6], [1] Präzisionswerte neigen dazu, Besonders Kurze Texte besser zu bewerten, da die Länge des zu bewertenden Textes im Nenner steht. Während bei ROUGE der Recall-Wert hinzugezogen wird, um solche Fälle zu identifizieren, versucht BLEU sich auf einen Wert zu beschränken. Um dennoch kurze Texte nicht fälschlicherweise zu favorisieren, wird eine sogenannte Brevity Penalty BP eingeführt. Sie wird zum Präzisionswert hinzumultipliziert, und ist kleiner, je kürzer der zu bewertende Text im Vergleich zu den Referenztexten ist. [6] empfiehlt die Brevity Penalty nach folgender Formel zu definieren:

$$BP = \begin{cases} 1, & |R_j| < |C| \\ e^{1-\frac{|R_j|}{|C|}}, & |R_j| \geq |C| \end{cases} \tag{3}$$

[6], [1]

Ausblick

Die Fertigstellung und Auswertung der Anwendung befinden sich noch in Arbeit, sodass noch keine umfassenden Nutzerdaten vorliegen. Jedoch konnte in Testläufen festgestellt werden, dass die Anwendung akzeptable Resultate erzielt.
Es lässt sich dennoch bereits jetzt sagen, dass die vorgestellten Metriken eine Textnormalisierung erfordern, da sonst Synonyme als unterschiedliche Wörter betrachtet würden.
Für zukünftige Arbeiten wäre es interessant zu untersuchen, ob Metriken auch unterschiedliche Wortarten unterschiedlich gewichten könnten. Beispielsweise scheinen Verben und Substantive wesentlicher für den Inhalt eines Satzes zu sein als Artikel. Ein Beispiel zur Verdeutlichung: Der Satz „Die Katze jagt die Maus“ behält seinen wesentlichen Inhalt, selbst wenn die Artikel entfernt werden („Katze jagt Maus“), während das Entfernen von Verben und Substantiven den Satz unverständlich macht („Die die“). Eine solche Differenzierung könnte die Metriken aussagekräftiger darüber machen, wie Präzise der Inhalt eines generierten Textes ist. Allerdings wäre zu erwarten, dass der Fokus auf eine korrekte Grammatik dadurch verloren geht.

Literatur und Abbildungen

[1] Fabio Chiusano. Two minutes NLP — Learn the BLEU metric by examples. https://medium.com/nlplanet/two-minutes-nlp-learn-the-bleu-metric-by-examples-df015ca73a86, 01 2022.

[2] Fabio Chiusano. Two minutes NLP — Learn the ROUGE metric by examples. https://medium.com/nlplanet/two-minutes-nlp-learn-the-rouge-metric-by-examples-f179cc285499, 01 2022.

[3] Dr. Aleksandra Klofat. Wie funktionieren Transformer? Definition und Praxis. https://www.informatik-aktuell.de/betrieb/kuenstliche-intelligenz/wie-funktionieren-transformer-definition-und-praxis.html, 05 2023.

[4] Michael A. Nielsen. Neural Networks and Deep Learning. http://neuralnetworksanddeeplearning.com/index.html, 2015.

[5] Addy Osmani. Island Architecture. In *Learning JavaScript Design Patterns*, pages 243–245. Sebastopol, CA, USA, O'Reilly Media Inc., 2 edition, 2023.

[6] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318. Philadelphia, Association for Computational Linguistics, 2002.

[7] Fred Schott and Nate Moore. Introducing Astro: Ship Less JavaScript. https://astro.build/blog/introducing-astro/, 06 2021.

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. https://arxiv.org/pdf/1706.03762.pdf, 08 2023.

[9] Elena Voita, David Talbot, Fedor Moiseev, Rico Sennrich, and Ivan Titov. Analyzing Multi-Head Self-Attention: Specialized Heads Do the Heavy Lifting, the Rest Can Be Pruned. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5797–5808. Florence, Association for Computational Linguistics, 2019.

Design und Implementierung einer automatisierten Pipeline zur Erzeugung von Trainingsdaten für semantische Netze

Steve Fredy Kana Meka **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Motivation und Kontext

Im Zeitalter fortschreitender Digitalisierung und der Expansion autonomer Systeme stellt das it:movES-Projekt der Hochschule Esslingen einen bedeutenden Schritt hin zur Realisierung intelligenter Verkehrslösungen dar. Innerhalb dieses Projekts arbeitet ein Team von Studierenden eng mit Professoren zusammen, um Hardware und Software für autonome Modellfahrzeuge im Maßstab 1:10 zu entwickeln.
Eines der Ziele dieses Projekts ist es, semantische Netze auf eingebetteten Systemen zu implementieren, um den maximalen Informationsgehalt aus Kamerabildern zu extrahieren. Dies soll dazu beitragen, andere Sensortechnologien im Fahrzeug überflüssig zu machen. Die Bedeutung semantischer Netze in diesem Zusammenhang ist unbestritten, ihre Effizienz hängt jedoch stark von der Verfügbarkeit und Qualität der Trainingsdaten ab.
Die im Projekt it:movES verwendeten Kamerabilder müssen präzise gelabelt werden, um den Netzen zu ermöglichen, Objekte und Situationen korrekt zu identifizieren und zu klassifizieren. Derzeit verfügbare Trainingsdaten reichen entweder in Quantität und Qualität nicht aus oder sind nicht spezifisch genug für die in der Bosch Future Mobility Challenge geforderten Verkehrsszenarien. Zudem erfordert die Anpassung dieser Daten an die spezifischen Anforderungen des Modellmaßstabs besondere Aufmerksamkeit.

Zielsetzung und -Fragen der Arbeit

Diese Bachelorarbeit zielt darauf ab, eine Pipeline zur automatisierten Generierung und zum Labeling von Bildern für das Training semantischer Netze zu konzipieren und zu implementieren. Dabei liegt der Fokus auf der präzisen Verarbeitung von Verkehrsszenarien. Ziel ist es, eine Lösung zu entwickeln, die durch die Synthese von Simulationsdaten und realen Bildern die Erzeugung einer hohen Anzahl an qualitativ hochwertigen, gelabelten Trainingsbildern ermöglicht. Dies soll erreicht werden, indem das hochmoderne Segment-Anything-Netz von Meta [4] für die Bildsegmentierung (siehe Abbildung 1) und das fortschrittliche YOLOv8-Modell [3] zur Erzeugung von Labels in einem integrierten System eingesetzt werden.

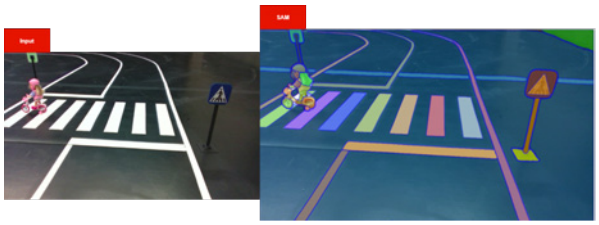

Abb. 1: SAM Output [1]

Die Forschungsfragen, die diese Arbeit antreibt, umfassen:

- Wie kann eine Pipeline gestaltet werden, die eine effiziente Synthese von Simulations- und Realbilddaten ermöglicht und gleichzeitig die Anforderungen an präzise gelabelte Bilder für semantische Netze erfüllt?
- Welche technischen und algorithmischen Herausforderungen müssen bei der Implementierung des Segment Anything-Netzes und des YOLOv8-Modells gemeistert werden, um eine hohe Labelgenauigkeit zu gewährleisten?

Methodik der Arbeit

Das folgende Diagramm (siehe Abbildung 2) stellt die Methode unserer automatisierten Bildannotation dar. Es zeigt, wie Input-Bilder durch eine Reihe von vordefinierten Verarbeitungsschritten fließen, beginnend bei der Dateneingabe bis hin zur finalen Erzeugung von segmentierten und beschrifteten Bildern. Diese Methode gliedert sich in mehrere Phasen, wobei jede Phase einen spezifischen Schritt im Gesamtprozess abbildet.

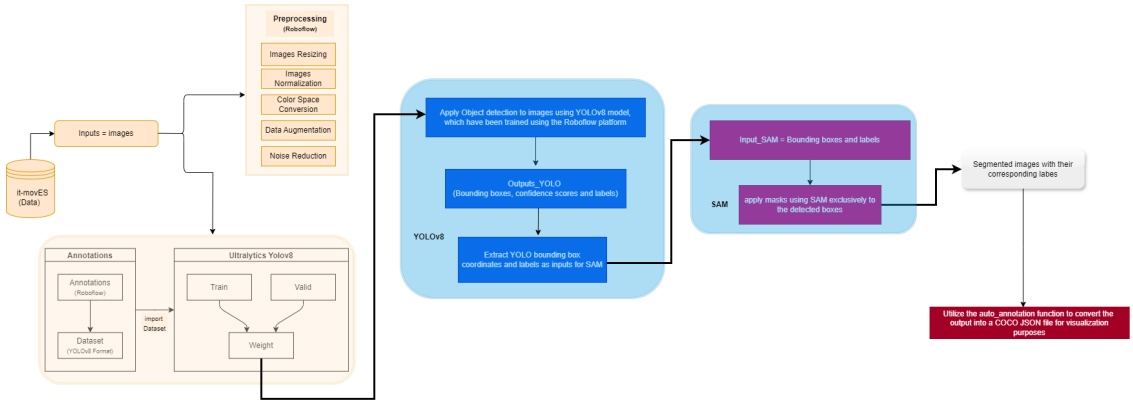

Abb. 2: Design Pipeline [1]

- Phase 1 – Objekterkennung mit YOLOv8: Zu Beginn der Pipeline dienen Bilder als Input. Diese werden durch das YOLOv8-Modell verarbeitet, welches auf der Roboflow-Plattform [2] trainiert wurde. YOLOv8 führt eine Objekterkennung auf den Eingabebildern durch, um Objekte zu lokalisieren und zu klassifizieren. Die Ergebnisse dieser Phase umfassen Bounding Boxes, die die Positionen der Objekte im Bild markieren, sowie Konfidenzwerte und Labels, die die Genauigkeit der Vorhersage des Modells und die Art des erkannten Objekts angeben.
- Phase 2 – Extraktion und Vorbereitung für SAM: Im Anschluss an die Objekterkennung werden die Koordinaten und Labels der Bounding Boxes aus den Ausgaben von YOLOv8 extrahiert. Diese Informationen werden als Eingabe für das SAM-Modell aufbereitet, welches die Segmentierung der Objekte übernehmen wird.
- Phase 3 – Segmentierung mit SAM: Das SAM (Segment Anything Model) verarbeitet die Bounding Boxes und Labels, die von YOLOv8 geliefert wurden. Anschließend wendet es Segmentierungsmasken exklusiv auf die detektierten Objekte an. Dadurch werden die Objekte nicht nur identifiziert, sondern auch innerhalb des Bildes klar abgegrenzt und segmentiert.

Ausblick

In Bezug auf die Weiterentwicklung und Verbesserung unserer Pipeline steht die Optimierung von SAM im Zentrum. Eine potenzielle Erweiterung konnte die Anreicherung von SAM mit der Fähigkeit umfassen, eigenständig genauere Labels aus den generierten Masken zu extrahieren. Dies würde es ermöglichen, qualitativ hochwertige Trainingsdaten effizienter zu produzieren und die Abhängigkeit von externen Labeling-Modellen zu verringern. Die Eingliederung von Zero-Shot Object-Detection-Modellen in SAM könnte die Generierung von Trainingsdaten für semantische Netze revolutionieren. Mit solchen Modellen könnte SAM ohne zusätzliche Trainingsdaten neue Objektklassen identifizieren und klassifizieren. Dies würde die Skalierbarkeit und Anpassungsfähigkeit des Systems erheblich verbessern und die Erstellung von Trainingsdaten für semantische Netze effizienter gestalten. Damit könnte SAM eine Schlüsselrolle in der zukünftigen Entwicklung von KI-Systemen spielen und dazu beitragen, die Grenzen der automatisierten Datenannotation neu zu definieren.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] B. Dwyer et al. Roboflow (Version 1.0)[Software]. https://roboflow.com, 2024.

[3] G. Jocher et al. Ultralytics YOLO (Version8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics, 2023.

[4] Alexander Kirillov et al. Segment anything. *Meta AI Research*, 2023.

Lokalisierung technischer Bauteile für Pick-and-Place Anwendungen mit künstlicher Intelligenz

Wissam Kasti **Thao Dang**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Robert Bosch GmbH, Stuttgart

Einleitung

Künstliche Intelligenz (KI) wird in der Vision häufiger genutzt, den die KI ermöglicht es Computern und Systemen, aussagekräftige Informationen aus digitalen Bildern, Videos und anderen visuellen Eingaben zu extrahieren und auf Grundlage dieser Informationen Maßnahmen zu ergreifen oder Empfehlungen zu geben. Ähnlich wie das menschliche Sehvermögen, das über viele Lebenszeiten hinweg trainiert wurde, um Objekte zu unterscheiden, Entfernungen abzuschätzen, Bewegungen zu erkennen und Unregelmäßigkeiten in Bildern zu identifizieren, trainiert die künstliche Intelligenz Maschinen, diese Fähigkeiten in wesentlich kürzerer Zeit zu erlernen. Anstelle von Netzhaut, Sehnerven und visuellem Cortex nutzen diese Systeme Kameras, Daten und Algorithmen.

Motivation

In der industriellen Produktion gibt es mehrere Projekte, die klassische, kantenbasierte Bildverarbeitung nutzen, um komplexe Bauteile zu erkennen und ihre Position sowie Ausrichtung in 2D zu bestimmen. Der Konfigurations- und Pflegeaufwand für die Modelle, die zur Identifikation dieser Teile verwendet werden, ist jedoch hoch. Für die Segmentierung wurde YOLO (You Only Look Once) eingesetzt, um die Effizienz und Genauigkeit moderner neuronaler Netzwerke zu untersuchen. Ein Vergleich mit klassischen Bildverarbeitungsmethoden ist vorgesehen, um die jeweiligen Vorteile und Einschränkungen der beiden Ansätze aufzuzeigen. Die Bachelorarbeit widmet sich den folgenden Fragen:

1. Wie präzise lassen sich Position und Ausrichtung solcher Objekte mit KI Modellen bestimme?
2. Wie hoch ist der Rechen- und Zeitaufwand für die Detektion?
3. Wie universell können solche Modelle gestaltet werden?
4. Wie einfach oder schwierig ist das Training und Pflege dieser Modelle?

YOLOv8 (You only look once)

YOLOv8 (You Only Look Once Version 8) ist die neueste Iteration einer hochmodernen Echtzeit-Objekterkennungstechnologie, die in der Computer Vision und künstlichen Intelligenz einen bedeutenden Fortschritt darstellt. Ursprünglich von Joseph Redmon und Ali Farhadi entwickelt, hat sich das YOLO-Framework kontinuierlich weiterentwickelt, um immer präzisere und effizientere Modelle zu bieten. YOLOv8 setzt diese Tradition fort und bietet Verbesserungen in Genauigkeit, Geschwindigkeit und Benutzerfreundlichkeit.

Im Kern funktioniert YOLOv8 durch die Implementierung eines neuronalen Netzwerks, das darauf trainiert ist, Objekte in Bildern und Videos zu erkennen und zu lokalisieren. Das Modell betrachtet ein Bild nur einmal, um verschiedene Objekte zu identifizieren und ihre Positionen innerhalb des Bildes festzulegen. Diese Ein-Schritt-Methode unterscheidet YOLO von anderen Algorithmen, die ein Bild mehrfach analysieren müssen, um ähnliche Ergebnisse zu erzielen.

YOLOv8 verwendet eine Kombination aus fortschrittlichen Convolutional Neural Networks (CNNs) und Anker-Box-Techniken, um präzise Vorhersagen zu treffen. CNNs sind speziell darauf ausgelegt, Bilddaten zu verarbeiten, indem sie Muster wie Kanten, Formen und Texturen erkennen. Anker-Boxen dienen als Bezugspunkte, die das Modell verwendet, um die wahrscheinlichen Positionen von Objekten zu bestimmen.

Ein wesentliches Merkmal von YOLOv8 ist seine Effizienz. Durch die Reduzierung der Anzahl der erforderlichen Berechnungen und die Optimierung der Netzarchitektur erreicht YOLOv8 eine hohe Verarbeitungsleistung, was es ideal für Echtzeitanwendungen macht. Diese Effizienz ist besonders nützlich in Bereichen wie autonomes Fahren, Überwachungssysteme

und Augmented Reality, wo schnelle und genaue Objekterkennung entscheidend ist.
Zusammenfassend lässt sich sagen, dass YOLOv8 eine bahnbrechende Technologie in der Objekterkennung darstellt, die sowohl durch ihre Präzision als auch durch ihre Geschwindigkeit beeindruckt. Mit kontinuierlichen Verbesserungen und Anpassungen setzt YOLOv8 neue Maßstäbe für Anwendungen, die auf zuverlässige und schnelle Bildverarbeitung angewiesen sind. [1]

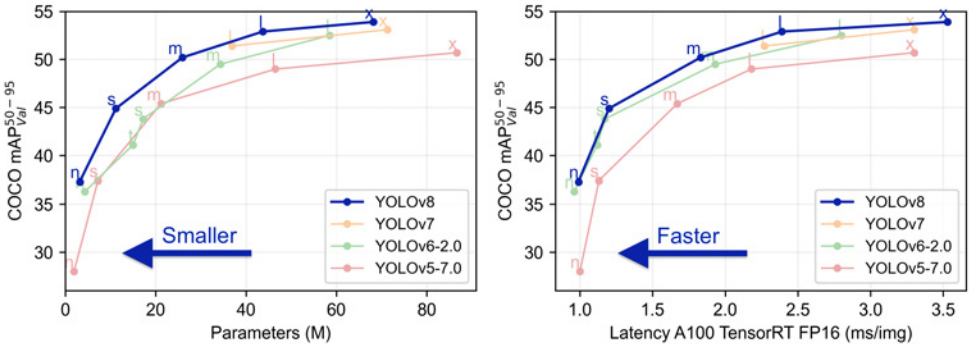

Abb. 1: Vergleich mit älteren Versionen [1]

Segmentierung

Die Instanzsegmentierung erweitert die Objekterkennung, indem sie nicht nur einzelne Objekte in einem Bild identifiziert, sondern sie auch vom Rest des Bildes trennt.
Ein Instanzsegmentierungsmodell liefert als Ausgabe eine Reihe von Masken oder Konturen, die jedes Objekt im Bild umreißen, zusammen mit Klassenbeschriftungen und Vertrauenswerten für jedes Objekt. Diese Methode ist besonders nützlich, wenn es darauf ankommt, nicht nur die Position von Objekten im Bild zu bestimmen, sondern auch ihre genaue Form zu erfassen. [3]

Abb. 2: Beispiel [2]

Literatur und Abbildungen

[1] Fatih Akyon, Ayush Chaurasia, Burhan Qaddoumi, and Glenn Jocher. YOLOv8. https://docs.ultralytics.com/models/yolov8/#citations-and-acknowledgements, 2023.

[2] Rayan Potter. Beispiel. https://medium.com/anolytics/image-segmentation-the-deep-learning-approach-1e48035dfade, 2022.

[3] Burhan Qaddoumi and Glenn Jocher. Segmentation. https://docs.ultralytics.com/de/tasks/segment/, 2023.

Analyse und Modellierung von Serviceprozessen in der Industriehydraulik mit BPMN 2.0 zur Automation mit einer Low-Code Workflow-Engine

Daniel Kaul **Thomas Rodach**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Bosch Rexroth AG, Lohr am Main

Einleitung

Kleinserien und variantenreiche Ausführungen einzelner Produkte führen im Service der Industriehydraulik zu einer Vielzahl von individuellen Prozessen. Diese Variabilität erzeugt ein hohes Informationsaufkommen und setzt ein umfangreiches Know-how der Servicemitarbeiter voraus. Besonders im Umgang mit digitalisierten Produkten führt die Kalibrierung der Elektronik zu vielen Teilschritten, was in einer hohen Komplexität resultiert. Hinzu kommt, dass für die meisten Prozesse bislang keine zentrale Dokumentation vorliegt, wodurch zusätzliche Produktivitätseinbußen entstehen. Um der steigenden Komplexität, die durch die zunehmende Digitalisierung und Elektrifizierung der Produkte entsteht, entgegenzuwirken, soll für den Service eine Plattform entstehen, welche durch teilautomatisierte Workflows den Benutzer entlastet.

Zielsetzung

Im Rahmen dieser Arbeit sollen die Serviceprozesse erfasst und mit BPMN 2.0 sowie CMMN modelliert werden. Anschließend folgt eine Analyse der Prozesse, bei welcher der Fokus auf der Identifikation und Herausarbeitung von modularen Prozessbausteinen liegt. Ziel dieser Arbeit ist die Schaffung wiederverwendbarer Prozesskomponenten, mit denen zukünftig der Aufwand für die Abbildung neuer Service-Workflows reduziert werden kann. Im Zusammenspiel mit der grafischen Low-Code-Plattform „Flowable“ sollen Fachexperten ohne Programmierkenntnisse befähigt werden, digitalisierte Service-Workflows auf Basis ihres Expertenwissens zu modellieren. Abschließend erfolgt die Übertragung der dokumentierten Prozesse in die Workflow-Engine, wo sie mit zusätzlichen Elementen angereichert werden, um ausführbare Modelle zu generieren.

BPMN 2.0 Grundlagen

Business Process Model and Notation, kurz BPMN, beschreibt einen Standard für die Modellierung von Geschäftsprozessen, welcher durch die Object Management Group (OMG) verwaltet wird. Das Ziel der OMG ist die Schaffung einer Notationsform, die für alle Nutzergruppen im Unternehmen verständlich und leicht zugänglich ist. Des Weiteren umfasst BPMN nicht nur die grafische Gestaltung von Geschäftsprozessen, sondern definiert auch die Notation im XML-Format, wodurch die Ausführbarkeit mittels Workflow-Engines gewährleistet wird [4]. Das Konzept der sogenannten Basiselemente schafft die Balance zwischen Komplexität und Übersichtlichkeit der Modelle. Um eine schnelle Einfindung in neue Prozessmodelle zu erleichtern, beruhen die zahlreichen BPMN-Symbole auf wenigen grafischen Grundelementen, wodurch die Zuordnung zu bestimmten Symbolklassen vereinfacht wird (siehe Abb. 1). Dadurch ermöglichen BPMN dem Leser eine schnelle Einarbeitung in neue Modelle, ohne dabei die Gestaltungsmöglichkeiten einzuschränken.

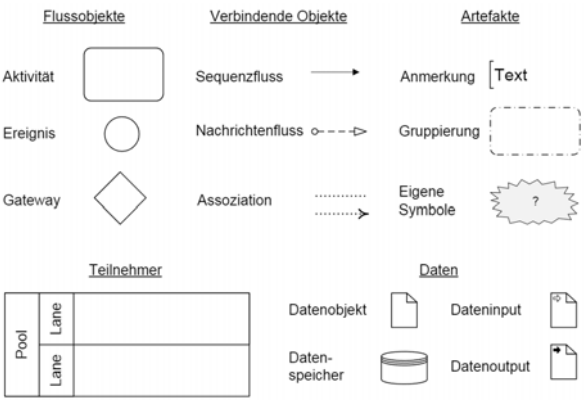

Abb. 1: BPMN Basiselemente [2]

BPMN in der Praxis

Obwohl sich BPMN über die Jahre als Standard etabliert hat, sind die kritischen Stimmen nicht verstummt. In diesem Zusammenhang wird vor allem die Eignung der Notationsform für den Unternehmenskontext hinterfragt. Häufig wird kritisiert, dass BPMN zwar als Brücke zwischen IT- und Fachexperten konzipiert wurde, diese Ansprüche jedoch nicht erfüllt. Die zahlreichen Aspekte, die sich eindeutig der Modellierung ausführbarer Prozesse zuordnen lassen, stellen einen wesentlichen Grund dafür dar. Dies führt dazu, dass BPMN bei fachlich orientierten Modellierern auf wenig Akzeptanz stößt [1]. Besonders für nicht IT-versierte Nutzer können Begriffe wie „Error handling“ oder „throw“ und „catch“ im Kontext der Prozessmodellierung schwer zuordenbar sein [2]. Des Weiteren sehen viele Experten den hohen Einarbeitungsaufwand als Nachteil, da die Notation und Regeln bei vollständigem Einsatz der Symbolpalette als sehr komplex wahrgenommen werden. Viele Experten weisen daher auf die Notwendigkeit hin, die Elemente für die fachliche Anwendung einzugrenzen [3]. Die Eingrenzung hat zur Konsequenz, dass der Anspruch von BPMN, sowohl fachliche als auch ausführbare Prozessmodelle zu erzeugen, verloren geht, da die fachlichen Modelle zusätzlich mit Komponenten angereichert werden müssen. Auf der anderen Seite wird die klar definierte Semantik von einigen auch als Vorteil angesehen, da sie eine disziplinierte Analyse und Dokumentation erzwingt [1]. In der Abschließenden Betrachtung lässt sich festhalten, dass eine Evaluierung des Einsatzes im Unternehmen für den eigenen Verwendungszweck erforderlich ist. Es ist daher abzuwägen, ob die Möglichkeit, ausführbare Prozesse zu erzeugen, die damit einhergehenden Nachteile übersteigt. Trotz möglicher Defizite bleibt BPMN 2.0 ein mächtiges Werkzeug für die Prozessmodellierung. Darüber hinaus hat sich in der Praxis gezeigt, dass ein bewusster Umgang mit den Nachteilen der Notation dazu beitragen kann, die Schwächen zu minimieren.

Potenziale einer Workflow-Engine

Eine Workflow-Engine, auch Process-Engine genannt, beschreibt eine zentrale Software-Komponente, welche für die automatisierte Ausführung von Prozessmodellen zuständig ist. Die Kernaufgabe des Systems besteht in der Orchestrierung der in den Modellen definierten Abläufe sowie in der Bereitstellung einer Schnittstelle für die Kommunikation mit anderen Anwendungen [5]. Für diese Arbeit wurde das Workflow-Management-System des Softwareherstellers „Flowable“ eingesetzt. Die Flowable-Workflow-Engine basiert auf den Modellierungsstandards BPMN, CMMN sowie DMN und erlaubt dadurch eine mehrdimensionale Modellierung komplexer Geschäftsprozesse. Mittels einer grafischen Low-Code-Modellierungsumgebung können Nutzer per Drag-and-Drop schnell und flexible Anwendungen entwickeln sowie verwalten. Im Rahmen dieser Arbeit sind insbesondere die Prozessbausteine „Call Activity” sowie „Service Registry Task” von Bedeutung. Erstgenannter ermöglicht die Aufrufe autarker Prozesse, während Letzterer die Kommunikation zu den Produkten mittels einer REST-API sicherstellt. Ein weiterer wesentlicher Aspekt ist die Möglichkeit der Wiederverwendung von Formularen, welche die Interaktion mit den Prozessverantwortlichen realisieren. Die genannten Funktionen der Flowable-Plattform stellen die essenzielle Grundlage für die Gestaltung modularer Prozessbausteine dar.

Ausblick

Im weiteren Verlauf der Arbeit soll die Palette der modularen Prozesskomponenten auf weitere Produktfamilien ausgedehnt werden. Die Überführung in die Flowable-Plattform offenbarte jedoch frühzeitig, dass sich der Benutzerkreis aufgrund der doch hohen Komplexität der Modellierungsumgebung auf wenige geschulte Mitarbeiter beschränken wird. Dennoch kann die Nutzung der Low-Code-Plattform als erfolgreich angesehen werden, da auf diese Weise der permanente Austausch zwischen IT- und Fachexperten entfällt.

Literatur und Abbildungen

[1] Thomas Allweyer. Eignet sich BPMN für das Business? https://www.kurze-prozesse.de/2010/03/19/eignet-sich-bpmn-fur-das-business/, 03 2010.

[2] Jakob Freund and Bernd Rücker. *Praxishandbuch BPMN mit Einführung in CMMN und DMN*. Carl Hanser Verlag, 5 edition, 2017.

[3] Andreas Gadatsch. *Geschäftsprozesse analysieren und optimieren*. Springer Vieweg, 2 edition, 2022.

[4] Object Management Group. Business Process Model and Notation - Version 2.0. https://www.omg.org/spec/BPMN/2.0/PDF, 01 2011.

[5] Frank Leymann and David Schumm. Process Engine. https://wirtschaftslexikon.gabler.de/definition/process-engine-52692/version-275810, 02 2018.

Generierung von Offline-HD-Karten für autonomes Fahren mithilfe von maschinellem Lernen

Noah Koehler **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Mercedes-Benz AG, Böblingen

Einleitung

Im Bereich des autonomen Fahrens ist die Erstellung von hochauflösenden (HD) Karten entscheidend für eine sichere und zuverlässige Navigation. Traditionelle Ansätze wie VectorMapNet [2] und InstaGraM [3] basieren auf bildbasierten Eingabedaten und erfordern eine rechenintensive und auch fehleranfällige Vorverarbeitung. Mit einem alternativen Ansatz welcher geometrische Eingabedaten, die aus aufgezeichneten Autofahrten gewonnen werden, verwendet kann diese Notwendigkeit jedoch umgangen werden, was zu einem präziseren und effizienteren Prozess führen soll. In dieser Thesis soll eine innovative End-to-End-Pipeline für das Offline-HD-Kartenlernen unter Verwendung geometrischer Eingabedaten entwickelt werden. Durch den Einsatz moderner tiefer neuronaler Netzarchitekturen und die Nutzung der inhärenten Struktur geometrischer Daten zielt dieser Ansatz darauf ab, die Einschränkungen bildbasierter Methoden zu überwinden und einen direkteren Weg zur HD-Kartenerstellung zu bieten.

Hintergrund und verwandte Arbeiten

HD-Karten für autonomes Fahren HD-Karten sind wesentliche Komponenten von autonomen Fahrsystemen. Sie liefern präzise und aktuelle Informationen über die Straßenumgebung, einschließlich Fahrbahnmarkierungen, Verkehrsschildern und anderen relevanten Merkmalen. Diese Karten dienen als wichtiger Input für Wahrnehmungs-, Lokalisierungs- und Pfadplanungsalgorithmen und ermöglichen eine sichere und effiziente Navigation. Herkömmliche Methoden zur Erstellung von HD-Karten beruhen häufig auf bildbasierten Eingabedaten, wie LiDAR-Punktwolken oder Kamerabildern. Diese Methoden beinhalten in der Regel die Projektion der Eingabedaten auf eine Birds-Eye-View-Ebene (BEV), gefolgt von Segmentierung, Merkmalsextraktion und Kartenerstellung (vgl. [2], [3]), zu sehen in Abb. 1.

Der Prozess kann jedoch sehr rechenintensiv sein und aufgrund der projektiven Transformation zu Verzerrungen führen.

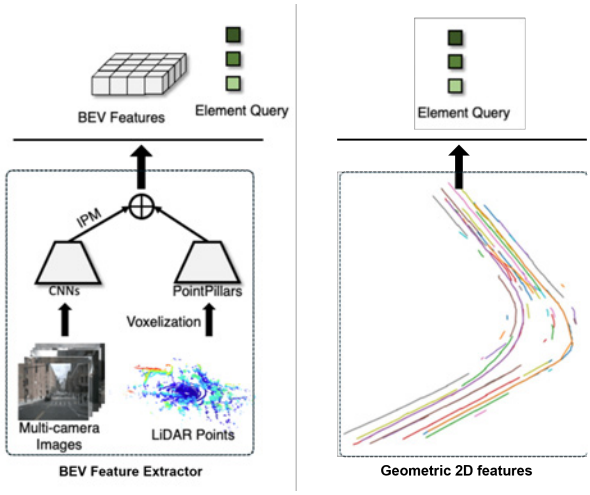

Abb. 1: Vergleich von Bild- (links) zu geometrischen (rechts) Eingangsdaten [2]

Geometrische Eingabedaten für HD-Map Learning Geometrische Eingabedaten, wie die Vektordarstellung, bieten eine alternative Darstellung der Straßenumgebung. Diese Datenquellen erfassen von Natur aus die Struktur der Szene und bieten einen direkteren und effizienteren Ansatz für das Lernen von HD-Karten (siehe Abb. 1). Allerdings haben sich bis heute nur wenige Studien auf die Entwicklung von End-to-End-Pipelines speziell für das Offline-Lernen von HD-Karten unter Verwendung geometrischer Eingabedaten konzentriert.

Vorgeschlagener Ansatz

Der vorgeschlagene Ansatz verwendet eine neuartige Technik, um geometrische Eingabedaten aus der Fahrzeugerkennung und ihrer vektorisierten Form in einem strukturierten Format darzustellen, das für

Deep-Learning-Modelle geeignet ist. Die wichtigsten Komponenten dieses Ansatzes sind:

Geographische Kacheln Die Eingabedaten werden mit Hilfe des H3-Kachelsystems von Uber [1], modelliert in Abb. 2, in abgegrenzte geographische Gebiete unterteilt. Dieser Schritt stellt sicher, dass das Modell überschaubare Regionen als Eingabe erhält, was eine effiziente Verarbeitung sowie Skalierbarkeit ermöglicht.

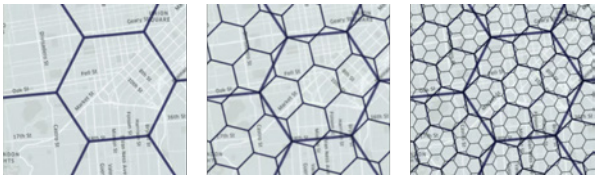

Abb. 2: Vergleich unterschiedlicher Auflösungen des H3-Kachel-Systems von Uber [1]

Graph-basierte Datendarstellung Die geometrischen Daten in den einzelnen Kacheln werden in eine Graph Struktur umgewandelt, wobei die Knoten die Koordinaten und die Kanten die logischen Verbindungen zwischen ihnen darstellen. Durch diese Darstellung können die Daten sowohl als Graph-basierte als auch als sequenzielle Eingaben verarbeitet werden, was Flexibilität für verschiedene Deep-Learning-Architekturen bietet.

Deep-Learning-Modelle Fortgeschrittene Deep-Learning-Modelle, wie z.B. Graph-Neural-Networks oder Sequence-to-Sequence-Modelle, werden verwendet, um Repräsentationen aus der Graph-basierten Datendarstellung zu lernen. Diese Modelle sind darauf ausgelegt, strukturierte Daten effektiv zu verarbeiten und aus ihnen zu lernen, indem sie die inhärenten Beziehungen zwischen Knoten und Kanten nutzen.

Abstandsbasierte Evaluierung Die Ausgabe des Modells, welche die vorhergesagte HD-Karte darstellt, kann mit abstandsbasierten Metriken wie dem euklidischen Abstand, oder auch der Manhattan-Distanz evaluiert werden, was eine quantitative Bewertung der Modellleistung ermöglicht und iterative Verbesserungen erleichtert. Durch die Verwendung des geographischen Kachelsystems und der Graph-basierten Datendarstellung zielt der vorgeschlagene Ansatz darauf ab, die Struktur der geometrischen Eingabedaten effizient zu erfassen. Die Kombination dieser Techniken zusammen mit Deep-Learning-Modellen, die auf die Verarbeitung strukturierter Daten zugeschnitten sind, ermöglicht einen direkteren und effizienteren Weg zum Offline-Lernen von HD-Karten, der die Grenzen herkömmlicher bildbasierter Methoden überwinden soll.

Schlussfolgerung

Die vorgeschlagene End-to-End-Pipeline für das Offline-HD-Kartenlernen unter Verwendung geometrischer Eingabedaten stellt einen innovativen Ansatz dar, um die Herausforderungen der HD-Kartenerstellung für das autonome Fahren zu bewältigen. Durch die Ausnutzung der inhärenten Struktur geometrischer Daten und den Einsatz von Machine-Learning zielt diese Pipeline darauf ab, die Notwendigkeit rechenintensiver Vorverarbeitungsschritte zu umgehen und einen direkteren Weg zur HD-Kartenerstellung zu bieten.

Literatur und Abbildungen

[1] Isaac Brodsky. H3: Uber's Hexagonal Hierarchical Spatial Index. https://www.uber.com/en-DE/blog/h3/, 06 2018.

[2] Yicheng Liu. VectorMapNet: End-to-end Vectorized HD Map Learning. https://arxiv.org/abs/2206.08920, 06 2023.

[3] Juyeb Shin. InstaGraM: Instance-level Graph Modeling for Vectorized HD Map Learning. https://arxiv.org/abs/2301.04470, 01 2023.

Entwurf und Implementierung eines Code-Generators für OPC UA Field Level Communication (FLC)

Cedric Kolarik **Michael Scharf**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Steinbeis Embedded Systems Technologies GmbH, Esslingen

Einleitung

Die Relevanz von Open Platforms Communication (OPC) Unified Architecture (UA) in der Automatisierungstechnik nimmt zu. In Form der OPC UA Field eXchange (FX) Erweiterung wird derzeit daran gearbeitet, OPC UA auch auf der Feldebene einsetzen zu können. Bei der Erarbeitung der Spezifikation ist es von entscheidender Bedeutung, die entwickelten Konzepte in Form von Prototypen zu testen. Da sich diese Prototypen häufig ändern können, ist es von Vorteil, Teile des notwendigen Programmcodes generieren zu können. Der beschleunigte Entwicklungsprozess ermöglicht die zeitnahe Erprobung neuer Konzepte.

OPC UA

Die OPC Foundation bietet mit OPC UA ein standardisiertes Kommunikationsmodell, das in enger Zusammenarbeit mit der Industrie entwickelt wird. Der Standard beinhaltet Informationsmodellierung, Kommunikationswege, Protokollanbindungen und ein Sicherheitsmodell. Für verschiedene Branchen werden individuelle Spezifikationen erarbeitet, die die speziellen Anforderungen der jeweiligen Industrie berücksichtigen.

Informationsmodellierung

Die Informationsmodellierung stellt einen wesentlichen Aspekt von OPC UA dar. Es ist von essentieller Bedeutung, die Daten, Beziehungen und Funktionen von Geräten für den Austausch adäquat darstellen zu können. Jeder OPC UA Server enthält einen „AddressSpace“, der die Objekte des Servers hierarchisch strukturiert. Das Metamodell definiert den Aufbau und Modellierungsregeln für den „AddressSpace“. Informationen werden darin mit Hilfe von „Nodes“ und „References“ modelliert. „Nodes“ können Objekte, Variablen und Methoden repräsentieren oder Typen für diese definieren. „References“ modellieren Beziehungen zwischen „Nodes“. Alle weiteren OPC UA Modelle verwenden den „AddressSpace“ zur Definition ihrer eigenen Informationsmodelle. Um Informationsmodelle auszutauschen, gibt es XML-basierte Nodeset-Dateien, die beispielsweise in einem Modellierungstool erstellt und von Servern eingelesen werden können. [4]

PubSub-Kommunikation

OPC UA bietet die Möglichkeit die Kommunikation zwischen Geräten mithilfe eines Publish-Subscribe (PubSub) Modells zu realisieren. Informationen werden dabei in Form von Nachrichten von einem OPC UA Publisher an ein oder mehrere Subscriber gesendet. Für den Transport der Nachrichten wird eine Middleware verwendet, die entweder Broker-basiert oder Broker-los sein kann. Im Falle der Broker-losen Kommunikation wird die Netzwerkinfrastruktur zur Middleware. [2]

OPC UA auf Feldebene

Die OPC UA FX-Spezifikation beschreibt ein Informationsmodell, das speziell für die Modellierung von Automatisierungsprozessen auf der Feldebene ausgelegt ist. Dazu werden Automatisierungskomponenten definiert, die von den Anwendern individuell nach ihren Anforderungen mit Hard- und Softwareelementen bestückt werden können. Die Funktionalität wird in „FunctionalEntities“ abgebildet, die mit Hilfe des PubSub-Kommunikationsprotokolls Verbindungen zu anderen „FunctionalEntities“ herstellen. Dabei verarbeiten sie „Inputs“, die von anderen „FunctionalEntities“ bereitgestellt werden, und geben über „Outputs“ Daten an andere „FunctionalEntities“ ab. [3]

Aufgaben und Zielsetzung

Für die Prototyping Plattform soll es möglich sein mithilfe von Konfigurationsdateien OPC UA Code zu generieren. Dafür muss vor der Implementierung ein geeignetes Konfigurationsformat gefunden und anhand

der relevanten Informationen definiert werden. Zur Umsetzung benötigt es eine Programmiersprache, die es möglich macht Parser für Konfigurations- und Nodeset-Dateien zu erstellen, die entnommenen Informationen zu verarbeiten und daraus Code zu generieren. Die Anwendung sollte so entworfen werden, dass Nutzer mit technischem Hintergrund die Konfiguration mithilfe des bereitgestellten Schemas durchführen können. Der Code kann nach der Erstellung vom Nutzer weiter manuell angepasst werden.

Programmablauf

Das Diagramm 1 zeigt den Ablauf der Codegenerierung aus Sicht des Anwenders mit Hilfe der Anwendung.

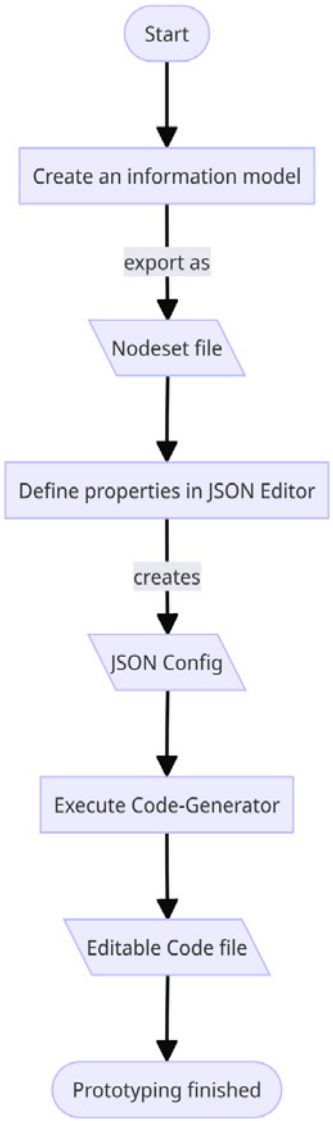

Abb. 1: Workflow des Nutzers [1]

Der Nutzer erstellt zunächst ein Informationsmodell für seine Geräte mithilfe eines Modelling Tools, welches anschließend als Nodeset exportiert werden kann. Die Konfiguration wird anhand des erarbeiteten Schemas erstellt . In der Konfiguration werden die im Nodeset definierten Automatisierungskomponenten um Informationen für die Server, Netzwerkkonfigurationen sowie das Verbindungsmanagement zwischen „Functional-Entities" zur ergänzt. Die Anwendung verarbeitet die bereitgestellten Informationen und generiert aus diesen mithilfe von Code-Templates anpassbaren Quellcode.

Implementierung

Die Architektur der Anwendung ist im Diagramm 2 zu sehen.

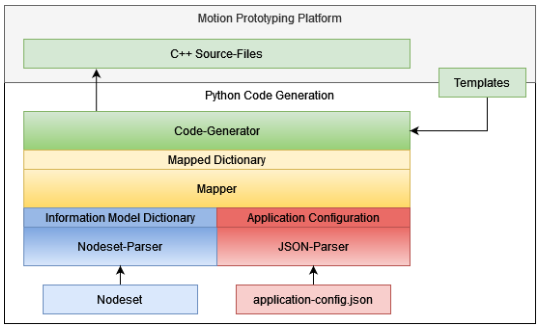

Abb. 2: Architektur der Anwendung [1]

Die Anwendung wird mit der Programmiersprache Python umgesetzt, welche unter anderem einen guten Parser-Support, eine einfache String-Behandlung sowie eine Templating-Engine für die Erzeugung von Code bietet. Die dafür benötigten Templates werden anhand der von Steinbeis entwickelten Prototyping-Plattform erstellt. Die Plattform abstrahiert den offenen OPC UA Stack open62541 und ist in C++ geschrieben. Das Parsen der JSON-Datei erfolgt mithilfe des nativen JSON-Parsers, welcher die Datei in ein Dictionary einliest. Auf diese Weise können die Informationen weiterverarbeitet werden. Für die Verarbeitung der Nodesets wird die „asyncua"-Bibliothek verwendet, welche einen Nodeset-XML-Parser bereitstellt. Die eingelesenen Informationen werden mithilfe eines Mappers mit der Konfiguration verknüpft. Die relevanten Nodes werden in der Konfiguration über den „BrowseName" referenziert, sodass eine Entnahme der entsprechenden Node aus dem Nodeset möglich ist. Nach der Verknüpfung wird das Code-Template mit den Informationen befüllt und der Quellcode erzeugt.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] OPC Foundation. *OPC Unified Architecture Part 14: PubSub*. OPC Foundation, 2024.

[3] OPC Foundation. UAFX Part 81: Connecting Devices and Information Model. https://opcfoundation.org/developer-tools/documents/view/193, 2024.

[4] Wolfgang Mahnke, Stefan-Helmut Leitner, and Mathias Damm. *OPC Unified Architecture*. Springer Berlin / Heidelberg, 2009.

Neuentwicklung eines kompakten und modularen Testsystems mit Single-Pair-Ethernet Multidrop (10BASE-T1S) für Zugriff über Ethernet

Tibor Lederer **Clemens Klöck**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Steinbeis Embedded Systems Technologies GmbH, Esslingen

Einleitung

Im industriellen Kommunikationsumfeld kommen viele verschiedene Feldbusse zum Einsatz, die sich oft schon auf physikalischer Ebene, spätesten aber auf Protokoll-Ebene unterscheiden [3]. Um diese zu verbinden, sind Gateways oder Busknoten nötig, um zwischen den verschiedenen Feldbussen zu übersetzen und einen Datenaustausch zwischen den jeweiligen Systemen zu ermöglichen. Ungünstigerweise gibt es dabei auch Unterschiede im Prinzip, wie Daten auf den jeweiligen Feldbussen ausgetauscht werden, die oft nicht direkt zueinander kompatibel sind und daher aufwändige Sonderlösungen innerhalb der Gateways erfordern.
Besonders interessant wird diese Problematik auf der unteren Ebene der Automatisierungspyramide, auf der Kommunikationsprotokolle genau für ihren speziellen Einsatzzweck konzeptioniert werden. Diese sind oft durch Steckverbindungen limitiert, da Jeder Pin Bauraum im Kabel und auf Platinen benötigt. Oft geht es dabei aber auch einfach nur um die daraus resultierenden Kosten.
Genau diese Problematik wird im Automotive-Bereich seit Jahren, primär durch Kosten motiviert, behandelt und mit Protokollen wie dem Controller-Area-Network (CAN) erfolgreich angegangen. Durch immer weiter steigende Datenmengen durch kamera-basierte Assistenzsysteme kommen die traditionellen Automotiven-Bussysteme schnell an ihre Grenzen. Hier kristallisiert sich aufgrund der hohen Bandbreite und vollständigen, nicht proprietären, Standardisierung durch das Institute of Electrical and Electronics Engineers (IEEE) immer weiter Ethernet (IEEE 802.3) als führende Technik heraus [1]. Der Nachteil an den „Klassischen“ Ethernet-Standards, die man auch im Consumer-Bereich findet, wie 100BASE-T und 1000BASE-T sind dabei die große Anzahl der Kontakte/Adern so wie die Punk-zu-Punkt-Topologie, die diese für den Automotive-Bereich unattraktiv machen.
Die Lösung der IEEE liegt dabei in den neueren Standards, die unter dem Namen Single-Pair-Ethernet zusammengefasst werden können. Darunter ist auch ein Standard namens 10BASE-T1S zu finden, der im Gegensatz zu den anderen Ethernet-Standards eine Multidrop-Topologie unterstützt.
Diese Eigenschaft macht den Standard auch für die Industrielle Kommunikation interessant, um Bussysteme wie RS-485 abzulösen. Dadurch wird es immer weiter möglich bis auf die untersten Schichten herunter mittels Ethernet zu kommunizieren und unnötige Busknoten und Gateways zu eliminieren, was eine Reduzierung der Komplexität und Steigerung der Interoperabilität zwischen den Systemen ermöglicht [5].

Motivation und Zielsetzung

Im Rahmen der Thesis wird ein existierendes modulares Testsystem vom Grunde auf neu für die Verwendung von 10BASE-T1S konzeptioniert und entwickelt. Dabei werden Aspekte wie automatische Arbitrierung für die PLCA Funktionalität so wie Software definierte Terminierung und physikalische Anforderungen für kompakte und modulare Tragschienengehäuse untersucht. Hierbei fällt der Fokus auf die Entwicklung eines einfachen IO-Moduls auf Basis des ARM M0+ basierenden Raspberry Pi RP2040 Mikrocontrollers und dem, per SPI angeschlossenen, LAN8651 MAC-PHY von Microchip. Der Treiber soll hierbei speziell auf langsameren Mikrocontrollern durch Nutzung der ARM-üblich eingebauten DMA-Controller beschleunigt werden, um CPU Zyklen für andere Aufgaben freizuhalten. Mit den fertig entwickelten IO-Modulen soll im Zusammenspiel mit einem Gateway das Verhalten bei einer größeren Menge an Bus-Teilnehmern charakterisiert werden.

Abb. 1: IO-Module des modulares Testsystems [2]

Arbitrierung

Der 10BASE-T1S Standard beinhaltet ein eigenes Token-Ring-Verfahren namens PHY-Level Collision Avoidance (PLCA), um den Buszugriff der einzelnen Teilnehmer zu steuern. Um diese Funktion aktivieren zu können, muss jedes Gerät auf dem Bus eine lokal einzigartige und fortlaufende Node-ID erhalten. Der PLCA-Master bekommt dabei nur die Anzahl der Nodes auf dem Bus und verteilt darauf hin „Transmit Opportunities“ mittels „Beacon“-Signalen [4].
In unveränderlichen Systemen, wie einem Bus im Fahrzeug, bekommt dafür jeder Bus-Teilnehmer manuell eine ID bei der Konzeptionierung zugewiesen. Bei modularen Systemen muss jedoch eine neue Vergabe der IDs (Arbitrierung) erfolgen, wenn sich am Aufbau etwas ändert. Nicht selten wurde das bei Systemen mit RS-485 mittels DIP-Schaltern realisiert, die der Anwender manuell konfigurieren muss.
Eine automatisierte Arbitrierung resultiert in reduziertem Konfigurationsaufwand und kann mithilfe eines weiteren Kommunikationssystems erfolgen. Für diesen Anwendungsfall kommt oft eine Linientopologie zum Einsatz. Diese ermöglicht den physikalischen Aufbau der Module auszunutzen, um eine fortlaufende Adressierung zu gewährleisten.
Für das neu entwickelte System kommt ein einfacher asynchroner UART zum Einsatz, der seriell durch alle Module geschleift wird und eine gemeinsame Leitung, die parallel an alle Teilnehmer angeschlossen wird. Mit ein paar Pullup- und Pulldown-Widerständen, 4 GPIOs und ein wenig Software wird daraus ein voll automatisiertes Arbitrierungssystem das nach Start des SPE Busses auch für andere Funktionen wie dem Verteilen eines SYNC/Burst Signals zur zeitlichen Synchronisierung aller Teilnehmer genutzt werden kann.

10BASE-T1S Physik

Physikalisch zeichnet sich 10BASE-T1S durch zwei Signalleitungen aus, die differenziell mit 100 Ohm Impedanz gekoppelt sind und Signalpegel von +-500mV um eine gemeinsame Masse besitzen. Im Gegensatz zu üblichen Ethernet Standards wie 100BASE-T wird das Signal nicht durch Transformer gekoppelt, sondern durch Kondensatoren. Gegen besseres Abblocken von Störungen wird bei größeren Distanzen eine Stromkompensierte Drossel (Common Mode Choke) zwischen den Kondensatoren und dem PHY verbaut [4].
An beiden Enden des Busses muss dieser terminiert werden, um Reflexionen des Signals zu vermeiden. Dies geschieht im Falle von 10BASE-T1S oft durch einen 50 Ohm Widerstand von jedem der beiden Signalleitungen zur gemeinsamen Masse. Soll diese Terminierung nun Software-Definiert erfolgen, so müssen diese Widerstände zugeschaltet oder getrennt werden können. Eine elegante Lösung für diesen Zweck sind analoge Schalter, bei dessen Auswahl die -500mV Signalpegel beachten werden müssen, da diese oft nur von Rail-to-Rail (0V und VCC in diesem Fall) gehen.

LAN8651 SPI DMA Treiber

Der Treiber für die Kommunikation zwischen RP2040 und dem LAN8651 MAC-PHY besteht zum größten Teil aus mehreren Ringpuffern und einer geschickten Verkettung bestehend aus vier DMA-Kanälen.
Nach der initialen Konfiguration erwartet der LAN8651 für die Applikationsdaten (Ethernet Frames) einen Header (4 Bytes) gefolgt von einem TX-Datenblock (64 Bytes). Zeitgleich schickt dieser einen RX-Datenblock (64 Bytes) gefolgt von einem Footer (4 Bytes), zurück. Der Header enthält verschiedene Bitfelder, die dem MAC-PHY mitteilen, um welche Daten es sich im folgenden Datenblock handelt. Mittels DataValid (Block enthält Daten), StartValid (Block enthält den Start eines Frames) und EndValid (Block enthält das Ende eines Frames), kann ein komplettes Ethernet Frame übertragen werden [4].
Die 64 Byte Block-Aufteilung bedeutet, dass Ethernet Frames beim schicken erst einmal in diskrete Blöcke aufgeteilt werden muss, bevor diese übertragen werden können. Dafür gibt es verschiedene Ringpuffer, die als Transferpuffer agieren, Header und TX-Daten sind dabei genau so getrennt wie RX-Daten und Footer. Diese Trennung ist wichtig, weil teils leere Daten geschickt werden müssen und dafür verschiedene Header benötigt werden. Um den Rechenaufwand zwischen den Transfers minimal zu halten, werden diese jedoch generiert, so bald ein Ethernet-Frame im Ringpuffer abgelegt wird.
Der eigentliche Transfer wird per DMA-Kanälen durchgeführt, diese laufen während dem Transfer komplett

eigenständig und blockiert damit nicht die CPU. Eine Verkettung von je zwei DMA-Kanälen erlaubt dann einen komplett autonomen Transfer pro Block.
Nach jedem Transfer wird in einem Interrupt der Footer analysiert und entschieden, ob im nächsten Zyklus nur Daten geschickt werden (RX-Transferpuffer voll), nur Daten gelesen werden (TX-Transferpuffer leer oder MAC-PHY TX Puffer voll), gelesen und geschickt oder gar nichts transferiert wird. Die Interrupt-Funktion ist dabei auf Laufzeit optimiert, um möglichst viel Zeit im DMA-Transfer zu verbringen und die Zeit dazwischen zu reduzieren. Ist die Analyse abgeschlossen, werden die Source und Destination Pointer der DMA-Kanäle neu konfiguriert und die Transfers wieder gestartet.
Dieser Transfer-Loop läuft damit dauerhaft und bekommt mehr als 13 Mbit (98 % SPI-Auslastung), bei weniger als 1 % CPU-Auslastung, übertragen. Weiter optimiert könnte diese Loop bei Inaktivität pausiert werden und bei neuen TX-Daten oder RX-Daten (Rückmeldung über IRQ Pin) fortgesetzt werden.

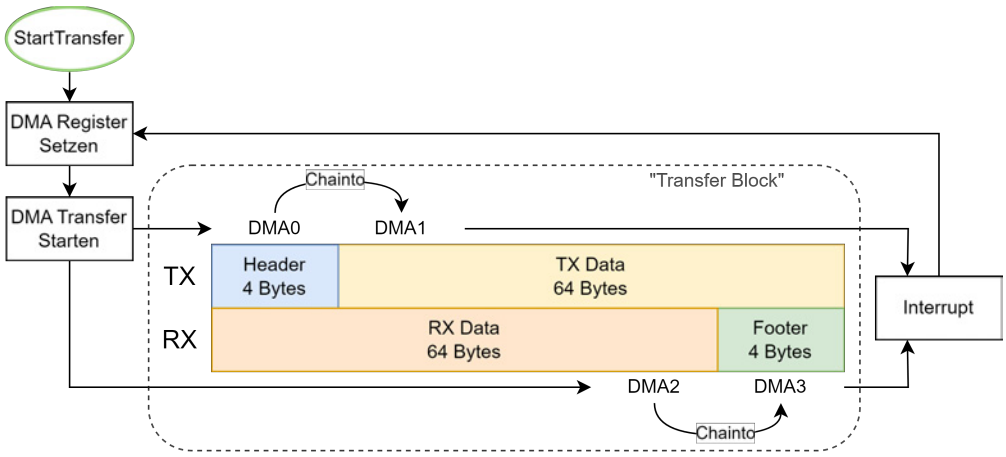

Abb. 2: SPI DMA-Transfer System [2]

Ausblick

Wie bei dem neuen Testsystem können auch andere Systeme in der Zukunft mit Ethernet ausgerüstet werden und damit eine Durchdringung der IEEE 802.3 in allen Schichten der Automatisierungspyramide erzielt werden. Das reduziert unnötige Gateways und steigert die Interoperabilität der Systeme. Die Profite dieser vertikalen Integration liegen in einfacherem Zugriff auf Prozessdaten und einfacherer Integration verschiedenster Systeme [5]. So können einfacher, mehr Daten übertragen und zum Beispiel für Statistiken/Predictive Maintenance verwendet werden.

Literatur und Abbildungen

[1] Amir Bar-Niv and Mark Davis. The Right Stuff: A Past and Future History of Automotive Connectivity. https://www.marvell.com/blogs/the-right-stuff-a-past-and-future-history-of-automotive-connectivity.html, 2022.

[2] Eigene Darstellung.

[3] Jeffrey Hibbard. 5 Real-Time, Ethernet-Based Fieldbuses Compared. https://www.automate.org/tech-papers/5-real-time-ethernet-based-fieldbuses-compared, 2016.

[4] Microchip Technology Inc. Microchip LAN8650/1 Datenblatt. https://ww1.microchip.com/downloads/aemDocuments/documents/AIS/ProductDocuments/DataSheets/LAN8650-1-Data-Sheet-60001734.pdf, 2024.

[5] Fiona Treacy. Accelerating the Transition to Industry 4.0 with Industrial Ethernet Connectivity. https://www.analog.com/en/signals/thought-leadership/acceler-trans-to-ind-4-pt-0-with-ind-ethernet-connect.html, 2020.

Konzeption und Realisierung eines Dashboards für die Auswertung und Visualisierung von Maschinendaten

Julius Liebherr **Catharina Kriegbaum-Kling**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Albert Herz GmbH, Altusried

Einleitung

Die Albert Herz GmbH ist ein Abpackspezialist für Käsereiprodukte. Die Gesellschaft hat ihren Sitz in Altusried im Oberallgäu und ist ein vollständiges Tochterunternehmen des Molkereikonzerns Bayernland eG. Auf mehreren Produktionslinien wird Rohware portioniert und verpackt. Des Weiteren gehört die Reifung von Käselaiben zur Geschäftstätigkeit der Albert Herz GmbH. Neben der regional verbreiteten Eigenmarke wird ein großer Teil für Kunden aus dem Groß- und Einzelhandel produziert. Um ein besseres Verständnis für den Produktionsprozess zu erhalten, soll mithilfe dieser Arbeit eine Möglichkeit geschaffen werden, die Performance einer Produktionslinie zu überwachen und in einem Dashboard visuell darzustellen.

Problemstellung

Industrielle Maschinen erzeugen eine Vielzahl an Daten. Trotzdem werden diese häufig nicht genutzt und damit wertvolle Potenziale verschwendet. Moderne Informationssysteme bieten die Möglichkeit der einfachen und effizienten Aufbereitung und Analyse von Daten. Oft bieten Hersteller von Industriemaschinen dafür dazugehörige Software an, mit der die Maschinendaten ausgewertet werden können. Da die hier betrachtete Produktionslinie jedoch aus Maschinen mehrerer Hersteller besteht, kann ein derartiges System nicht eingesetzt und eine andere Lösung muss implementiert werden. Um Probleme bei der Produktion identifizieren zu können und Lösungen auszumachen, ist eine tiefgehende Kenntnis des Produktionsprozesses notwendig. Alleine die fachliche Expertise der zuständigen Mitarbeiter kann das nicht in einer ausreichenden Komplexität leisten. Subjektive Eindrücke können Faktoren auslassen, verzerren oder vorhandene Muster übersehen. Hierfür kann eine umfassende Datenauswertung neue Blickwinkel schaffen und die Eindrücke mit Zahlen und Daten untermauern.

Industrial Internet of Things

Das Industrial Internet of Things (IIoT) ist die Bezeichnung für das Internet der Dinge mit der Beschränkung auf die industrielle Anwendung. Das Internet der Dinge beschreibt die Vernetzung von Geräten, Sensoren und anderen Gegenständen untereinander. Durch diese Vernetzung werden alltägliche Gegenstände zu „intelligenten“ Teilnehmern eines Netzwerks bzw. des Internets. Der Vorteil liegt im Informationsaustausch und der damit einhergehenden Automatisierung. In der Industrie wird diese Thematik häufig im Kontext der Vernetzung von Maschinen, Sensoren und anderen physischen Systemen benutzt, wobei in Abgrenzung zum Internet der Dinge andere Faktoren, wie Sicherheit oder Verfügbarkeit eine höhere Rolle spielen [6].

OPC-UA

Für die industrielle Kommunikation existieren Schnittstellen, die dafür sorgen, dass mehrere Systeme einander verstehen. Einer der am weitesten verbreiteten Schnittstellenstandards nennt sich OPC-UA, Kurzform für Open Platform Communication – Unified Architecture [3]. Er definiert die Spezifikationen für eine stabile, sichere und unabhängige Verbindung zwischen Entitäten. Eingesetzt wird dieser in der Industrie vorwiegend zur Vernetzung und Automatisierung von Maschinen. Entwickelt und gewartet wird der Standard von der OPC-Foundation. Durch die Herstellerunabhängigkeit wird eine Vernetzung über einzelne Hersteller oder Betriebssysteme hinaus gewährleistet. Schon der Vorgänger von OPC-UA, OPC Classic, findet sich in der gesamten Industrie. Dieser war jedoch nicht mehr zeitgemäß und zusätzlich von einer externen Technologie abhängig und wurde daher von OPC-UA abgelöst [4].

OEE

Die Overall Equipment Effectiveness, kurz OEE, ist eine der bedeutendsten Kennzahlen in Industriebetrieben.

Sie gibt prozentual den Anteil der Zeit, in der Gutteile produziert werden, im Verhältnis zu der maximal verfügbaren Produktionszeit an. Dabei werden die Faktoren Qualität, Leistung und Verfügbarkeit betrachtet. Durch die Einbeziehung der drei Faktoren werden die wichtigsten Performanceindikatoren aufgeschlüsselt und Ursachen können genauer bestimmt werden [2].

Durchführung

Zu Beginn des Projektes steht die Anforderungsanalyse. Hierfür muss erarbeitet werden, was von der Anwendung erwartet wird und wie diese aufgebaut sein soll. In diesem Fall bestand das vorrangig in der Definition der Daten, welche für die spätere Auswertung angedacht sind. Jede Maschine hat ihre speziellen Kennzahlen, die für diese relevant sind.

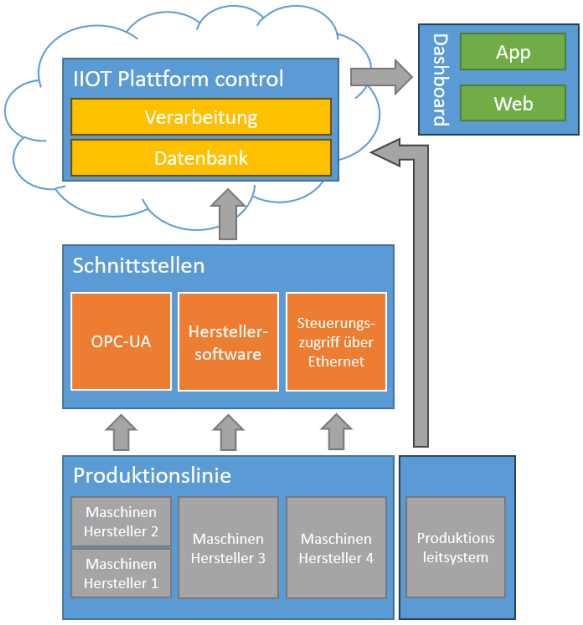

Abb. 1: Aufbaulogik System [1]

Um die Maschinendaten aus den Maschinen zu erhalten, bieten Industriemaschinen verschiedene Schnittstellentechnologien an. Aufgrund mehrerer Hersteller müssen hierbei mehrere Ansätze gewählt werden. Wie in Abbildung 1 zu sehen, werden für das Projekt drei verschiedene Methoden benutzt. Einmal über die Ethernet Schnittstelle, indem auf die Steuerung der Maschine direkt zugegriffen wird. Eine weitere ist die Benutzung einer Herstellereigenen Software, über die die Daten abgegriffen werden können. Die anderen Hersteller bieten eine OPC-UA Schnittstelle an. Über die Schnittstellen werden die Daten in das System geladen und dort in einer Datenbank gespeichert. Dazu kommen weitere Daten aus dem Produktionsleitsystem direkt in die Datenbank. Von hier aus können die Maschinendaten verarbeitet und in der gewünschten Art in den Dashboards dargestellt werden. Für das System wurde eine IIoT Plattformlösung der Eberle Automatische Systeme GmbH & Co KG gewählt. Diese entspricht den Anforderungen an das System und ermöglicht eine passgenaue, aber dennoch professionelle Lösung. Dadurch kann sichergestellt werden, dass sowohl Sicherheit als auch Verfügbarkeit auf einem adäquaten Stand sind. Vor allem Sicherheit ist eine wichtige Thematik, da die Anwendung in der Cloud betrieben wird und auch extern, z.B. über das Smartphone, erreichbar und daher besonders vor fremden Zugriff geschützt werden muss, um keine sensiblen Firmendaten offenzulegen. In Abbildung 2 ist eine beispielhafte Darstellung eines Dashboards innerhalb der gewählten Plattform control zu sehen.

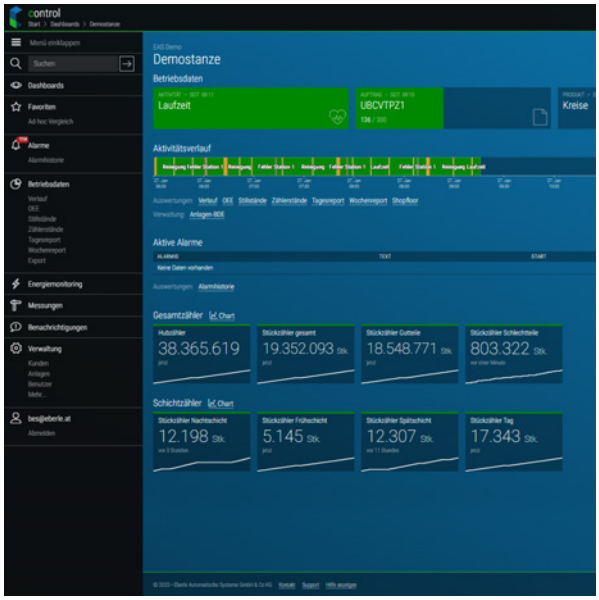

Abb. 2: Beispiel Dashboard control [5]

Ausblick

Im weiteren Verlauf der Bachelorarbeit wird das System aufgesetzt und die Dashboards gemäß den Anforderungen erstellt. Sobald das System in Betrieb genommen wurde und fehlerfrei arbeitet, kann damit begonnen werden, die abgebildeten Daten genauer zu untersuchen. Damit wird es möglich sein, ein besseres Verständnis für den Produktionsprozess zu entwickeln und Erkenntnisse zu erlangen, wo und wodurch, Probleme und Leistungsdefizite verursacht werden. Sofern sich das System mit der Zeit beweisen kann und seinen angedachten Zweck erfüllt, kann über eine Ausweitung der Datenanalyse auf die weiteren Produktionslinien des Unternehmens nachgedacht werden.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Markus Focke and Jörn Steinbeck. *Steigerung der Anlagenproduktivität durch OEE-Management*. Springer Gabler, 2018.

[3] OPC Foundation. Unified Architecture. https://opcfoundation.org/about/opc-technologies/opc-ua/, 2024.

[4] OPC Foundation. What is OPC? https://opcfoundation.org/about/what-is-opc/, 2024.

[5] Eberle Automatische Systeme GmbH und Co KG. IoT Plattform control. https://www.eberle.at/de/produkte/control/, 2024.

[6] Ljiljana Stojanovic and Olaf Sauer. Industrial Internet of Things (IIoT). https://www.iosb.fraunhofer.de/de/geschaeftsfelder/automatisierung-digitalisierung/anwendungsfelder/iiot.html, 2024.

Evaluierung der Leistungsfähigkeit von ChatGPT in der Laborübung Programmieren: Sicherstellung des Kompetenzerwerbs mit KI-basierter Unterstützung

Johannes Loser **Reiner Marchthaler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

In der heutigen digitalen Ära hat die künstliche Intelligenz (KI) eine bemerkenswerte Entwicklung erfahren und ist zu einem integralen Bestandteil unseres Alltags geworden. Eine faszinierende Anwendung dieser Technologie ist die Generierung menschenähnlicher Texte, wie sie in Chatbot-Systemen wie ChatGPT von OpenAI demonstriert wird. Diese Abschlussarbeit untersucht, inwieweit das neueste Sprachmodell GPT-4 dazu geeignet ist, die Laboraufgaben des Moduls "Programmieren"korrekt zu lösen und wie die Qualität der generierten Quellcodes bewertet werden kann. Zudem wird analysiert, wie häufig Studierende der Hochschule Esslingen KI-basierte Chatbots zur Unterstützung im Studium nutzen und ob der Einsatz dieser Hilfsmittel den Aufbau ihrer Programmierkompetenz positiv oder negativ beeinflusst. Auf Grundlage dieser Ergebnisse sollen Vorschläge zur Anpassung der Laboraufgaben gemacht werden.

Grundlagen

ChatBots wie ChatGPT basieren auf der revolutionären Transformerarchitektur, die 2017 von Vaswani et al. [3] in der Forschungsarbeit mit dem Titel „Attention is all you need“ vorgestellt wurde. Diese Architektur zeichnet sich durch die Verwendung von Aufmerksamkeitsmechanismen aus. Dies ermöglicht dem Modell, sich auf relevante Teile der Eingabe zu konzentrieren, wodurch die Leistung im Vergleich zu früheren Architekturen erheblich verbessert wird. Traditionelle sequenzielle Modelle wie LSTM (Long short-term memory) und GRU (Gated recurrent Unit) haben Schwierigkeiten, lange Abhängigkeiten effektiv zu modellieren, was zu Informationsverlust und schlechter Leistung bei der Verarbeitung von Textdaten führen kann.
Die Transformer-Architektur hat diese Architekturen übertroffen und wird nun als Standardansatz im Bereich des Natural Language Processing (NLP) verwendet.
Es gibt verschiedene Varianten des Transformer-Modells, darunter BERT (Bidirectional Encoder Representations from Transformers) von Google, das eine encoder-basierte Architektur verwendet, und ChatGPT, das auf der GPT-Architektur (Generative Pre-trained Transformer) basiert und Reinforcement Learning from Human Feedback (RLHF) einsetzt.

Beschreibung des Labors

Das Labor besteht aus 10 einzelnen Abgaben, die thematisch auf die Vorlesungsinhalte abgestimmt sind (z. B. Laboraufgabe 01: Ein- und Ausgabe, Laboraufgabe 02: Rechnen und Verzweigungen, Laboraufgabe 03: Schleifen). Die Studierenden haben jeweils zwei Wochen Zeit, um die Labore in Moodle abzuschließen. Jede Laborabgabe besteht aus vier bis zehn einzelnen Aufgaben, darunter offene Programmieraufgaben, Multiple-Choice-Fragen als auch Fehlerkorrektur-Aufgaben.
Insgesamt umfasst das Labor 83 (Programmier-)Aufgaben, die folgendermaßen verteilt sind:

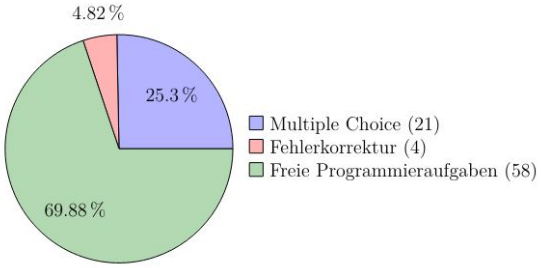

Abb. 1: Übersicht der Laboraufgaben [1]

Analyse der Performance von ChatGPT

Im Rahmen der Auswertung wurden die Aufgabentypen Multiple Choice, Fehlerkorrektur und offene Programmieraufgaben getrennt voneinander analysiert. Diese differenzierte Betrachtung ermöglichte eine präzise Bewertung der Leistungsfähigkeit von ChatGPT in den verschiedenen Aufgabenbereichen.

Aufgabentyp	Durchschnittlicher Schwierigkeitsgrad (1-5)	Korrektheit (%)
Multiple Choice	2,3	96,66
Fehlerkorrektur	1,4	100
Offene Programmieraufgaben	3,0	95,4

Abb. 2: Performance von ChatGPT bei den unterschiedlichen Aufgabentypen [1]

Die Auswertung zeigte, dass ChatGPT unabhängig vom Schwierigkeitsgrad der Laboraufgaben eine sehr gute Performance bietet und fast alle Aufgaben perfekt lösen kann. 58 freie Programmieraufgaben wurden zu 95,4% korrekt gelöst, 21 Multiple Choice Fragen zu 96,66% richtig bearbeitet und 4 Fehlerkorrektur Aufgaben zu 100% richtig gelöst. Ein statistischer Zusammenhang zwischen dem Schwierigkeitsgrad der Aufgaben und der Performance von ChatGPT konnte nicht signifikant nachgewiesen werden. Dies deutet darauf hin, dass GPT-4o unabhängig von der Komplexität der Aufgaben stets eine hohe Leistung erbringt. Auch die Codequalität war in puncto Modularität, Lesbarkeit, Klarheit und Effizienz stets von hoher Qualität.

Nutzungsverhalten der Studenten

Im Rahmen der Abschlussarbeit wurde das Nutzungsverhalten von KI-Chatbots wie ChatGPT oder Google Gemini von Studierenden im 1. und 2. Semester untersucht. Bei einer Umfrage gaben 60% der Studierende an, diese Tool bereits regelmäßig im Rahmen des Studiums einzusetzen. Überraschenderweise würden jedoch lediglich 20% der Studierenden KI-Chatbots zum Lösen von Programmieraufgaben verwenden, wenn dies hypothetisch als Hilfsmittel erlaubt wäre. Der Großteil der Studierenden würde diese Tools vorwiegend zum Debugging, zur Fehlersuche oder zum besseren Verständnis von Programmierkonzepten nutzen. Diese Ergebnisse lassen sich darauf zurückführen, dass 80% der Studierenden befürchten, die regelmäßige Nutzung von KI-Chatbots als Hilfsmittel zur Bearbeitung der Laboraufgaben könnte zu einer Abhängigkeit von diesen Tools und zur Beeinträchtigung ihrer Problemlösefähigkeiten im Programmieren führen.

Ausblick

Ob die Verwendung von KI-Chatbots die Entwicklung der Programmierfähigkeiten positiv oder negativ beeinflusst, hängt maßgeblich vom Nutzungsverhalten und den individuellen Vorkenntnissen ab. Der Einsatz von KI-Tools wie ChatGPT zur Bearbeitung von Programmieraufgaben kann problematisch sein, da er die aktive Auseinandersetzung, Integration neuen Wissens, Entwicklung komplexer Schemata und Problemlösungskompetenzen der Studierenden hemmt, wodurch wichtige Lernprozesse und -phasen nach der Cognitive Load Theory [2](CLT) übersprungen werden. Basierend auf den Umfrageergebnissen besteht diese Gefahr jedoch lediglich für eine Minderheit der Studierenden. Diese Erkenntnisse sollen für die Umstrukturierung und Verbesserung der Laboraufgaben genutzt werden.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] M. Sweller. Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17:147–177, 2005.

[3] Ashish Vaswani et al. Attention Is All You Need. https://arxiv.org/abs/1706.03762, 2017.

Adversary Emulation zum Vergleich des Sicherheitsniveaus verschiedener Systemkonfigurationen und -versionen in Windows-Umgebungen

Julian Mayer **Martin Mink**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma ERNW Enno Rey Netzwerke GmbH, Heidelberg

Motivation

Adversary Emulation, oftmals auch als Red Teaming bezeichnet [5], zielt darauf ab, einen Angreifer zu imitieren. Dabei werden reale Taktiken, Techniken und Prozeduren von Angreifern verwendet, um die Sicherheit von IT-Systemen einer Organisation gegenüber realen Angriffen zu testen. Die Sicherheit eines Systems wird maßgeblich von dessen Aktualität und Konfiguration beeinflusst. Windows-Betriebssysteme sind weltweit am weitesten verbreitet [4], und die meisten Organisationen verwenden diese zusammen mit dem Active Directory-Verzeichnisdienst. Der hohe Verbreitungsgrad, zusammen mit komplexen Konfigurationsmöglichkeiten, macht diese Systeme sehr interessant für Angreifer mit den unterschiedlichsten Motiven. Daher ist eine sichere Konfiguration (typischerweise auch als „Hardening“ bezeichnet) dieser Systeme wesentlich für den zuverlässigen Geschäftsbetrieb. Die spezifische Konfiguration dieser Systeme ist von Organisation zu Organisation individuell. Es existieren unterschiedliche Handlungsempfehlungen, sog. Best Practices oder auch Hardening Guides, um die Systeme sicher zu konfigurieren. Der Unterschied des Sicherheitsniveaus zwischen verschiedenen Hardening-Konfigurationen ist oft nicht trivial zu ermitteln, da meist eine unübersichtlich große Anzahl an Einstellungen vorhanden und außerdem tiefes technisches Verständnis notwendig ist. Auch die Auswirkung einer einzelnen Konfigurationseinstellung auf die Systemsicherheit ist somit häufig nicht direkt ersichtlich, begünstigt durch teilweise fehlende oder unvollständige Dokumentation der Einstellungen. In der Praxis kommt es regelmäßig vor, dass IT-Administratoren einzelne Konfigurationsänderungen vornehmen. Zudem können durch neue Systemversionen neue Konfigurationsoptionen auftauchen. Dadurch können Empfehlungen und Hardening Guides schnell veraltet sein bzw. es wird aufgrund bestimmter Anforderungen im Unternehmen von diesen abgewichen. Die Möglichkeit festzustellen, wie sich das Sicherheitsniveau durch einzelne Konfigurationseinstellungen oder neue Systemversionen verändert, bringt insbesondere für IT-Administratoren einen erheblichen Mehrwert, um eine Organisation zuverlässig gegen Angriffe abzusichern.

Ziele der Arbeit

Ziel der Arbeit ist es, eine Methodik zu entwickeln, die Adversary Emulation verwendet, um Sicherheitsunterschiede zwischen verschiedenen Systemkonfigurationen und -versionen festzustellen. Der Fokus soll hierbei auf Windows-Betriebssystemen in einer Active Directory-Umgebung liegen. Dabei wird davon ausgegangen, dass es sich um Domain-joined Systeme ohne zusätzlich installierte Software handelt. Die Methodik soll beinhalten, welche Angriffstechniken nötig und geeignet sind, um die Sicherheit solcher Systeme bzw. deren Konfiguration umfänglich zu prüfen. Basis hierfür soll das MITRE ATT&CK Framework [1] sein.

Es sollen verschiedene Systeme bzw. Versions- oder Konfigurationsstände angegriffen und die Ergebnisse verglichen werden. Die Methodik soll beschreiben, wie man die Ergebnisse vergleichen und interpretieren kann. Zudem soll ein schrittweiser Ansatz erarbeitet werden, mit dem geprüft werden kann, wie sich einzelne Konfigurationseinstellungen (z. B. Einstellungen eines Hardening Guides) auf die Sicherheit eines Systems auswirken.

Zuletzt soll die zuvor beschriebene und erarbeitete Methodik implementiert und in einem Laborversuch angewandt werden. Dabei soll ein Adversary Emulation Tool verwendet werden, um die Angriffe automatisiert auszuführen. Bei Bedarf soll dieses Tool erweitert werden, z. B. durch die Implementierung fehlender Angriffstechniken. Die Ergebnisse dieses Tools sollen, wenn möglich, automatisiert verglichen und interpretiert werden, um das Sicherheitsniveau zweier Systeme festzustellen und zu unterscheiden. Zudem soll für das

gewählte Tool ein schrittweiser Ansatz implementiert werden, um die Auswirkungen einzelner Konfigurationseinstellungen anhand eines Beispielszenarios zu untersuchen.

MITRE ATT&CK Framework

Das MITRE ATT&CK Framework [1] ist eine Wissensbasis bestehend aus Angriffstechniken, die in der Realität beobachtet wurden. Ein zentraler Bestandteil ist die MITRE ATT&CK Matrix, die eine Übersicht über die verschiedenen Angriffstechniken bietet und diese in Taktiken einordnet. MITRE ATT&CK führt zu jeder Angriffstechnik Beispiele, Empfehlungen zur Mitigation und Erkennungsmöglichkeiten auf. Diese Informationen helfen, um die Methoden von Angreifern besser zu verstehen und somit die IT-Sicherheit zu verbessern. Eine wichtige Komponente, um die Sicherheit von IT-Systemen zu überprüfen, ist die Simulation realer Angriffsszenarien. Dafür dient das MITRE ATT&CK Framework oft als Grundlage.

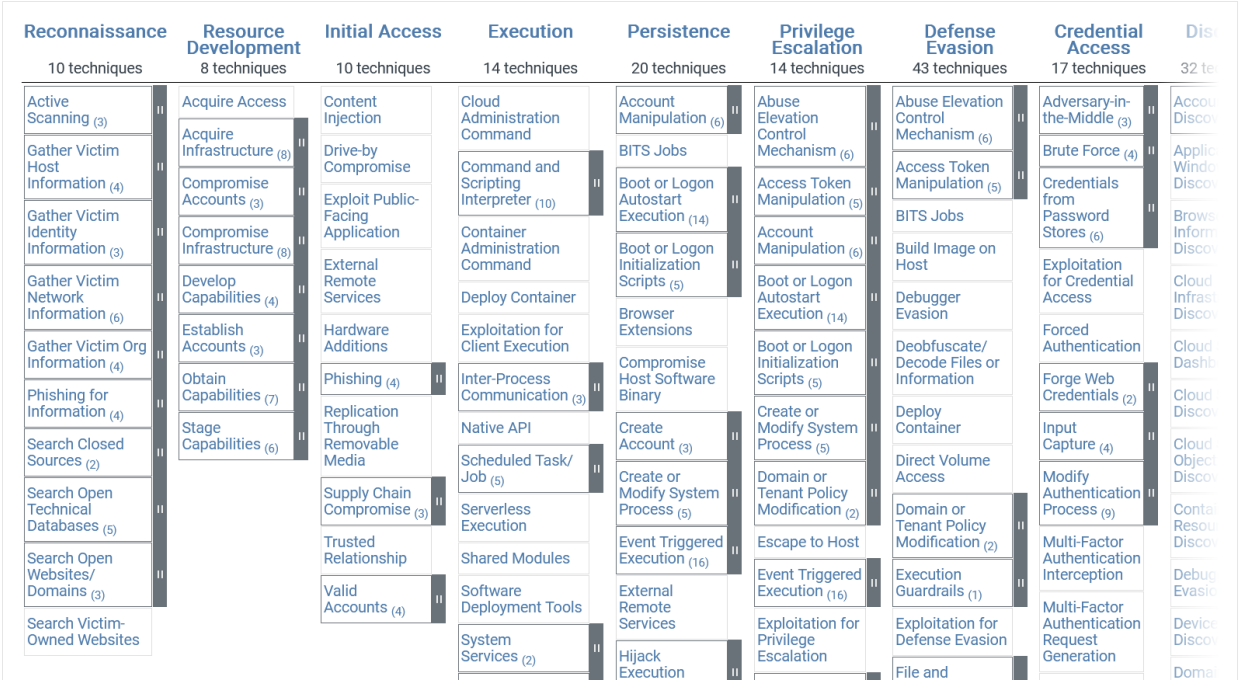

Abb. 1: Auszug aus der MITRE ATT&CK Enterprise Matrix [3]

Ausblick

Da sich die Arbeit noch im Anfangsstadium befindet, können noch keine Ergebnisse präsentiert und nur ein Ausblick gegeben werden.

Es werden nicht alle Angriffstaktiken und -techniken, die in realen Angriffen vorkommen, geeignet sein, um die Ziele zu erreichen. Zum Beispiel beschäftigen sich bestimmte Taktiken bzw. Techniken mit der Umgehung von AV-Lösungen oder IDS-Systemen, in der Arbeit soll aber nur das Betriebssystem selbst angegriffen bzw. geprüft werden. Auch Techniken, die bekannte Softwareschwachstellen ausnutzen, sind in der Regel nicht zielführend, da die Schwachstellen bereits gepatcht sind und somit nur geprüft werden würde, ob das System die neuesten Updates enthält. Zudem gibt es Angriffstechniken, die Betriebssystem Features nutzen, die durch eine Systemkonfiguration nicht beeinflusst bzw. eingeschränkt werden können.

Um eine Systemkonfiguration umfänglich prüfen zu können, wird ein Ansatz nötig sein, der zwei verschiedene Perspektiven berücksichtigt. Zum einen werden Angriffe lokal auf dem System ausgeführt, also als hätte ein Angreifer schon initialen Zugriff zum System. Demnach ist der Test nicht von der Eindringung in das System abhängig. Zum anderen wird dann das System von außen angegriffen, um Lateral Movement Szenarien abzudecken. Somit werden die Ergebnisse von Angriffen innerhalb des Systems und von außen gegen das System kombiniert.

Nach ersten Vergleichen verschiedener Adversary Emulation Tools scheint MITRE Caldera [2] das am besten zur Umsetzung geeignete Tool zu sein. Caldera baut auf dem MITRE ATT&CK Framework auf und enthält bereits einige Implementierungen von Techniken als sogenannte *Abilities*. Diese können in einem *Adversary Profile* gesammelt und kombiniert werden. Die Profile können in *Operations* gegen Ziele ausgeführt werden. Dazu wird auf den Zielen ein Agent

installiert, der die Befehle und Angriffe entgegennimmt und ausführt. Nach Beendigung einer *Operation*, gibt es eine Zusammenfassung über die Ergebnisse der einzelnen Angriffe. Diese können zur Auswertung weiterverarbeitet werden. Zudem ist Caldera durch eigene *Abilities* und Plugins erweiterbar und besitzt eine umfangreiche REST API. Somit kann später der Vergleich verschiedener Systeme und der schrittweise Ansatz zur Ermittlung der Auswirkungen einzelner Konfigurationseinstellungen automatisiert werden.

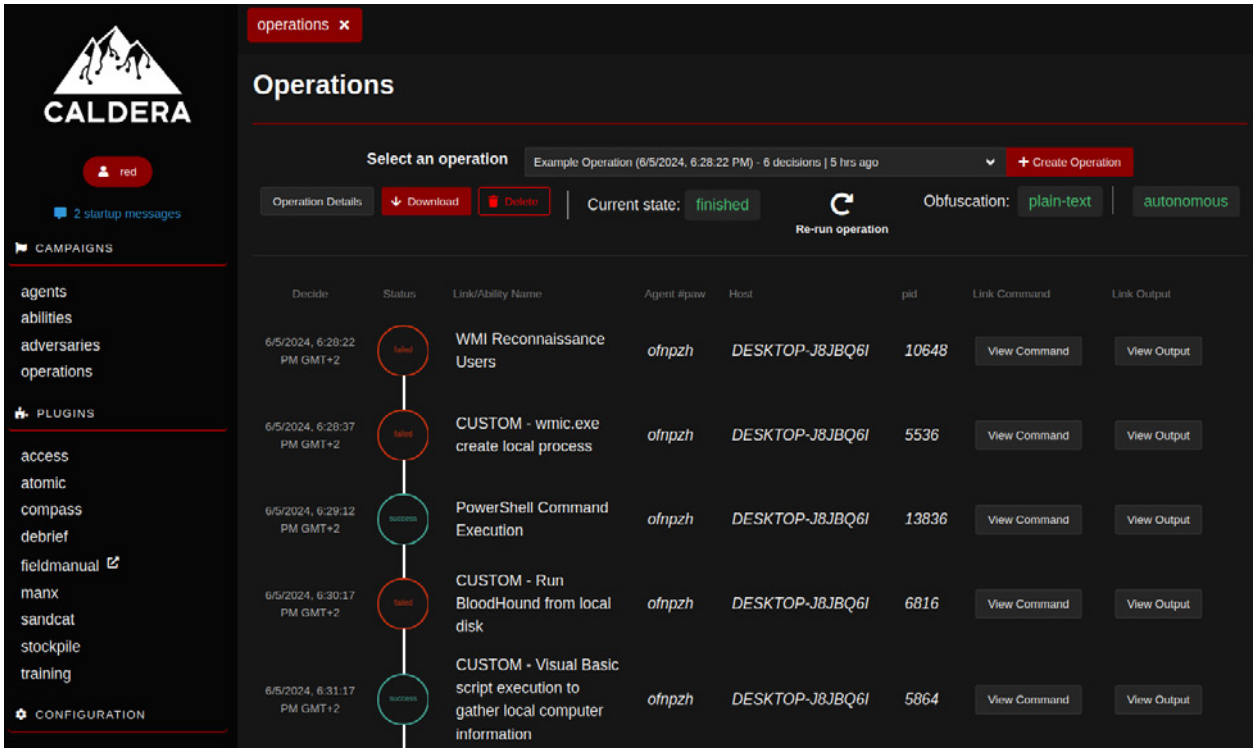

Abb. 2: Beispielhafte Operation in der Caldera Weboberfläche [3]

Literatur und Abbildungen

[1] MITRE Corporation. MITRE ATT&CK®. https://attack.mitre.org/, 2024.

[2] MITRE Corporation. MITRE Caldera. https://caldera.mitre.org/, 2024.

[3] Eigene Darstellung.

[4] Statcounter GlobalStats. Desktop Operating System Market Share Worldwide. https://gs.statcounter.com/os-market-share/desktop/worldwide, 2024.

[5] Blake Strom, Tim Schulz, and Katie Nickels. Getting Started with ATT&CK: Adversary Emulation and Red Teaming. https://medium.com/mitre-attack/getting-started-with-attack-red-29f074ccf7e3, 07 2019.

Gegenüberstellung von generischem und konkretem Entwicklungsansatz und deren jeweiligen Auswirkungen bei der Entwicklung von CRM-Systemen

Matthias Meier **Dieter Morgenroth**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Promotive GmbH, Böblingen

Einleitung

Das 21. Jahrhundert ist in der Welt der Softwaretechnik geprägt von kontinuierlich optimierten Konzepten in der Softwareentwicklung. Microservices sind hierbei in den 2010er Jahren eine bahnbrechende Neuheit. 2009 erstmals von Netflix als Alternative zu einer monolithischen Architektur deklariert, wurde der Begriff „Microservice“ im Jahr 2012 bei der „33rd Degree“ Konferenz in Krakau vorgestellt. Mittlerweile sind neben Netflix auch Amazon, Spotify und viele weitere Weltkonzerne auf Microservice-Architekturen umgestiegen [1]. Jedoch sind es nicht nur Konzepte in der Softwareentwicklung, sondern auch die entwickelten Endprodukte, die regelmäßig optimiert werden. In den letzten 15 Jahren stieg dabei die Beliebtheit von Customer Relationship Management-Systemen (CRM-Systemen). Diese werden heutzutage in fast allen Branchen benutzt, beispielsweise im Einzelhandel, Banken oder im Gesundheitswesen und umfasst dabei die in Abbildung 1 dargestellten Features.

Abb. 1: Features von Customer Relationship Management Systemen [4]

Problemstellung

Die sich ständig ändernden, branchenspezifischen Kundenanforderungen an ein CRM-System, zusätzlich zum gewünschten Umfang des Systems, lassen für die Entwickler verschiedenste Implementierungsmöglichkeiten zu. Hierzu zählt die Wahl zwischen generischen Datentypen und konkreten Datenobjekten, die bereits beim Entwurf der Spezifikation grundlegende Unterschiede aufweisen. Oberflächlich machen sich die Unterschiede ohnehin bemerkbar, da die Nutzung eines generischen Ansatzes mit den Grundsätzen einer NoSQL-Datenbank lediglich einen Nutzer beinhaltet, der bei erster Nutzung der Anwendung jegliche Attribute und Tabellen, sowie die darin gespeicherten Objekte, vollends selbst erstellen kann. Der konkrete Ansatz stellt wiederum eine gewöhnliche Webanwendung dar, die bereits während der Erstellung der Architektur eine enge Zusammenarbeit mit dem Kunden erfordert, da die Tabellen, Attribute und Entitäten schon vor der ersten Nutzung erstellt werden müssen. Eine bildliche Darstellung für die verschiedenen Datenbanken bietet Abbildung 2.

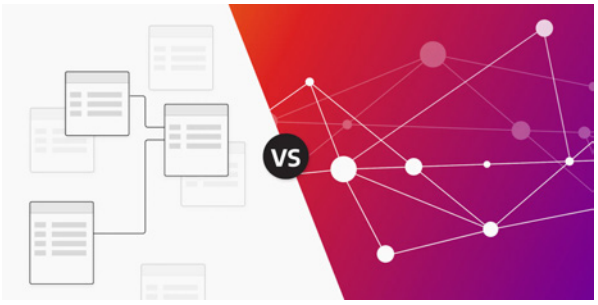

Abb. 2: SQL vs. NoSQL [3]

Für die Backend-Entwicklung ist die Wahl zwischen beiden Ansätzen jedoch einflussreicher, da neben den Auswirkungen auf Skalierbarkeit, Wartbarkeit und

Wiederverwendbarkeit auch auf der Sicherheit ein Hauptaugenmerk liegen muss. Diese ganzen Unterschiede wirken sich auf die Komplexität und Ressourcen aus. Im Endeffekt also auch auf die Kosten für die Entwicklung und Instandhaltung der Software.

Stand der Forschung

Die Beliebtheit von CRM-Systemen steigt über die letzten 15 Jahre stetig [5], während die Nutzung von Microservice-Architekturen in größeren Softwareanwendungen über die letzten 10 Jahre ohnehin einen enormen Anstieg verzeichnet [2]. Aufgrund der Aktualität des wachsenden Zusammenhangs gibt es zu vielen Methoden und deren Auswirkungen auf Performance und Entwicklungsaufwand noch keine Forschung. Dazu zählt auch die Wahl der Herangehensweise, ob man eine Anwendung mit generischen Datentypen oder mit konkreten Datenobjekten entwickeln möchte. Die individuellen und ständig wechselnden Kunden- beziehungsweise Nutzer-Anforderungen stellen die Komplexität der Wahl in der Herangehensweise dar. Ob man die richtige Wahl getroffen hat, lässt sich erst nach mittel- bis langfristiger Nutzung einer Anwendung beantworten.

Zielsetzung

Das Ziel dieser Arbeit ist die Beantwortung der Frage, ob die Entwicklung von CRM-Systemen, basierend auf einer Microservice-Architektur, mit generischen Datentypen als sinnvolle Alternative zu einer Entwicklung mit konkreten Datenobjekten darstellt und wann dies der Fall wäre. Der Fokus wird dabei auf die zu beachtenden Maßstäbe bei der Architekturentwicklung gelegt.

Verfahren

Der Vergleich soll mittels zweier Minimum Viable Products (MVPs) durchgeführt werden. Dabei werden die Besonderheiten beider Ansätze, parallel zur Entwicklung, hinsichtlich der Architekturbeschreibung und der Implementierung ausgewertet. Nach Fertigstellung beider MVPs gilt es, die Auswirkungen beider Ansätze auf die Qualitätsmerkmale von Software mit variierenden Datenmengen bei Ausführung der CRUD-Methoden zu messen. Eine Veranschaulichung, wie die Auswertung der Messungen aussehen könnte, zeigt Abbildung 3.

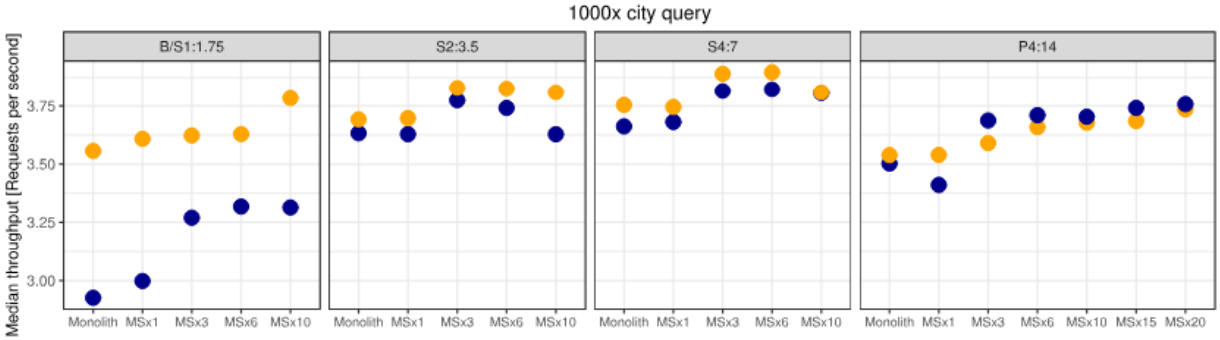

Abb. 3: Beispiel für die Messung der vertikalen Skalierung bei unterschiedlichen Anzahlen von Microservices im Vergleich zu einem Monolithen [1]

Literatur und Abbildungen

[1] Grzegorz Blinowski, Anna Ojdowska, and Adam Przybyłek. Monolithic vs. Microservice Architecture: A performance and scalability evaluation. *IEEE Access*, pages 1, 11, 2022.

[2] Marek Gajda. State of Microservices 2020. https://tsh.io/state-of-microservices/#developers, 2020.

[3] Matea Pesic. SQL vs NoSQL Databases. https://memgraph.com/blog/sql-vs-nosql-databases, 06 2023.

[4] PerfectView CRM PVCRM. Was ist CRM? https://www.perfectviewcrm.de/was-ist-crm/, 2024.

[5] Mark Taylor. 18 CRM Statistics you need to know for 2023 (and beyond). https://www.superoffice.com/blog/crm-software-statistics, 2022.

Praxisbezogener Leitfaden zur Implementierung eines Data Governance Frameworks

Andreas Menzel **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Seit mehr als 15 Jahren nimmt das Datenvolumen in Unternehmen exponentiell zu. Durch Big Data und das Internet of Things (IoT) hat sich dieses Wachstum weiter beschleunigt und mit den Fortschritten im Bereich der künstlichen Intelligenz anschließend einen erneuten Auftrieb bekommen. Zeitgleich hat auch die Vielfalt der Daten durch neue Datentypen und -Formate und durch die Zusammenführung aus einer steigenden Anzahl unterschiedlicher Quellen weiter zugenommen. Diese Datenflut zu kontrollieren und zu verwalten, stellt für die Datenverwaltung oft ein enormes Problem dar und erschwert es, eine gleichbleibend hohe Qualität dieser Daten zu erreichen. Das ist insbesondere deshalb problematisch, da Unternehmen ihre Datenbestände als wertvolle Ressource betrachten und sie für faktengestützte Entscheidungen unverzichtbar sind. Somit müssen sie eine hohe Qualität haben, um zur Wettbewerbsfähigkeit beizutragen. Eine weitere, wenn nicht sogar gravierendere Konsequenz ergibt sich aus der größeren Angriffsfläche und den steigenden potenziellen Schwachstellen, weshalb die Datensicherheit schwieriger umzusetzen ist. Die Verbreitung von Remote-Arbeit und Cloud-Computing haben in diesem Zusammenhang weitere Herausforderungen geschaffen. Der durch die aufgezählten Faktoren entstehende Druck auf Unternehmen erhöht sich letztendlich nochmals durch regulatorische Anforderungen, die im Rahmen von Gesetzen wie der Datenschutz-Grundverordnung (DSGVO) festgelegt werden. Aus diesen Entwicklungen resultiert, dass das Thema Data Governance für Unternehmen immer mehr an Bedeutung gewonnen hat und mittlerweile als unverzichtbar gilt. Bei einer aktuellen Studie von BARC aus dem Jahr 2023 wurden 2398 Daten-, BI- und Analytics-Spezialisten danach befragt, welche Themen aus den Bereichen Data, BI und Analytics sie für das Jahr 2024 als wichtigste Trends wahrnehmen. [1] Dabei landete Data Governance nach „Data security and privacy“ und „Data quality management“ auf dem dritten Platz, gefolgt von 17 weiteren Themen. Abbildung 1 zeigt die Ergebnisse und die Entwicklung derselben Umfrage aus den vorherigen Jahren. Es lässt sich erkennen, dass in den Prognosen für die Jahre 2020 bis 2023 die Datenqualität mit dem ersten Platz durchgehend als wichtigster Erfolgsfaktor angesehen wurde. Dass die Data Governance während dieser Zeit so gut in den Umfragen abgeschnitten hat, lässt sich dadurch erklären, dass sie nicht isoliert vom Datenqualitätsmanagement betrachtet werden kann. Aus der Prognose für das Jahr 2024 kann man durch den neu hinzugekommenen Aspekt Datensicherheit und Datenschutz an erster Stelle davon ausgehen, dass auch für die kommenden Jahre der Trend für Data Governance weiter anhalten wird. Der Zusammenhang zwischen diesen drei Bereichen lässt sich kurzum wie folgt beschreiben: Die Data Governance setzt beim Datenmanagement an und spielt eine übergeordnete Rolle, indem sie den Ordnungsrahmen in Form von Standards, Richtlinien und Prozessen vorgibt, um die Anforderungen an die Datenqualität, Datensicherheit und den Datenschutz im ganzen Unternehmen und über den gesamten Datenlebenszyklus, also vom Zeitpunkt der Erstellung bis zur endgültigen Löschung, einheitlich festzulegen. Für die Umsetzung und fortlaufende Kontrolle werden Rollen, wie beispielsweise Datenadmins oder Datenverwalter festgelegt und Zuständigkeiten zugeteilt. Im Laufe der Zeit haben sich sogenannte Data Governance Frameworks etabliert, die eine strukturierte Herangehensweise für die Anwendung des Konzepts der Data Governance in einem Unternehmen darstellen und als Vorlage für die Umsetzung auf strategischer Ebene dienen. Bei einem Ansatz ohne ein solches Framework, das die Data Governance zentral regelt, besteht das Risiko, dass die Standards und Richtlinien unterschiedlich festgelegt werden und beispielsweise jede Abteilung unterschiedliche Schwerpunkte legt, die nicht mit der übergeordneten Unternehmensstrategie übereinstimmen und die Zusammenarbeit erschweren.

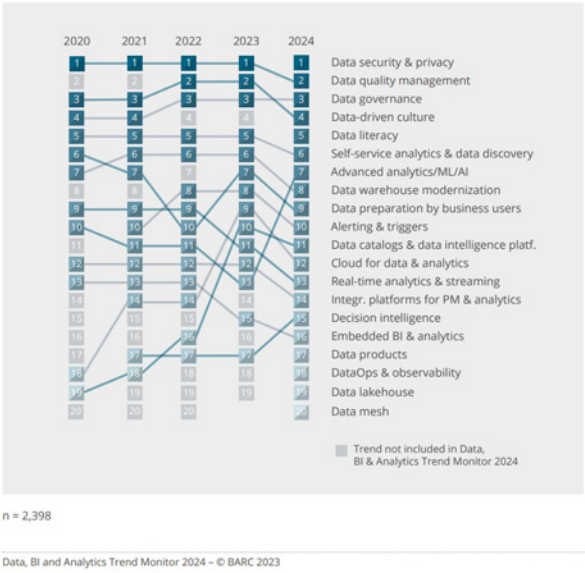

Abb. 1: Entwicklung der Trends [1]

Ziel der Arbeit

Im Rahmen dieser Bachelorarbeit soll ein Überblick darüber gegeben werden, wie bestehende Data Governance Frameworks implementiert und in ein Unternehmen integriert werden können und welche Faktoren dabei zu berücksichtigen sind, wobei die einzelnen Implementierungsschritte ausführlich besprochen werden. Der Schwerpunkt soll dabei auf den Voraussetzungen, Erfolgsfaktoren und Herausforderungen liegen, die mit dem Implementierungsprozess einhergehen. In diesem Zusammenhang soll auch ein eigenes Modell eines Frameworks entworfen und vorgestellt werden, das sich speziell an die Bedürfnisse von mittelständischen Unternehmen richtet. Die Motivation dazu ist, dass sich die meisten gängigen Frameworks nicht für kleine und mittelständische Unternehmen eignen, da sie zu komplex sind und die Ressourcen in finanzieller, personeller und infrastruktureller Hinsicht oft nicht ausreichen, was die Praxis bestätigt hat. [2] Zum Schluss der Arbeit soll mithilfe von Fallstudien ein Einblick darüber gegeben werden, wie die bewährten Frameworks in verschiedenen Branchen zum Einsatz kommen und wie sich die Ergebnisse aus der Realität mit den theoretischen Annahmen und Vorstellungen decken.

Implementierung von Data Governance Frameworks

Vor der Auswahl eines geeigneten Frameworks gibt es zwei Hauptaufgaben, die in jedem Fall durchgeführt werden sollten. Eine davon ist, über ein sogenanntes Maturity-Assessment den aktuellen Reifegrad des Unternehmens zu bestimmen, um herauszufinden, ob die Voraussetzungen in Schlüsselbereichen wie Prozessstruktur und Organisationsstruktur überhaupt gegeben sind, wofür sich mehrere verschiedene Modelle mit jeweils anderem Fokus bewährt haben. [5] Die zweite Aufgabe besteht darin, eine Strategie zu entwickeln, die definiert, welche Aspekte in welchem Umfang abgedeckt werden sollen und welche Ziele und Prioritäten sich daraus für den Governance-Ansatz ableiten lassen. [3] Somit kann der Schwerpunkt je nach Unternehmen und Branche anders liegen und beispielsweise im Gesundheitswesen hohe Anforderungen an den Aspekt Compliance stellen oder für den Finanzsektor eine hohe Datenqualität erfordern. Daraus ergibt sich dann im Anschluss ein Gesamtbild, das die Entscheidungsgrundlage für ein bestimmtes Framework legt.

Beispiel am Self-Data-Governance-Framework

Um das weitere Vorgehen zu beschreiben, wird an dieser Stelle das Self-Data-Governance-Framework (siehe Abbildung 2) ausgewählt. Dieses eignet sich besonders gut, um den Aufbau eines solchen Frameworks zu erklären, da es zum einen modular und zum anderen schrittweise aufgebaut ist. [6]
Im ersten Modul wird festgelegt, welche Datenthemen über das Framework gesteuert werden sollen. Wie bereits zuvor erwähnt, sollte ein Schwerpunkt jedoch trotzdem bereits vor der Wahl des Frameworks gesetzt werden, da einige branchenspezifisch sind oder den Schritt der Aspektwahl nicht beinhalten. Das zweite Modul stellt die nächste Steuerungsebene dar und beschäftigt sich mit der Eingrenzung bestimmter Unteraspekte der Themen wie beispielsweise die Datengenauigkeit oder -vollständigkeit im Rahmen des Themas Datenqualität. Dafür wird der aktuelle Ist-Zustand ermittelt und ein Soll-Zustand vorgegeben, wobei die Zielsetzung und Evaluation dieser Zustände dann zum Beispiel über Metriken erfolgen kann. Danach werden Rollen und Verantwortlichkeiten zugewiesen und ein Zeitrahmen festgelegt, bis wann der Soll-Zustand erreicht werden sollte. Beim dritten Modul geht es darum, welche Maßnahmen getroffen werden müssen, um den geplanten Soll-Zustand in der vorgegebenen Zeitspanne in der Praxis umzusetzen, wobei auch eine Reflexion mit verwandten Themenbereichen stattfindet. Das vierte und letzte Modul umfasst das Prüfen und Bewerten der Maßnahmenumsetzung und ermittelt einen prozentualen Wert für den Umsetzungsgrad. Sollte dieser zu gering sein oder es zu Änderungen bei der Vorgehensweise oder der Umsetzung kommen, wird der gesamte Zyklus erneut durchlaufen. Sobald die Ist-Zustände nach dem Reifegradmodell einer anderen Stufe entsprechen,

kann dies angepasst werden, wodurch ein ständiger Indikator für den momentanen Umsetzungserfolg der Data Governance Initiative gegeben ist. [4]

Abb. 2: Flussdiagramm des Self-Data-Governance-Frameworks [4]

Ausblick

Um den wachsenden Datenmengen sowie steigenden Anforderungen an Datenqualität, Datenschutz und Datensicherheit auch in Zukunft standhalten zu können, werden Unternehmen aller Größen in Data-Governance-Maßnahmen investieren müssen. Für große Unternehmen bedeutet das, dass flexible und modulare Ansätze an Bedeutung gewinnen werden und eine ständige Weiterentwicklung und Anpassung ihrer Data Governance Frameworks darüber entscheiden wird, ob sie weiterhin wettbewerbsfähig bleiben können. Insgesamt wird die Zukunft durch eine verstärkte Kombination von Modulen verschiedener Frameworks geprägt sein, um auf die Dynamik reagieren zu können. Was kleine und mittelständige Unternehmen betrifft, ist davon auszugehen, dass sie in dieser Hinsicht verstärkt auf Frameworks angewiesen sein werden, um ihre Bemühungen strukturiert zu organisieren.

Literatur und Abbildungen

[1] Business Application Research Center. Data, BI and Analytics Trend Monitor 2024. https://barc.com/de/trend-monitor-2024/, 2023.

[2] Carolyn Begg and Tom Caira. Exploring the SME Quandary: Data Governance in Practise in the Small to Medium-Sized Enterprise Sector. *The Electronic Journal Information Systems Evaluation*, 2012.

[3] Pradeep Kutty and Paul Christensen. A step-by-step guide to setting up a data governance program. https://www.ibm.com/blog/a-step-by-step-guide-to-setting-up-a-data-governance-program/, 2023.

[4] iRights. Lab. Wie lässt sich mit dem Framework arbeiten? – Flussdiagramm und Erläuterung. https://cloud-irights.oopen.de/index.php/s/eefp4Zkkw53LWWK, 2023.

[5] Rupa Mahanti. *Data Governance Success*. Springer Singapore, 2021.

[6] Henry Steinhau. Mit dem Self-Data-Governance-Framework für Schutz, Qualität und Ethik von Daten sorgen. https://emmett.io/article/self-data-governance-framework, 2023.

Prototypische Entwicklung einer Videostreaming Anwendung als Progressive Web App unter Einsatz von Astro

Christoph Merck **Jürgen Koch**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma pep.digital GmbH, Esslingen

Einleitung

Webanwendungen sind in der heutigen Zeit weit verbreitet und aus der digitalen Landschaft nicht mehr wegzudenken. Sie ermöglichen eine plattformübergreifende Bereitstellung interaktiver Softwarelösungen über das Internet im Kontext eines Browsers.
Bei der Entwicklung von Webanwendungen wird nicht mehr direkt mit den herkömmlichen Mitteln wie HTML, CSS und JavaScript gearbeitet. Stattdessen wird mittels etablierter JavaScript-Frameworks wie Angular, React oder Vue gearbeitet. Diese bieten verschiedene Vorteile, um die Entwicklung von komplexen Webanwendungen zu erleichtern. Jedoch bringen diese Frameworks große Mengen JavaScript mit sich, welche sich negativ auf die Ladegeschwindigkeit einer Webanwendung auswirken.
„There's a simple secret to building a faster website – just ship less." [7] Das Web-Framework Astro widmet sich dieser Problematik und möchte das Problem durch einen innovativen Ansatz lösen. Dabei setzt das Web-Framework auf einen statischen Ansatz und auf eine neuartige Architektur, die eine geringere Menge an JavaScript verspricht. Somit soll weniger JavaScript an einen Client gesendet und eine schnelle Ladezeit gewährleistet werden.
Die folgende Abbildung 1 verdeutlicht, welche Mengen an JavaScript durch das Web-Framework Astro eingespart werden können. Dazu wird der Median der übertragenen JavaScript-Dateigrößen mit etablierten Web-Frameworks wie React, Vue, Next, Angular und Nuxt verglichen.

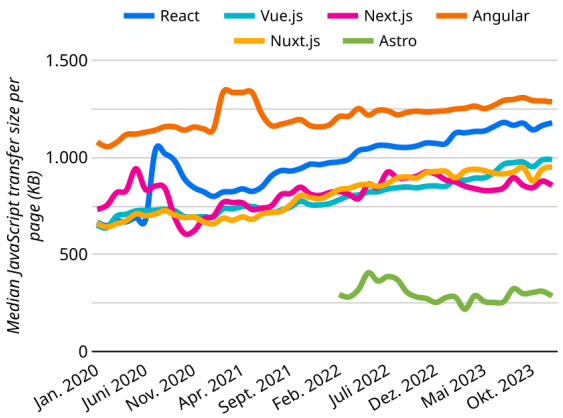

Abb. 1: Entwicklung der übertragenen JavaScript-Dateigrößen von Web-Frameworks im Zeitraum von 2020 bis Ende 2023 [4]

Zielsetzung

Im Rahmen dieser Arbeit werden die Grenzen und Möglichkeiten des Web-Frameworks Astro ermittelt. Dazu wird das inhaltsorientierte Web-Framework für die prototypische Implementierung einer Audio- und Video-Streaming-Anwendung eingesetzt. Diese soll zusätzlich als eine Progressive Web App (PWA) zur Verfügung gestellt werden, um die Benutzerinnen und Benutzer mittels Push-Nachrichten über eingehende Anrufe zu informieren.
Das Web-Framework Astro, welches für die Benutzeroberfläche als auch für die Anwendungslogik verantwortlich ist, nutzt für die Echtzeitkommunikation den WebRTC-Standard. Dieser benötigt zusätzlich einen Signaling-Server, der eine initiale Verbindung zwischen den Benutzerinnen und Benutzern aufbaut und die benötigten Informationen überträgt. Ein weiterer Server soll ein Application Programming Interface

(API) zur Verfügung stellen, über das Daten in einer Datenbank gespeichert werden können.
Durch die Implementierung dieser Anwendung und die Beantwortung der Forschungsfragen soll eine fundierte Entscheidung über den Einsatz des Web-Frameworks Astro in Bezug auf die Entwicklung einer Webanwendung getroffen werden.

Astro

Astro ist ein modernes Web-Framework, das sich durch die Generierung von schnellen und inhaltsfokussierten Webseiten auszeichnet. Dabei greift Astro weitestgehend auf serverseitiges Rendern zurück, um schnelle Webseiten mit einer verbesserten Search Engine Optimization (SEO) zu bieten. [3] Das primäre Hauptmerkmal des Web-Frameworks ist dabei die Island Architecture, in welcher das Web-Framework pioniert und diese Architektur popularisiert. [2] Astro erzeugt statische und leichtgewichtige HTML-Seiten, basierend auf Komponenten. Diese Komponenten können sowohl mit Astro als auch mit anderen Web-Frameworks wie React, Preact, Svelte oder Vue entwickelt werden. Benötigt eine Komponente clientseitiges JavaScript, bietet Astro durch die Island Architecture die Möglichkeit, das JavaScript mit den zugehörigen Abhängigkeiten individuell und nach Bedarf herunterzuladen. [6]

Progressive Web App

Der Terminus Progressive Web App (PWA) bezeichnet eine Webseite, die durch das Implementieren bestimmter Technologien und das Einhalten von Prinzipien einer nativen Anwendung ähnelt. Dabei befindet sich die Anwendung zu jedem Zeitpunkt im Kontext eines Browsers und kann somit plattformübergreifend, ohne den Drang für einen App-Store verwendet werden. PWAs zeichnen sich dabei besonders durch schnelle Ladezeiten, eine Offline-Funktionalität und die Möglichkeit, die Webanwendung nativ zu installieren, aus. Entwickelt wird eine PWA mittels HTML, CSS und JavaScript und wird durch einen Service Worker und ein Web App Manifest erweitert. Der Service Worker agiert dabei als ein Proxy, der ein- und ausgehende Anfragen abfängt. Anschließend können diese verarbeitet und zwischengespeichert werden, sodass bei einer fehlenden Netzwerkverbindung weiterhin eine Antwort geliefert werden kann. Das Web App Manifest hingegen beinhaltet Informationen bezüglich der Anwendung, wie beispielsweise den Namen, verschiedene Icons oder die Farbe des Startbildschirms. [5]

WebRTC

Web Real-Time Communication (WebRTC) ist ein offener Standard, der Echtzeitkommunikation innerhalb eines Webbrowsers ermöglicht, ohne dass zusätzliche Plugins oder Software erforderlich sind. Dabei können Audio-, Video-Streams oder Daten übertragen werden. Eingesetzt wird WebRTC von großen Konferenz-Kollaborationssystemen auf mobilen sowie Desktop-Geräten. [1]

Vorgehen

Bevor mit der Implementierung der Anwendung begonnen wurde, ist ein ausführlicher Plan erstellt worden. Im Zuge dessen wurden die funktionalen und nichtfunktionalen Anforderungen, die an die Anwendung gestellt werden, detailliert definiert. Diese dienen während der Implementierung als ein Leitfaden, um eine funktionierende und qualitativ hochwertige Anwendung zu gewährleisten. Darauf folgt das Entwerfen eines Systemarchitekturdiagramms, das die Struktur und Beziehungen zwischen den einzelnen Komponenten visualisiert. Im Anschluss wurde mittels Wireframes die grundlegende Benutzeroberfläche der Anwendung gestaltet, um eine intuitive Benutzererfahrung zu gewährleisten.
Im Anschluss an die Planungsphase konnte mit der Entwicklung der Anwendung begonnen werden. Dabei wurde, basierend auf den zuvor definierten Vorgaben, ein Astro-Projekt aufgesetzt. Um die Entwicklung von Astro-Inseln zu ermöglichen, erfolgte eine Integration einer Erweiterung für das Vue-Framework. Den Einstieg bildete die statische Start- und Hilfeseite, gefolgt von einer Unterseite, die mittels einer Vue-Island einen WebRTC-Client implementiert. Für die initiale Verbindung der Clients wurde ein Signaling-Server mit dem WebSocket-Protokoll entworfen, der die WebRTC-Daten an den jeweiligen Client überträgt. Somit ist die Anwendung bereits in der Lage, eine Audio- und Videoverbindung zwischen Clients zu erzeugen.

Ausblick

Im weiteren Verlauf der Arbeit soll die Webanwendung mit einem Service Worker und einem Web App Manifest versehen werden, um den Vorgaben der PWA-Technologie zu entsprechen. Durch die Vorteile, die mit der PWA-Technologie einhergehen, sollen Benutzerinnen und Benutzer durch Push-Benachrichtigungen über eingehende Anrufe informiert werden.
Basierend auf den Erkenntnissen, die während der Implementierung gesammelt werden, können die Forschungsfragen der Arbeit beantwortet werden. Diese spiegeln die Eignung des Web-Frameworks für die Entwicklung von interaktiven Webanwendungen wider.

Literatur und Abbildungen

[1] Niklas Blum, Serge Lachapelle, and Harald Alvestrand. WebRTC - Realtime Communication for the Open Web Platform. *ACM queue*, 19:77–93, 2021.

[2] Astro Docs. Astro Islands. https://docs.astro.build/en/concepts/islands, 2024.

[3] Astro Docs. Why Astro? https://docs.astro.build/en/concepts/why-astro, 2024.

[4] HTTP Archive HTTP Archive. Core Web Vitals Technology Report (cwvtech.report). https://lookerstudio.google.com/u/0/reporting/55bc8fad-44c2-4280-aa0b-5f3f0cd3d2be/page/M6ZPC?params=%7B%22df44%22:%22include%25EE%2580%25800%25EE%2580%2580IN %25EE%2024.

[5] Christian Liebel. *Progressive Web Apps: das Praxisbuch*. Rheinwerk Verlag, 1 edition, 2019.

[6] Addy Osmani. *Learning JavaScript design patterns: a JavaScript and react developer's guide*. O'Reilly Media, Inc., 2 edition, 2023.

[7] Fred Schott and Nate Moore. Introducing Astro: Ship Less JavaScript. https://astro.build/blog/introducing-astro, 06 2021.

Interaktive Visualisierung von Requirements mit Augmented Reality: Eine Analyse der Usability und Effektivität

Kyle Mezger **Andreas Rößler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma IT-Designers GmbH, Esslingen

Einleitung

Mit der immer weiter steigenden Komplexität moderner Systeme und Produkte wird auch das klare Formulieren von Anforderungen an das System oder Produkt immer wichtiger. Gibt es Unklarheiten und Missverständnisse in den vor und während dem Entwicklungsprozess gestellten Anforderungen, so kann das gravierende Folgen auf die Dauer und Kosten der Entwicklung haben. Daher ist heutzutage das Requirements-Engineering (Anforderungsanalyse) ein essenzieller Bestandteil der Entwicklung komplexer Systeme und Produkte. Dabei soll Requirements-Engineering laut der Autorin Chris Rupp folgende vier Hauptaufgaben erfüllen [3]:

- Wissen vermitteln
- Gute Anforderungen herleiten
- Anforderungen vermitteln
- Anforderungen verwalten

Diese Bachelorarbeit befasst sich vor allem mit der Hauptaufgabe des Vermittelns von Requirements. Diese Aufgabe ist essenziell um eine klare Kommunikation zwischen Stakeholdern, also bspw. Auftraggebern, und Entwicklern zu fördern. Gibt es sehr viele Requirements wird es vor allem für Stakeholder, die meist nicht täglich in die Entwicklung involviert sind, schwer einen Überblick über die große Anzahl von Requirements zu behalten und Probleme zu erkennen.

Zeitgleich zur wachsenden Komplexität von Systemen und der Menge derer Requirements ergeben sich durch den momentanen Fortschritt der Technik immer mehr Möglichkeiten zur Darstellung verschiedenster Daten. Augmented Reality Technologien, also Technologien, die virtuelle Daten in die echte Welt einblenden, werden durch neue Endgeräte wie das VR- und AR-Headset Meta Quest 3 immer zugänglicher und bieten neue Möglichkeiten Daten zu visualisieren und zu präsentieren. Das Headset ist in der Lage mithilfe von Inside-Out-Tracking, also Tracking mit nur Kameras und Sensoren, die am Headset selbst befestigt sind [2], seine Position im Raum zu bestimmen und so AR-Inhalte darzustellen. Daher soll untersucht werden, welche Interaktionskonzepte für die Darstellung von Requirements in AR denkbar wären und ob sie einen tatsächlichen Mehrwert bieten.

Zielsetzung

Es sollen Interaktionskonzepte zur Visualisierung von Requirements in Augmented Reality ausgearbeitet und in Prototypen beispielhaft und darstellend implementiert werden. Die Prototypen sollen dabei mit WebXR für das VR- und AR-Headset Meta Quest 3, welches in Abbildung 1 gezeigt ist, erstellt werden und über den internen Browser der Meta Quest 3 aufgerufen werden können. Die Interaktionskonzepte sollen dann auf ihre Usability, Effektivität und Umsetzbarkeit bewertet werden, wobei ein Fokus darauf liegen soll, ob die Visualisierungen durch ihre Darstellung in AR einen Mehrwert zur Darstellung auf traditionellen 2D-Displays bieten.

Abb. 1: VR- und AR-Headset Meta Quest 3 [1]

Interaktionskonzept „Explodierende Bauteile"

Das erste erstellte Interaktionskonzept zur Darstellung von Requirements bezieht sich vor allem auf physische, also anfassbare Produkte. Dabei wird von dem Produkt ein virtuelles 3D-Modell erstellt, bei dem jedes Bauteil sein eigenes Objekt ist. Dann wird eine Animation des Modells erstellt, in der sich alle Bauteile des Produkts voneinander wegbewegen. So entsteht ein Modell, in dem jedes Bauteil alleinsteht und vollständig betrachtet werden kann. In diesem explodierten Modus werden dann, wie in Abbildung 2 zu sehen ist, jedem Bauteil seine Requirements zugeordnet und als schwebende Panels angezeigt. Die Explosion kann auch rückgängig gemacht werden, wodurch die Requirement-Panels wieder verschwinden und das Produkt in seiner Gänze betrachtet werden kann.

Abb. 2: Konzept der explodierenden Bauteile dargestellt an einem Tetris-Block [1]

Um bei komplexen 3D-Modellen, wie beispielsweise Autos, die Komplexität der Animation gering zu halten, werden nicht alle Bauteile animiert. Bei Autos wären bspw. die Räder zunächst nur eine gemeinsame Komponente. Jedoch soll auch eine Detailansicht existieren, in der sich z.B. ein Reifen auswählen lässt, um dann für den Reifen eine eigene vergrößerte Explosionsanimation zu ermöglichen. So wird der Nutzer bei einer großen Übersicht nicht überflutet von Requirements, aber der Detailgrad ist trotzdem noch gegeben, wenn gewünscht.
Das Problem bei diesem Konzept ist die technische Umsetzbarkeit. Für die Implementierung muss es ein akkurates 3D-Modell des Produkts geben, welches dann animiert werden muss. Die Existenz eines 3D-Modells ist zwar bei vielen Entwicklungen schon gegeben, doch die Animationen müssen für konstante Ergebnisse ohne Clipping von Hand erstellt werden. Dadurch ist das Konzept sehr zeit- und damit kostenaufwändig zu implementieren.

Interaktionskonzept Wolken von Anforderungen

Das zweite Interaktionskonzept ist, im Gegensatz zu den „Explodierenden Bauteilen", auch bei nicht-physischen Systemen wie beispielsweise Softwaresystemen anwendbar. Hierbei sollen die Anforderungen in Clustern vorliegen, die dann jeweils als Wolke von Anforderungs-Panels dargestellt werden. So bedeutet eine räumliche Nähe zu anderen Requirements eine Relation zwischen den beiden. Zudem sollen beispielsweise einzelne Requirements ausgewählt werden können, um nach Requirements zu filtern die mit dem ausgewählten zusammenhängen. Ein Problem bei der Darstellung ist es jedoch, Requirements lesbar zu machen, die durch die Darstellung hintereinander sind.
Der Vorteil dieses Konzepts ist die technische Umsetzbarkeit und theoretische Automatisierbarkeit. Der Prozess des Umwandelns von Requirements in Clustern zu Requirement-Panels in Wolken benötigt fast keine Handarbeit und könnte sich theoretisch voll automatisieren lassen. Dadurch wäre der Kosten- und Zeitaufwand nach Erstellung eines initialen Prozesses deutlich geringer als beim Konzept der explodierenden Bauteile. Jedoch ist auch das interaktive Potential bei diesem Konzept deutlich geringer als beim ersten Konzept. Die Wolken sind nur eine Veränderung der Anordnung der Requirements und können auf den ersten Blick wenig Hilfe beim Überblick und der Navigation der Requirements bieten. Zudem ist bei diesem Konzept auch denkbar und eventuell von Vorteil, es als 2D-Anwendung zu implementieren, wodurch die meisten Vorteile der 3D-Visualisierung auch gegeben wären.

Fazit und Ausblick

Das untersuchte Konzept der Wolken an Requirements erfährt durch die Darstellung in AR wenig Verbesserung

der Usability. Daher ist es, auch wenn es theoretisch stark automatisiert werden könnte, weniger geeignet die Vermittlung von Requirements im Prozess des Requirements-Engineerings zu verbessern. Die Visualisierung von Requirements in Augmented Reality hat in Konzepten wie den Explodierende Bauteilen jedoch Potenzial, ein neues Verständnis der Daten zu schaffen und die Vermittlung der Anforderungen durch eine intuitivere Darstellung zu verbessern. Jedoch ist die Implementierung und Automatisierung für solche aufwendigen Konzepte aufgrund der wenigen Automatisierbarkeit sehr zeit- und dadurch kostenintensiv. Könnten jedoch Teile des Prozesses der Implementierung automatisiert werden, könnte der Mehrwert der Darstellung die Kosten der Implementation überwiegen.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Michael J. Gourlay and Robert T. Held. Head-Mounted-Display Tracking for Augmented and Virtual Reality. *Information Display*, 33:6–10, 2017.

[3] Chris Rupp. *Requirements-Engineering und -Management*. Carl Hanser Verlag GmbH, 7 edition, 2020.

Chatbots als Katalysator für digitale Veränderungen in der Lebensversicherung: Konzeption und prototypische Implementierung für verbesserte Kundeninteraktionen

Valmir Molliqaj **Anke Bez**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma adesso SE, Stuttgart

Einleitung

Nachdem ChatGPT im November 2022 vorgestellt wurde, ist das Thema künstliche Intelligenz und Chatbots noch stärker in den Fokus von Unternehmen, Forschungseinrichtungen und der allgemeinen Öffentlichkeit gerückt. Im Laufe der letzten Jahre stieg die Nutzung und Einführung von Chatbots erheblich, und viele Unternehmen integrierten Chatbots aufgrund ihrer vielfältigen Möglichkeiten, die Effizienz und Kundenkommunikation zu verbessern. Viele Unternehmen hoffen zudem auf einen Wettbewerbsvorteil durch die Einführung. Die Einführung eines Chatbots in einem Unternehmen ermöglicht, dass die Kundenkommunikation rund um die Uhr verfügbar ist. Das Unternehmen kann daher auf Kundenanfragen schnell reagieren und präzise Informationen liefern sowie Routineaufgaben automatisieren. Ebenso können Chatbots Fragen zu Produkten beantworten und somit als Kundenberater dienen. Die Versicherungsbranche hat erkannt, dass Unternehmen von Chatbots profitieren können. Daher haben bereits einige Versicherungsunternehmen Chatbots eingeführt.

Ziel der Arbeit

Das Ziel dieser Bachelorarbeit ist es, die Potenziale von Chatbots für die Lebensversicherungsbranche zu analysieren und ein Konzept für einen Chatbot zu entwickeln, der die Kundeninteraktionen verbessert. Die Ziele tragen dazu bei, ein tieferes Verständnis für den Einsatz von Chatbots in der Lebensversicherung zu erhalten und die digitale Transformation voranzutreiben.

Abb. 1: Mögliche Anwendungen von Chatbots [2]

Die Abbildung 1 zeigt, dass ein Chatbot allgemein in vielen Bereichen eingesetzt werden kann. Am häufigsten wurde die Vereinbarung von Terminen (44 %) ausgewählt, während der Abschluss einfacher Verträge (19 %) am seltensten ausgewählt wurde. Weiterhin sind 17 % der Meinung, dass keine der genannten Zwecke eine mögliche Anwendung darstellt.

Konzept

Die Entwicklung eines Chatbots kann sehr umfangreich sein und daher müssen bei der Planung eines Chatbots einige Punkte beachtet werden. Die Planung muss ständig angepasst und optimiert werden. Meistens wird die Entwicklung eines Chatbot unterschätzt, da die verschiedenen Aspekte und Kompetenzen eng aufeinander abgestimmt werden müssen, um eine erfolgreiche Entwicklung und Implementierung zu gewährleisten [3].

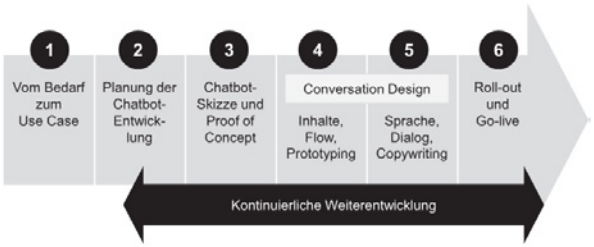

Abb. 2: Die sechs Schritte der Chatbot-Entwicklung [1]

Bei der Konzeptionierung müssen die sechs Schritte, die in der Abbildung 2 zu sehen sind, beachtet werden. Zu Beginn wird der Bedarf identifiziert, um die häufigsten Probleme und Bedürfnisse in der Kundenkommunikation zu ermitteln. Basierend auf diesen Erkenntnissen wird ein Use Case definiert, den der Chatbot abdecken soll. Der Use Case dient für die weitere Entwicklung als Grundlage. Nachdem der Use Case definiert wurde, folgt die Planung der Chatbot-Entwicklung. Dies umfasst die Auswahl der technischen Plattform, die Zuweisung von Ressourcen und einen detaillierten Entwicklungsplan. In der nächsten Phase wird eine Skizze des Chatbots erstellt, die den grundlegenden Aufbau und die wichtigsten Interaktionen darstellt. Die Skizze dient für den Proof of Concept als Ausgangspunkt, um den Nutzen und die Machbarkeit zu überprüfen.

Im Conversation Design wird der Gesprächsablauf, mit dem Ziel natürliche und hilfreiche Interaktionen zu gestalten, entwickelt. Des Weiteren wird das Design des Chatbots entwickelt, um die Interaktion mit dem Chatbot intuitiv und kundenfreundlich zu gestalten. Im letzten Schritt soll der Chatbot eine letzte Qualitätssicherung durchlaufen und das Feintuning wird vorgenommen. Sobald der Chatbot diverse Tests erfolgreich absolviert hat, wird der Chatbot zunächst nur für eine Pilotgruppe freigegeben. Die Pilotgruppe ermöglicht einen tieferen Einblick in die Anzahl der Sessions, an welchen Tagen und Uhrzeiten und wie viele Personen den Chatbot wieder benutzen oder welche Fragen der Chatbot nicht zufriedenstellend beantwortet hat. Diese Erkenntnisse werden genutzt, um Verbesserungen für die breite Masse vorzunehmen, bevor der Chatbot endgültig live gehen kann [1].

Ausblick

Das Konzept des Chatbots kann künftig zur Verbesserung der Kundenkommunikation in der Lebensversicherungsbranche beitragen und somit die Effizienz und die Effektivität der Unternehmen erhöhen. Aufbauend auf dem hier vorliegenden Konzept können zukünftige Arbeiten den Chatbot mit neuen Funktionen erweitern und daher kann das Konzept als Grundlage für eine Entwicklung eines Chatbots dienen.

Literatur und Abbildungen

[1] Beate Bruns and Cäcilie Kowald. *Praxisleitfaden Chatbots*. Springer Gabler Verlag, 2023.

[2] Thomas Donath. Verbraucher offen für die Nutzung intelligenter Chatbots im Kundenservice. https://www.nordlight-research.com/de/publikationen/presse/presse-detail/verbraucher-offen-fuer-die-nutzung-intelligenter-chatbots-im-kundenservice.html, 2024.

[3] Andreas Kohne, Philipp Kleinmanns, Christian Rolf, and Moritz Beck. *Chatbots Aufbau und Anwendungsmöglichkeiten von autonomen Sprachassistenten*. Springer Vieweg, 2020.

Simulation eines Zweitaktmotors durch KI

Selina Moritz **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma STIHL AG, Waiblingen

Einleitung

Der Zweitaktmotor zeichnet sich durch seine Robustheit und hohen Leistungsfähigkeit aus. Er wird in vielen Arbeitsbereichen eingesetzt; von Rasenmähern über Motorräder bis hin zu Kettensägen [4]. Im Rahmen dieser wissenschaftlichen Arbeit wird speziell ein Zweitaktmotor in einer Kettensäge betrachtet. Trotz seiner vergleichsweisen technischen Einfachheit ist die (Weiter-) Entwicklung eines solchen Zweitaktmotors sehr zeit- und kostenintensiv und belastet die Umwelt. Hier kommt künstliche Intelligenz (KI) ins Spiel. KI bezeichnet im Allgemeinen Maschinen, die intelligentes Verhalten zeigen können. Heutzutage ist KI allgegenwärtig und wird in den unterschiedlichsten Arbeitsbereichen eingesetzt. Sei es in der Medizin, im Finanzwesen, in der Technik, oder im Marketing und Vertrieb. Insbesondere in einem Teilbereich der KI, dem Machine Learning (ML) und dort wiederum im Deep Learning (DL) mit neuronalen Netzen wurden zahlreiche Fortschritte erzielt [3]. Dies soll nun auch im Themenfeld der (Weiter-) Entwicklung von Zweitaktmotoren in Kettensägen Anwendung finden.

Ziel

Die Gründe für die (Weiter-) Entwicklung können vielfältig sein. Sie reichen von der Schadstoffreduzierung über die Lärmreduktion bis hin zur Verbesserung der Leistung und der Wartungsfreundlichkeit. Jede (Weiter-) Entwicklung erfordert eine Vielzahl von Signalmessungen, womit das Maschinenverhalten validiert werden kann. Zwei dieser Signale sind zur Veranschaulichung vereinfacht in Abbildung 1 über dem Kurbelwinkel dargestellt.

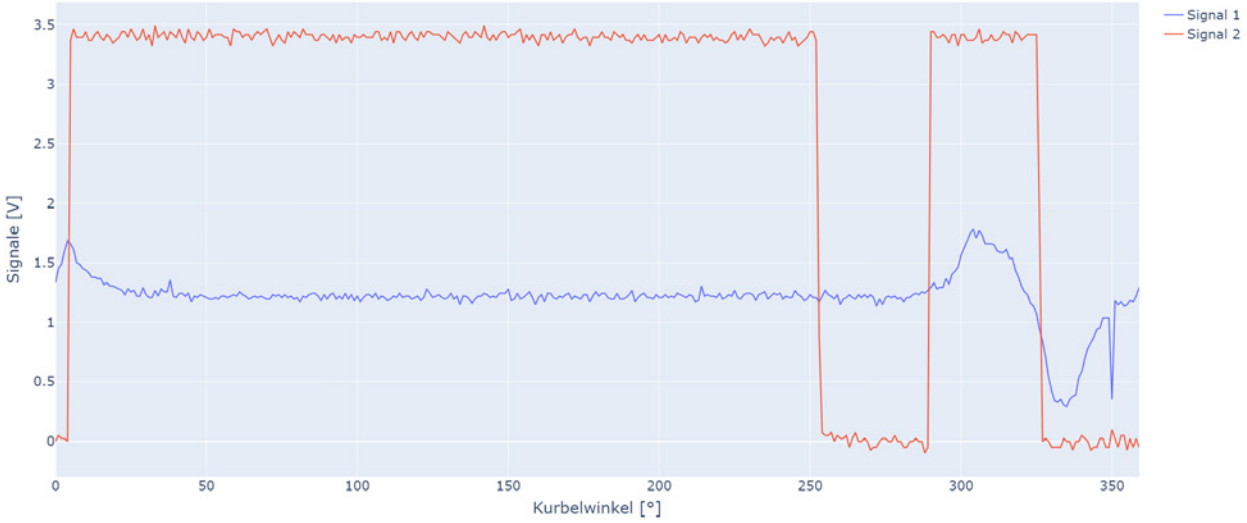

Abb. 1: Verlauf von zwei charakteristischen Signalen [2]

Diese Messungen müssen nun an realen Maschinen ausgeführt werden, um die nötigen Signale zu erhalten. Es müssen dabei verschiedene Betriebszustände sowie eine Vielzahl von Umgebungsbedingungen wie z. B. Gemischbildung oder Luftdruck berücksichtigt werden. Die genannten Nachteile des Zeit- und Kostenauf-

wandes sowie der Umweltbelastung bei der (Weiter-) Entwicklung, ergeben sich allerdings genau aus dieser Vielzahl von Messungen an realen Maschinen. Für jede Messung müssen die Maschine und die erforderlichen Messinstrumente vorbereitet werden. Die Messung selbst muss sorgfältig durchgeführt werden. Um zuverlässige Ergebnisse zu erhalten, müssen dieselben Messungen häufig wiederholt werden, was den Zeit- und Kostenaufwand nochmals erhöht. Außerdem werden bei jeder Messung Abgase ausgestoßen und es kann zu erhöhten HC (Kohlenwasserstoff)-Emissionen kommen, z. B. durch unvollständige Verbrennung oder Wandablagerungen im Motor, was zu Luftverschmutzung führt und sich dadurch negativ auf die Umwelt auswirkt [1]. Diese negativen Auswirkungen sollen verringert werden, indem die Ausgabe der Signale für die Validierung des Maschinenverhaltens durch KI-Simulationen erzeugt werden. Langfristig soll jedes Szenario der (Weiter-) Entwicklung mit jeder möglichen Kombination von Umgebungsbedingungen und Betriebszuständen simuliert werden können. Mit dieser Arbeit soll dafür der Grundstein gelegt werden. Es wird speziell die Optimierung des Laufverhaltens von Maschinen betrachtet, wofür unter anderem der Zündwinkel geändert werden muss, was mit vielen Messungen an realen Maschinen einhergeht. Messungen mit verschiedenen Zündwinkeln müssen zum Beispiel bei der Einführung eines neuen Serientriebwerkes durchgeführt werden, da dort der Zündwinkel festgelegt werden muss. Dies erfordert viele Messungen im Vorhinein, um herauszufinden, welche Einstellung des Zündwinkels ein gutes Laufverhalten der Maschine unter verschiedenen Betriebs- und Umgebungsbedingungen gewährleistet. Auch bestehende Triebwerke können durch Änderung des Zündwinkels optimiert und an gegebene Randbedingungen angepasst werden. Darüber hinaus können Algorithmen angepasst oder neu entwickelt werden, die zur Auswertung der Signale der Messungen benötigt werden. Diese Algorithmen müssen ebenfalls anhand von Messungen mit verschiedenen Zündwinkel getestet werden. Ziel dieser Arbeit ist es also, die Signalmessungen, die bei der Änderung des Zündwinkels momentan an realen Maschinen durchgeführt werden müssen, durch KI zu simulieren. Die negativen Aspekte, die sich aus Messungen an realen Maschinen ergeben, sollen minimiert werden. Hierbei wird für den Anfang von idealen Umgebungsbedingungen ausgegangen sowie mit nur einem Betriebszustand gearbeitet.

LSTM

Die Signale, die bei den Messungen erzeugt werden, sind lange Zeitreihendaten. Hierfür eignen sich besonders rekurrente neuronale Netze (RNN). Herkömmliche RNNs können zwar frühere Informationen für zukünftige Outputs nutzen, jedoch lässt diese Fähigkeit beim Lernen von längeren Abhängigkeiten nach. Deswegen wurde sich für ein spezielles RNN entschieden, dem Long Short-Term Memory (LSTM). Dies zeichnet sich dadurch aus, auch längere Abhängigkeiten in Zeitreihendaten repräsentieren zu können, indem es bestimmte Zellzustände nutzt. Alle Daten werden durchlaufen und währenddessen können den Zellzuständen Informationen hinzugefügt oder entfernt werden. Welche Informationen gespeichert werden und wie lange eine Information behalten wird, wird durch die drei sogenannten Gates entschieden. Der berechnete hidden State S(t-1) des vorherigen LSTM-Blocks sowie ein Input Xt werden in dem aktuellen LSTM-Block zusammengeführt. Zuerst wird durch das Forget Gate ft entschieden, welche Teile der Informationen von dem Zellzustand entfernt werden können. Danach entscheidet das Input Gate it, welche neuen Informationen in dem Zellzustand gespeichert werden sollen. Es folgt die Aktualisierung des alten Zellzustandes C(t-1) in den neuenZellzustand Ct. Zum Schluss legt das Output Gate ot fest, welche Teile des Zellzustandes als Output St produziert werden. Solch ein LSTM-Block ist Abbildung 2 zu entnehmen [5]

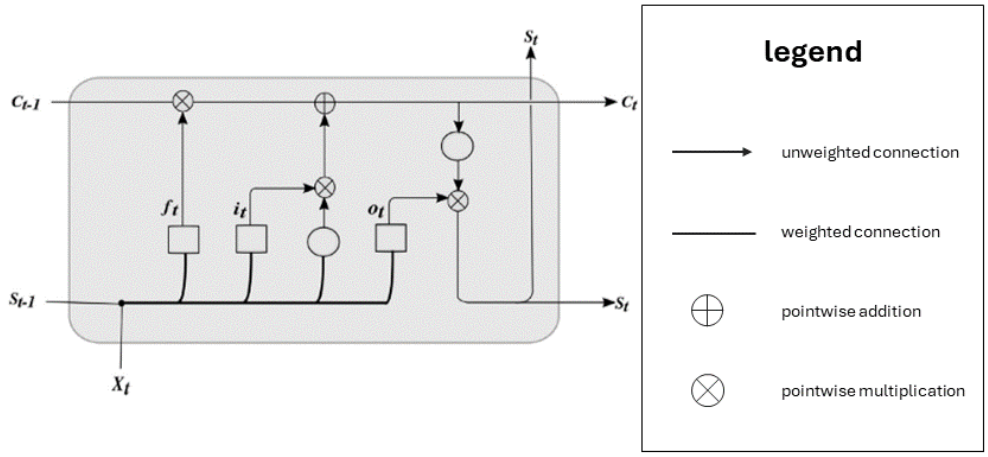

Abb. 2: LSTM Block [5]

Vorgehensweise

Um die Signale nun mit einem LSTM zu simulieren, erfolgte zuerst die Messung der Signale an einer realen Maschine mit einem bestimmten Zündwinkel. Um eine konsistente Datenbasis zu schaffen, musste ein Skript implementiert werden, um diese Signale entsprechend aufzubereiten und an den Anwendungsfall anzupassen. Basierend auf diesen Signalen wurden drei verschiedene Test-Cases mit verschiedenen Kombinationen aus Hyperparametern festgelegt. Der erste Test-Case deckt den „Worst-Case“ ab und ist für den Einstieg gedacht. Hierbei werden verschiedene LSTM-Modelle unabhängig voneinander implementiert und trainiert. Die charakteristischen Signale für jeweils einen bestimmten Zündwinkel sollen auf der Grundlage der vergangenen Signale dieses Zündwinkels vorhergesagt werden. Bei dem zweiten Test-Case werden Signalmessungen mit anderen Zündwinkeln simuliert, die aber noch zeitlich voneinander abhängen. Es folgt die Implementierung eines einzigen LSTM-Modells. Dieses wird nun mit den Signalmessungen bestimmter Zündwinkel trainiert, um die charakteristischen Signale für einen anderen, unbekannten Zündwinkel vorherzusagen. Der Ablauf des letzten Test-Cases ist identisch mit dem Ablauf von Test-Case zwei. Allerdings werden die Messungen der charakteristischen Signale mit anderen Zündwinkeln an einer realen Maschine durchgeführt. Während die Signale der simulierten Messungen zeitlich voneinander abhängen, sind die tatsächlich gemessenen Signale nicht voneinander abhängig.

Ausblick

Die Netze der Test-Cases werden im weiteren Verlauf der Arbeit implementiert, validiert und optimiert. Da sich die simulierten Messungen der Signale von Test-Case zwei aufgrund ihrer Abhängigkeit untereinander von den Messungen aus Test-Case drei unterscheiden, wird entsprechend den Ergebnissen mit dem LSTM ein anderes Modell in Betracht gezogen. Des Weiteren können in Zukunft verschiedene Betriebszustände oder Umgebungsbedingungen miteinbezogen werden.

Literatur und Abbildungen

[1] Richard Basshuysen and Fred Schäfer. *Handbuch Verbrennungsmotor: Grundlagen, Komponenten, Systeme, Perspektiven*. Springer, 2015.

[2] Eigene Darstellung.

[3] Patrick Krauss. *Künstliche Intelligenz und Hirnforschung: Neuronale Netze, Deep Learning*. Springer, 2023.

[4] Günter Merker and Rüdiger Teichmann. *Grundlagen Verbrennungsmotoren: Funktionsweise und alternative Antriebssysteme Verbrennung, Messtechnik und Simulation*. Springer, 9 edition, 2019.

[5] Alaa Sagheer and Mostafa Kotb. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing*, 323:203–213, 2019.

Evaluierung und Erweiterung eines Software Assurance Maturity Model für den sicheren Produktentwicklungslebenszyklus

Jan Mueller **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Festo SE & Co. KG, Esslingen

Motivation

In der heutigen Industrie, die zunehmend auf vernetzte Systeme und digitale Technologien angewiesen ist, hat die Produktsicherheit von industriellen Komponenten an Relevanz gewonnen. Während traditionelle Sicherheitsaspekte wie mechanische Zuverlässigkeit und physische Robustheit nach wie vor von entscheidender Bedeutung sind, hat die Produktsicherheit einen ebenso wichtigen Stellenwert eingenommen. [2]
Industrielle Komponenten, die in kritischen Infrastrukturen, Produktionsanlagen und anderen Schlüsselbereichen eingesetzt werden, sind zunehmend Zielscheiben für Cyberangriffe. Diese Angriffe können gravierende Auswirkungen haben, einschließlich Produktionsausfällen, Datenverlusten und erheblichen finanziellen Schäden. [2]
Durch den voraussichtlich 2027 in Kraft tretenden Cyber Resilience Act (CRA), wird die Sicherheit der Produkte für große Firmen verpflichtend. [5] Sichere Produktentwicklungslebenszyklen (SDLC – Secure Development LifeCycle) sind eine Möglichkeit, die Sicherheit der Produkte bereits bei der Entwicklung zu gewährleisten.

Ziel der Arbeit

Im Rahmen dieser Arbeit soll der aktuelle Reifegrad der einzelnen Bereiche erfasst werden, die Ergebnisse evaluiert und daraus Handlungsempfehlungen abgeleitet werden. Des Weiteren soll eine kontinuierliche Nutzung des Reifegradmodells gewährleistet werden.

Grundlagen

Es gibt verschiedene Arten von Reifegradmodellen, die evaluiert werden. Es gibt beschreibende, vorschreibende sowie vergleichende Modelle. [6] In der Informationstechnologie wird vor allem das Capability Maturity Model Integration (CMMI) verwendet. [6]

Building Security in Maturity Model (BSIMM) ist ein vergleichendes Capability Maturity Model der Firma Synopsys. [4] Es vergleicht 126 Aktivitäten, von über 130 teilnehmenden Organisationen, die in acht verschiedenen Industriesektoren zu Hause sind. BSIMM ist ein Reifegradmodell, das durch Dritte ausgeführt wird. Dadurch fallen Kosten für externe Gutachter an, sowie ein zusätzlicher Zeitaufwand entsteht.
OWASP SAMM (Software Assurance Maturity Model) ist ein Reifegradmodell von der Non-Profit Organisation Open Worldwide Application Security Project. OWASP SAMM ist ein vorschreibendes Capability Maturity Model, das mithilfe eines selbst ausgefüllten Fragebogens den Reifegrad bestimmt. Es bietet zudem die Möglichkeit sich anonym mit anderen Nutzern zu vergleichen. Der Fragebogen besteht aus 90 Fragen, die in 5 Bereiche unterteilt sind (siehe Abbildung 1). Die Fragen behandeln Security-Themen aus den Bereichen Verwaltung, Entwurf, Implementierung, Überprüfung und Betrieb. [7]

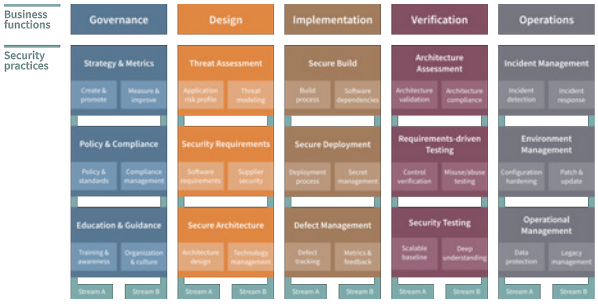

Abb. 1: Aufbau OWASP SAMM [7]

OWASP SAMM bietet zusätzliche Features, wie das Abbilden auf andere Reifegradmodelle oder eine Online-Version des Fragebogens. Durch externe Firmen wird bereits eine Software-as-a-Service Lösung angeboten. OWASP SAMM erfüllt die gestellten Anforderungen, ist Open-Source und bietet Anleitungen für eine

erfolgreiche Durchführung. [7]

Evaluierung

Nach dem vollständigen Ausfüllen des Fragebogens bekommen Nutzer die aktuelle Reifegradbewertung für ihren jeweiligen Bereich. Mithilfe eines Netzdiagramms (siehe Abbildung 2) lassen sich so schnell Schwachstellen aufzeigen.

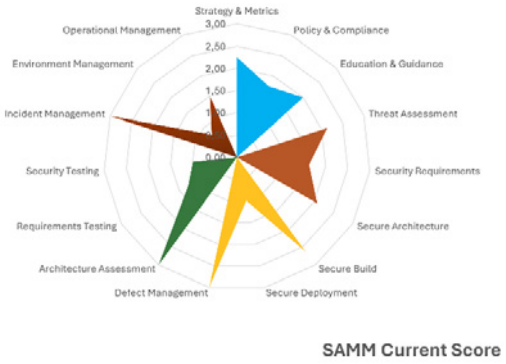

Abb. 2: Netzdiagramm SAMM Current Score [1]

Der Reifegrad der einzelnen Aktivitäten kann zwischen null und drei variieren. Null stellt das nicht erfüllen der Aktivität dar, drei die Automatisierung, sowie eine ständige Verbesserung der Aktivität (siehe Abbildung 3). Der Reifegrad ist die zentrale Bewertung des Prozesses. Nicht jede Aktivität muss ein Reifegrad von drei erreichen. Bereits bei einem Reifegrad von zwei ist der Prozess standardisiert. [7]

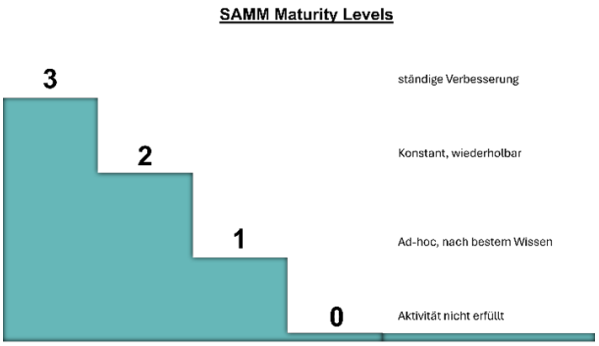

Abb. 3: SAMM Maturity Levels [1]

Die Evaluierung der Ergebnisse erfordert eine umfassendere Betrachtung als lediglich den Reifegrad. Zu diesem Zweck werden Interviews durchgeführt, um potenzielle Ursachen zu identifizieren, die durch das Reifegradmodell nicht berücksichtigt werden. Fehlt es bei einer Aktivität zum Erfüllen an einer Methodenbeschreibung zur Umsetzung, gibt es zu wenig Kapazität oder ist die Aktivität nicht erfüllbar. Dies kann vorkommen, da der Fragebogen generisch ist und z.B. nicht auf eingebettete Systeme ausgelegt. [7] Durch einen Vergleich der einzelnen Ergebnisse, lassen sich Handlungsempfehlungen für die einzelnen Bereiche, aber auch das Unternehmen ableiten.

Ausblick

Nach abgeschlossener Evaluation der Ergebnisse und Feedback der Bereiche wird eine Anpassung des Fragebogens vorgenommen. Es werden weiterführende Tools analysiert, wie das DevSecOps-Maturity-Model (DSOMM), das tiefer in die Softwareentwicklung blickt. [3]

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Arne Dreißigacker, Bennet von Skarczinski, and Gina Rosa Wollinger. *Cyberangriffe gegen Unternehmen in Deutschland*. Kriminologisches Forschungsinstitut Niedersachsen E.V., 2020.

[3] OWASP DSOMM. OWASP Devsecops Maturity Model. https://owasp.org/www-project-devsecops-maturity-model/, 2024.

[4] Synopsys Inc. Building Security in Maturity Model. https://www.synopsys.com/software-integrity/software-security-services/bsimm-maturity-model.html, 2024.

[5] Europäische Komission. VERORDNUNG DES EUROPÄISCHEN PARLAMENTS UND DES RATES über horizontale Cybersicherheitsanforderungen für Produkte mit digitalen Elementen und zur Änderung der Verordnung (EU). https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52022PC0454, 2022.

[6] Jens Poeppelbuss and Maximilian Roeglinger. What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management. *19th European Conference on Information Systems*, 2011.

[7] OWASP SAMM. Software Assurance Maturity Model. https://owaspsamm.org, 2024.

Interkulturelles Projektmanagement in Scrum – Ein strategischer Ansatz zur Verbesserung deutsch-indischer Teamzusammenarbeit

Johannes Niebel **Anke Bez**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Mercedes-Benz AG, Stuttgart

Einleitung

In der heutigen globalisierten Wirtschaft sind Kooperationen über Ländergrenzen hinweg eine Notwendigkeit geworden, insbesondere im IT-Sektor. Indien, als einer der führenden IT-Exporteure weltweit, ist ein zentraler Partner für viele deutsche Unternehmen. [4] Diese Zusammenarbeit findet oftmals über die Nutzung agiler Methoden, wie Scrum, statt und ermöglicht damit eine flexible und effektive Projektumsetzung. Jedoch führen kulturelle Unterschiede häufig zu Herausforderungen, die die Effizienz und Effektivität dieser Projekte beeinträchtigen können.

Ziele der Arbeit

Diese Arbeit zielt darauf ab, die interkulturellen Herausforderungen zwischen deutschen und indischen Scrum-Teams zu identifizieren und Maßnahmen für das Scrum-Framework und darüber hinaus vorzuschlagen, um die Zusammenarbeit zu verbessern. Ein besonderes Augenmerk liegt auf der Analyse und dem Management dieser kulturellen Herausforderungen durch praktische Anpassungen der Scrum-Methodiken.

Vorgehensweise

Die Untersuchung basiert auf einer Kombination aus Literaturrecherche und empirischen Studien, einschließlich Interviews mit Projektmanagern und Teammitgliedern aus deutsch-indischen IT-Projekten. Die erhobenen Daten werden genutzt, um kulturelle Herausforderungen zu identifizieren und effektive Strategien für deren Management zu entwickeln.

Interkulturelle Herausforderungen nach Hofstede's Kulturdimensionen-Modell

Abb. 1: Identifizierte Herausforderungen geclustert in Hofstede's Kulturdimensionen-Modell [1]

Abbildung 1 zeigt die identifizierten kulturellen Herausforderungen geclustert nach 4 Dimensionen des Hofstede's Kulturdimensionen-Modell's, welche folgend beschrieben werden:

1. **Machtdistanz:** Beschreibt das Maß, in dem weniger mächtige Mitglieder einer Gesellschaft eine ungleiche Machtverteilung akzeptieren. In Gesellschaften mit hoher Machtdistanz werden Hierarchien stark betont und Respekt vor Autoritäten ist tief verankert, was in Arbeitsumgebungen zu einer zentralisierten Entscheidungsfindung führen kann. [3]

2. **Individualismus vs. Kollektivismus:** Individualismus kennzeichnet Gesellschaften mit losen zwischenmenschlichen Verbindungen, wo jeder für sich selbst sorgt. Im Gegensatz dazu betont Kollektivismus enge, lebenslange Gruppenbindungen und bedingungslose Loyalität, was die Teamarbeit und Gruppenkohäsion fördert. [3]

3. **Unsicherheitsvermeidung:** Misst, wie stark eine Gesellschaft durch Regeln, Planung und Ordnung Unsicherheit und Unbekanntes zu minimieren versucht. Gesellschaften mit starker Unsicherheitsvermeidung bevorzugen klare Richtlinien und strukturierte Prozesse. [3]
4. **Genuss:** Beschreibt, inwieweit eine Gesellschaft die freie Befriedigung grundlegender und natürlicher menschlicher Bedürfnisse nach Lebensfreude und Spaß zulässt. Kulturen mit hoher Ausprägung in dieser Dimension fördern eine Atmosphäre, die Lebensqualität und persönliches Wohlbefinden betont. [3]

Abbildung 1 verdeutlicht die kulturellen Unterschiede zwischen Deutschland und Indien und zeigt, wie kulturelle Werte Arbeitsstile und zwischenmenschliche Beziehungen beeinflussen. Die Grafik betont spezifische Arbeitsaspekte beider Länder in diesen Bereichen. Durch gezielte Anpassungen im Scrum-Framework können Missverständnisse reduziert und die Zusammenarbeit effizienter gestaltet werden. Zu den identifizierten Herausforderungen gehören:

1. **Hierarchiedenken**: In Indien sind Hierarchien stärker ausgeprägt, was die Kommunikation beschränkt und mit den Prinzipien von Scrum kollidiert.
2. **Fehlerkultur und Feedback**: Deutsche Teammitglieder sind oft direkter im Feedback, was zu Verzögerungen, Missverständnissen und Konflikten führen kann.
3. **Kulturelle Sensibilität und Teambuilding**: Indien legt großen Wert auf Teambuilding und kulturelle Trainings, die in Deutschland weniger verbreitet sind.
4. **Arbeitshaltung und Priorisierung**: Unterschiedliche Auffassungen von Quantität und Qualität zwischen den Teammitgliedern führen zu Konflikten, da unterschiedliche kulturelle Prägungen die Erwartungen an die Arbeitsleistung beeinflussen.
5. **Standards und Qualität**: Unterschiede in den Qualitätsstandards und Erwartungen zwischen den Teammitgliedern führen zu erhöhtem Abstimmungsbedarf und Arbeitsaufwand.
6. **Sprachbarrieren**: Die Verwendung von Englisch als Fremdsprache erschwert die Kommunikation über komplexe IT- und fachspezifische-Themen erheblich.

Maßnahmen zur Verbesserung der Zusammenarbeit in Scrum-Teams

In der Bachelorarbeit werden sowohl Maßnahmen innerhalb des Scrum-Frameworks als auch begleitende Maßnahmen diskutiert, um die interkulturelle Zusammenarbeit in deutsch-indischen IT-Projekten zu optimieren. Diese Maßnahmen zielen darauf ab, die kulturellen Herausforderungen abzubauen und die Effizienz sowie Effektivität der Teams zu steigern:

- **Scrum-Retrospektive und Trompenaars Modell der interkulturellen Kompetenz**: Diese Maßnahme integriert die Prinzipien der interkulturellen Kompetenz in die Scrum-Retrospektiven, um die kulturelle Sensibilität und Zusammenarbeit in den Teams zu verbessern. [2]
- **Sprintbewertungsmethodik**: Die Anwendung einer spezifischen Methodik zur Bewertung und Reflexion der Sprints unterstützt die kontinuierliche Verbesserung und fördert eine Kultur des offenen Feedbacks, die für interkulturelle Teams besonders wichtig ist.
- **Aufbau persönlicher Beziehungen durch Reisen**: Diese Maßnahme fördert das gegenseitige Verständnis und das Vertrauen durch persönliche Interaktionen, was wiederum die Zusammenarbeit und Kommunikation in verteilten Teams stärkt.
- **Kulturspezifisches Reporting**: Ein an die kulturellen Bedürfnisse angepasstes Reporting-System verbessert die Kommunikation und Koordination zwischen den Teams in Deutschland und Indien.
- **Überwinden von Sprachbarrieren durch Tool-Einsatz**: Der Einsatz spezieller Kommunikationstools hilft, Sprachbarrieren zu überwinden und fördert eine effizientere und effektivere Kommunikation innerhalb der Teams.

Ausblick

Die vorgeschlagenen Maßnahmen stellen strategische Ansätze dar, um die identifizierten Herausforderungen der interkulturellen Zusammenarbeit in deutsch-indischen Teams zu adressieren. Der nächste Schritt besteht darin, diese Maßnahmen zu validieren und systematisch zu implementieren, um ihre Wirksamkeit in der Praxis zu überprüfen.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Charles Hampden-Turner and Fons Trompenaars. *Riding the Waves of Culture: Understanding Diversity in Global Business (English Edition)*. John Murray Press, 2020.

[3] Geert Hofstede, Gert Jan Hofstede, and Michael Minkov. *Lokales Denken, globales Handeln: Interkulturelle Zusammenarbeit und globales Management*. Beck-Wirtschaftsberater, 2017.

[4] World Integrated Trade Solution World Bank. ICT service exports By Country, in BoP, current US$. https://wits.worldbank.org/CountryProfile/en/country/by-country/startyear/LTST/endyear/LTST/indicator/BX-GSR-CCIS-CD, 2024.

Mapping of IEC 62443 Product Requirements to Security Gateway Functionalities

Jelle Pichl **Tobias Heer**

Department of Computer Science and Engineering, Esslingen University

Work carried out at Department of Computer Science and Engineering, Esslingen

Motivation and Problem

Ensuring the secure development of products is relevant to industries worldwide. Many countries have regulatory frameworks in place to ensure that products meet the required security standards. In industrial settings, the IEC 62443 series is important for the security of industrial automation and control systems (IACS), defining various requirements in areas such as development processes and product specifications.
The continuous development of IACS presents both challenges and opportunities. The advent of digitalisation has led to an increase in the number of devices connected to the network. In particular, the integration of legacy devices into the network presents companies with significant challenges. Legacy devices, such as programmable logic controllers, often lack sufficient security features. The replacement of these legacy devices or the implementation of improved security features is often hindered by cost and certification restrictions. Due to insufficient security features, the integration of these legacy devices into networks requires a careful approach to minimise potential security risks and ensure the security and integrity of the overall system.
One approach to securely operate a legacy device within an existing network infrastructure is to use a security gateway. The gateway provides functionalities such as authentication, VPN encryption, and firewall to retrofit security for communication towards the network. Such a gateway is suitable for deployment in restricted sectors such as manufacturing and process industries, where compliance with relevant technical and organisational standards is crucial.
Compliance with product requirements from security standards is crucial for the development of devices used in industrial domains. Therefore, we want to ensure that security standards relevant to the development of a gateway are complied with. To guarantee this, we want to determine product requirements for the gateway functionalities authentication, VPN, and firewall. We derive the product requirements from IEC 62443 and related standards and use them to build a mapping table. Using this mapping table, we check the suitability of software solutions used within the development of the gateway.

Design

In this work, we proceed in three steps to achieve the above mentioned objectives. Each step is visualized in Figure 1 and discussed in detail below.

1. Literature research of relevant security standards that addresses requirements for network or embedded devices.
2. Identification of relevant product requirements for the functionalities authentication, VPN, and firewall.
3. Mapping of requirements to existing software solutions.

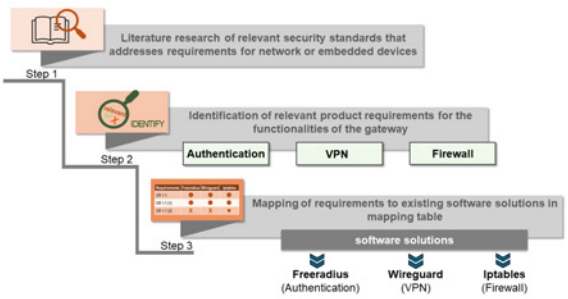

Fig. 1: Illustration of the approach to the work [6]

In the first step, we conduct a literature research. The initial focus of this research is the IEC 62443 series of standards, with emphasis on two important sub-standards: IEC 62443-3-3 and IEC 62443-4-2. IEC 62443-3-3 specifies requirements for systems to

meet certain security levels. These security levels range from level 1 to level 4 and are used to define and evaluate the security level of components. IEC 62443-4-2 specifies requirements and recommendations for security measures throughout the life cycle of components. The standard also provides categories of requirements. For example, network access control for different types of devices such as embedded devices or network devices. We aim to identify further relevant standards that contain product requirements for network or embedded devices, including standards in other related sectors, such as the medical and building automation domains. For example, ISO 27033 provides requirements for network security, including guidelines for VPNs or gateways. By incorporating findings from various standards that relate to industry-specific regulations, we ensure that the product requirements identified represent aspects and challenges in the industrial, medical, and building automation sectors. This ensures that requirements relevant to product development are covered. The difficulty of the first step is to ensure a comprehensive and complete collection of all relevant standards. All standards and their associated sectors that are taken into consideration are shown in Figure 2.

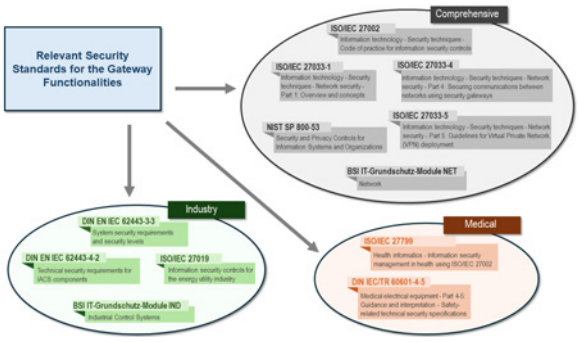

Fig. 2: Standards overview [6]

In the second step, relevant product requirements for the gateway functionalities firewall, VPN encryption, and authentication are gathered from the identified and considered standards. We analyze each standard and extract requirements that are applicable to the functionalities. A further challenge could be comparing and reconciling the different standards with each other. It is also important to extract the weighting in the form of verbal forms used by the product requirements specified in the standard. These weightings can, for example, correspond to a "must", "may" or "should". In this way, we consider the importance and relevance of the requirements for the gateway functionalities. We then map all requirements with security level and weighting in a table.

In the third step, we map selected software solutions to these derived requirements. To effectively measure fulfilment by different software solutions, we create a comprehensive requirements table. This table lists all identified requirements, with each requirement assigned a weighting previously taken from the standard. By systematically organizing the requirements in the table, we can assess each software solution. For example, FreeRadius could be used as an implementation for 802.1X authentication, iptables for firewalls, and WireGuard for VPN. This evaluation allows a thorough examination of the extent to which available software solutions meet these security requirements and the importance of each requirement. Ultimately, this categorisation and evaluation process helps to assess the suitability of the software solution and identify gaps.

Evaluation

We set requirements for completeness of derived requirements to ensure that relevant standards are fully considered. We check whether standards relevant to our domains have been included in the assessment. By conducting the above-described literature research, we ensure comprehensive consideration of relevant standards.

We evaluate the extent to which the derived requirements for the functionalities of the gateway are covered by software solutions. Ideally, a software solution would meet all product requirements given by the requirements table. Given the complexity of real-world use, this is unlikely. However, the mapping process is an effective way to determine whether a software solution meets requirements and to identify requirements that cannot be met. It is important to consider how these requirements can be met by compensatory means and the consequences of not meeting them. This mapping process will help us to identify software solutions that come closest to meeting requirements, even if none of them fulfil all the requirements sufficiently.

The weighting defined in the standards allows us to go beyond a purely quantitative assessment of the requirements met. It allows for a qualitative assessment to determine which software solution not only fulfills the most requirements but also the most critical ones. This is crucial, as a software solution that meets numerous less important requirements but does not meet essential requirements may be less suited than a software solution that prioritizes these requirements. If a requirement cannot be met, it is important to determine whether this is acceptable. This comprehensive assessment will enable us to make a decision that optimises the match between our requirements and the selected software solution.

Related Work

The following related works address standards that have been considered in the development of security gateway solutions and describe mapping processes from standards to requirements. We present these work and discuss how this work differs.

T. Frauenschläger and J. Mottok [2] present a security gateway architecture for use in critical infrastructure, considering the IEC 61850 and IEC 62351 standards during development. IEC 61850 standardizes the communication protocols used in substations and power systems, while IEC 62351 specifies security aspects for associated communications protocols. E. Andrade et al. [1] presents the development of a security gateway consisting of a security library and a bridging device. The security library implements security mechanisms according to the standards IEC 61850 and IEC 62351. The bridging device provides security features such as authentication and confidentiality for legacy devices. In this work, we focus on requirements given by standards in other sectors, like the medical sector. We further focus on VPN and firewall functionalities and the requirements given for these functionalities.

S. Fritzsch et al. [3] [4] describe how relevant requirements can be identified from IEC 62443-4-2. They present two use cases. The first use case [3] identifies relevant requirements for industrial firewalls, while the second use case [4] identifies requirements for security gateways. The gateway enables secure data exchange in the International Data Spaces System. The extracted requirements are then assigned to the corresponding security levels. In contrast to these works, we investigate further functionalities, such as authentication and VPN. In addition, we consider not only the requirements of IEC 62443-4-2 but also the extent to which software solutions meet these requirements.

K. Niemann and J. Puls [5] describe a modelling process for a secure development lifecycle for embedded devices. They describe how security requirements can be derived from IEC 62443-4-1 and then implemented and tested. We want to transfer this to the standards that we consider and extract the requirements for the functionalities.

Result

For the development of products in critical infrastructures, it is necessary to meet the relevant product requirements defined by regulatory standards. In this work, we consider relevant product requirements for a security gateway with authentication, VPN encryption, and firewall functionalities for use with legacy devices in critical infrastructures. We aim to provide a mapping table of requirements derived from standards in different domains to determine the suitability of software solutions for use with different functionalities of the gateway. The table can be used for products in other related sectors in the future.

References and figures

[1] Eduardo Andrade, Jorge Granjal, João P Vilela, and Carlos Arantes. A Security Gateway for power distribution systems in open networks. https://www.sciencedirect.com/science/article/pii/S0167404821003163, 12 2021.

[2] Tobias Frauenschläger and Jürgen Mottok. Security-Gateway for SCADA-Systems in Critical Infrastructures. https://ieeexplore.ieee.org/abstract/document/9920070, 09 2022.

[3] Sebastian Fritsch, Andreas Fufl, Josef Güntner, Lutz Jänicke, Stefan Menge, Holger Mühlbauer, Siegfried Müller, Dan-Mihai Negrea, Steffen Pfendtner, and Marc Schmierer. TeleTrusT - IEC 62443-4-2 Use Case Industrial Firewall. https://www.teletrust.de/fileadmin/user_upload/2021-TeleTrusT-IEC_62443-4-2_Use_Case_Industrial_Firewall.pdf, 2021.

[4] Sebastian Fritsch, Andreas Fufl, Josef Güntner, Lutz Jänicke, Stefan Menge, Holger Mühlbauer, Siegfried Müller, Dan-Mihai Negrea, Steffen Pfendtner, and Marc Schmierer. TeleTrusT - IEC 62443-4-2 Use Case Security Gateway. https://www.teletrust.de/fileadmin/user_upload/2021-TeleTrusT-IEC_62443-4-2_Use_Case_Security_Gateway.pdf, 2021.

[5] Karl-Heinz Niemann and Jan-Niklas Puls. Musterprozess für einen sicheren Produkt-Entwicklungslebenszyklus nach IEC 62443-4-1. https://serwiss.bib.hs-hannover.de/frontdoor/index/index/docId/2935, 09 2023.

[6] Own representation.

Performante Darstellung von Diagrammen bei großen Datenmengen im Webkontext

Johannes Pungier **Dieter Morgenroth**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Digital Building Industries AG, Böblingen/Esslingen

Abstract

Zur Zeit entwickelt die Digital Building Industries AG an einem Softwareprodukt namens „Berta & Rudi“. Mit dieser Anwendung soll die Energie- und die Energiekostenrechnung von Gebäuden und Liegenschaften mithilfe von Künstlicher Intelligenz vereinfacht und beschleunigt werden. Für Berta & Rudi sucht die Digital Building Industries AG außerdem nach einer Softwarelösung für performante Diagramme in Webanwendungen. Die Diagramme werden voraussichtlich verwendet für die Darstellung von Lastgängen von Energiequellen in Gebäuden, oder für Wetterdaten, über den zeitlichen Verlauf von einem Jahr. Je nach stündlicher oder sogar minütlicher Auflösung können diese Lastgang-Datensätze groß werden.
Es wurde bereits eine kleine Beispielintegration mit der Javascript Bibliothek „Kendo UI“ [12] im Vue.js Framework der Version 2 entwickelt. Diese entspricht performancetechnisch nicht den Erwartungen.
Das Laden der Diagramme mit den großen Datensätzen im Browser kann lange dauern. Außerdem soll ein Nutzer im Diagramm filtern und zoomen können. Diese Aktionen sind ebenfalls rechnungsaufwändig und werden mit der derzeitigen Kendo UI Implementierung mit wenig Bildern pro Sekunde ausgeführt, was die Nutzererfahrung unangenehm macht. Außerdem stürzt der Browser bei zu großen Datensätzen ab.

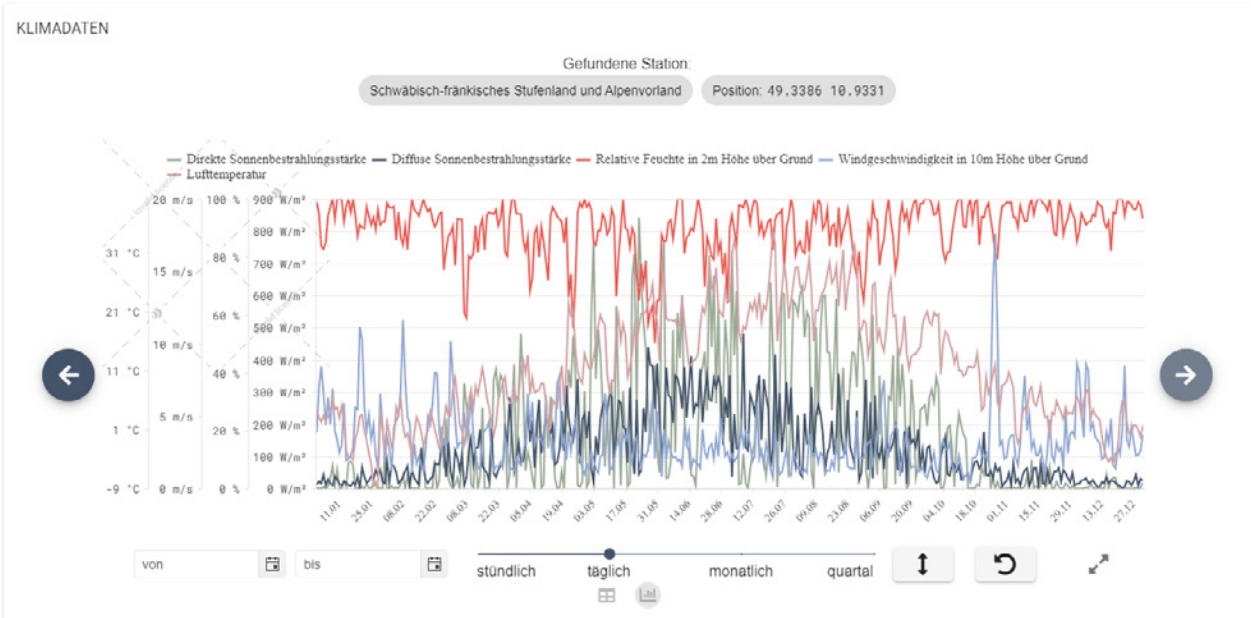

Abb. 1: Wetterdaten Diagramm implementiert mit Kendo UI in Berta & Rudi [4]

Zielsetzung

Ziel der Thesis ist es eine bestmögliche Lösung für die Diagramme im Webkontext zu finden. Im Optimalfall lässt sich eine Javascript Bibliothek für Diagramme mit großen Datenmengen, welche ebenfalls gut dokumentiert ist, finden. Mit dieser soll dann Kendo UI ersetzt werden. Außerdem sollen nebenbei weitere Praktiken,

welche sich positiv auf die Leistung einer Webseite auswirken, gefunden werden. Anschließend soll eine fertige Webanwendung programmiert werden, welche dann in andere Webanwendungen integriert werden kann.

Stand der Forschung

Zur Zeit wird üblicherweise Javascript in Kombination mit Hypertext Markup Language (HTML) und Cascading Style Sheets (CSS) verwendet um Webanwendungen und Webseiten zu entwickeln. [13] Für Javascript existieren eine Vielzahl an Javascript-Bibliotheken für Diagramme. Die meisten Diagramm-Bibliotheken verwenden als Technologiegrundlage Scalable Vector Graphics (SVG) [3], HTML Canvas [9] oder Web Graphics Library (WebGL) [6]. Zum Beispiel verwendet „Kendo UI“ primär SVG, dies ist leider die am wenigsten performante Technologie unter den drei Gelisteten. Dies liegt unter anderem daran, dass SVG oft auf das Document Object Model (DOM) zugreift, was viel Leistung kostet. Die Stärken von SVG sind die Auflösungsunabhängigkeit und die Kompatibilität mit CSS. Das heißt, für das Problem der Thesis müssen andere Technologien/Bibliotheken hinzugezogen werden.
Canvas und WebGL hingegen sind eher dafür bekannt eine bessere Leistung zu haben. Vor allem WebGL, da diese Technologie unter anderem für das Zeichnen von vielen Knotenpunkten und Polygonen im 2D- und 3D-Bereich gemacht ist.

Aufbau der Methodik

Zuerst wird per Internetrecherche nach populären Javascript-Bibliotheken für Diagramme gesucht. Zudem werden bei manchen Suchanfragen Schlüsselwörter eingefügt um nach populären und leistungsstarken Bibliotheken für Diagramme zu suchen. Dann werden vielversprechend klingende Kandidaten vorerst in eine Tabelle eingetragen. Dort wird verglichen auf welcher Technologie die Kandidaten basieren, mit welchen Funktionen und Eigenschaften die Bibliotheken beworben werden und zuletzt ob sie kostenlos und/oder open-source sind, falls sie nicht kostenlos sind: wie hoch die Gebühren sind. Besonders Bibliotheken, bei welchen beworben wird, dass sie leistungsstark sind, gelten als interessant.
Um die Performanceaspekte unserer Technologie Kandidaten zu testen, sollen Beispiel-Anwendungen geschrieben werden. Diese Beispiel-Anwendungen sollen Diagramme und die Interaktion (Filtern und Zoomen etc.) mit diesen beinhalten. Die Implementierung dieser sollte nur Grundfunktionen und Funktionen, welche voraussichtlich großen Einfluss auf Performance haben, beinhalten um nicht zu viel Zeit für die Entwicklung der Beispiel-Anwendungen aufzuwenden. Hier kann auch möglicherweise festgestellt werden, ob sich gewisse Technologien implementierungstechnisch eher eignen. Zum Beispiel, Modularfähigkeit oder Integrierbarkeit in existierende Anwendungen könnten eine Rolle spielen. Eine gut geschriebene und umfangreiche Dokumentation mit zahlreichen Beispielen ist ebenfalls ein großer Pluspunkt.
Es soll ein sinnhaftes Metrikkonzept konzipiert werden um die Beispiel-Anwendungen, und um auch die finale Implementierung, zu messen und zu bewerten. Fokus ist die Performance in Bezug zur Benutzererfahrung. Verschiedene Aspekte der Performance werden gemessen und gegenübergestellt. Die Beispielprogramme werden dann mit den Metriken gemessen.
Sobald die Testergebnisse feststehen wird abgewogen, welche Bibliothek für unseren Anwendungsfall am geeignetsten ist. Anschließend wird mit der ausgewählten Bibliothek weiterentwickelt.
Es folgt die Implementierung der gewünschten Diagramm-Webanwendung mit der ausgewählten Bibliothek. Diese Webanwendung wird direkt in Berta & Rudi integriert, und es wird untersucht ob es eine Auswirkung auf die Performance von Berta & Rudi gibt.

Implementierung von Beispielprogrammen

Die Diagramme in den Beispielprogrammen sollen folgende Funktionen enthalten:

- Zoomfunktion
- Verschiebungs- bzw. Panfunktion
- Annotationen welche den ausgewählten Datenpunktwert anzeigen

Jedes Diagramm soll 4 Datenreihen, mit je 100, 1.000, 10.000, 100.000 und 1.000.000 für die jeweilige Messung, darstellen. Gemessen werden:

- Ladezeit
- Random Acces Memory (RAM)-Auslastung ohne Nutzerinteraktion
- RAM-Auslastung beim Zoomen und Verschieben
- Video Random Acces Memory (VRAM)-Auslastung ohne Nutzerinteraktion
- VRAM-Auslastung beim Zoomen und Verschieben
- Central Processing Unit (CPU)-Last ohne Nutzerinteraktion
- CPU-Last beim Zoomen und Verschieben

- Graphics Processing Unit (GPU)-Last ohne Nutzerinteraktion
- GPU-Last beim Zoomen und Verschieben
- Bilder pro Sekunde des Diagramms beim Zoomen und Verschieben
- Bilder pro Sekunde des Diagramms beim Navigieren der Maus über Datenpunkte

Vor allem die Ladezeit und die Bilder pro Sekunde beim Interagieren mit den Diagrammen ist relevant, da dies wichtig für eine angenehme Nutzererfahrung ist. Speicherauslastungen und CPU- und GPU-Last werden gemessen um zu sehen ob die jeweils betroffene Hardware überlastet wird, und um vielleicht Flaschenhälse in der Leistung aufzuspüren.

Gewählte Kandidaten

Für SVG-basierte Bibliotheken wurden Kendo UI und „Chartist.js“ [1] gewählt. Bei Canvas-basierten Bibliotheken ist die Entscheidung auf „Chart.js“ [2] und „uplot“ [11] gefallen. Und für WebGL wurden „Timechart“ [7], „Scichart“ [10] und „LightningChart“ [8] als Kandidaten ausgewählt.

Messungen und Vergleich

Wenn es um die Anzahl der Bilder pro Sekunde geht, schneiden die Bibliotheken, welche auf WebGL basieren, besser ab als die die auf Canvas oder SVG basieren siehe Abb. 2. Im Beispielprogramm für Kendo UI konnte bei einer Datenreihengröße von 1.000.000 kein Wert gemessen werden, da bei so vielen Datenpunkten der Browser nicht genügend Speicher hat. Im Beispielprogramm für Chartist.js konnten für das Zoomen und Pannen keine Werte gemessen werden, da die Zoomfunktion dort grundlegend anders funktioniert, und die Panfunktion nicht vorhanden ist.

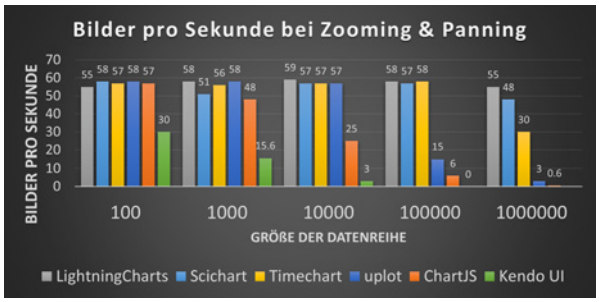

Abb. 2: Bilder pro Sekunde bei Zooming & Panning mit verschiedenen Diagramm Bibliotheken [4]

Die WebGL Diagramme behalten mehr Bilder pro Sekunde bei größeren Datenreihen bei. Bei der Ladezeit schneiden die Canvas Bibliotheken teilweise besser als die WebGL Bibliotheken ab, vor allem uplot hat die kürzeste Ladezeit bei allen Datenreihengrößen. Die Ladezeit für die SVG Bibliotheken steigt bei jeder Messung fast exponentiell an, siehe Abb. 3.

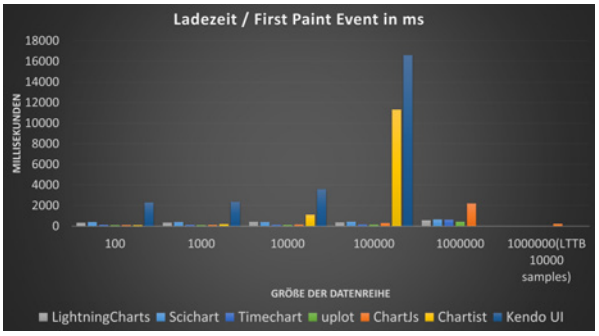

Abb. 3: Ladezeiten der verschiedenen Bibliotheken [4]

Resultierend kann man sagen, dass die WebGL Bibliotheken die geeigneteren Kandidaten sind, da die längere Ladezeit nur marginal größer als die der Canvas Bibliotheken ist, und weil die Leistung bei der Nutzerinteraktion die der anderen Bibliotheken deutlich übersteigt.

Implementierung in Berta & Rudi

Für die Implementierung wurde Scichart gewählt, da sie gut bei den Metriken abgeschnitten hat und weil die Dokumentation der Bibliothek gut und umfangreich ist. Eine Alternative dazu ist LightningChart, da auch hier gute Metrikergebnisse und gute Dokumentation präsent sind. Timechart hat leider einen signifikanten Fall von Bildern pro Sekunde bei einer Datenreihengröße von 1.000.000 Datenpunkten, zudem ist die Dokumentation nicht so umfangreich wie bei den anderen beiden WebGL Kandidaten.

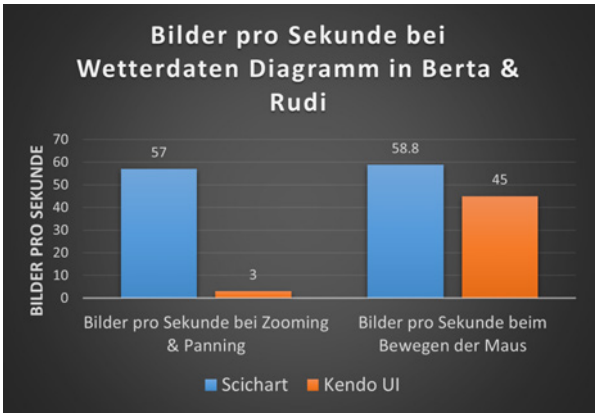

Abb. 4: Bilder pro Sekunde des Wetterdaten Diagramms basierend auf Scichart und auf Kendo UI [4]

Um den Unterschied zwischen der alten Implementierung mit Kendo UI und der Neuen mit Scichart zu vergleichen, wurde ein neue Messung der Bilder pro Sekunde der Diagramme innerhalb der Berta & Rudi Anwendung durchgeführt siehe Abb. 4. Hier sieht man eine deutliche Verbesserung der Menge an Bilder pro Sekunde, und somit auch eine Verbesserung der Nutzererfahrung.

Fazit

Das Ziel eine neue Lösung für die Diagramme von Berta & Rudi zu finden wurde erreicht. Es wurden zwei Bibliotheken gefunden welche auf den Anwendungsfall passen, eine gute Dokumentation haben und angenehm zum Entwickeln sind. WebGL als grundlegende Technologie für Javascript Bibliotheken hat einen großen Leistungsvorteil.
Der Grund warum dennoch Diagramm Bibliotheken basierend auf SVG oder Canvas entwickelt werden liegt daran, dass WebGL Programme mit Graphics Library Shader Language (GLSL) geschrieben werden. [5] Dies ist eine Shader Code Programmiersprache und setzt, für den Umgang, Wissen über Computergrafik und Matrix-Mathematik voraus. Zudem muss man in der Sprache viele Aktionen und Operationen, welche in der Canvas und SVG Application Programming Interface (API) bereits abstrahiert sind, selbst programmieren. Deshalb ist WebGL Programmierung viel aufwendiger und für viele Entwickler weniger lohnenswert. Glücklicherweise braucht man als Nutzer der WebGL Javascript Bibliotheken kein Wissen über WebGL Programmierung.

Literatur und Abbildungen

[1] ChartistJS ChartistJS. CHARTIST - SIMPLE RESPONSIVE CHARTS. https://gionkunz.github.io/chartist-js/, 2024.

[2] ChartJS ChartJS. Simple yet flexible JavaScript charting library for the modern web. https://www.chartjs.org/, 2024.

[3] World Wide Web Consortium. About SVG. https://www.w3.org/Graphics/SVG/About, 2004.

[4] Eigene Darstellung.

[5] WebGL Fundamentals. WebGL Shaders and GLSL. https://webglfundamentals.org/webgl/lessons/webgl-shaders-and-glsl.html, 2015.

[6] Khronos Group. LOW-LEVEL 3D GRAPHICS API BASED ON OPENGL ES. https://www.khronos.org/webgl/, 2024.

[7] Weiwen Hu. Time Chart. https://github.com/huww98/TimeChart, 2022.

[8] LightningChart LightningChart. LightningChart. https://lightningchart.com/js-charts/, 2024.

[9] Mozilla Developer Network. Canvas API. https://developer.mozilla.org/en-US/docs/Web/API/Canvas_API, 2024.

[10] SciChart SciChart. SciChart. https://www.scichart.com/, 2024.

[11] Leon Sorokin. uPlot. https://github.com/leeoniya/uPlot, 2023.

[12] Kendo UI. Comprehensive Vue UI Component Library. https://www.telerik.com/kendo-vue-ui, 2024.

[13] Lionel Sujay Vailshery. Most used programming languages among developers worldwide as of 2023. https://www.statista.com/statistics/793628/worldwide-developer-survey-most-used-languages/, 2024.

Analyse und Handlungsempfehlung für die Einführung eines zentralen und toolgestützen Testdatenmanagement bei der Hugo Boss AG

Julian Raach **Thomas Rodach**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Hugo Boss AG, Metzingen

Einleitung

Viele Unternehmen verfügen bereits über ein effektives Testmanagement, dabei wird das Testdatenmanagement oft vernachlässigt. Das Testdatenmanagement spielt eine wichtige Rolle bei der Qualitätssicherung von Software. Es trägt dazu bei, potenzielle Risiken zu minimieren und die Effizienz der Testprozesse zu steigern. Gemäß einer Umfrage von Softwaretestern verwenden mehr als 30 % ihre Testzeit damit, sich mit fehlerhaften Testdaten auseinanderzusetzen. Dies bedeutet, dass mindestens ein ganzer Tag pro Woche damit verschwendet wird, Testdaten bereitzustellen. Ein effizientes Testdatenmanagement entlastet die Tester, indem es sicherstellt, dass die benötigten Testdaten leicht verfügbar sind und den Testanforderungen entsprechen. Dies beschleunigt den gesamten Testprozess und verbessert die Qualität der Testergebnisse. [4]

Zielsetzung

Im Rahmen dieser Arbeit erfolgt eine Analyse des Testdatenmanagements bei der Hugo Boss AG. Bei der Betrachtung des Testdatenmanagementprozesses werden Anforderungen und Funktionen abgeleitet, die als Grundlage dienen. Anschließend werden verschiedene Tools auf dem Markt betrachtet, um festzustellen, welche Tools den Anforderungen entsprechen und am besten in den Testdatenmanagementprozess der Hugo Boss AG integriert werden können.

Testdatenmanagement

Das Testdatenmanagement umfasst alle methodischen, konzeptionellen, organisatorischen und technischen Maßnahmen und Verfahren zur effektiven Nutzung der Ressource Testdaten. Dabei liegt das Ziel darin, die Testdaten unter Berücksichtigung von Datenschutzanforderungen optimal in die Testprozesse einzubinden. Es soll sichergestellt werden, dass die Testdaten effizient genutzt werden können und damit die Anforderungen an die Testfälle erfüllen. Der Testdatenmanagementprozess umfasst die Verwaltung von Testdaten, einschließlich der Definition von Standards, Aufgaben, Rollen und Verantwortlichkeiten. Im Testdatenentwicklungsprozess werden die Anforderungen an die Testdaten festgelegt und die Erstellung sowie Wartung der Testdaten durchgeführt, um sicherzustellen, dass sie in der erwarteten Qualität, Menge und Form zur Verfügung stehen. [3]

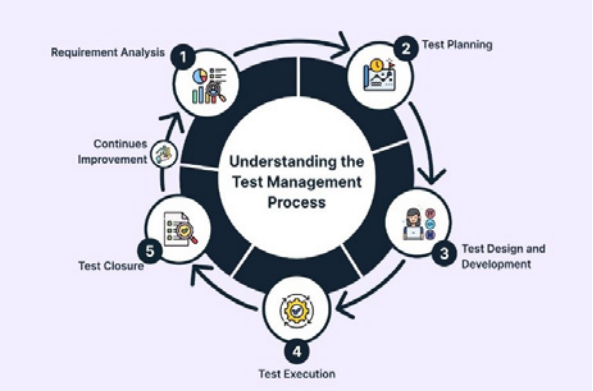

Abb. 1: Testmanagement Prozess [2]

Das Testdatenmanagement wird frühzeitig in das Testmanagement integriert, indem es als wesentlicher Bestandteil des gesamten Testprozesses betrachtet wird. Die Anforderungen an die Testdaten werden in der Requirement Analysis Phase definiert und im Test Planning (siehe Abb. 1) in den Zeitplan des Testmanagements eingearbeitet. Dadurch wird sichergestellt, dass die erforderlichen Testdaten rechtzeitig verfügbar sind und den Qualitätsanforderungen entsprechen. Nach der Verwendung der Testdaten werden in der Test Closure Phase Berichte erstellt. Dabei werden die Qualität der Testdaten bewertet und mögliche Optimierungspotenziale identifiziert. Zudem

wird geprüft, ob und wie die Testdaten in zukünftigen Testzyklen effizient wiederverwendet werden können, um Ressourcen zu schonen und den Testprozess zu verbessern. Das Testdatenmanagement muss sich verschiedenen Herausforderungen stellen:

- Datenqualität
- Datenschutz
- Datenintegrität
- Verfügbarkeit der Testdaten
- Testdaten Automationen
- Wiederverwendbarkeit von Testdaten

Diese Herausforderungen muss das Testdatenmanagement berücksichtigen, um eine effektive Durchführung des Testprozesses sicherzustellen und qualitativ hochwertige Testergebnisse zu erzielen. Ein Testdatenmanagement kann diese Herausforderungen durch verschiedene Elemente bewältigen. In Abbildung 2 sind die unterschiedlichen Elemente eines Testdatenmanagements dargestellt.

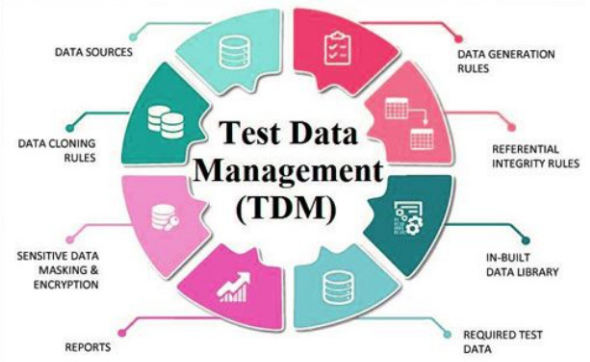

Abb. 2: Testdatamanagement Framework [1]

Tools im Testdatenmanagement

Die Tools im Bereich des Testdatenmanagements bieten eine Vielzahl von Funktionen. Sie können dazu beitragen, die Datenqualität und -verfügbarkeit durch den Einsatz von Datenbanken zu gewährleisten. Darüber hinaus verfügen die meisten Tools über integrierte Funktionen zur Anonymisierung von Daten, um den Datenschutz zu gewährleisten. Einige Tools bieten auch die automatische Erstellung von synthetischen Daten, um den Testprozess zu unterstützen. Dadurch können realistische Testdaten generiert werden, ohne auf echte Daten zurückgreifen zu müssen. Einige Tools bieten auch die Möglichkeit, Produktivsysteme zu klonen und deren Daten zu replizieren, um ein realistisches Testumfeld zu schaffen. Dadurch können Tests in einer Umgebung durchgeführt werden, die der Produktionsumgebung ähnlich ist, was die Genauigkeit und Aussagekraft der Tests erhöht.

Ausblick

Im weiteren Verlauf der Bachelorarbeit werden die Anforderungen an die Tools für das Testdatenmanagement erarbeitet. Diese Tools sollen dazu dienen, die aktuellen Herausforderungen im Testdatenmanagement zu bewältigen oder zu vereinfachen. Anschließend werden die verschiedenen auf dem Markt verfügbaren Tools evaluiert, um eine fundierte Handlungsempfehlung aussprechen zu können.

Literatur und Abbildungen

[1] G. Appsierra. Testdatamanagement Framework. https://www.appsierra.com/blog/test-data-management-strategies, 11 2023.

[2] S. Browser. Testmanagement Prozess. https://www.browserstack.com/test-management/what-is-test-management, 2023.

[3] Klaus Franz, Tanja Tremmel, and Eckehard Kruse. *Basiswissen Testdatenmanagement*. d.punkt.verlag, 2018.

[4] R. Priyanka. New Research on continous quality assurance. https://www.capgemini.com/us-en/new-research-on-continuous-quality-assurance/, 03 2020.

KI in der Logistik: ML Methoden und deren Integration in BI-Systeme

Aida Reci **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Die dynamische Entwicklung der Künstlichen Intelligenz (KI) und des Machine Learning (ML) hat in den vergangenen Jahren zahlreiche Branchen tiefgreifend transformiert. Besonders die Logistikbranche profitiert erheblich von diesen technologischen Fortschritten. Die Integration von ML-Methoden in Business Intelligence (BI)-Systeme ermöglicht es, logistische Prozesse zu optimieren, Kosten zu senken und die Effizienz zu steigern. Trotz des klaren Fortschritts in der Digitalisierung bleibt die vollständige Ausschöpfung der Potenziale, die KI und ML bieten, in vielen Unternehmen unzureichend. Diese unzureichende Integration führt oft zu einer suboptimalen Nutzung der verfügbaren Datenressourcen und verhindert, dass Unternehmen die maximal möglichen Vorteile erzielen.
In der Thesis wird die Bedeutung der KI in der Logistik eingehend untersucht, wobei der Fokus auf den verschiedenen ML-Methoden und deren Integration in BI-Systeme liegt. Die Arbeit soll detailliert die theoretischen Grundlagen der KI und des ML erörtern, die spezifischen Anwendungen und Vorteile in der Logistik analysieren sowie die Herausforderungen und Limitationen diskutieren, die mit der Integration dieser Technologien einhergehen. Darüber hinaus werden Fallstudien erfolgreicher Implementierungen vorgestellt, um praxisnahe Einblicke zu gewähren und die strategische Bedeutung der Integration dieser Technologien zu unterstreichen. Ziel der Arbeit ist es, fundierte Handlungsempfehlungen zu entwickeln, die Führungskräften und IT-Experten helfen, die Vorteile von KI und ML voll auszuschöpfen und somit zur nachhaltigen Steigerung der Unternehmensperformance beizutragen. Wie können also Unternehmen sicherstellen, dass sie das volle Potenzial dieser revolutionären Technologien nutzen?

Bedeutung von KI für die Logistik

Im Bereich der KI haben logistische Anwendungen von Anfang an großes Interesse geweckt. Bereits in den frühen Tagen der Robotik und Bildverarbeitung konzentrierte sich die Forschung auf die Unterstützung und Automatisierung von Produktionsprozessen. Ein klassisches Beispiel für Planungsanwendungen ist die Produktionsplanung und -steuerung, die Teil des Informationssystems ist. ML und Data Mining spielten schon früh eine entscheidende Rolle bei der Analyse umfangreicher Datensätze, um Wissen zu extrahieren und strategische Entscheidungsprozesse in der Logistik zu unterstützen. Die Koordination, sowohl über Systemebenen hinweg als auch innerhalb einer Ebene, wurde bereits in den Anfängen der verteilten Künstlichen Intelligenz erforscht. Diese Beispiele verdeutlichen, wie die Logistik wertvolle Fragestellungen und Anregungen für die Entwicklung von KI-Theorien und -Methoden bietet und wie die KI ihrerseits signifikante Beiträge zur Bewältigung logistischer Herausforderungen leistet [4].

Die Rolle des ML in der Logistik

In der heutigen schnelllebigen Welt ist ML zum Eckpfeiler der Logistikbranche geworden, indem es entscheidend zur Modernisierung und Effizienzsteigerung beiträgt. ML analysiert große Datenmengen, um präzisere Prognosen zu erstellen und die Ressourcenverwaltung zu optimieren, was zu einer erheblichen Kostenreduktion führt. Diese Technologie senkt nicht nur die Frachtkosten und minimiert Lieferausfälle, sondern verbessert auch die Genauigkeit der Nachfrageprognosen durch das Lernen aus historischen Daten. Ein weiterer Vorteil von ML in der Logistik ist die Optimierung der Routenplanung. Durch Anpassungen der Lieferwege in Echtzeit minimiert das System Verzögerungen und reduziert gleichzeitig den Kraftstoffverbrauch. Zudem unterstützt ML die Mustererkennung in Lieferdaten, was die Liefergenauigkeit verbessert und das Risiko von Fehllieferungen verringert. Durch die Automatisierung von Planungs- und Dispositionsprozessen steigert ML die Gesamteffizienz und ermöglicht es Unternehmen, potenzielle Störungen

in der Lieferkette frühzeitig zu erkennen und proaktiv darauf zu reagieren.
Machine Learning hat sich als unverzichtbares Instrument etabliert, das die logistische Leistungsfähigkeit signifikant verbessert und Unternehmen einen entscheidenden Wettbewerbsvorteil verschafft [2].

Überführung zu spezifischen Lernmethoden

Um diese umfassenden Vorteile zu realisieren, nutzt Machine Learning spezifische Methoden, die sich in zwei Hauptkategorien unterteilen lassen: überwachtes und unüberwachtes Lernen. Diese Methoden ermöglichen es ML-Systemen, aus den vorhandenen Datenmengen zu lernen und intelligente, datengetriebene Entscheidungen zu treffen.

Überwachtes Lernen

Beim überwachten Lernen, wie in Abbildung 1 illustriert, werden Modelle mit vordefinierten Antworten trainiert, die es ermöglichen, Muster und Zusammenhänge in den Daten zu erkennen und vorherzusagen.

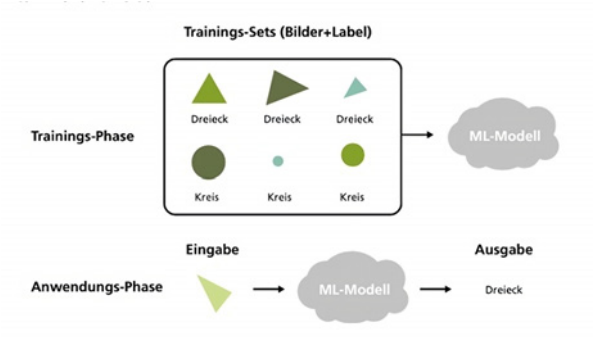

Abb. 1: ML mit überwachtem Lernen [3]

Unüberwachtes Lernen

Unüberwachtes Lernen, dargestellt in Abbildung 2, erforscht ohne spezifische Vorgaben Daten, um verborgene Strukturen oder Muster selbstständig zu entdecken. Diese Methode fördert die Entwicklung autonomer Systeme.

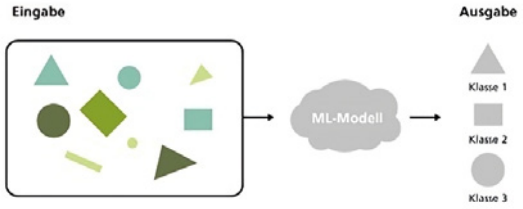

Abb. 2: ML mit unüberwachtem Lernen [3]

Diese fortschrittlichen ML-Methoden eröffnen neue Perspektiven für die Technologiewelt, indem sie intelligentere und autonomere Systeme ermöglichen [3].

Integration von ML in BI-Systeme

Die Integration von ML in BI hat das Potenzial, die Effizienz und Effektivität der Datenanalyse erheblich zu steigern. Fortschritte in der Rechenleistung haben es ermöglicht, dass auch kostengünstige Plattformen nun komplexe ML-Modelle unterstützen, was den Zugang zu diesen Technologien erweitert hat. In neuerer Zeit haben viele BI-Anbieter begonnen, ML-Technologien zu implementieren, um verborgene Datenmuster besser aufzudecken. Es wird erwartet, dass BI-Plattformen, die ML-Technologien intuitiv nutzen, bald zur Norm werden. Benutzer dieser Plattformen werden dadurch in der Lage sein, tiefergehende Einsichten zu gewinnen und schneller auf Basis fundierter Informationen zu handeln.
Die Automatisierung durch ML erleichtert die Identifikation kritischer, aber versteckter Geschäftsinformationen. Gleichzeitig können Datenanalysten sich anspruchsvolleren Aufgaben widmen, da repetitive Analysetätigkeiten reduziert werden. Es stellen sich jedoch Herausforderungen, wie die Notwendigkeit, Nutzern zu helfen, die durch ML erzeugten Daten und Visualisierungen zu verstehen. Vertrauen in die Ergebnisse ist essenziell, was Transparenz in den verwendeten Algorithmen und eine effektive Filterung relevanter Informationen erfordert.
Zusammenfassend lässt sich sagen, dass die Kombination von BI und ML signifikante Vorteile für Unternehmen bietet, indem sie tiefere Einblicke und eine verbesserte Entscheidungsfindung ermöglicht [1].

Zusammenfassung und Ausblick

Die Integration von KI und ML hat die Logistikbranche revolutioniert, indem sie Prozesse optimiert und Entscheidungsfindungen auf Basis großer Datenmengen verbessert. Die Verfügbarkeit leistungsstarker und kostengünstiger Technologien hat dazu geführt, dass ML-Modelle zunehmend in BI-Systeme integriert

werden, was die Erwartung schafft, dass solche fortschrittlichen Analysewerkzeuge bald zum Standard werden. Dieser Trend wird von einem gesteigerten Vertrauen in die Technologie und einer verbesserten Benutzerfreundlichkeit unterstützt.
Mit weiteren Fortschritten in KI und ML wird die Effizienz der Datenverarbeitung weiter steigen, was präzisere Vorhersagen und eine verbesserte Automatisierung ermöglicht. Die Herausforderung bleibt, die Transparenz und das Vertrauen in die Systeme zu sichern und ihre Bedienung benutzerfreundlich zu gestalten. Die Kombination von BI und ML wird weiterhin eine zentrale Rolle in der Entwicklung der Logistik spielen, indem sie Unternehmen hilft, effizienter und wettbewerbsfähiger zu agieren. Unternehmen müssen bereit sein, in diese Technologie zu investieren und die notwendigen Anpassungen vorzunehmen, um das volle Potenzial der KI- und ML-Integration auszuschöpfen.

Literatur und Abbildungen

[1] Nico Litzel. So beeinflusst Machine Learning die Business Intelligence. https://www.bigdata-insider.de/so-beeinflusst-machine-learning-die-business-intelligence-a-676786/, 02 2018.

[2] Daniel Mahnken. Digitale Logistik: Machine Learning in der Logistik – Beispiele & Hauptanwendungsfälle. https://www.saloodo.com/de/blog/digitale-logistik-machine-learning/, 2022.

[3] Anike Murrenhoff, Martin Friedrich, and Dr. Markus Witthaut. Künstliche Intelligenz in der Logistik. In *Künstliche Intelligenz in der Logistik*, pages 5–7. Frauenhofer Institut, 2021.

[4] Ingo J. Timm and Andreas D. Lattner. Künstliche Intelligenz in der Logistik. In *Künstliche Intelligenz in der Logistik*, pages 99–103. Springer Verlag, 2010.

Digital Analytics - Den Erfolg digitaler Produkte messbar machen und datenbasiert entscheiden

Fabian Sanzi **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma BMW AG, München

Einleitung

Schon früh mit Beginn des Internets und dem Aufkommen von Webseiten war es im Interesse der Betreiber das Nutzungsverhalten zu analysieren. Bald darauf rückten auch wirtschaftliche Fragestellungen hinsichtlich des Erfolgs der Webseiten in den Vordergrund. Dieser sollte durch das Auswerten von Daten gezielt gesteigert werden. Zu Beginn setzte man dabei auf die Analyse der Logfiles der Server. Durch das Aufkommen von Tools wurde dann auch das clientseitige Erfassen von Daten möglich. Die Einführung von Javascript erweiterte diese Möglichkeiten erheblich, sodass fast jede Interaktion des Nutzers erfasst werden kann. Nebenher wurden aus anfangs reinen Informationswebseiten immer komplexere Web Applikationen. Um diese neuen Dimensionen abzubilden, spricht man daher heutzutage meist von Digital Analytics. [3]

Problemstellung und Zielsetzung

Der Praxispartner BMW stellt den Händlern Webapplikationen zur Verfügung. Bisher gibt es dort im Frontend der Applikationen keine Implementierung einer Digital Analytics Lösung. Ziel der Arbeit ist es daher, ein entsprechendes Analytics-Konzept zu entwerfen. Dafür müssen die nötigen Messgrößen identifiziert und definiert werden. Schlussendlich sollen die erfassten Daten zur Verbesserung des Produktes beitragen. Daher wird in der Arbeit auch auf die datengetriebene Entscheidungsfindung eingegangen, um zu erfassen, inwiefern Daten bisher im Entwicklungsprozess genutzt werden und welche Aspekte im Sinne einer datengetriebenen Kultur bedeutsam sind.

Vorgehen

In der Literatur sind dabei Vorgehensmodelle für eine Implementierung von Web Analytics zu finden. Eine Ausarbeitung schlägt das Durchlaufen von fünf Phasen vor (siehe Abbildung 1). Im Rahmen der Anforderungsermittlung sollte grundlegend die Frage geklärt werden, welche Inhalte die Applikation bereitstellt und welche Ziele aus Sicht des Unternehmens und des Benutzers verfolgt werden. In der nächsten Phase sollte man Metriken definieren und entsprechend festlegen welche Tools die Anforderungen erfüllen. Ein wichtiger Aspekt ist dabei auch der Datenschutz, der seit Inkrafttreten der Datenschutzgrundverordnung 2018 neue Hürden erhält. Nach Betrachtung und Auswahl der Tools erfolgt anschließend im Rahmen der dritten Phase die Implementierung und somit die Datensammlung. Im Anschluss daran, sollte während der vierten Phase in den gesammelten Daten Muster erkannt werden und insgesamt nützliche Einblicke gewonnen werden. Aufbauend darauf sollen auf Grundlage dieser Einblicke Verbesserungen erarbeitet und umgesetzt werden. Die letzte Phase sieht dann die Evaluation der implementierten Änderungen vor. Die letzten beiden Phasen erfolgen dabei iterativ. [4]

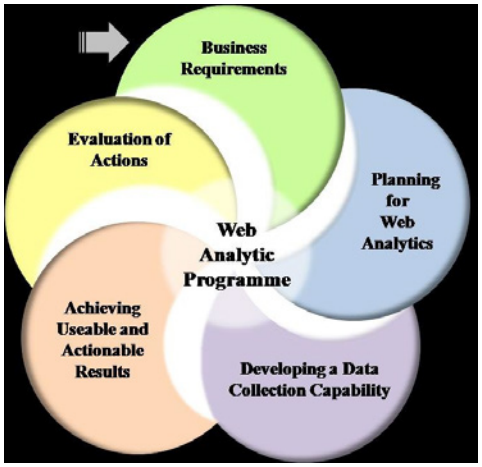

Abb. 1: Vorgehensmodell Web Analytics [4]

Aktueller Stand Digital Analytics

Web Analytics hat sich über die Jahre hinweg stark entwickelt. Insbesondere in den Jahren zwischen

2012 und 2018 wurden die von Google und anderen Unternehmen angebotenen Lösungen weiterentwickelt. Neben Webseiten konnten nun auch Nutzer in Native Apps getrackt werden. Insgesamt konnten von nun an sämtliche digitale Touchpoints entlang der User Journey besser erfasst werden, weshalb man auch vermehrt von Event Tracking spricht. Die Daten, die meist in der Cloud der jeweiligen Anbieter gespeichert sind, konnten nun einfacher in Datenbanken anderer Systeme geladen werden. In den letzten Jahren zeigte sich vermehrt auch die Nutzung dieser digitalen Daten für Personalisierung oder auch predictive Modeling. Fast 50% geben an, dass Digital Analytics hilft, die Ziele des digitalen Geschäfts zu erreichen. [1]

Datengetriebene Entscheidungsfindung

Entscheidungen im Allgemeinen ist die Lösung, zu der eine entscheidende Person beziehungsweise ein entscheidender Personenkreis nach Betrachtung und Einschätzung der gegebenen Situation und Abwägung sämtlicher Handlungsoptionen und Alternativen findet. Diese ziehen stets Konsequenzen nach sich. Innerhalb eines Unternehmens lassen sich Entscheidungen hinsichtlich ihrer Häufigkeit und der Komplexität in strategische, taktische und operative Entscheidungen einteilen (siehe Abbildung 2). [5]

Abb. 2: Klassifizierung Entscheidungen im Unternehmen [5]

DevOps

DevOps wird beim Praxispartner grundlegend in der Softwareentwicklung eingesetzt. DevOps ist ein agiler Ansatz, der die kollaborative Zusammenarbeit sonst getrennt arbeitender Teams fordert. Klassisch bezieht es sich auf die Zusammenarbeit von Entwicklern und für den Betrieb verantwortliche IT-Kräfte. Wichtig sind dabei einige kulturelle Aspekte. Besonders wichtig sind eine gesteigerte Transparenz und enge Kommunikation. Ebenso sind beide Teams gleichermaßen für den Erfolg verantwortlich. Im Hinblick auf Entscheidungsprozesse ist die Autonomie hervorzuheben. DevOps zielt auch auf schnellere Entwicklung ab. Die Teams müssen daher gemeinsam ohne langwierige Prozesse Entscheidungen treffen. Zur Kultur von DevOps zählt auch frühes Feedback, um das Produkt schnell und kontinuierlich zu verbessern. Dafür notwendig ist ein Monitoring. [2]

Ausblick

Die Bachelorarbeit gibt Einblick in die Erarbeitung eines Konzeptes für Digital Analytics einer Verkaufslösung. Dabei wird sich an den im Framework vorgestellten Phasen und den Aspekten orientiert. Im Hinblick auf den Informationsbedarf wird untersucht, welche Metriken aussagekräftige Einblicke ermöglichen. Das Tool, welches eingesetzt werden soll, ermöglicht Event Tracking. Somit wird auch eine Eventstruktur erarbeitet, die die wichtigen Schritte der durch die Applikation unterstützten Prozesse abbildet. Im Hinblick auf die letzten beiden Phasen wird geprüft, inwiefern vom Phasenmodell abgewichen werden muss, da DevOps eine eigene Logik im Hinblick auf das Monitoring und die Einarbeitung von Feedback vorsieht.

Literatur und Abbildungen

[1] Sara D'Onofrio, Andreas Meier, et al. *Big Data Analytics Grundlagen, Fallbeispiele und Nutzungspotenziale*. Springer Viehweg, 2021.

[2] Tom Hall. What is DevOps Culture? https://www.atlassian.com/devops/what-is-devops/devops-culture, 2024.

[3] Marco Hassler. *Digital und Web Analytics*. mitp Verlags GmbH & Co. KG, 2019.

[4] Verena Hausmann, Petra Schubert, and Susan. Williams. Developing a Framework for Web Analytics. In *eDependability: Reliable and Trustworthy eStructures, eProcesses, eOperations and eServices for the Future*. Bled eConference, 2012.

[5] Miquel Sànchez-Marrè. *Intelligent Decision Support Systems*. Springer Nature Switzerland AG, 2023.

Vergleich von React mit Svelte anhand einer Kata-Plattform

Fabio Saupp **Jörg Nitzsche**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma IT-Designers GmbH, Esslingen

In der heutigen Welt der Webentwicklung sind reaktive Benutzeroberflächen mittels eines Javascript-Frameworks der Standard. Frameworks wie React, Angular und Vue.js dominieren die Szene und bieten Entwicklern mächtige Werkzeuge zur Erstellung komplexer Webanwendungen [12]. Ein aufstrebendes Framework welches sich zunehmend an Beliebtheit erfreut, ist Svelte. Im Gegensatz zu React, das mittels einer Runtime arbeitet, kompiliert Svelte Code zu effizientem Vanilla Javascript, was zu einer potenziell verbesserten Performance und einer reduzierten Bundlegröße führt [11] [4].
Diese Arbeit untersucht die Unterschiede zwischen React und Svelte anhand einer Kata-Plattform, welche es Mitarbeitern der IT-Designers GmbH und Studierenden verschiedener Hochschulen ermöglicht, Programmieraufgaben, sogenannte Kata zu bearbeiten und von Reviewern bewerten zu lassen.

Motivation

Die Motivation für den Vergleich der beiden Frameworks liegt in der Optimierung der Benutzererfahrung der Kata-Plattform. Eine performante Webanwendung führt zu einer höheren Zufriedenheit der Benutzer, sowohl bei den Studierenden als auch den Reviewern. [13] Besonders relevant ist die Performance in Umgebungen mit begrenzter Bandbreite und Rechenleistung wie beispielsweise bei mobilen Geräten. Da die Kata-Plattform auf einem einzelnen Server deployt wird, ist hier eine Skalierung nur begrenzt möglich. Eine optimierte Webanwendung kann dazu beitragen, die Plattform auch bei hoher Nutzerlast performant zu halten.
Neben der Performance soll auch die Entwicklererfahrung von Svelte mit React verglichen werden, sodass eine Empfehlung für einen Umstieg auf Svelte oder beibehalten des Status quo ausgesprochen werden kann.

Methodik

Um die Performance von React und Svelte zu vergleichen, werden zwei Ansätze verfolgt:

- Synthetische Benchmarks: Es werden verschiedene DOM-Manipulationen, wie das Einfügen, Bearbeiten und Entfernen von Elementen in beiden Frameworks implementiert und die Ausführungszeit gemessen.
- Real-World Benchmarks: Ausgewählte Views der Kata-Plattform werden in Svelte nachgebaut und mit der bestehenden React-Version verglichen. Die Performance wird anhand von Metriken wie First Contentful Paint, Largest Contentful Paint und Total Blocking Time bewertet, welche mit dem Tool Lighthouse ermittelt werden.

Zusätzlich zur Performance wird die Entwicklererfahrung in beiden Frameworks anhand von Kriterien wie der Codekomplexität, der Verfügbarkeit von Tutorials sowie der Größe und Aktivität der Community bewertet.

Ergebnisse des synthetischen Benchmarks

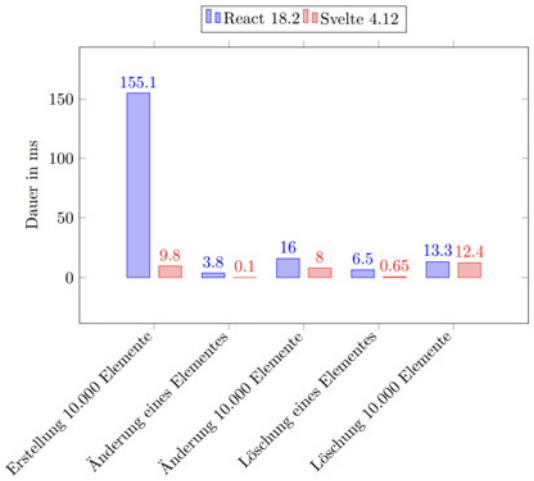

Abb. 1: Ergebnisse des synthetischen Benchmarks [3]

Die synthetischen Benchmarks zeigen, dass Svelte bei den meisten DOM-Manipulationen eine deutlich bessere Performance als React aufweist. Besonders beim Einfügen und Bearbeiten von Elementen ist Svelte um ein Vielfaches schneller. Besonders schnell ist Svelte bei der Änderung eines Elementes, da hier eine direkte DOM-Manipulation vorgenommen wird, welche in React Best Practices verletzen würde [11].

Ergebnisse des Real-World Benchmarks

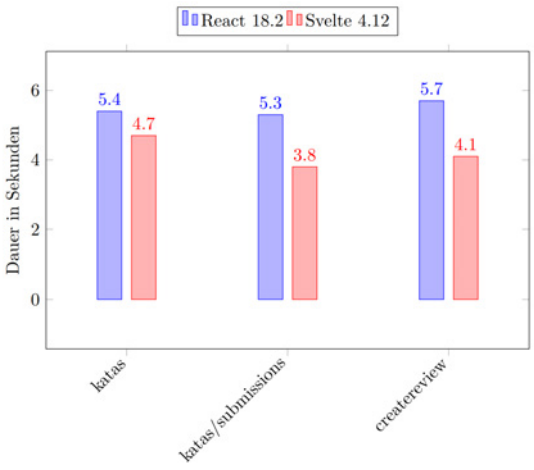

Abb. 2: Ladezeiten der Views (LCP) [3]

Auch bei den Real-World Benchmarks zeigt Svelte eine bessere Performance als React. Die Ladezeiten der nachgebauten Views in Svelte sind durchweg kürzer, was sich in einem höheren Lighthouse Score widerspiegelt.

Evaluation

Neben der Performance werden weitere Aspekte der Entwicklererfahrung bewertet:

- **Codekomplexität:** Die Analyse des Codes mithilfe des Tools Lizard [15] ergibt, dass React zwar mehr Codezeilen und Funktionen benötigt, die Funktionen in Svelte jedoch tendenziell etwas komplexer sind.
- **Tutorials:** Svelte bietet ein interaktives Tutorial mit gut abgegrenzten Themenbereichen. Die Community-Tutorials zu beiden Frameworks sind qualitativ hochwertig und detailliert [7] [14].
- **Community:** React hat mit einer deutlich höheren Anzahl an Downloads eine größere Community als Svelte [6]. Die Beliebtheit von Svelte wächst jedoch schneller an und das Framework erfreut sich einer hohen Zufriedenheit unter den Entwicklern [12].
- **Ökosystem:** React hat ein größeres Ökosystem als Svelte [1] [2] [9] [10]. Es ist jedoch möglich generische Javascript-Bibliotheken einfach in Svelte zu integrieren, wodurch dieser Abstand verringert wird [8] [5].

Fazit

Svelte bietet gegenüber React eine verbesserte Performance und eine reduzierte Bundlegröße. Dies führt zu einer schnelleren Ladezeit und einer besseren Benutzererfahrung, was besonders in Umgebungen mit begrenzter Bandbreite und Rechenleistung relevant ist. Die Entwicklererfahrung in Svelte ist ebenfalls positiv, mit einem interaktiven Tutorial und einer wachsenden Community [14] [6]. Das Ökosystem von React ist zwar größer, Svelte bietet jedoch für die meisten Anwendungsfälle bereits eine Lösung oder ermöglicht die Verwendung generischer Javascript-Bibliotheken.

Empfehlung für die Kata-Plattform

Für die Kata-Plattform wird empfohlen, bei React zu bleiben. Laut den Lighthouse-Tests welche ein Low-End Mobilgerät emulieren, ist Svelte zwar schneller, jedoch reicht dies nicht aus, um in eine neue Geschwindigkeitskategorie zu gelangen, sodass die Performance immer noch schlecht ist. Zudem werden Abhängigkeiten mit einem hohen Footprint, wie etwa der Monaco-Editor für das Editieren des Codes benötigt, wodurch die Einsparungen durch Verwendung eines leichtgewichtigeren Frameworks

insignifikant werden. Außerdem ist die Entwicklung in React momentan schon weit fortgeschritten, sodass der zusätzliche Aufwand des Wechsels des Frontend-Tech-Stacks schwer zu rechtfertigen ist.

Erkenntnisse für andere Projekte der IT-Designers GmbH

Die Ergebnisse dieser Arbeit zeigen deutlich auf, worin die Stärken von Svelte und React liegen. Svelte ermöglicht es hocheffiziente Web-Apps in einer Javascript-nahen Sprache zu entwickeln, welche besonders in Umgebungen mit begrenzten Ressourcen punkten. React hat hingegen einen höheren Footprint, bietet jedoch das größere Ökosystem, mehr Stabilität und eine größere Community, was React zu einer besseren Wahl für komplexere Projekte machen kann.

Literatur und Abbildungen

[1] brillout brillout. GitHub - brillout/awesome-react-components: Curated List of React Components & Libraries. https://github.com/brillout/awesome-react-components, 2024.

[2] chentsulin chentsulin. chentsulin/awesome-react-renderer: Awesome list of React Renderer. https://github.com/chentsulin/awesome-react-renderer, 2024.

[3] Eigene Darstellung.

[4] Tan Li Hau. *REAL-WORLD SVELTE supercharge your apps with Svelte 4 by mastering advanced web development concepts*. Packt Publishing Ltd., 1 edition, 2023.

[5] Garnier Julian. Anime.js - Sphere animation ▪ REPL ▪ Svelte. https://svelte.dev/repl/ea1c82d07d4e4e0191664abdeef5dad5?version=4.2.16, 2024.

[6] Team Npmtrends. react vs svelte | npm trends. https://npmtrends.com/react-vs-svelte, 2024.

[7] Team React. Tutorial: Tic-Tac-Toe-React. https://react.dev/learn/tutorial-tic-tac-toe, 2024.

[8] Team Reactanime. plus1tv/react-anime: (´ ´) *: A super easy animation library for React! https://github.com/plus1tv/react-anime?tab=readme-ov-file, 2024.

[9] Team Rocketlaunchr. rocketlaunchr/awesome-svelte. https://github.com/rocketlaunchr/awesome-svelte, 2024.

[10] Mudit Somani. TheComputerM/awesome-svelte. https://github.com/TheComputerM/awesome-svelte, 2024.

[11] Sebastion Springer. *React: the comprehensive guide*. Rheinwerk Publishing, 2023.

[12] Team Stackoverflow. Stack Overflow Developer Survey 2023. https://survey.stackoverflow.co/2023/?utm_source=social-share&utm_medium=social&utm_campaign=dev-survey-2023, 2023.

[13] Wiktor Stadnik and Ziemowit Nowak. *The Impact of Web Pages' Load Time on the Conversion Rate of an E-Commerce Platform*. Borzemski,Leszek;++, 2018.

[14] Team Svelte. Welcome to Svelte ▪ Svelte Tutorial. https://learn.svelte.dev, 2024.

[15] Terry Yin. terryyin/lizard. https://github.com/terryyin/lizard, 2024.

Robust Template Matching for 6-DoF Pose Estimation based on DL Feature Points

Florian Schaal **Markus Enzweiler**

Department of Computer Science and Engineering, Esslingen University

Work carried out at Fraunhofer IPA, Stuttgart

Introduction

In both industry and retail, processes are increasingly being automated through the use of robots. Robust optical detection is therefore essential for various robotic handling tasks, such as picking retail items or handle metal components. In this thesis, a welding process is automated using an experimental setup. The process sequence envisages that the welding robot can independently detect, localize and handle the components in the system. Since new components often have to be added on the fly in the teach-in process, a template-based approach is to be used instead of an offline training. The use of Deep-Learning Feature Points promises a good trade-off between speed and accuracy. The goal of the bachelor thesis is to compare different deep learning approaches and develop a robust matching method. Despite the challenges of smooth and metallic surfaces, which are prone to reflections, the presented approach should provide a robust matching result which can be used for 6-DoF pose estimation.

Traditional Template Matching

Template Matching is a basic method in image processing and pattern detection. A template can be designed, modeled or shaped in any way, serving as a pattern to create, or in this case to detect certain objects [2]. Such a template is used for the localization of objects in target images and scenes. The matching process is based on the comparison of an image section (template) with different areas of the target image to find the location of the template or to identify instances within the image. To do so the template is moved across the image like a sliding window in which differences between them are quantified. At each step, a metric is calculated for the respective position, which reflects the quality of the match at exactly that location. The result is then stored in a matrix and can be used to match the template within the image [6]. The goal is either to minimize the distance metric to find the point of least dissimilarity or to maximize the similarity metric to find the point of the highest similarity. Traditional feature point-based matching methods include SIFT and SURF [8] [1] . These are algorithms that extract features from the target image and the template and describe these features. To do this, keypoints are searched for in both images and a descriptor is calculated for each of these keypoints, which describes the essential characteristics of the point and its surroundings in the form of a vector. These descriptors are invariant to scaling and rotation and are used to find matches in the target image and template. For this purpose, distance measures such as the Euclidean distance or the Hamming distance can be used, depending on the type of descriptor. Algorithms such as k-nearest neighbor or FLANN (Fast Library for Approximate Nearest Neighbors)are often used to find similar descriptors.

Deep Learning-Based Approaches

Deep learning-based approaches to template matching have significantly advanced the field by improving accuracy and efficiency in various applications. Traditional template matching methods, often based on cross-correlation techniques, struggle with high noise and require extensive manual adjustments and expertise. Deep learning models, such as Convolutional Neural Networks (CNN), offer robust alternatives by learning hierarchical features directly from the data, improving the ability to deal with variations and noise in the input images [5]. There are several possible approaches here. SuperPoint, for example, extracts robust point features and is used in combination with SuperGlue to find precise point-to-point matches using Graph Neural Networks [3] [11]. LightGlue, on the other hand, is a new, improved version of this and uses transformer-based mechanisms for adaptive feature matching. It is particularly efficient in latency-sensitive applications such as 3D reconstruction, as the features are compared iteratively and the width and depth of

the network are adjusted depending on the difficulty of the image pairs to optimize computation time [12]. GlueStick extends these approaches by integrating line features which, in combination with the point features, enable robust image matching. Based on the connectivity information of the points and lines, the robustness against structural changes and textureless areas is increased as well as the effectiveness in case of illumination changes [9].

6-DoF Pose Estimation

Six-degrees-of-freedom (6DoF) pose estimation refers to the determination of the position and orientation of an object in three-dimensional space. The six degrees of freedom include the three translational movements (forward/backward as x, left/right as y, up/down as z) and the three rotational movements (pitch as θy, yaw as θz, roll as θx) as shown in Fig 1. 6-DoF pose estimation occupies a central role in robotics, augmented reality, virtual reality and autonomous driving. The 6-DoF pose of an object is represented by a vector that describes the position (x,y,z) and orientation (θx, θy, θz) of the object. Mathematically, the orientation can be represented by Euler angles, a rotation matrix or quaternions [4].

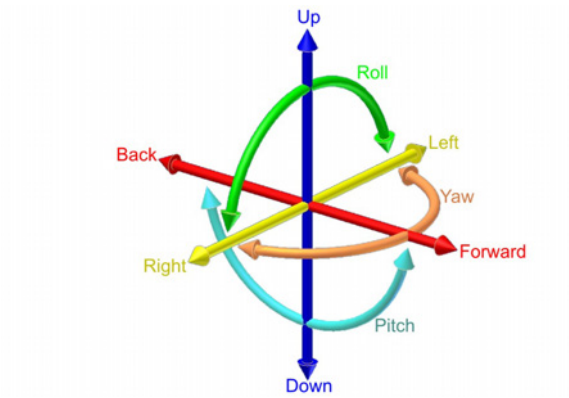

Fig. 1: The six degrees of freedom: forward/back, up/down, left/right, yaw, pitch, roll [7]

Generation of a Dataset

A small dataset in the Bop format was created for the template matching process using CAD files (.stl). A Python script was implemented for this purpose, which loads the objects into Blender, centers them there, adds a metallic surface to them and ensures adequate lighting. For the rendering, the camera is moved around the object in a sphere with a fixed radius and equal steps to map all possible perspectives of the object. Finally, the rgb and depth-images plus the mask are saved. In addition, the camera parameters and the ground truth data are stored in JSON files.

Implementation of 2D Template Matching

For the 2D matching procedure, a class was implemented that uses the GlueStick model to extract and match the feature points and lines. Since the extracted feature points exhibit a high degree of similarity at many points in the images, especially on the edges, due to the metallic surface, the alignment of the template cannot be matched in many cases. This could be overcome by filtering the keypoints according to their similarity. To ensure robust matching, a number of different perspectives of the templates are used instead of a single template. In the first step, the matching process is carried out for each template and a score is calculated, indicating how many of the total number of points and lines found were actually matched. In the second step, the template with the highest combined score of matched points and lines is used to calculate the final homography to match the template with the target image. The homography can be used to determine the 2D points in the target image, which are then needed for the 3D matching process. Fig. 2 shows the transformed 2D points in the target image, in this case the shape of the polygon, sorted by the highest IOU (Intersection over Union).

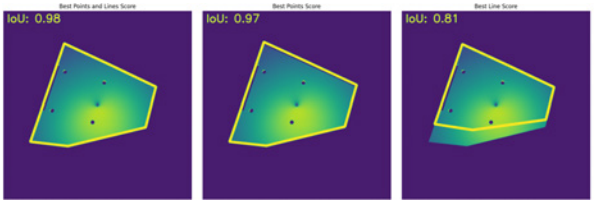

Fig. 2: Matched Shape in the target Image with IOU [10]

By using multiple templates and the combination of points and lines robust matching can be ensured even for textureless metallic surfaces and different viewing angles.

Implementation of 3D Matching

The 3D matching process uses the result of the 2D matcher, in which the homography matrix is calculated to describe the relationship between two planes in image coordinates. 3D matching is now used to determine the pose of an object in three-dimensional space. This involves matching 2D image points with their corresponding 3D world coordinates. This enables the calculation of the 6-DoF pose. A common method for solving this issue is the Perspective-n-Point (PnP)

problem. PnP algorithms calculate the position and orientation of a camera based on a number of 3D points and their 2D projections in the image. Robust methods such as RANSAC (Random sample consensus) are often used to improve the accuracy of the pose estimation and to eliminate outliers. By applying these techniques, the 3D position of an object in the world coordinate system can be determined precisely.

Results and Future Work

The current results show that the matching procedure delivers good results despite the difficulties associated with metallic surfaces. In the 2D matching procedure, the method achieves an average IOU value of 0.96 with the generated data set. The lowest measured value is 0.41 and the maximum is 0.99, indicating a robust matching. However, the results depend heavily on the number of templates used, which is why a balance must be struck between performance and runtime. This, in turn, depends on the hardware in which the matching process is executed. Good results can therefore also be expected for the evaluation of 6-DoF pose estimation, but these still have to be evaluated as part of the thesis. In the future, the procedure must be tested and evaluated on the robot's real system in order to validate the results of the simulated data.

References and figures

[1] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In *Lecture Notes in Computer Science (LNCS)*, volume 3951. Springer, 2006.

[2] Roberto Brunelli. *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley, 1 edition, 2009.

[3] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperPoint: Self-Supervised Interest Point Detection and Description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2018.

[4] Sabera Hoque, Md. Yasir Arafat, Shuxiang Xu, Ananda Maiti, and Yuchen Wei. A Comprehensive Review on 3D Object Detection and 6D Pose Estimation With Deep Learning. In *IEEE Access*, volume 9. IEEE, 2021.

[5] Thomas Hossler. Where's Waldo? A Deep Learning approach to Template Matching, 2017.

[6] Ana Huaman. Template Matching. https://docs.opencv.org/4.x/de/da/tutorial_template_matching.html, 2023.

[7] Horia Ionescu. The six degrees of freedom: forward/back, up/down, left/right, yaw, pitch, roll. https://commons.wikimedia.org/w/index.php?curid=10878582, 2010.

[8] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 2004.

[9] R. Pautrat, I Suárez, Y Yu, M Pollefeys, and V. Larsson. GlueStick: Robust Image Matching by Sticking Points and Lines Together. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 2023.

[10] Own representation.

[11] Paul-Edouard Sarlin, Daniel DeTone, Tomas Malisiewicz, and Andrew Robinovich. SuperGlue: Learning Features Matching with Graph Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2020.

[12] Paul-Edouard Sarlin, Daniel DeTone, Tomas Malisiewicz, and Andrew Robinovich. LightGlue: Local Feature Matching at Light Speed. In *arXiv Preprint*. arXiv, 2023.

Analyse und Aufbau Frontendtests bei #NETZlive

Lennard Schatz **Jörg Nitzsche**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Netze BW GmbH, Stuttgart

Einleitung

"Testautomatisierung ist heutzutage der Standard in der Softwareentwicklung, besonders im agilen Umfeld. Bei sich – gewollt – ständig ändernden und neuen Anforderungen ist ein stabiles Gerüst aus Tests die Versicherung gegen ungewollte Seiteneffekte bei der Weiterentwicklung des Softwaresystems" [5]
Frontendtests sind wesentliche Komponenten moderner Softwareentwicklung, die sicherstellen, dass Webanwendungen zuverlässig und benutzerfreundlich funktionieren. Im digitalen Zeitalter, in dem User experience (UX) und Performance eine zentrale Rolle spielen, ist es unerlässlich, die Benutzeroberflächen regelmäßig zu testen. Frontendtests helfen dabei, Fehler frühzeitig zu erkennen, die Stabilität der Anwendung zu gewährleisten und die Entwicklungszeit durch frühzeitige Fehlerbehebung zu verkürzen. Durch den gezielten Aufbau von Frontendtests können Entwickler die Qualität ihrer Anwendungen erheblich steigern und so den Erfolg und die Zufriedenheit der Endnutzer sicherstellen.

Konzept und Zielsetzung

Zu Beginn muss der vorhandene Bestand an Frontendtests gründlich geprüft werden, um eine genaue Bestandsaufnahme zu erstellen. Auf Basis dieser Analyse werden dann gezielte Verbesserungen vorgenommen, um die Testabdeckung zu erhöhen. Durch die Überprüfung des bestehenden Testbestands werden Lücken und Schwachstellen identifiziert, welche eine potenzielle Gefahr für die Stabilität und Qualität der Anwendung darstellen. Anschließend gilt es, diese Lücken durch die Entwicklung zusätzlicher Testfälle zu schließen und vorhandene Tests zu optimieren. Ziel ist es, eine umfassende und effektive Testabdeckung sicherzustellen, die alle wesentlichen Funktionalitäten der Anwendung abdeckt. Dadurch soll die Zuverlässigkeit der Software gesteigert und eine solide Basis für zukünftige Entwicklungen geschaffen werden.

Umsetzung und Probleme

Zur Umsetzung der Zielsetzung wurde zunächst der bestehende Bestand an Frontendtests detailliert analysiert, um die aktuelle Testabdeckung zu bewerten. Auf Basis dieser Analyse wurden Lücken identifiziert, die durch zusätzliche Tests geschlossen werden mussten. Neue Testfälle wurden entwickelt und bestehende Tests optimiert, um eine umfassendere Abdeckung sicherzustellen. Diese Tests wurden anschließend in die CI/CD-Pipeline integriert, um eine kontinuierliche Prüfung zu gewährleisten.

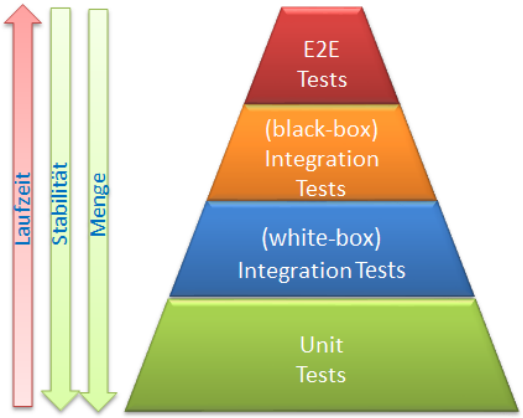

Abb. 1: Testpyramide [4]

Die Auswahl geeigneter Tools und Frameworks war zeitaufwendig und erforderte gründliche Evaluierung. #NETZlive ist ein Projekt, welches in der funktionalen Programmiersprache Elixir geschrieben ist. Wallaby ist ein Tool welches speziell für Elixir entwickelt ist. Wallaby ist dadurch leicht in bestehende Elixir-Projekte zu integrieren. Andere Tools die üblich sind für Tests gegen die graphische Oberfläche sind Selenium und Cypress. [2]
Mit diesen 3 Tools werden die fehlenden Tests geschrieben. Wallaby wird für einfache Tests innerhalb des Elixir-Ökosystems verwendet. Selenium eignet sich dagegen besser zum Durchführen von komplexen Testszenarien. Ein Vorteil gegenüber Wallaby ist, dass

Selenium in Python geschrieben werden kann, was vielen leichter fällt als Elixir. Cypress wird verwendet um schnelle und zuverlässige Tests durchzuführen. Ein großer Vorteil von Cypress ist, dass es einfach zu verwenden ist und gleichzeitig alles gut dokumentiert. Dies macht es für den Fachbereich einfacher die Tests nachzuvollziehen oder sogar mit geringer Kenntnis selbst welche zu erstellen.
Ein weiterer wichtiger Punkt bei der Auswahl der Tools war, was für Arten von Tests geschrieben werden müssen . Ein Integrationstest kann zum Beispiel im Gegensatz zu Unit-Tests die Datenbank als Ganzes mit einbeziehen. Bei Unit-Tests arbeitet man oft mit Mockdaten oder fängt die Testung erst nach einer idealerweise vorhanden Datenbankabstraktionsschicht an. Unit-Tests haben dafür eine geringere Laufzeit als Integration Tests (vgl. Abbildung 1).

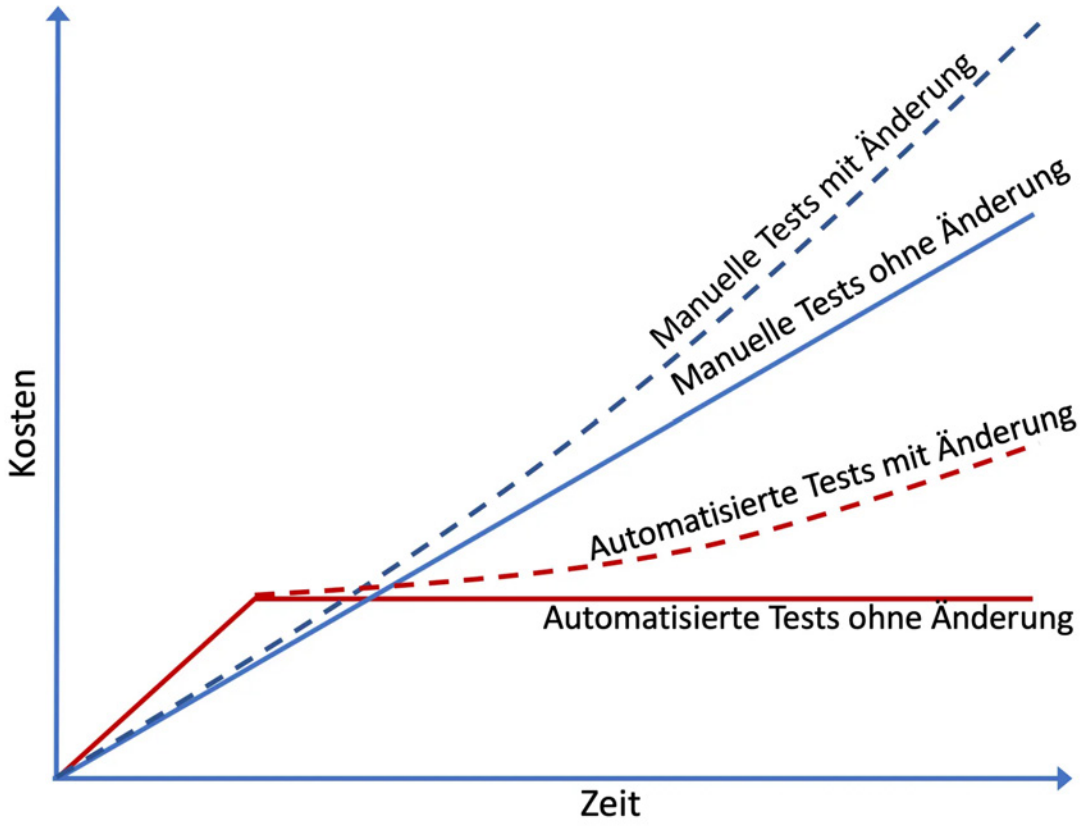

Abb. 2: Kosten von automatisierten Tests gegen Kosten von manuellen Tests [1]

Ausblick

Ein großes Thema wird daher die Verbesserung der Testabdeckung sein. Ein weiteres Thema wird sein Tests, die aktuell noch manuell ausgeführt werden, zu automatisieren. "Manuelle Tests werden von einem Menschen durchgeführt, der sich durch die Anwendung klickt oder mithilfe der richtigen Tools mit der Software und den APIs interagiert. Dies ist sehr teuer, da jemand eine Umgebung einrichten und die Tests durchführen muss. Außerdem entstehen leicht Fehler, wenn sich der Tester vertippt oder versehentlich Schritte aus dem Testskript auslässt." [3] Automatisierte Tests haben zusätzlich den Vorteil, dass diese auf lange Sicht kostengünstiger sind und öfters ausgeführt werden können (vgl. Abbildung 2). Auch wenn automatisierte Tests sehr viele Vorteile mit sich bringen kann auf manuelles Testing nicht ganz verzichtet werden. Deshalb wird auch noch betrachtet inwieweit automatisiert werden kann und an welchen Stellen es sinnvoller ist manuell zu testen.

Literatur und Abbildungen

[1] Roman Anger. Test-Ziele – Warum wir automatisiert testen. https://pentacor.de/test-ziele-warum-wir-automatisiert-testen, 2020.

[2] Gerd Beneken, Felix Hummel, and Martin Kucich. *Grundkurs agiles Software-Engineering*. Springer Vieweg, 2022.

[3] Sten Pittet. Unterschiedliche Arten von Softwaretests. https://www.atlassian.com/de/continuous-delivery/software-testing/types-of-software-testing, 2024.

[4] Waldemar Siebert. Unit Tests. https://www.testautomatisierung.org/lexikon/unit-testing, 2020.

[5] Dehla Sokenou. Cypress überall – Ein einziges Automatisierungswerkzeug für alle Teststufen?! In *Softwaretechnik-Trends*, volume 43, page 1. Gesellschaft für Informatik e.V., 2023.

Möglichkeiten automatisierter Integration von eventbasierten Prozessen

Marvin Schatz **Jörg Nitzsche**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Netze BW GmbH, Karlsruhe

Einleitung

In modernen Softwarearchitekturen gewinnen Microservices immer mehr an Bedeutung, da sie flexibler und besser skalierbar sind als große Softwaremonolithen. Trotz der vielen Vorteile bringen sie aber auch Herausforderungen mit sich. Als kleine, unabhängige Dienste müssen sie häufig mit Events über Schnittstellen kommunizieren. Das kann einige Risiken mit sich bringen. Wenn Microservices nicht erreichbar sind oder Nachrichten nicht verarbeitet werden können, können Nachrichten verloren gehen und zu Fehlern in den Prozessen führen. Um das zu verhindern, werden Systeme benötigt, die für die Übertragung der Events zuständig sind. Leider ist die Konfiguration und Instandhaltung dieser Systeme oft sehr kompliziert und zeitintensiv, da sich die Anzahl der registrierten Services sowie die Verbindungen und Events zwischen den Microservices häufig ändern können. Das führt dazu, dass es sehr schwierig ist, den Überblick über den Kommunikationsfluss zwischen den Systemen zu behalten und Entwickler viel Zeit kostet, die bestehende Infrastruktur zu verstehen.

Zielsetzung der Arbeit

Diese Arbeit befasst sich damit, verschiedene Services in einer Microservice-Systemlandschaft über eine zentrale Plattform zu verbinden. Die Konfiguration einer solchen Plattform ist oft sehr aufwendig und fehleranfällig. Deswegen soll die Plattform mithilfe eines Integration-Tools automatisiert konfiguriert werden und alle Elemente bereitstellen, die für die Kommunikation über die Plattform benötigt werden. Hierfür registrieren sich die Services mithilfe von API-Beschreibungen bei der Plattform und stellen somit die nötigen Informationen für den Aufbau der Plattform und deren Schnittstellen bereit und minimieren damit den Integrationsaufwand für die Entwickler.
Die Plattform sorgt dafür, dass alle Nachrichten der verschiedenen Services zuverlässig übertragen werden. Zudem muss die Reihenfolgetreue der Events garantiert werden können. Das bedeutet, dass die Events in der gleichen Reihenfolge beim Empfänger ankommen, wie sie beim Sender abgesendet wurden. Events, die nicht verarbeitet werden können, weil benötigte Daten von vorherigen Events fehlen, werden zurückgehalten und zu einem späteren Zeitpunkt erneut ausgeliefert.
Zusätzlich soll eine automatisierte Dokumentation die Verbindungen zwischen den Services detailliert darstellen und damit für einen guten Überblick über die registrierten Services und ausgetauschten Events auf der Plattform sorgen.

Grundlagen

In Zeiten, in denen vermehrt auf verteilte Systeme und immer komplexere Kommunikationsstrukturen gesetzt wird, gewinnt auch die asynchrone Kommunikation zunehmend an Bedeutung. Die asynchrone Kommunikation unterscheidet sich von der synchronen Kommunikation in einem wesentlichen Punkt: Hier kann der Empfänger entscheiden, wann er die Nachricht verarbeitet. Das heißt, er muss dabei nicht immer für den Sender erreichbar sein.
Um das zu erreichen, müssen Nachrichten in Warteschlangen zwischengespeichert werden, damit sie zum gewünschten Zeitpunkt vom Empfänger abgerufen werden können. Dafür gibt es zwei relevante Kommunikationsmodelle, die hierfür verwendet werden können.

Point-to-Point

Point-to-Point ist ein Kommunikationsmodell, bei dem Nachrichten in einer Warteschlange abgelegt werden, bis sie vom Empfänger verarbeitet werden. Bei dieser Form der Übertragung wird sichergestellt, dass eine Nachricht immer nur an einen Empfänger gesendet wird. Für den Fall, dass ein Nachrichtenkanal mehrere Empfänger hat, die eine Nachricht empfangen wollen, stellt die Point-to-Point-Verbindung sicher, dass nur einer dieser Empfänger die Nachricht verarbeiten kann.

Es ist auch möglich, dass innerhalb eines Kanals mehrere Empfänger mehrere Nachrichten gleichzeitig empfangen können, dabei wird jedoch immer eine bestimmte Nachricht an einen expliziten Empfänger gesendet. [3]

Abb. 1: Point-to-Point [4]

Pub/Sub-Messaging

Für Nachrichten, die für mehrere Empfänger bestimmt sind, kann das Publish/Subscribe-Pattern verwendet werden. Diese Übertragungsform bietet deutlich mehr Flexibilität, denn hier können mehrere Abonnenten Nachrichten gleichzeitig und asynchron empfangen. Hierfür werden Broker verwendet, die beliebig viele Kopien einer Nachricht erstellen und an die verschiedenen Empfänger verteilen. Jeder Output-Channel des Brokers hat einen Subscriber, der eine Nachricht maximal einmal empfangen kann. Das bedeutet, jeder Empfänger erhält jede Nachricht, die für ihn bestimmt ist, einmal. [3]

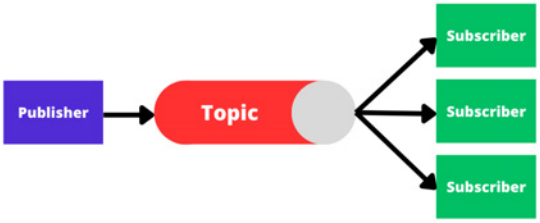

Abb. 2: Publish/Subscribe-Pattern [4]

Konzept

Für den Aufbau dieser Plattform soll der Azure Service Bus verwendet werden. Die Services können sich mithilfe von AsyncAPI-Dateien [2] bei der Plattform registrieren und somit die erforderlichen Informationen für die Konfiguration des Service Bus bereitstellen. Die Dateien werden von dem Integrations-Tool eingelesen und daraus die benötigten Elemente für die Kommunikation über den Service Bus erzeugt. Das Tool erzeugt im Service Bus Namespace ein Topic, auf das die Sender ihre Nachrichten legen können, sowie jeweils eine Subscription pro Subscriber.
Sender senden ihre Events auf das Topic, dort können die einzelnen Events mithilfe von Topic-Strings und Nachrichtenfiltern den entsprechenden Subscriptions der Empfänger zugeordnet werden. Jede Subscription verfügt über eine eigene Queue, in der die Events zwischengespeichert werden und vom Empfänger ausgelesen werden können.
Die Verbindungen der Services werden durch das Tool dokumentiert und der Kommunikationsfluss in einem detaillierten Netzplan bildlich dargestellt, um die Übersichtlichkeit des Netzwerks jederzeit zu gewährleisten.

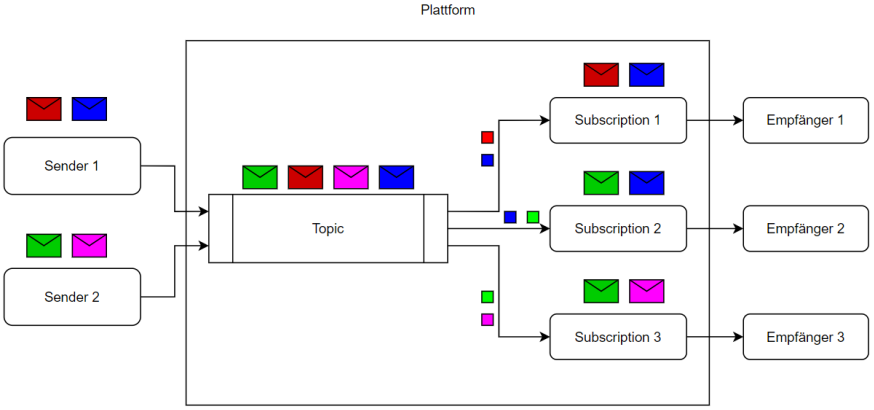

Abb. 3: Kommunikation über ein Topic [1]

Ausblick

In der vorliegenden Bachelorthesis werden bereits die Grundkonzepte für die eventbasierte Nachrichtenübertragung ausgearbeitet und die automatisierte Konfiguration für den Azure Service Bus implementiert. Mithilfe von Nachrichtenfiltern können Nachrichten und Events den registrierten Services zugeordnet und zuverlässig übertragen werden. In der verbleibenden Zeit soll zusätzlich zur Erweiterung des Integrations-Tools eine automatisierte Dokumentation für die Plattform erstellt werden. Die in der Thesis beispielhaft implementierte Plattform soll im Anschluss an die Arbeit in das Projektumfeld integriert und weiter ausgearbeitet werden, um eine vollumfängliche Integrationsplattform zu erhalten. Das beinhaltet ein zuverlässiges Monitoring und Logging, damit die Kommunikation zwischen den Services überwacht werden kann und Fehler durch aussagekräftige Logeinträge schnell behoben werden können. Außerdem muss ein Sicherheitskonzept entworfen werden, um eine zuverlässige Authentifizierung und eine sichere Kommunikation zwischen den Services zu gewährleisten, damit der Angriffsvektor so gering wie möglich gehalten wird.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Dunith Dhanushka. Understanding AsyncAPIs with a Practical Example. https://www.asyncapi.com/blog/understanding-asyncapis, 03 2021.

[3] Gregor Hohpe and Bobby Woolf. *Enterprise Integration Patterns: Designing, Building, and Deploying Messages Solutions*. Addison-Wesley, 2004.

[4] O. D. Romain. Pub/Sub vs. PTP: Understanding Messaging Patterns in Event-Driven Architecture. https://www.code-review.tech/pub-sub-vs-ptp-understanding-messaging-patterns-in-event-driven-architecture/, 03 2023.

Evaluation of Web Components for the Creation of Single Page Applications

Leonie Schick **Harald Melcher**

Department of Computer Science and Engineering, Esslingen University

Work carried out at pep.digital GmbH, Esslingen am Neckar

Introduction

Using a framework to develop user interfaces (UIs) and web applications has many benefits, such as helping to implement common use cases in web applications, for example routing, creating HTML templates and managing states. However, using a framework such as Angular, React, or Vue also introduces more dependencies. On the one hand, frameworks depend on the developers who support and maintain them, who may one day decide to stop development and move on to other projects. On the other hand, they depend on the libraries and projects on which they are built. Using such a framework may therefore result in the inclusion of libraries and modules that are not required by the project itself, which may subsequently become a source of vulnerability. Especially for smaller applications, it can also lead to an increase in overhead and dependency on tools that are otherwise not needed.
A possible alternative to the use of a framework is the use of web components, which allow developers to create their own custom elements. These custom elements can be used in HTML like native tags and provide some of the features offered by frameworks like Angular, such as scoped styling and easy reusability of components.
This thesis evaluates the potential of web components for the development of web applications, using the implementation of a social media client for the Mastodon platform as a case study. It considers the role of web components and other recently introduced JavaScript features, such as import maps, and assesses their suitability as competitors to frameworks such as Angular or React. The library Lit is used as an abstraction layer to facilitate the creation of components.

Mastodon

Mastodon is a microblogging server developed by Mastodon gGmbH, which was founded by Eugen Rochko in 2021. It is a decentralized social media platform and allows users to post and share statuses, short text messages that can contain media attachments like images or videos.
The Mastodon server supports the ActivityPub protocol, allowing it to speak to other Mastodon servers as well as servers of different platforms with compatible ActivityPub implementations as seen in figure 1. When users sign up to Mastodon, they choose which server they want to join and where their data will be stored. Because the project is Open Source, anyone can choose to host their own Mastodon Servers, and users are encouraged to join smaller instances. Mastodon supports the creation of client applications by providing an API documentation and listing existing third-party apps on their official website. [1]

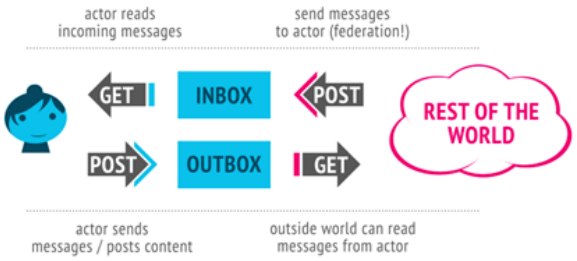

Fig. 1: Sending messages via ActivityPub [2]

Web Components

Web Components are a way to create custom, reusable HTML Elements. The Web Component Specification consists of four parts: Custom Elements, the Shadow DOM, HTML Templates, and ES Modules. [6] Custom Elements are a mechanism that can be used to create new, custom HTML tags, while the Shadow DOM enables developers to encapsulate a web component and protect it from outside DOM manipulation and global CSS rules. Templates are used to define markup inside of HTML that will not render until it is activated

and are utilized to describe how a web component should be rendered. JavaScript modules can then be used to share these web components in a project across multiple files.

Lit

Lit is a JavaScript library that can be used to build web components. It provides a component base class, called LitElement, that provides reactive states, a declarative template system, and styles that are scoped by default. Because every Lit component is also a standard web component, Lit components can be used in projects independently of what framework they were built with, as well as projects that were built with no framework at all. [4]
To create a Lit component, the component class must extend LitElement, the component base class provided by the Lit library. The render method of a component is then used to declare the template of the Lit component, which defines how the component should be rendered. The template can include expressions, which are placeholders for dynamic content. [3] Styles applied to a component are rendered to the Shadow Root of the component, and automatically scoped to it as well.

Requirements and Design

Since Mastodon is a decentralized social media platform, the client application must work with different Mastodon servers and allow users to choose which server they want to communicate with. Users should be able to log in, view posts and user profiles, and create new text posts. Additionally, they should be able to edit text posts and delete their posts if necessary. The app should also provide a search functionality.
The design of the application was inspired by the layout and UI of other popular microblogging platforms like Twitter as well as the official Mastodon client, see figure 2 for an example. Users can access the main functionalities over the header, including a dialog that allows them to choose the server they want to communicate with, and a button that redirects them to this server to authenticate and log in. The application should provide a mobile and desktop view.

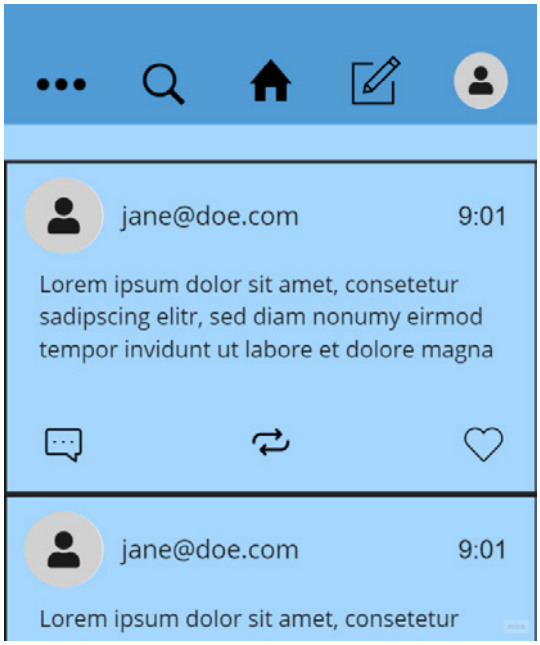

Fig. 2: Mock-up of the home feed on mobile devices [5]

Evaluation

The client application was successfully implemented using web components and Lit. The initial workload to get the application running was relatively high, but once the basic architecture and components were created, it was relatively simple to extend the application and create new modular components. However, the configuration of a web server and implementation of routing require more work and additional libraries.
For developers familiar with JavaScript, using web components and plain JavaScript might be a good alternative when implementing smaller applications with a limited scope. For more complex use cases, it may be beneficial to make use of a package manager like NPM and to make use of Lit in TypeScript for better typing and more convenient component definitions.

References and figures

[1] Mastodon gGmbH. Mastodon documentation. https://docs.joinmastodon.org/, 2024.

[2] Christine Lemmer-Webber, Jessica Tallon, Erin Shepherd, Amy Guy, and Evan Prodromou. AcitivityPub. https://www.w3.org/TR/2018/REC-activitypub-20180123/, 2018.

[3] Google LLC. Rendering - Lit. https://lit.dev/docs/components/rendering/, 2023.

[4] Google LLC. What is Lit? - Lit. https://lit.dev/docs/, 2023.

[5] Own representation.

[6] Carlos Rojas. *Building Native Web Components*. Apress Berkeley, CA, 2020.

Konzeption und prototypische Implementierung einer Visualisierung von Verkehrsdaten in nahezu Echtzeit

Ertugrul Sevgili **Mirko Sonntag**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Anwendungszentrum KEIM des Fraunhofer Instituts für Arbeitswirtschaft und Organisation (IAO), Esslingen

Einleitung und Problemstellung

In städtischen Gebieten, wo eine hohe Verkehrsdichte und viele verschiedene Verkehrsteilnehmer aufeinandertreffen, werden die Herausforderungen hinsichtlich der Verkehrssicherheit immer komplexer. Darüber hinaus erfordert die Optimierung des Verkehrsflusses eine präzise und schnell zugängliche Visualisierung von Verkehrsdaten, um Entscheidungsträgern und Wartungspersonal mehr Transparenz und Nachvollziehbarkeit zu bieten.
Oft basieren traditionelle Verkehrssysteme auf zeitlich begrenzten und ungenauen Daten, was die Fähigkeit zur schnellen und präzisen Reaktion auf sich ändernde Verkehrsmuster einschränkt. In diesem Kontext stellt die 5G-Technologie einen bedeutenden Fortschritt dar, da sie im Vergleich zu vorherigen Mobilfunkgenerationen deutlich höhere Datenübertragungsraten, geringere Latenzzeiten und eine verbesserte Netzwerkkapazität bietet. [4] Diese Eigenschaften sind entscheidend für die Umsetzung von Echtzeitanwendungen, die in Smart Cities eine zentrale Rolle spielen. [3] [6]

Projekt 5G-trAAffic

Das Projekt 5G-trAAffic ist ein wegweisendes Forschungsprojekt, das von der Stadt Aalen initiiert wurde, um die Verkehrssicherheit und den Verkehrsfluss durch den Einsatz der 5G-Technologie zu verbessern. Ein Hauptziel des Projekts ist eine proaktive Verkehrssteuerung und -sicherheit durch präzise Visualisierung von Echtzeitdaten, um Entscheidungsträgern effektive Informationen bereitzustellen und die Verkehrssicherheit zu erhöhen. [5]

Fallstudie: IST-Zustand

Im Rahmen des Projekts 5G-trAAffic wurden folgende drei spezifische Kreuzungen in der Stadt Aaalen ausgewählt. An den ausgewählten Kreuzungen befinden sich Multisensoren, die sowohl Lastkraftwagen (LKW) als auch Personenkraftwagen (PKW) erfassen können, sowie spezielle Sensoren, die nur auf die Erfassung von PKWs ausgerichtet sind.

- Knoten 130: Stuttgarter Strasse / Friedrichstrasse: Insgesamt sind an diesem Standort 15 Sensoren installiert, davon sind vier Multisensoren und 11 Sensoren für PKW.
- Knoten 320: Friedrichstrasse / Gartenstrasse: Insgesamt sind an diesem Standort 13 Sensoren installiert, davon sind zwei Multisensoren und 11 Sensoren für PKW.
- Knoten 330: Friedrichstrasse / Friedhofstrasse: Insgesamt sind an diesem Standort 20 Sensoren installiert, davon sind ein Multisensor und 19 Sensoren für PKW. (siehe Abbildung 1)

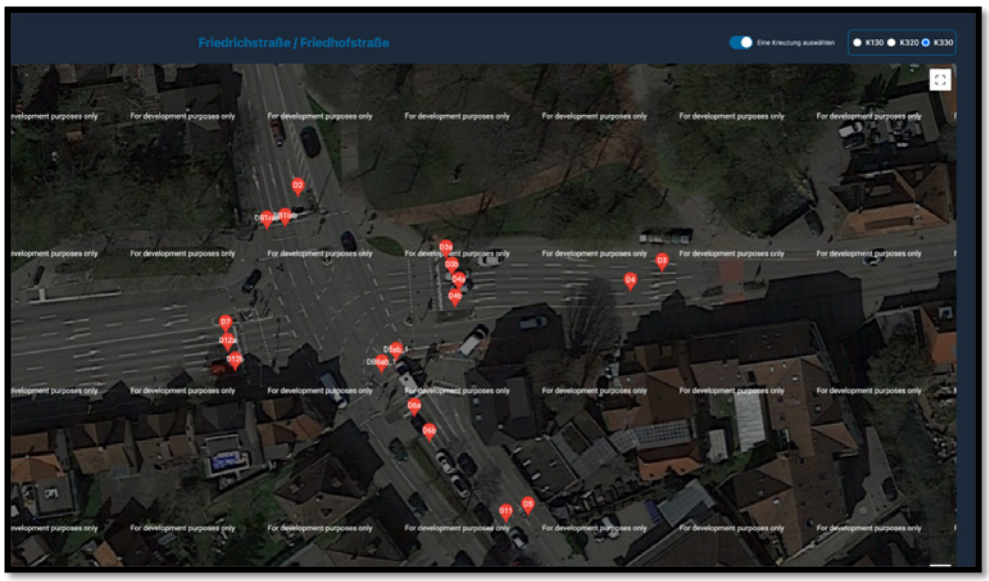

Abb. 1: Markierte Sensoren auf Kreuzung K330 [2]

Die Vielzahl der installierten Sensoren an den drei untersuchten Knotenpunkten generiert ein enormes Datenvolumen. Diese Daten bilden die Grundlage für die Entwicklung fortschrittlicher Verkehrsmanagementlösungen, die im Rahmen des Projekts 5G-trAAffic erprobt und optimiert werden sollen.

Ziel der Arbeit

Das Hauptziel dieser Bachelorarbeit, durchgeführt im Rahmen des Projekts 5G-trAAffic, ist die Entwicklung eines Systems zur Erhöhung der Verkehrssicherheit durch verbesserte Warnmechanismen und die Optimierung des Verkehrsflusses. Der Schwerpunkt liegt dabei auf der effektiven Visualisierung von Echtzeitdaten, um sowohl Entscheidungsträgern als auch Verkehrsteilnehmern relevante Informationen über kritische Verkehrssituationen bereitzustellen.

Konzept

Die Architektur des Projekts besteht aus mehreren Schlüsselkomponenten, die in einem integrierten Rahmen zusammenarbeiten, um eine präzise Informationsdarstellung in nahezu Echtzeit zu ermöglichen.

- Datenquellen: Die primären externen Datenquellen des Systems sind Sensoren, die an strategischen Verkehrs- und Parkplatzstandorten installiert sind. Zusätzlich sollten Wetterdaten von meteorologischen Diensten bezogen werden, die das Verkehrsaufkommen und die Parkplatznutzung beeinflussen können.

  Um eine effiziente und automatisierte Datenintegration zu gewährleisten, sieht das Konzept vor, dass diese Datenquellen über eine API automatisch angebunden werden. Dadurch können Clients die notwendigen Daten direkt abrufen und in das System eingeben.

- Echtzeit-Benachrichtigung und -Aktualisierung: Das Backend verwendet moderne Echtzeit-Kommunikationstechnologien wie WebSockets, um das Frontend über neue Daten zu informieren. Sobald eine solche Benachrichtigung empfangen wird, aktualisiert das Frontend umgehend die angezeigten Informationen ohne Benutzerinteraktion, um die neuesten Daten widerzuspiegeln.

- Frontend-Visualisierung: Die visualisierten Daten werden über das Dashboard und die Kartenansicht dargestellt. Das Dashboard bietet interaktive Diagramme und Echtzeitanalysen. Die Kartenansicht zeigt geografische Daten an, wodurch Nutzer Verkehrsinformationen in Bezug auf physische Standorte interpretieren können.

Implementierung

Das Projekt nutzt eine Vielzahl moderner Technologien und Frameworks, um eine zuverlässige und skalierbare Lösung zu entwickeln:

- Für das Backend werden Java in Kombination mit dem Spring Framework eingesetzt. Die Echtzeit-Kommunikation wird durch den Einsatz von WebSockets und STOMP realisiert.

- Die Datenpersistenz erfolgt über PostgreSQL, wobei Liquibase zur Verwaltung von Datenbankmigrationen dient.

- Das Frontend basiert auf Next.js und Tailwind CSS.
- Zudem sind alle Services und die Datenbank in jeweils separaten Docker-Containern untergebracht.

Testen

Echtzeitsysteme werden oft in harte und weiche Echtzeitsysteme unterteilt. Harte Systeme erfordern strikte Zeitvorgaben, während weiche Systeme flexiblere Zeitgrenzen erlauben. Die im Rahmen dieser Bachelorarbeit entwickelte Anwendung zur kontinuierlichen Darstellung von Verkehrsdaten auf einem Dashboard kann ein weiches Echtzeitsystem betrachetet werden.
Um die Echtzeitfähigkeit des Systems für Parkplatzdaten zu testen, wurde es eine POST-Anfrage mit dem Daten, die mit einem Zeitstempel vom 30. Mai 2024, 14:05:07.52244Z. versehen, für ein Parkhaus gesendet. Aus denen Abbildungen (siehe Abbildung 2, Abbildung 3) geht hervor, dass die neuen Daten innerhalb von 131 Millisekunden vom Backend empfangen, verarbeitet und im Frontend angezeigt wurden.
Dies entspricht einer Reaktionszeit von etwa 0,1 Sekunden, was in vielen weichen Echtzeitsystemen als akzeptabel gilt und die Erwartungen an eine zeitnahe Informationsverarbeitung und -darstellung erfüllt. [1]

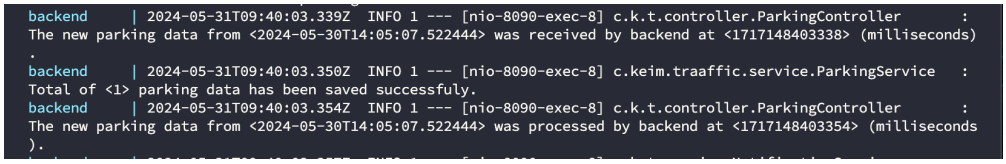

Abb. 2: Der Zeitraum des empfangenen Parkplatz-Datensatzes vom Backend [2]

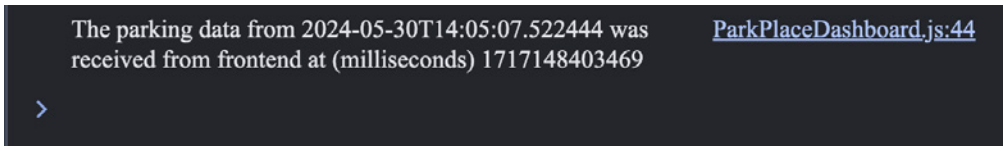

Abb. 3: Der Zeitraum des empfangenen Parkplatz-Datensatzes vom Frontend [2]

Ausblick

Das entwickelte System im Rahmen des Projekts 5G-trAAffic in der Stadt Aalen hat gezeigt, dass es möglich ist, Verkehrs-, Parkplatz- und Wetterdaten in Echtzeit zu erfassen, zu verarbeiten und visuell ansprechend darzustellen.

Derzeit werden Daten aus externen Quellen manuell über Swagger POST-Operationen in die Datenbank eingegeben. Eine wichtige Weiterentwicklung wäre die Automatisierung dieses Prozesses in einem Partnersystem, um die Effizienz des Systems zu steigern und die Datenaktualität sowie die Reaktionsgeschwindigkeit des Systems zu verbessern.

Literatur und Abbildungen

[1] Wikipedia contributors. Real-time computing. https://en.wikipedia.org/w/index.php?title=Real-time_computing&oldid=1217334040, 2024.

[2] Eigene Darstellung.

[3] Reply Digital Services. Low Latency: what makes 5G different. https://www.reply.com/en/telco-and-media/low-latency-what-makes-5g-different, 2024.

[4] Fraunhofer Gesellschaft. 5G – Die Zukunft im Netz. https://www.fraunhofer.de/de/forschung/aktuelles-aus-der-forschung/5g-die-zukunft-im-netz.html, 2024.

[5] Fraunhofer IAO. 5G trAAffic. https://www.keim.iao.fraunhofer.de/de/projekte/5G.html, 2024.

[6] Bundesamt für Sicherheit in der Informationstechnik. Was versteht man unter 5G? https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Technologien_sicher_gestalten/5G/5g-was-versteht-man-darunter.html, 2024.

Data Mesh: Herausforderungen und Lösungen skalierbarer Datenarchitekturen

Barsan Shemari **Anke Bez**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Mercedes-Benz Group AG, Stuttgart

Einleitung

Daten sind in der heutigen Zeit unverzichtbar und allgegenwärtig. Mit der fortschreitenden Automatisierung und dem verstärkten Einsatz von IT-Systemen gewinnen sie zunehmend an Bedeutung [5]. Ihr gesamter Lebenszyklus wird durch eine Datenarchitektur beschrieben, welche den Entwurf dafür liefert, wie sie durch Speichersysteme fließen, verarbeitet und genutzt werden. Datenarchitekturen umfassen Datenverwaltungssysteme wie Datenbanken, Data Warehouses (DWH) und Data Lakes (DL) [3]. Traditionelle Datenarchitekturen, die auf zentrale Datenteams setzen, stoßen jedoch angesichts des Hypes um Big Data und Künstliche Intelligenz an ihre Grenzen. Um diesen Herausforderungen zu begegnen, hat Zhamak Dehghani das Konzept des Data Mesh (dt. Datennetz) eingeführt. Data Mesh stellt einen Paradigmenwechsel dar, indem es von der kostspieligen und zeitaufwändigen Verwaltung monolithischer DWHs und DLs sowie der hohen Belastung zentraler Datenteams wegführt. Es adressiert den Konflikt zwischen organisatorischen Anforderungen und bestehenden Architekturen und ermöglicht eine effizientere, dezentralisierte Datenverwaltung [1].

Ziel der Arbeit

Das grundlegende Ziel dieser Arbeit ist es, die Grenzen herkömmlicher Datenarchitekturen wie DWHs und DLs aufzuzeigen und das Konzept des Data Mesh als innovative Lösung vorzustellen. Die Untersuchung konzentriert sich dabei insbesondere auf die ersten beiden zentralen Prinzipien des Data Mesh: Domain Ownership und Data as a Product. Diese Prinzipien sind besonders wichtig, da sie die Datenverantwortung auf domänenspezifische Teams verlagern und Daten als eigenständige Produkte behandeln. Dies führt zu einer höheren Skalierbarkeit, Agilität und Effizienz im Datenmanagement, sodass Unternehmen besser auf dynamische Anforderungen und wachsende Datenmengen reagieren können.

Theoretische Grundlagen

Das Konzept des DWH wurde eingeführt, um autorisierten Personen zuverlässige und verständliche Unternehmensdaten zur Entscheidungsunterstützung bereitzustellen [6]. Ein DWH integriert Informationen aus verschiedenen Datenquellen und ist themenorientiert, zeitbezogen und nicht-flüchtig [2]. Mit dem Aufkommen von Big Data und der wachsenden Datenmenge im Internet stieg der Bedarf an neuen Lösungen zur Speicherung und Analyse großer Mengen an strukturierten, semi-strukturierten und unstrukturierten Daten. Hier kommt das Konzept des DL ins Spiel. DLs sind zentralisierte, skalierbare Speicher für große Mengen an rohen, unverarbeiteten Daten im Originalformat. Diese Art der Datenhaltung ermöglicht es Unternehmen, bessere, datengestützte Entscheidungen zu treffen [7]. Während ein DWH wie eine Flasche Wasser ist – gereinigt, verpackt, strukturiert und bereit zum Verzehr – kann ein DL als natürlicher See betrachtet werden, in dem jeder das Wasser untersuchen und beliebig nutzen kann [4]. Eine Gegenüberstellung der wesentlichen Merkmale von DWHs und DLs zeigt Abbildung 1.

Eigenschaften	DWH	DL
Sinn & Zweck	Für die Datenabfrage zur Entscheidungsunterstützung optimiert	Speicherung vielfältiger Daten
Datenverarbeitung	Hochverarbeitete Daten	Hauptsächlich Rohdaten
Datenhaltung	Strukturiert, tabellarisch	Unstrukturierte, semistrukturierte oder strukturierte Daten
Flexibilität	Relativ unflexibel, da Daten für bestimmten Zweck vorgesehen und entsprechend aufbereitet sind	Daten können dank Rohformat flexibel genutzt und an verschiedene Bedürfnisse angepasst werden
Nutzer	Breites Spektrum an möglichen Benutzern	Aufgrund der Komplexität hauptsächlich Experten
Speicher	Teuer und performant	Kostengünstig, skalierbar
Schema	Schema-on-writing	Schema-on-reading
Sicherheit	Hohe Datenkontrolle	Niedrige Datenkontrolle
Aufnahme neuer Daten	Aufwändiger ETL-Prozess	Schnelle Ablage

Abb. 1: Vergleich zwischen DWH und DL, eigene Darstellung in Anlehnung an [7]

Die ersten beiden Generationen von Datenarchitekturen für Unternehmen, das DWH und der DL, sind zentralisiert, monolithisch und domänenunabhängig. In diesen erfolgen Datenspeicherung, -transformation, -manipulation, -konsum und -management in einem einzigen System. Diese Architekturen sind jedoch nur begrenzt skalierbar und neigen dazu, bei wachsendem Datenvolumen langsamer, teurer und komplizierter zu werden – ein Resultat zentraler Flaschenhälse. Im Bereich der Softwareentwicklung hat bereits ein Paradigmenwechsel von monolithischen Architekturen hin zu modularen Microservices stattgefunden. Ein vergleichbarer Wechsel hin zu dezentralisierten, verteilten und domänenabhängigen Architekturen steht im Datenmanagement noch aus [1].

Data Mesh

Data Mesh bietet einen dezentralisierten Ansatz für Datenarchitekturen und legt den Fokus auf domänenspezifische Datenverantwortung. Es verspricht Skalierbarkeit und Agilität sowie einen schnelleren Zugriff auf hochwertige Daten und adressiert gleichzeitig die Herausforderungen herkömmlicher Architekturen. Data Mesh betrachtet Daten als Produkt und weist die Verantwortung für diese den jeweiligen Domänen zu [1].

Domain Ownership

Domain Ownership ist das erste von insgesamt vier zentralen Prinzipien des Data Mesh. Es zielt darauf ab, zentrale Flaschenhälse durch die Verlagerung der Datenverantwortung auf einzelne Bereiche, sogenannte Domänen (engl. domains), zu beseitigen. Diese Domänen, die ihre eigenen Daten produzieren und nutzen, verstehen sie am besten und sind somit in der Lage, sie am effektivsten zu beeinflussen und zu kontrollieren. Daher werden sie als Eigentümer (engl. owner) dieser definiert. So könnte der After-Sales-Bereich von Mercedes-Benz eine Domäne darstellen, die für Fahrzeugdaten nach dem Verkauf verantwortlich ist, einschließlich Wartungs- und Reparaturhistorien, Garantieinformationen und Kundenfeedback. Durch die Übernahme der Datenverantwortung kann die After-Sales-Domäne die Datenqualität und -verfügbarkeit verbessern und effizientere Serviceprozesse gewährleisten. Zudem können Domänen in weitere Subdomänen aufgeteilt werden, um spezifischere Verantwortungsbereiche abzudecken. Durch die Eigenverantwortung der Domänen sinkt die Notwendigkeit eines zentralen Datenteams. Unternehmen können die Bereitstellung ihrer Daten entsprechend ihrem Wachstum skalieren und neue Datenquellen, zusätzliche Datenkonsumenten sowie Anwendungsfälle leichter integrieren. Dieser dezentrale Ansatz reduziert die Notwendigkeit teamübergreifender Abstimmungen und fördert so Agilität und Dynamik. Indem die Verantwortung für Daten und deren Bereitstellung auf die Domänen übertragen wird, sinkt zudem die Gefahr der Entstehung veralteter und verwaister Kopien. Diese Koppelung von Datenverantwortung und Domänen ermöglicht eine kontinuierlich hohe Datenqualität und stellt sicher, dass die Daten stets aktuell und relevant bleiben [1].

Data as a Product

Ein weiterer Grundbaustein des Data Mesh ist das sogenannte Product Thinking, welches durch das zweite Kernprinzip Data as a Product eingeführt wird. Dabei werden Daten von ihren Domänen als Produkte behandelt und ihren Nutzern beziehungsweise Kunden entsprechend angeboten. Diese Datenprodukte müssen, ähnlich wie andere Produkte, bestimmte Anforderungen erfüllen, um die Zufriedenheit ihrer Kunden sicherzustellen, in diesem Fall Praktikabilität, Nützlichkeit und Benutzerfreundlichkeit (siehe Abb. 2). Datenprodukte sollen die Realität so genau wie möglich darstellen und gleichzeitig flexibel genug sein, um auf neue Bedürfnisse reagieren zu können. Daher hat Zhamak Dehghani acht unverzichtbare Usability-Attribute definiert, die jedes Datenprodukt aufweisen muss: auffindbar, adressierbar, verständlich, vertrauenswürdig, nativ zugreifbar, interoperabel, eigenständig nützlich und sicher. Diese Attribute stellen sicher, dass Datenprodukte unternehmensweit geteilt werden können und ihre Qualität regelmäßig gepflegt wird. Das Prinzip Data as a Product verändert die Beziehung zwischen Teams und Daten, indem der Erfolg von Datenprodukten anhand der Anzahl ihrer Nutzer und deren Zufriedenheit gemessen wird, anstatt anhand der Menge der in DWHs und DLs verfügbaren Datensätze. Data Mesh sieht vor, dass Datenprodukte unternehmensweit zur Verfügung gestellt werden, um eine Datendemokratisierung zu erreichen und die Entstehung privater Datensilos zu verhindern [1].

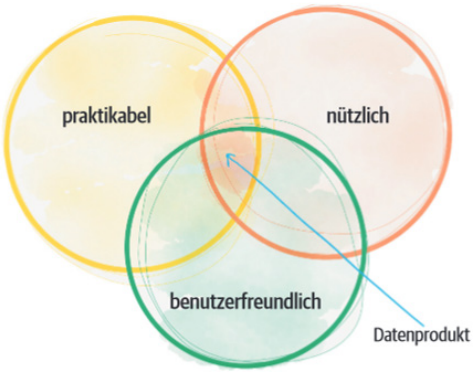

Abb. 2: Anforderungen an Datenprodukte [1]

Das große Ganze sehen

Während das Prinzip der Self-Serve Data Platform die Bereitstellung und das Anbieten von Datenprodukten effizient und unabhängig von anderen Domänen ermöglicht, behandelt das Prinzip der Federated Computational Governance das Thema Data Governance. Im Rahmen von Data Mesh wird die Datenverwaltung nicht an eine zentrale Instanz übergeben, sondern von einem föderalen Team übernommen, das aus Domänenvertretern besteht und von Experten unterstützt wird. Diese föderale Governance wird computergestützt und automatisiert ausgeführt, um eine konsistente und skalierbare Datenverwaltung sicherzustellen [1]. Abbildung 3 zeigt das Zusammenspiel aller vier Data Mesh Prinzipien.

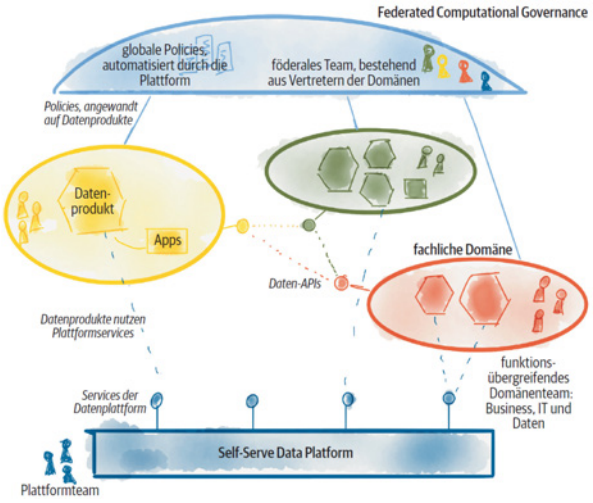

Abb. 3: Zusammenspiel der Data Mesh Prinzipien [1]

Ein Paradigmenwechsel

Data Mesh baut auf den bestehenden Konzepten von DWH und DL auf, ersetzt diese jedoch nicht. Stattdessen zielt es darauf ab, sie zu optimieren, bestehende Probleme zu lösen und neue Potenziale für Unternehmen zu erschließen. Data Mesh kann als soziotechnisches Paradigma klassifiziert werden, das neue Prinzipien einführt und das Zusammenspiel von Menschen und technischen Architekturen in komplexen Organisationen neugestaltet. Es bietet einen innovativen Ansatz zur Definition einer unternehmensweiten Datenstrategie und zielt darauf ab, das Datenmanagement so zu transformieren, dass sowohl Datenproduzenten als auch -konsumenten eine verbesserte Benutzererfahrung erleben können. In einem Data Mesh werden bereits vorhandene Architekturen kombiniert und zu einer verteilten Architektur, einem Datennetz, vereint. Es stellt somit eine Weiterentwicklung der traditionellen Datenarchitekturen dar, die den steigenden Anforderungen an moderne Datennutzung gerecht wird [1].

Ausblick

Im weiteren Verlauf der Arbeit wird eine empirische Untersuchung zum Stand, Reifegrad und den Herausforderungen der Implementierung von Data Mesh innerhalb der Mercedes-Benz Group AG durchgeführt. Hierzu werden qualitative Interviews mit Schlüsselpersonen im Unternehmen geführt, um tiefere Einblicke in die aktuelle Umsetzung und spezifische Hürden zu gewinnen. Zusätzlich wird exemplarisch ein Datenprodukt entwickelt. Die Ergebnisse der empirischen Untersuchung sowie die Erfahrungen und Erkenntnisse aus dem Entwicklungsprozess werden genutzt, um praktische Verbesserungsvorschläge zu entwickeln und Empfehlungen für die zukünftige Implementierung und Optimierung von Data Mesh in Unternehmen zu formulieren.

Literatur und Abbildungen

[1] Z. Dehghani. *Data Mesh: eine dezentrale Datenarchitektur entwerfen*. O'Reilly, 2023.

[2] K. Farkisch. *Data-Warehouse-Systeme kompakt: Aufbau, Architektur, Grundfunktionen*. Springer, 2011.

[3] IBM Deutschland GmbH. Was ist eine Datenarchitektur? https://www.ibm.com/de-de/topics/data-architecture, 2024.

[4] S. Gupta and V. Giri. *Practical Enterprise Data Lake Insights: handle data-driven challenges in an Enterprise Big Data Lake*. Apress, 2018.

[5] S. Helmis and R. Hollmann. *Webbasierte Datenintegration: Ansätze zur Messung und Sicherung der Informationsqualität in heterogenen Datenbeständen unter Verwendung eines vollständig webbasierten Werkzeuges*. Vieweg + Teubner Research, 2009.

[6] J. Holthuis. *Der Aufbau von Data Warehouse-Systemen: Konzeption, Datenmodellierung, Vorgehen*. Dt. Univ.-Verl., 2001.

[7] A. Nambiar and D. Mundra. An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management. *Big Data and Cognitive Computing*, 2022.

Partial automation of vulnerability management in Product Security using Natural Language Processing

Sungeeta Singh **Jürgen Koch**

Department of Computer Science and Engineering, Esslingen University

Work carried out at Festo SE & Co. KG, Esslingen

Motivation and problem formulation

Nowadays a hardware or software product can have thousands of additional third party components which lead to higher exposure to vulnerabilities [14]. Hence, to reduce risks and potential damage vulnerability management tasks can be automatized. As a starting point, vulnerability management encompasses the search for potential known vulnerabilities in a product using vulnerability scanning tools. For this automatized task, a Software Bill of Materials (SBOM) [18] containing information on the product components such as unique identifiers and databases that contain correlating product information with vulnerability information are required [17]. One possible database for vulnerabilities is the National Vulnerability Database (NVD) which is a central, searchable repository maintained by National Institute of Standards and Technology (NIST) [9] with global unique vulnerability identifiers named Common Vulnerabilities and Exposures (CVEs) [10] and unique product identifiers called Common Platform Enumerations (CPEs) [2]. Figure 1 provides an example for a CVE and CPE pair for an Apache product [13]. A vulnerability scanning tool using NVD matches the CPEs from the SBOM to the CPEs in the NVD to get all the corresponding CVEs returning a list of known potential vulnerabilities.

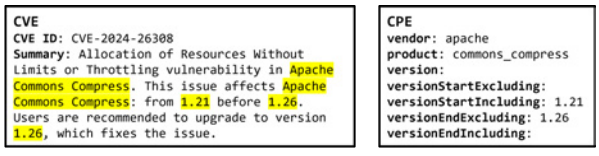

Fig. 1: Example for a CVE and CPE pair [13]

When new vulnerabilities emerge in a product, the MITRE and other authorized organizations take responsibility for issuing new CVEs [3]. NIST manually assigns metadata such as CPEs of the affected products to a CVE and is supported by further organizations since the number of CVEs is rapidly rising [15]. The manual process also includes generating a new CPE if one does not already exist for the affected product [20]. However, this manual assignment causes inconsistencies within NVD. Such inconsistencies include multiple existing CPEs for the same product. For instance, MySQL Server is affected in CVE-2023-21836 [11] and CVE-2023-21875 [12]. The product name "mysql_server" is assigned for the CPE linked to CVE-2023-21836 and "mysql" for the CPE linked to CVE-2023-21875. Additionally, the manual process leads to a median time lag of 35 days till a CPE is assigned to a CVE [5]. In order to automatize this process, this work focuses on extracting information from CVE descriptions. This information can then be used to enrich the database and to guess the CPEs for newly disclosed CVEs. This approach can reduce the median time lag and enable vulnerability scanning tools to find CVEs.

Data Preparation

This work is based on the approach of Wareus and Hell [20] which describes the information extraction as a Named Entity Recognition (NER) task. NER is a Natural Language Processing task aimed at identifying and classifying key information within text into predefined categories [19]. Therefore, the CVE description is split into individual words called tokens. Each of these input tokens are classified by a NER model. Because of this classification problem, an approach to prepare and label CVEs is described in the following. The example of the CVE and CPE pair, see Figure 1, shows that vendor, product and the various version types in the CPE can also be found in the summary while ignoring the case and special characters. This information which should be extracted by the NER task is marked yellow in the CVE summary in Figure 1. The values in the CPE are used as labels for the summary. All the summary and CPE pairs are disregarded that cannot be linked as described. The application of this mapping on the dataset with

CVEs from 1999 to 2019 leads to a reduction from around 130,000 [14] CVEs to approximately 23,000 CVEs. However, this mapping enables an automated labeling of the dataset for the NER task. Due to inconsistencies in CPEs, this labeling method does not accurately work for all data samples. To address this issue, the data preparation process is further enhanced by addressing some of the inconsistencies. For instance, some summaries provide information on a version range as the following format "1.3.0-1.5.0". The tokenizer considers any word between whitespaces as a single token. Consequently, the version range is considered as one token, and the labeling algorithm is not able to determine whether it represents the starting or the ending version. To improve the data preparation process, the version range is divided into three tokens: "1.3.0", "-", and "1.5.0". This modification corrects the automated labeling for this case.

Transfer Learning using Transformers for Named Entity Recognition

Wareus and Hell apply a Bidirectional Long Short-Term Memory (BLSTM) model with a Conditional Random Field (CRF) [7] for the NER task [20]. In contrast, this work investigates the application of Transfer Learning using pretrained transformer models for the NER task. Transformers like BERT can be trained on the entire set of the input sequence using the attention mechanism [4], meanwhile BLSTMs are trained on the ordered sequence from left-to-right and right-to-left. Therefore, BERT can learn the word context on surrounding words and not just the immediately preceding or following words. Models such as RoBERTa, SecureBERT and DeBERTa have been developed based on BERT. RoBERTa is an optimized version of BERT undergoing extensive training and using optimized hyperparameters [8]. For SecureBERT, RoBERTa has been trained on cybersecurity-related sources such as articles from Arxiv or websites of NIST and MITRE [1]. DeBERTa represents an enhanced model of BERT and RoBERTa using two novel techniques. Firstly, the disentangled attention is applied. For the attention weight calculation between two word pairs, the relative position of the pair is additionally considered next to the initially used content of the pair. Moreover, the pre-training technique which is used for BERT and RoBERTa is further enhanced. The used technique is called masked language modeling where the masked tokens are predicted in order to train the model. During the training, DeBERTa considers the absolute position of a word in contrast to the other two models [6]. Figure 2 illustrates the overall accuracy which is the percentage of the correctly annotated CVEs in the test dataset.

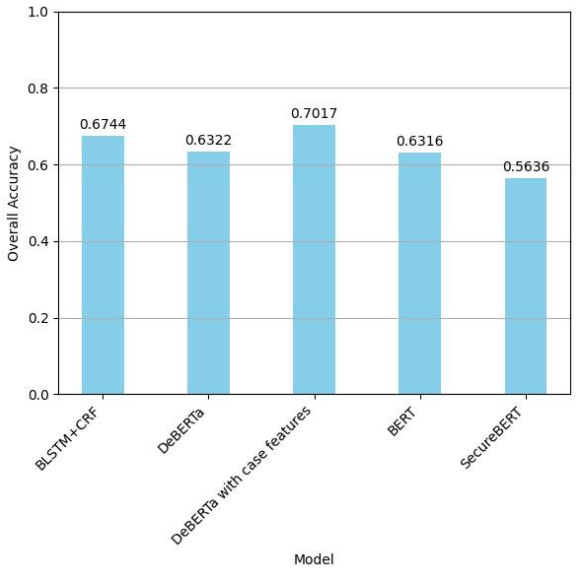

Fig. 2: Model performance of BLSTM with CRF, DeBERTa, BERT and SecureBERT measured by overall accuracy [16]

Best results are gained with DeBERTa after adding word level case features based on the features described in [20]. Since product and vendor typically are capitalized and versions fully or mostly contain character level digits, these features describe this information for each token and help to find vendor, product, and version [20].

Outlook

Future research is focused on further improving the automated dataset labeling. A specific focus lies in addressing more inconsistencies observed in version strings between the summary and CPE data. Certain cases show a disparity where the version range is presented as two version strings in the summary that are excluded from the version range, while the CPE represents the range with two inclusive version numbers. Furthermore, the transfer learning approach with DeBERTa and the word level case features delivers better results than the approach with a BLSTM model and a CRF [20]. Therefore, further experiments on BERT and SecureBERT can be performed with the added features. Additionally, the utilization of Large Language Models such as GPT 3.5 Turbo will be investigated for the NER task and for obtaining CPEs for a CVE in order to compare results of the transfer learning approach with pretrained transfer models.

References and figures

[1] Ehsan Aghaei et al. SecureBERT: A Domain-Specific Language Model for Cybersecurity. In *Security and Privacy in Communication Networks: 18th EAI International Conference, SecureComm 2022, Virtual Event, October 2022, Proceedings*. Springer, 2023.

[2] Brant Cheikes, David Waltermire, and Karen Scarfone. Common Platform Enumeration: Naming Specification Version 2.3. http://dx.doi.org/https://doi.org/10.6028/NIST.IR.7695, 2011.

[3] The MITRE Corporation. CVE Numbering Authorities (CNAs). https://www.cve.org/ProgramOrganization/CNAs, 2024.

[4] Jacob Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://arxiv.org/abs/1810.04805, 2019.

[5] Clément Elbaz, Louis Rilling, and Christine Morin. Automated Keyword Extraction from One-day Vulnerabilities at Disclosure. In *NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2020.

[6] Pengcheng He et al. DeBERTa: Decoding-Enhanced BERT with Disentangled Attention. In *2021 International Conference on Learning Representations*. Microsoft, 2021.

[7] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc., 2001.

[8] Yinhan Liu et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. https://arxiv.org/abs/1907.11692, 2019.

[9] National Institute of Standards and Technology NIST. General Information. https://nvd.nist.gov/general, 2022.

[10] National Institute of Standards and Technology NIST. Vulnerabilities. https://nvd.nist.gov/vuln, 2022.

[11] National Institute of Standards and Technology NIST. CVE-2023-21836 Detail. https://nvd.nist.gov/vuln/detail/CVE-2023-21836, 2023.

[12] National Institute of Standards and Technology NIST. CVE-2023-21875 Detail. https://nvd.nist.gov/vuln/detail/CVE-2023-21875, 2023.

[13] National Institute of Standards and Technology NIST. CVE-2024-26308 Detail. https://nvd.nist.gov/vuln/detail/CVE-2024-26308, 2024.

[14] National Institute of Standards and Technology NIST. Statistics Results. https://nvd.nist.gov/vuln/search/statistics?form_type=Basic&results_type=statistics&search_type=all&isCpeNameSearch=false, 2024.

[15] Kevin Poireault. RSAC: CISA Launches Vulnrichment Program to Address NVD Challenges. https://www.infosecurity-magazine.com/news/cisa-launches-vulnrichment-program/, 2024.

[16] Own representation.

[17] Luis Alberto Benthin Sanguino and Rafael Uetz. Software Vulnerability Analysis Using CPE and CVE. https://arxiv.org/abs/1705.05347, 2017.

[18] National Telecommunications and Information Administration NTIA. Framing Software Component Transparency: Establishing a Common Software Bill of Material (SBOM). https://www.ntia.gov/files/ntia/publications/framingsbom_20191112.pdf, 2019.

[19] Erik F. Tjong Kim Sang and Fien De Meulder. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*. -, 2003.

[20] Emil Wareus and Martin Hell. Automated CPE Labeling of CVE Summaries with Machine Learning. In *Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer International Publishing, 2020.

Entwicklung eines Single-Pair-Ethernet Gateways für ein modulares Testsystem

Georg Steinebrunner **Michael Scharf**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Steinbeis Embedded Systems Technologies GmbH, Esslingen

Einleitung

„Industrial Ethernet“ ist seit Ende der 90er Jahre ein Diskussionsthema in der Automatisierungstechnik. Denn es zeigt die Möglichkeit eines einheitlichen Kommunikationssystems für die gesamte Fabrikhalle, im Gegensatz zu der hohen Komplexität bei Feldbussen, die sich teils erheblich sowohl auf Signal- als auch auf Protokollebene unterscheiden. Hinzu kommt, dass diese Systeme keinem offenen Standard folgen, sondern auf proprietäre Kommunikationsprotokolle und physikalische Übertragungsverfahren setzen [4]. Ethernet ist hingegen vollständig durch das Institute of Electrical and Electronics Engineers (IEEE) standardisiert.
Seit wenigen Jahren gibt es interessante Ergänzungen zum IEEE 802.3 Ethernetstandard. Mit Single-Pair-Ethernet, das aus der Automobilindustrie hervorgegangen ist, sind Ethernetverbindungen über ein einzelnes verdrilltes Adernpaar (Twisted-Pair) möglich. Die Vorteile sind eine sehr einfache, kostengünstige und leichte Verkabelung. Trotzdem ist die Kommunikation sehr robust gegen elektromagnetische Störungen und die Übertragungsraten reichen von 10 Mbit/s bis 25 Gbit/s. Diese, durch den Einsatz im Auto motivierten Qualitäten, sind auch im Automatisierungsumfeld interessant.

10BASE-T1S Ethernet Standard

Ein Single-Pair-Ethernet Standard ist 10BASE-T1S. Entsprechend dem üblichen Namensschema sind 10 Mbit/s über ein einzelnes Twisted-Pair spezifiziert. Das *S* steht dabei für Short-Range. Direkte Verbindungen sind über mindestens 15 m möglich. Hervorzuheben ist aber die Multidrop Fähigkeit, mit der eine Busstruktur über mindestens 25 m mit mindestens acht Teilnehmern realisiert werden kann. Die Kommunikation erfolgt dann halb-duplex, wegen dem gemeinsam genutzten Medium kann es zu Kollisionen kommen. Deterministischen Latenzen sind trotzdem möglich. Denn der Standard bringt eine Kollisionsvermeidung mit (Physical Layer Collision Avoidance, PLCA), die auf ein zyklisches Kommunikationsverfahren setzt [3]. Bei dem Verfahren wird jedem Teilnehmer am Bus eine ID zugewiesen. Es gibt einen Koordinator, der mittels eines *Beacon* Signals für eine Synchronisierung sorgt. Anschließend erhält jeder Teilnehmer anhand seiner ID eine Sendemöglichkeit innerhalb eines Zyklus. Ein Busteilnehmer kann sein Sendefenster in Anspruch nehmen, indem er das *Commit* Signal benutzt. Eine vereinfachte Darstellung des Zyklus ist in Abbildung 1 zu sehen. Der minimale Zyklus wird erreicht, wenn kein Teilnehmer sendet.
Mit diesem Verfahren wird die Bandbreite unter den Teilnehmern aufgeteilt und Kollisionen werden vermieden. Ein Teilnehmer muss seine Sendemöglichkeit nicht wahrnehmen, er wird dann ausgelassen und der nächste Teilnehmer kann direkt senden. Die Menge der übertragenen Daten ist dynamisch, ein Teilnehmer kann die maximale Länge eines Ethernetframes ausreizen. Falls alle Teilnehmer die maximale Länge ausreizen, ergibt sich der maximale Zyklus. Das Verfahren ist somit nur fair hinsichtlich der Möglichkeit für jeden Teilnehmer zum Senden. Eine Fairness bezüglich der Datenmenge gibt es nicht. Somit bestehen Vor- und Nachteile gegenüber einem Zeitschlitzverfahren.

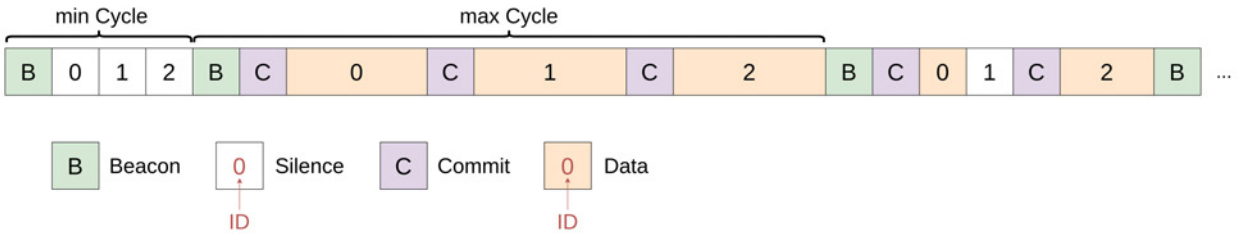

Abb. 1: Vereinfachter PLCA Zyklus [2]

Ziele

Ziel der Arbeit ist die Entwicklung eines Single-Pair-Ethernet Gateways für ein modulares Testsystem. Das System besteht dabei aus einem Gateway und mehreren Nodes. Dabei können die einzelnen Nodes unterschiedliche Schnittstellen bereitstellen. Die Kommunikation zwischen den einzelnen Punkten soll dabei mit 10BASE-T1S stattfinden. In einem bisher genutzten, bereits bestehenden System, wurde ein RS-485 Bus zur Kommunikation genutzt. Also soll evaluiert werden inwiefern sich Single-Pair-Ethernet mittels Multidrop hier als Ersatz, vielleicht sogar als Verbesserung, eignet. Die Kommunikation nach außen erfolgt vom Gateway aus über klassisches 100-Mbit/s-Ethernet, somit soll auch die Möglichkeit der direkten Kommunikation mit den Nodes von außerhalb betrachtet werden. Dies wird durch die vollständige Kommunikation mittels Ethernetstandards im gesamten System prinzipiell möglich, eine Übersetzung wie sie mit dem RS-485 Bus notwendig war könnte entfallen.
Das Gateway basiert auf einem STM32H7 Mikrocontroller, an diesen wird ein 10BASE-T1S fähiger MAC-PHY angebunden. Dieser Baustein vereint die Medium Access Control (MAC) mit dem Zugriff auf das physische Medium (PHY). Auf der vorherigen Version des Gateways kam das Echtzeitbetriebssystem NuttX zum Einsatz, dies soll beibehalten werden. Entsprechend ist ein Treiber für den MAC-PHY zu entwickeln, außerdem muss ein Protokoll zur Kommunikation zwischen Gateway und Nodes geplant werden. Dies betrifft auch die Zuordnung der PLCA IDs bei Systemstart. Schlussendlich muss eine Anpassung der bisher genutzten Platine des Gateways vorgenommen werden, um den MAC-PHY zu integrieren. Nach Inbetriebnahme des Systems kann das Kommunikationsverhalten im System untersucht werden.

Umsetzung

Der 10BASE-T1S MAC-PHY kommuniziert über ein Serial Peripheral Interface (SPI) mit dem Mikrocontroller. In einem ersten Prototyp wurde nur das korrekte Handling des Datenaustauschs zwischen den beiden Geräten betrachtet, ohne den Einsatz eines Echtzeitbetriebssystems. Der Testaufbau erfolgte mittels Entwicklungsboards, denn dies erleichtert den Einsatz eines Logikanalysators erheblich. Damit ist es möglich, die Kommunikation auf dem SPI-Bus direkt zu betrachten und somit Implementierungsfehler schnell zu erkennen. Für die Datenübertragung mit dem MAC-PHY gibt es ein Protokoll, das zwischen Kontroll- und Datentransaktionen unterscheidet. Erstere werden zur Konfiguration der Register im MAC-PHY genutzt. Bei den Datentransaktionen werden Ethernetframes übertragen. Bei beiden Verfahren sieht das Protokoll einen Header vor, der beispielsweise Informationen zur Art der Transaktionen, Registeradressen und auch ein Paritätsbit enthält.
Die Datentransaktionen werden in 64 Byte große Blöcke unterteilt. Vor jedem Block steht ein Header mit Kontrollinformationen, damit im MAC-PHY die Blöcke wieder korrekt zusammengesetzt werden. Aufgrund dieser ständigen Unterbrechung der Nutzdaten wurde im Treiberprototyp eine zweiteilige Ringpufferstruktur implementiert. Die Elemente im ersten Puffer beschreiben dabei die Position eines Ethernetframes im zweiten Datenpuffer. Mit dieser Struktur können zu Testzwecken viele Frames beliebiger Länge versendet werden. Der Vorteil der Ringpuffer ist eine sehr hohe Speichereffizienz, allerdings ist das Kopieren von Daten in beziehungsweise aus dem Puffer aufwändig.
Für das Senden und Empfangen von Testdaten wird auf der Gegenseite eine USB-Netzwerkkarte mit einem integrierten 10BASE-T1S PHY genutzt. Dadurch kann die Netzwerkkommunikation beispielsweise mittels des freien Softwaretools Wireshark analysiert werden. Außerdem ist es so möglich ein Pythonskript zu implementieren, dass sehr viele Ethernetframes beliebiger Länge generiert und versendet. Das ist sehr hilfreich beim Testen und kann auch für Durchsatzmessungen genutzt werden.
Im zweiten Schritt wurde ein MAC-PHY Treiber für das Echtzeitbetriebssystem NuttX entwickelt. NuttX ist streng nach dem Portable Operating System Interface (POSIX) implementiert und ist sehr modular aufgebaut [1]. Es gibt eine strikte Trennung zwischen einem Kernel und Software, die in einem weniger privilegierten Userspace ausgeführt wird. Der bestehende Prototyp konnte in kurzer Zeit als Userspacetreiber für NuttX adaptiert werden, da nur die Schnittstelle

zur SPI-Kommunikation geändert werden musste. Die Integration des Treibers in den NuttX Kernel war deutlich aufwändiger. Denn der NuttX Netztwerkstack ist prinzipiell so aufgebaut, dass der Empfang eines Ethernetframes mittels eines Interrupts signalisiert wird. Der bisherige Prototyp empfängt neue Frames, indem zyklische Anfragen an den MAC-PHY gestellt werden. Abbildung 2 zeigt, im Kontext eines 10BASE-T1S Multidrop-Netzwerks, die Rolle des Netzwerktreibers im Betriebssystem. Die Ringpuffer sind nicht mehr notwendig, da die Speicherverwaltung durch den Kernel übernommen wird.
Die Integration des MAC-PHYs auf der bestehenden Platine erfolgte im dritten Schritt. Zunächst wurden einige Komponenten entfernt, die in der neuen Schaltung nicht mehr notwendig sind, auch um Platz für den MAC-PHY zu schaffen. Neben dem Mikrocontroller ist ein Speicherbaustein vorhanden, der zur Zwischenspeicherung von Messdaten dient. Die Signalleitungen zu diesem Speicher müssen alle gleich lang sein, damit keine unterschiedlichen Signallaufzeiten entstehen. Um den Aufwand hier zu verringern, wurde das bestehende Layout teilweise übernommen. Auch die Führung der Single-Pair-Ethernet Datensignale ist recht kritisch. Beispielsweise müssen korrekte Abstände zu anderen Signalen eingehalten werden, um Übersprechen zu verhindern. Da die Größe der Platine durch ein vorhandenes Gehäuse vorgegeben ist, war trotz vierlagiger Platine viel Detailarbeit notwendig, um alle Leitungen unterzubringen.

Ausblick

Erste Ergebnisse zeigen bereits die Vorteile von PLCA schon beim Einsatz von drei Busteilnehmern. Da Kollisionen effektiv verhindert werden, kann die gesamte Bandbreite der Verbindung genutzt werden und die Aufteilung der Bandbreite erfolgt dynamisch unter den Teilnehmern. Nach Fertigung und Inbetriebnahme der Platine soll im weiteren Verlauf auch betrachtet werden, inwiefern das Verfahren mit deutlich mehr Teilnehmern skaliert, besonders dann, wenn die übertragende Datenmenge pro Teilnehmer sehr unterschiedlich ist. Bei vollständiger Implementierung des Projekts wird voraussichtlich die Kommunikationsfähigkeit in vertikaler Richtung der Automatisierungspyramide deutlich verbessert. Es wird eine durchgehende Ethernetkommunikation bis zu Mikrocontrollern in der Feldebene erzielt.

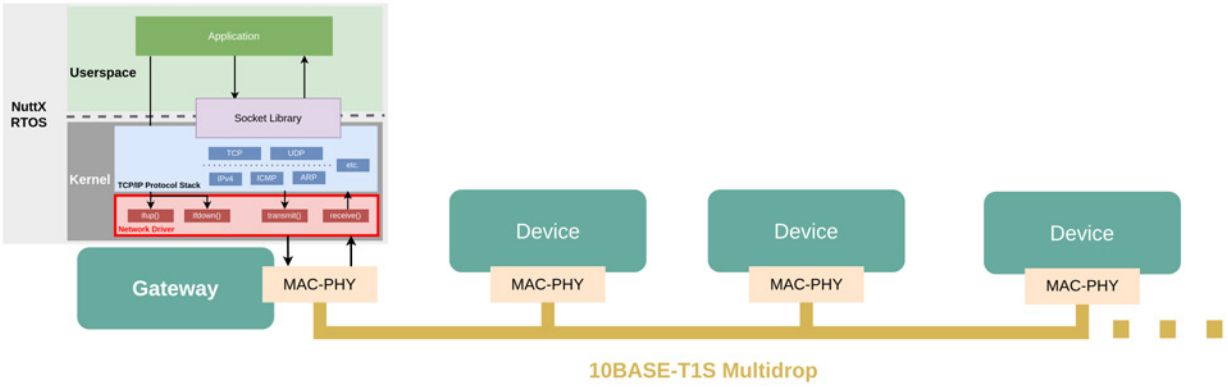

Abb. 2: Rolle des NuttX Netzwerktreibers im Kontext eines10BASE-T1S Multidrop-Netzwerks [2]

Literatur und Abbildungen

[1] The Apache Software Foundation. The Inviolable Principles of NuttX. https://nuttx.apache.org/docs/latest/introduction/inviolables.html, 2023.

[2] Eigene Darstellung.

[3] Institute of Electrical Engineers and Electronics. IEEE Standard for Ethernet - Amendment 5: Physical Layer Specifications and Management Parameters for 10 Mb/s Operation and Associated Power Delivery over a Single Balanced Pair of Conductors. *IEEE Std 802.3cg-2019 (Amendment to IEEE Std 802.3-2018 as amended by IEEE Std 802.3cb-2018, IEEE Std 802.3bt-2018, IEEE Std 802.3cd-2018, and IEEE Std 802.3cn-2019)*, 2020.

[4] Gerhard Schnell and Bernhard Wiedemann. *Bussysteme in der Automatisierungs- und Prozesstechnik*. Wiesbaden: Springer Vieweg, 9 edition, 2019.

Sichere und performante Integration eines NIDS als Docker-Container auf einer Industriefirewall: Strategien zur Netzwerktrafficweiterleitung

Leopold Stenger **Rainer Keller**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma ads-tec Industrial IT GmbH, Nürtingen

Einleitung und Motivation

Vor einigen Jahren noch als innovatives Konzept angesehen, ist Industrie 4.0 im Jahr 2024 zu einem integralen Bestandteil der Industrie- und IT-Welt geworden. Die Digitalisierung und Vernetzung von Maschinen und Anlagen sind in vielen Unternehmen bereits Standard. Aktuelle Studien bestätigen dies und zeigen, dass 48 Prozent der deutschen Industrieunternehmen sich selbst als führend bei digitalen Innovationen im internationalen Vergleich sehen [3].
Doch wie sieht es mit der Sicherheit aus? Trotz der offensichtlichen Vorteile von Effizienz- und Produktivitätssteigerung, darf die Thematik der Cybersecurity nicht vernachlässigt werden. Ein effektiver Ansatz zur Lösung dieses Problems ist die Integration einer Netzwerküberwachungslösung. Für die Vernetzung von Maschinen und Anlagen werden zahlreiche Industriefirewalls eingesetzt. Wieso also diese nicht direkt für die Netzwerküberwachung nutzen?
Eine passende Lösung bietet die ads-tec Industrial IT GmbH mit ihrer Produktreihe *IRF*. Diese Produktreihe wurde speziell für den Einsatz in der Industrie entwickelt und bietet eine sichere und zuverlässige Möglichkeit, Maschinen und Anlagen miteinander zu vernetzen. Für einen flexiblen Einsatz der Firewalls, bieten diese die Funktion, einen Docker-Daemon zu betreiben [4]. Dies ermöglicht es, eigene Anwendungen in einem isolierten Container laufen zu lassen. Dabei ist jedoch zu beachten, dass der Daemon im Rootless-Mode betrieben wird, um die Sicherheit und Integrität des Systems zu gewährleisten [2].

Zielsetzung

Ziel dieser Arbeit ist es, ein Network Intrusion Detection System (NIDS) als Docker-Container auf einer IRF-3000 Firewall zu betreiben. Der Fokus liegt dabei auf der Netzwerktrafficweiterleitung vom Host-System in den Container. Abbildung 1 zeigt, wo die Weiterleitung auf dem System stattfindet. Für die Weiterleitung des Netzwerktraffics gibt es verschiedene Methoden, die untersucht und bewertet werden müssen. Daraus resultieren folgende Forschungsfragen:
- Welche Methoden eignen sich am besten, hinsichtlich Performance und Sicherheit, um den Netzwerktraffic vom Host-System in einen Docker-Container zu leiten?
- Welche Methode lässt sich praktisch am besten in der bestehenden Umgebung implementieren und ermöglicht eine effiziente und sichere Lösung zum Betreiben eines NIDS?

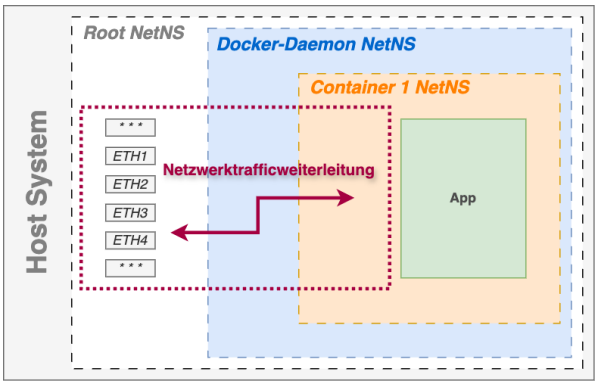

Abb. 1: Übersicht für die Netzwerktrafficweiterleitung [1]

Grundlagen und Vorgehensweise

Zur Beantwortung der Forschungsfragen müssen zunächst grundlegende Konzepte erarbeitet werden. Diese erstrecken sich über drei Ebenen: Anwendungsebene, Linux Kernel Feature Ebene und NIC Ebene.
Auf der *Anwendungsebene* müssen zwei Hauptkomponenten betrachtet werden: die Network Intrusion Detection Systeme und Docker selbst. Da in dieser Arbeit primär der Netzwerktraffic und dessen Weiterleitung im Fokus stehen, werden die NIDS nicht im

Detail betrachtet, sondern nur deren grundlegende Funktionsweisen und Anforderungen. Im Mittelpunkt steht die Betrachtung von Docker und dessen Funktionsweise. Hierbei ist insbesondere der Rootless-Modus von Interesse, da dieser die Grundlage für das Betreiben von Containern auf dem System darstellt. Es muss also untersucht werden, wie dieser Modus funktioniert und welche Einschränkungen und Möglichkeiten er bietet.
Eine weitere Ebene ist die *Linux Kernel Feature Ebene*. Hier gilt es die verschiedenen Konzepte des Linux Kernel zu erarbeiten, welche die Grundlage für die Containerisierung durch Docker und die Netzwerkweiterleitung bilden. Dazu gehören Namespaces, Cgroups, Capabilities und Seccomp. Diese Konzepte sind essentiell, um die Funktionsweise von Docker und die Einschränkungen des Rootless-Modus nachvollziehen zu können.
Die letzte zu betrachtende Ebene ist die *NIC Ebene*. Auf dieser müssen die Grundlagen des gesamten Netzwerkstacks verstanden werden. Dabei gilt es zu verstehen, wie der Netzwerktraffic auf der Firewall ankommt, verarbeitet wird und durch Bridges, TUN/TAP-Devices, Veth-Pairs etc. weitergeleitet wird. Mit dem Wissen aus diesen Grundlagen können dann die verschiedenen Methoden zur Netzwerkweiterleitung untersucht werden. In der Dokumentation von Docker-Rootless werden bereits einige Methoden aufgelistet [2]:

- *slirp4netns*
- *lxc-user-nic*
- *bypass4netns*

Neben diesen Ansätzen gibt es weitere Methoden, wie beispielsweise das Verschieben eines Netzwerkinterfaces in den Network Namespace des Containers. Diese Ansätze müssen anhand von Kriterien wie Performance, Sicherheit und Komplexität bewertet werden. Welche Methode lässt sich am einfachsten implementieren und bietet die beste Performance? Welche Methode ist am sichersten und bietet die geringste Angriffsfläche? Diese Fragen müssen beantwortet werden, um eine effiziente und sichere Lösung zu finden.

Ausblick

Abbildung 2 zeigt die Implementierung der Methode bypass4netns im Vergleich zum Standard-Netzwerkstack von Rootless-Docker. Diese im Paper von N. Matsumoto und A. Suda vorgestellte Methode bietet nach dem derzeitigen Stand der Forschung die besten Ergebnisse in Bezug auf Performance und Sicherheit. Es ist jedoch noch unklar, ob diese Methode auch für die Weiterleitung von Netzwerktraffic in ein NIDS geeignet ist. Grundlegend arbeitet die Methode mit dem *SECCOMP_IOCTL_NOTIF_ADDFD* Befehl, welcher in Version 5.9 des Linux Kernels eingeführt wurde. Zudem kommen *SOCK_STREAM* Sockets zum Einsatz, um den Netzwerktraffic weiterzuleiten [5].

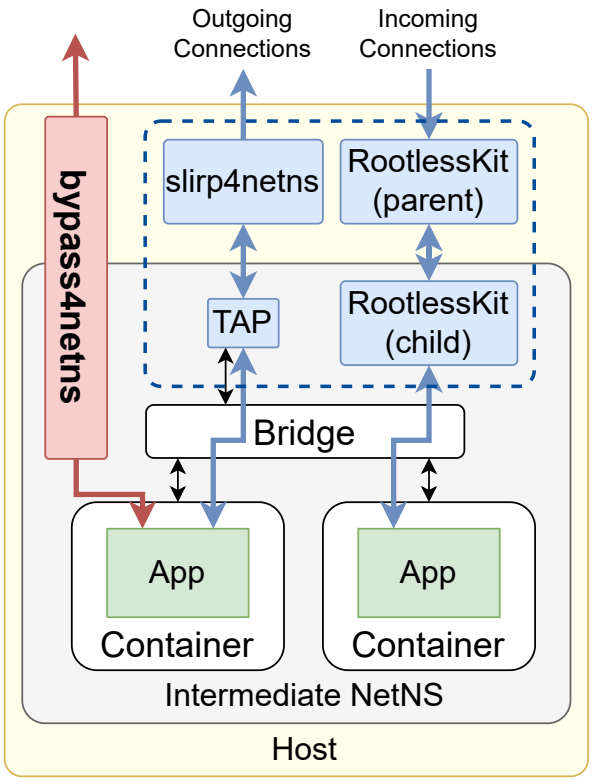

Abb. 2: Netzwerktrafficweiterleitung mit bypass4netns [5]

Um ein NIDS zu betreiben, müssen jedoch auch *SOCK_RAW* Sockets unterstützt werden. Nur diese enthalten alle Informationen des Netzwerktraffics, die für die Erkennung von Angriffen und somit für die Funktionalität eines NIDS essentiell sind. Ein möglicher Nachteil dieser Methode könnte jedoch die Komplexität der Implementierung darstellen. Zum Zeitpunkt der Erstellung dieses Artikels ist noch unklar, ob eine Anpassung von *bypass4netns* zur Unterstützung von *SOCK_RAW* Sockets im Rahmen dieser Arbeit möglich ist.
Schlussendlich soll die passendste Methode für die Netzwerktrafficweiterleitung gefunden werden. Also die Lösung, welche die Anforderungen an Performance und Sicherheit erfüllt und sich gleichzeitig am besten in die bestehende Infrastruktur der Industriefirewall integrieren lässt. Die Implementierung dieser Lösung soll dann in einem Proof-of-Concept demonstriert werden.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] docker docs. Run the Docker daemon as a non-root user (Rootless mode). https://docs.docker.com/engine/security/rootless/, 2024.

[3] bitkom eV. Digitale Innovationen: Hälfte der deutschen Industrie sieht sich vorn. https://www.bitkom.org/Presse/Presseinformation/Digitale-Innovationen-Haelfte-deutscher-Industrie-sieht-sich-vorn, 2023.

[4] ads-tec Industrial IT GmbH. IRF-3000 - Sicherer Fernzugriff und IIoT-Lösungen. https://www.ads-tec-iit.com/sicherer-fernzugriff-iiot-loesungen/firewall-router/irf-3000/, 2024.

[5] Naoki Matsumoto and Akihiro Suda. bypass4netns: Accelerating TCP/IP Communications in Rootless Containers. https://doi.org/10.48550/arXiv.2402.00365, 02 2024.

Prototypische Umsetzung einer Erkennung von Close-Cut-In Manövern bei Straßenbahnen

Michael Stober **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Bosch Engineering GmbH, Abstatt

Einleitung

Der vermehrte Einsatz von Fahrassistenzsystemen hat in den letzten Jahren zu einer deutlichen Verbesserung der allgemeinen Verkehrssicherheit geführt. So konnte bei Lastkraftwagen das Risiko von Kollisionen reduziert werden, wodurch die Anzahl der Unfälle laut einer Studie der Berufsgenossenschaft um 34% reduziert wurde [3]. Fahrassistenzsysteme kommen aber nicht nur auf den Straßen zum Einsatz. Sie halten auch bei Straßenbahnen Einzug. Dies ist wichtig, da beim Straßenbahnbetrieb in Großstädten häufig Mischverkehr und damit ein Unfallrisiko zwischen den Verkehrsteilnehmern entsteht. Für Fahrassistenzsysteme spielt die LiDAR-Technologie eine immer wichtigere Rolle. Sie liefert durch die präzise Umgebungserfassung und die hohe Auflösung der Laserscans einen hohen Informationsgehalt über die Umgebung. Der Einsatz kann somit wesentlich zur Reduzierung von Unfällen beitragen [1].

Motivation und Zielsetzung

Bei Unfällen zwischen Straßenbahnen und PKW besteht eine große Gefahr für Personenschäden, Sachschäden und in der Folge Ausfallzeiten des regulären Bahnbetriebs. Mehrheitlich trifft bei Unfällen im Straßenbahnbetrieb nicht den Tramführer die Schuld (16%), sondern den PKW-Fahrer [5]. Besonders Linksabbieger, welche die Fahrbahn der Straßenbahn schneiden, stellen ein erhöhtes Risiko dar. Eine Analyse der Unfallberichte (siehe Abb. 1) des Polizeipäsidiums Stuttgart belegt, dass Kollisionen mit Straßenbahnen besonders häufig auf Fehler von Linksabbiegern zurückzuführen sind [4].

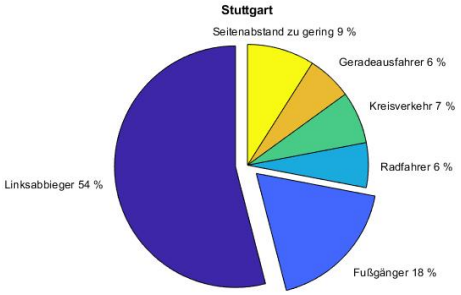

Abb. 1: Statistik über Unfalltyp Stuttgart [2]

Dieses Manöver lässt sich als Close-Cut-In (nachfolgend CCI) definieren. Hierbei ist im Allgemeinen ein nahes Einscheren oder Wenden nach links vor die Straßenbahn durch einen PKW gemeint.

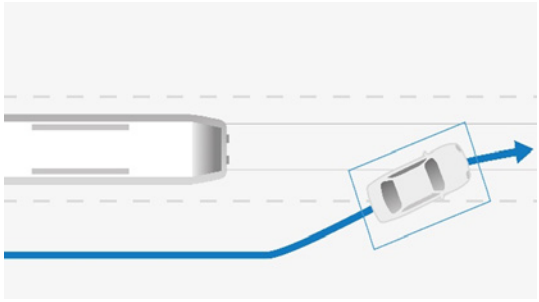

Abb. 2: Darstellung eines Close-Cut-Ins [2]

Um Aussagen über den Mehrwert der Verwendung von LiDAR zur Erkennung von CCI-Manövern treffen zu können, soll ein Proof of Concept erstellt werden. Hierfür soll eine Verarbeitungspipeline konzeptioniert und umgesetzt werden, sodass auf dieser Arbeit aufbauend mögliche Kollisionswarnstrategien untersucht werden können. Die Auswertung von öffentlichem Videomaterial zeigt, dass Kollisionen mit Linksabbiegern überwiegend bei geradem Schienenverlauf auftreten. Daher betrachtet diese Arbeit auf diesen kritischen

Bereich.

Datengenerierung

Realistische LiDAR Daten für Cut-In-Manöver insbesondere Close-Cut-Ins während der Fahrt können aufgrund der Gefahrensituation nicht nachgestellt und aufgenommen werden. Daher wurde für eine erste Analyse synthetische Daten für die Entwicklung der Pipeline generiert. Diese wurden anhand von Blender, einer Open-Source 3D-Grafiksoftware, erstellt.

Konzeptionierung

Für die Detektion von CCI Manövern wurde eine Pipeline aus drei wesentlichen Komponenten, wie in Abb. 3 dargestellt, entworfen. Die erste Komponente (A) ist für die Erkennung der Fahrzeuge in der unorganisierten Punktwolke zuständig. Hierbei soll ein neuronales Netz für die Verarbeitung der Punktwolke verwendet werden. Aus diesem sollen sogenannte Bounding Boxen zu detektierten Fahrzeugen resultieren. Bounding Boxen geben die Maße, Position und Maße von erkannten Fahrzeugen an. Anschließend werden die detektierten Fahrzeuge mit Hilfe eines Tracking-Algorithmus verfolgt (B). Daraus können wichtige Informationen über die Bewegung der Fahrzeuge gewonnen werden. Diese werden anschließend anhand von heuristischen Regeln auf Einscherverhalten geprüft (C).

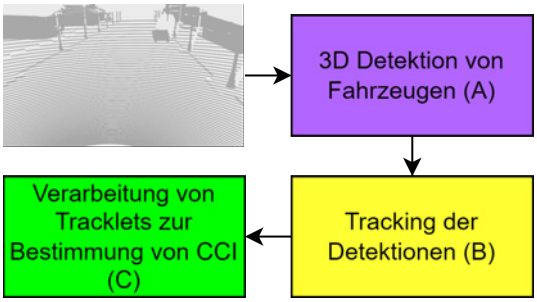

Abb. 3: Verarbeitungspipeline für die Bestimmung von CCIs [2]

Der Vorteil dieses modularen Systemaufbaus ist es, dass einzelne Module wie die Detektion sehr effizient verändert und verbessert werden können, da nur die Anforderungen an die Schnittstellen eingehalten werden müssen.

Detektion

Für die Detektion wurde das von Livox öffentlich zur Verfügung gestellte vortrainierte neuronale Netz verwendet. Dieser Detektor wurde von der ankerfreien CenterPoint Methode, wie in dem Paper [7] vorgestellt, inspiriert. Der Vorteil dieses Netzes ist, dass es sowohl in der Lage ist, gute Ergebnisse bei der Verarbeitung der synthetischen Daten zu erzielen als auch mit Realdaten des Livox HAP LiDARs sehr gute Ergebnisse zeigt. Aufgrund von fehlenden Trainingsdaten wurde auf eine Anpassung der Gewichte des neuronalen Netzes verzichtet. Das Ergebnis des Detektionsmoduls ist eine Menge von Detektionen, welche als Tupel gespeichert werden (x,y,z, theta, l,w,h, score).

Tracking

Ziel war es das Verfolgen von mehreren detektierten Fahrzeuge zu ermöglichen (Multiple Object Tracking). Zu diesem Zweck wurde ein 3D-Kalman-Filter mit einem konstanten Geschwindigkeitsmodell, das unabhängig von der Eigengeschwindigkeit ist, für die Zustandsschätzung verwendet. Für die Datenassoziation der Trajektorien und Detektionen wurde die Ungarischen Methode, wie in dem Paper [6] vorgeschlagen, verwendet. Dieser Ansatz ermöglicht eine sehr schnelle Datenverarbeitung von bis zu 207,4 FPS [6]. Für die Prädiktion des Zustandes wird ein Vector bestehend aus (x, y, z, theta, l, w, h, score ,vx, vy, vz) verwendet. Können prädizierte Trajektorien mit Detektionen assoziiert werden, werden diese pro Frame an das CCI Teilmodul weitergegeben.

Close-Cut-In Erkennung

Für die Erkennung von CCIs wurde ein heuristischer Ansatz gewählt. Hierbei wurden Regeln festgelegt, welche Hinweis auf ein einscherendes Verhalten geben. Dieses Erkennungsmodul beurteilt anhand von Schwellwerten für jedes stabil verfolgte Fahrzeug, ob eine relevante laterale Geschwindigkeit vorliegt, sich das Fahrzeug im unmittelbaren Gefahrenbereich aufhält, eine Mindestbewegung hat oder eine relevante Winkelveränderung besteht.

Ausblick

Die vorgestellte Pipeline wird auf ihre Perfomance untersucht, um weitere Verbesserungen der Linksabbiegerekennung zu erreichen. Für die Evaluierung von Detektion und Tracking kommt die CLEAR MOT Metrik zur Anwendung. Diese hat den Vorteil, anhand weniger Metriken die Perfomance von Algorithmen abschätzen zu können. Die CCI Detektion wird diese anhand der prozentualen zeitlichen Überschneidung von Ground Truth und angezeigter Detektion bewertet. Des Weiteren bietet die modulare Verarbeitungspipeline eine Basis in Zukunft weiterentwickelt zu werden und Untersuchungen zur Unfallverminderung durch Kollisionswarnstrategien zu ermöglichen.

Literatur und Abbildungen

[1] P. Annurag and A. Srinivaas. Advancements and Applications of LiDAR Technology in the Modern World: A Comprehensive Review. In *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–5. IEEE, 2023.

[2] Eigene Darstellung.

[3] J. Käsebier. Fahrerassistenzsysteme sorgen für mehr Sicherheit. *Arbeit & Gesundheit*, 2021.

[4] Polizeipäsidium Stuttgart. Unfallberichte. https://polizeiberichte-stuttgart.de/suche/Unfall, 2023.

[5] R. Thomas. Achtung Straßenbahn-Unfall! Verletzungsgefahr durch Schienenfahrzeuge. *Mobilitätsmagazin*, 2016.

[6] X. Weng, J. Wang, D. Held, and K. Kitani. 3D Multi-Object Tracking: A Baseline and New Evaluation Metrics. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020.

[7] T. Yin, X. Zhou, and P. Krähenbühl. Center-based 3D Object Detection and Tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021*. arXiv, 2021.

Entwicklung eines Multi-Task Learning Modells: Eine Integration von verschiedenen Methoden der Textanalyse

Pavithra Sureshkumar, Ralf Zeller **Gabriele Gühring**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma IT-Designers GmbH, Esslingen

Die Begriffe Künstliche Intelligenz (KI) und maschinelles Lernen (ML) gewinnen zunehmend an Bedeutung in verschiedenen Bereichen der Informatik, insbesondere wird dabei das Problem der Verarbeitung **natürlicher Sprache *(NLP = Natural Language Processing)*** angesprochen. Die Sprache, so wie wir sie kennen, ist selbstverständlich und kontextbasiert. Eine Maschine kann jedoch keinen Kontext wie wir Menschen erfassen und den Zusammenhang von Wörtern in einem Satz verstehen. Deswegen werden verschiedene Ansätze ermöglicht, damit die Maschine das, was wir als "Menschenverstand"verstehen, nachahmt. Die verschiedenen Sprachverarbeitungsmethoden in NLP ermöglichen es, große Mengen an Textdaten effizient und genau zu analysieren, daraus wertvolle Erkenntnisse zu gewinnen und den Kontext so nah wie möglich dem Modell beizubringen. Jedoch ist es heutzutage nicht mehr notwendig, einen großen Datensatz zu erstellen, sondern kann auf vortrainierte *(pre-trained)* Modelle zugreifen und kann diese auf die spezifische Aufgabe anpassen *(fine-tuning)*. In Abbildung 1 ist der Vorgang des *Fine-Tunings* eines *pre-trained* Modells dargestellt.

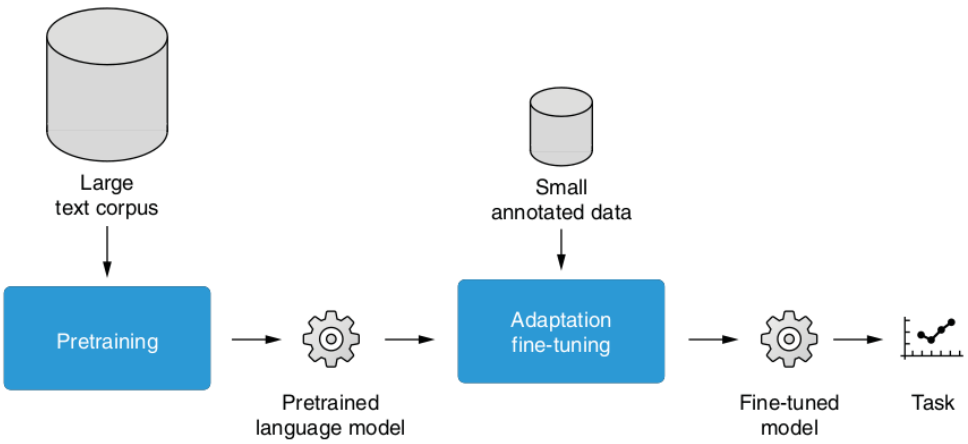

Abb. 1: Fine-tuning von einem pre-trained Modell [3]

Inhalt der Arbeit

Im Bereich der Anforderungsanalyse sind präzise und automatisierte Methoden zur Bewertung und Klassifikation von Anforderungssätzen von entscheidender Bedeutung um die Qualität von Softwareprojekten sicherzustellen. Die Hauptaufgabe einer Anforderungsanalyse ist es, die Genauigkeit und die Vereinheitlichung von den Anforderungen zu behalten, sodass keine Fehler bei der Softwareproduktion entstehen.
Diese Arbeit befasst sich mit der Entwicklung eines Modells für **Multi-Task Learning (MTL)**, das die Aufgaben eines **Multi-Label Text Klassifikation (MLTC)** Modells und eines **Named Entity Recognition (NER)** Modells vereinen soll.
Hierfür wird ein MLTC-Modell entwickelt, dass Anforderungssätze auf die Syntax der SOPHISTen-Anforderungsschablone und die Erkennung von Grammatikfehler prüft. Zusätzlich wird ein NER-Modell entwickelt, dass Anforderungen nach der detaillierten Anforderungsschablone der SOPHISTen taggt (siehe Abbildung 2).

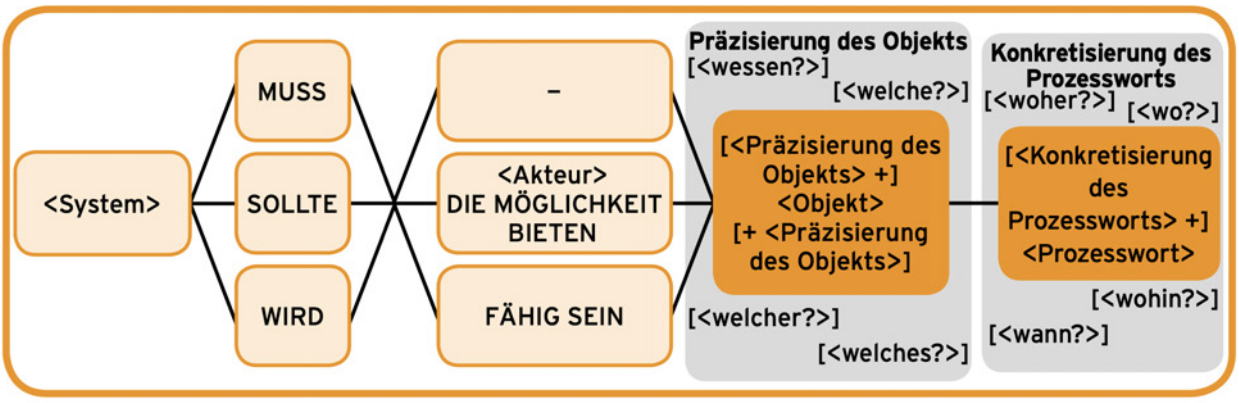

Abb. 2: Der detaillierte FunktionsMASTER [6]

Die SOPHIST-Anforderungen basieren auf der detaillierten Anforderungsschablone der SOPHISTEN, welche in der Anforderungsanalyse weit verbreitet ist (siehe Abbildung 2).

Transformer-basiertes Modell DeBERTa

Das Modell **BERT** ***(Bidirectional Encoder Representations from Transformers)*** wurde veröffentlicht, um die Vorgehensweise *fine-tuning* von *pretrained* Modellen zu verbessern [2]. Nach der Veröffentlichung von BERT wurden weitere BERT-ähnliche Modelle wie RoBERTa veröffentlicht, die die Performanz und Effizienz der BERT-Modell Architektur verbessern. Mit der Einführung von **DeBERTa** ***(Decoding-enhanced BERT with disentangled attention)*** wurde ein besseres Ergebnis mit dem Modell in der SuperGLUE Aufgabe erreicht als es mit dem *Baseline* Ergebnis für die menschliche Leistungsfähigkeit möglich war [4].
Für die Umsetzung des Modells wird hauptsächlich der Transformer DeBERTa verwendet, welcher durch seine fortschrittliche Architektur und leistungsstarke Sprachmodellierung herausragende Ergebnisse in der natürlichen Sprachverarbeitung erzielt. DeBERTa ist eines der modernsten Encoder-Transformer-Modelle in der heutigen Zeit und bietet sich für MLTC als ein guter Kandidat an. DeBERTa benutzt zwei besondere Techniken [4]:

1. ***disentangled attention mechanism***
2. ***enhanced mask decoder***

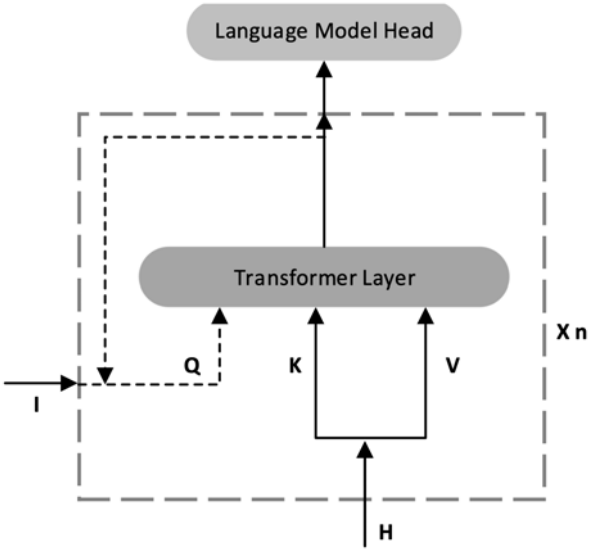

Abb. 3: Der enhanced masked decoder [4]

Disentangled attention mechanism stellt jedes Wort durch zwei Vektoren dar, die jeweils seinen Inhalt und seine Position kodieren. Die *attention* Gewichte zwischen den Wörtern werden dann mithilfe entkoppelter Matrizen berechnet, die sich jeweils auf die Inhalte und die relativen Positionen der Wörter beziehen. Jedes Wort im Satz hat zwei Informationsarten: den Inhalt und die Position des Wortes im Satz. Diese werden getrennt verarbeitet, um die Beziehung zwischen den Wörtern besser zu verstehen und um den Kontext zu erfassen [4].
Zweitens wird ein ***enhanced mask decoder*** verwendet, welcher in Abbildung 3 abgebildet ist. Dieser bezieht die absoluten Positionen in der Decoderebene ein, um die maskierten Tokens im *pre-training* des Modells vorherzusagen. Dabei wird nicht nur die relative Position, sondern auch die absolute Position ermittelt. Der Vorteil besteht darin, dass es genauere Vorhersagen der maskierten Felder beim Training des

Modells ermöglicht [4].
Darüber hinaus wird eine neue Methode des *virtual adversarial trainings* zur Feinabstimmung verwendet, um die Generalisierung der Modelle zu verbessern. Dabei wird das Modell robuster gemacht, um auf neue und unbekannte Daten zu reagieren [4]. Diese Techniken verbessern sowohl die Effizienz des *pretrainings* als auch die Leistung bei Aufgaben des *natural language understanding (NLU)* und der *natural language generation (NLG)* erheblich.
Man kann also mit DeBERTa und seinen Techniken nicht nur Sprache verstehen, sondern durch die Decodiertechnik auch Sprache erzeugen.
Durch die Anwendung von MLTC können für jeden Anforderungssatz mehrere Labels gleichzeitig vorhergesagt werden, was eine umfassende und differenzierte Analyse ermöglicht [5]. Die Ergebnisse der Experimente zeigen, dass der Einsatz von DeBERTa die Erkennungsgenauigkeit sowohl für SOPHIST-Anforderungen als auch für Grammatikfehler signifikant verbessert. Das Modell erreicht hohe Präzisions- und Recall-Werte und übertrifft damit herkömmliche Ansätze zur Anforderungsklassifikation und Grammatikprüfung. Als Vergleich wurden die Transformer-Modelle: BERT und RoBERTa, und andere getestet [4].

Der selbsterstellte Datensatz

MLTC wurde auf einem kleinen, selbsterstellten Datensatz von Anforderungssätzen angepasst. Da ein Transformer (hier DeBERTa) verwendet wird und dieser bereits auf einem großen Korpus vortrainiert wurde, ist die Größe des Datensatzes weniger entscheidend (siehe Abbildung 1). Der Datensatz ist eine Mischung aus Daten, die entweder keine SOPHIST-Anforderungen darstellen und keine Grammatikfehler enthalten, oder das Gegenteil zeigen. Es gibt auch Kombinationen von Daten, die keine SOPHIST-Anforderungen darstellen, aber Grammatikfehler enthalten, und umgekehrt. Die Grammatik wird nicht in einzelne Fehlerkategorien unterteilt, wie zum Beispiel ein Label für fehlende Wörter oder ein Label für falsche Zeitformen, sondern der Grammatikfehler wird nur als ein Label repräsentiert. Entweder ist also die Grammatik richtig oder falsch. Die Labels einer MLTC variieren von 0 bis 1. Als Beispiel, wenn das Label *is_grammar_correct* heißt und einen Wert von etwa 0.8 hat, ist die Grammatik mit hoher Wahrscheinlichkeit korrekt, da der Wert nahe bei 1 liegt. In Abbildung 4 ist ein Demo-Server zur Veranschaulichung des Modells dargestellt. Es wird für jedes Label ein eigener Score angezeigt. Der dargestellte Beispieltext lautet: ***"If a bell rings at midnight, the bell should indicate that the cats get their food on the blue plate."*** In diesem Fall handelt es sich, um einen SOPHIST-Anforderungssatz mit korrekter Grammatik, weil beide Label einen Score von etwas über 0.9 haben.

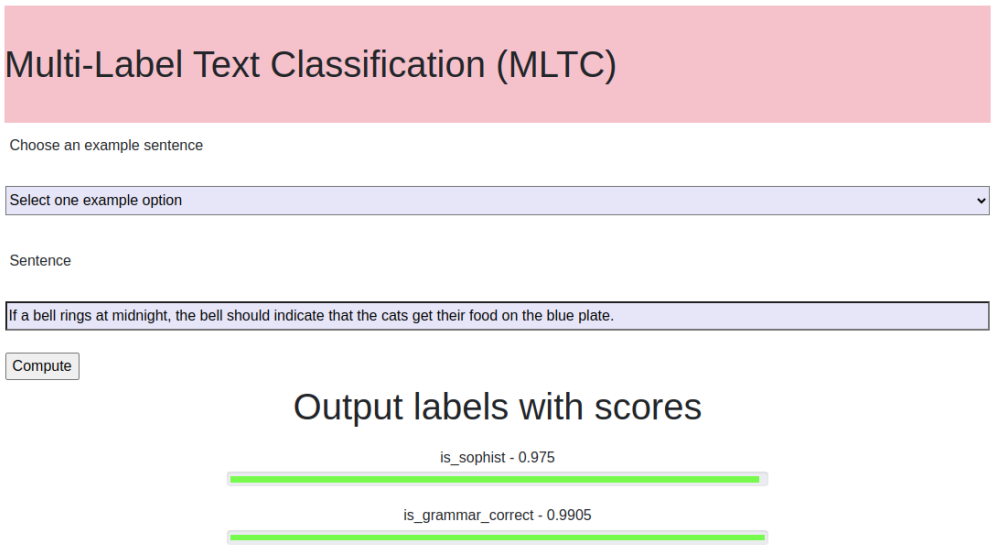

Abb. 4: Demo Multilabel Text Klassifikation [1]

Unterschied zwischen Binärer Klassifikation, Multiclass-Klassifikation, Multilabel-Klassifikation

In Abbildung 5 ist ein einfaches Beispiel für verschiedene Arten von Textklassifikationen in der Bildverarbeitung dargestellt. Bei einer **binären Klassifikation** wird nur eine Klasse behandelt. In diesem Beispiel erkennt das Modell, ob sich eine Katze auf dem Bild befindet oder nicht. Bei dem Bild in der

Mitte wird die **Multiclass-Klassifikation** dargestellt. Es werden zwar mehrere Klassen gegeben, aber als Ausgabe erhält man nur eine Klasse. Hier erkennt das Modell sowohl die Katze als auch den Hund, gibt aber nur eine einzige Ausgabe aus. Um mehrere mögliche Klassen als Ausgaben zu erhalten, wird die **Multilabel-Klassifikation (MLC)** verwendet. In diesem Beispiel gibt das Modell nicht nur die Katze als Klasse aus, sondern auch den Hund.

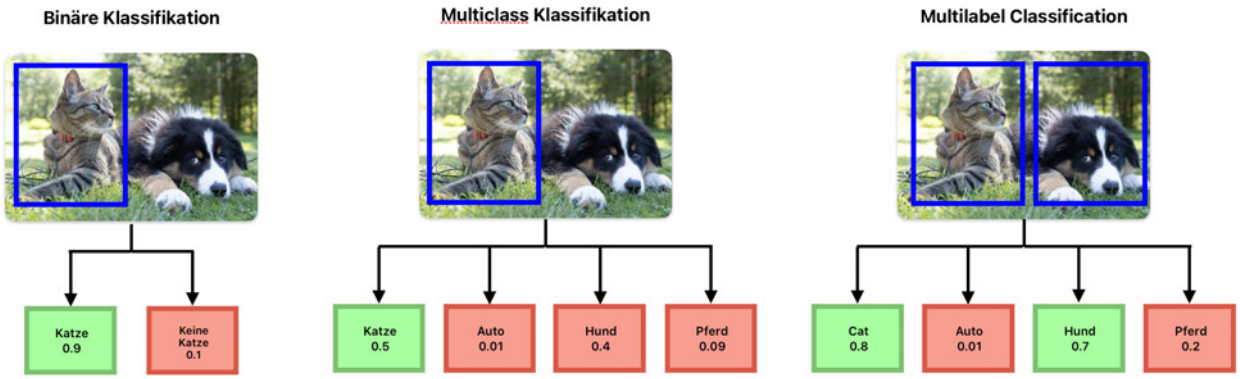

Abb. 5: Vergleich Multilabel Klassifikation in der Bilderkennung [1]

Multi-Task Modell

Beim MLT-Modell werden das NER-Modell und das MLTC-Modell in ein gemeinsames Modell zusammengeführt. Mit dem NER-Modell werden bei der Analyse von Anforderungssätzen, Tags für die Teile der Sophistenschablone 2) erzeugt. Das MLTC-Modell erzeugt die Labels *is_sophist* und *is_grammer_correct*. Mit den beiden Labeln kann man prüfen, ob eine Anforderung grammatikalisch richtig ist und ein Satz die Syntax der SOPHISTen-Schablone hat. Beide Modelle werden als eigenständige Modelle auf Basis der DeBERTa-Architektur [4] entwickelt.
Durch die Kombination der Modelle soll bei der Analyse eine erweiterte Analyse von Anforderungen ermöglicht werden. Das NER-Modell ist beschränkt auf das Taggen von aktiven Anforderungssätzen und gibt nicht aussagekräftige Ergebnisse für passive Sätze. Hierbei kann das MLT-Modell bei der Ausgabe der Analyse frühzeitig zeigen, ob die Anforderung nach der SOPHISTen-Schablone geschrieben und die Grammatik der Anforderung richtig ist.

Ausblick

Ziel dieser Arbeit ist die **Analyse von Anforderungssätzen** und die Verbesserung der **Bewertung von Anforderungen**, um die Genauigkeit bei Softwareprojekten zu erhöhen. Dies trägt dazu bei, die Anforderungsanalyse zu automatisieren und bildet eine Grundlage für weitere Forschungen in diesem Bereich sowie praktische Anwendungen. Die Integration fortschrittlicher NLP-Technologien wie DeBERTa zeigt das Potenzial, die Effizienz und Genauigkeit der Anforderungsbewertung signifikant zu steigern. Um das MLTC-Modell weiter zu verbessern, könnten zusätzliche Labels (wie zum Beispiel passive Anforderungen und Qualität) sowie weitere Daten in den Datensatz integriert werden. Zudem könnte die Optimierung der Modellparameter angepasst werden, um eine höhere Präzision der Label-Scores zu erreichen. Beim MLT-Modell können weitere Modelle hinzugefügt werden, um neue Features für die Anforderungsanalyse zu ermöglichen. Ein Beispiel ist Masked-Language Modeling, um für das Modalverb im Anforderungssatz Alternativen anzubieten.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*, pages 4171–4186. Association for Computational Linguistics, 2019.

[3] M. Hagiwara. *Real-World Natural Language Processing: Practical applications with deep learning*. Manning, 2021.

[4] Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Wei Chen. DeBERTa: Decoding-Enhanced BERT with Disentangled Attention. In *International Conference on Learning Representations*. ICLR, 2021.

[5] F. Herrera, F. Charte, A.J. Rivera, and M.J. Jesus. *Multilabel Classification: Problem Analysis, Metrics and Techniques*. Springer International Publishing, 2016.

[6] GmbH SOPHIST. *Die SOPHISTen Schablonen für alle Fälle (engl.: Patterns for all Purposes)*. SOPHIST GmbH, 5 edition, 2019.

Ausarbeitung der notwendigen IT Capabilities zur optimalen Unterstützung des Data Quality Management

Hilal Tarhan **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma ETAS GmbH, Stuttgart-Feuerbach

Einleitung

In der heutigen Geschäftswelt spielen Daten eine zentrale Rolle für den Erfolg eines Unternehmens. Daten bilden die Grundlage für strategische Entscheidungen, operative Effizienz und Wettbewerbsvorteile. Daher ist die Qualität dieser Daten von entscheidender Bedeutung. Gute Datenqualität sorgt dafür, dass die Informationen, die ein Unternehmen nutzt, genau, konsistent und verlässlich sind. Dies ist besonders wichtig in Zeiten der Digitalisierung, in denen immer mehr Prozesse automatisiert und datengetrieben werden. Schlechte Datenqualität hingegen kann zu Fehlentscheidungen, ineffizienten Prozessen und erheblichen Kosten führen. Das Management der Datenqualität stellt Unternehmen jedoch vor erhebliche Herausforderungen. Daten stammen aus verschiedenen Quellen, sind oft unterschiedlich strukturiert und unterliegen ständigen Veränderungen. Um diesen Herausforderungen zu begegnen, ist es unerlässlich, effektive Strategien und Technologien zur Sicherstellung der Datenqualität zu entwickeln und umzusetzen.

Ziel der Arbeit

Das Hauptziel dieser Arbeit ist es, die Vorteile und die Bedeutung einer guten Datenqualität zu verdeutlichen sowie die Herausforderungen und Probleme im Bereich des Datenqualitätsmanagements im Unternehmen aufzuzeigen. Weiterhin sollen die Anforderungen an die Datenqualität definiert und die Methoden sowie Prozesse des Datenqualitätsmanagements erläutert werden. Anhand dieser Informationen werden die erforderlichen IT-Fähigkeiten ermittelt, um das Datenqualitätsmanagement der Etas GmbH zu optimieren. Ein zentraler Bestandteil der praktischen Arbeit ist die Identifizierung und Testung geeigneter technologischer IT-Tools. Da heutzutage viele Prozesse automatisiert werden und es Tools gibt, die Zeit und Arbeit sparen, wiederholt eingesetzt werden können und leicht bedienbar sind, ist dies ein wichtiger Aspekt im Kontext der IT-Capabilities. Zunächst werden die derzeit bei der Bosch Group existierenden Tools, wie z.B. SAP, getestet und gegebenenfalls erweitert, um die Anforderungen an die Datenqualität zu erfüllen. Danach werden externe Datenqualitätsmanagement-Tools anhand von Datenqualitätskriterien und -zielen sowie den ermittelten IT-Fähigkeiten der Etas GmbH untersucht, ein Unternehmen der Bosch Group, das sich auf innovative Lösungen im Bereich der Automobiltechnologie spezialisiert hat. Die ermittelten Tools werden miteinander verglichen und getestet, um festzustellen, welche Tools benötigt werden um das Datenqualitätsmanagement zu unterstützen. Die am besten geeigneten Tools werden schließlich für die Etas GmbH zur Implementierung vorgeschlagen.

Datenqualität im Datenmanagement

Abb. 1: DAMA-DMBOK2 Data Management Framework [2]

Datenqualität ist ein wesentlicher Aspekt des Data Managements, das als die praktische Anwendung von Data Governance betrachtet wird (siehe Abbildung 1). Der Unterschied besteht darin, dass Data Governance die Ziele, Prozesse, Organisationsstrukturen und

Richtlinien für die Datenverwaltung definiert, während Data Management für die Implementierung dieser Rahmenbedingungen verantwortlich ist. Eine wichtige Unterscheidung ist, dass sich die Datenqualität oft nur auf die Struktur und den Inhalt der Daten bezieht und nicht auf die Qualität der Datenprozesse. Da jedoch die Qualität der Datenprozesse grundlegend für alle Aktivitäten im Bereich des Data Managements ist, muss sie im Kontext der Datenqualität berücksichtigt werden [6].

Bedeutung von Datenqualität und Dimensionen

Datenqualität bezieht sich sowohl auf die Eigenschaften, die hochwertige Daten auszeichnen, als auch auf die Prozesse, die zur Messung oder Verbesserung dieser Qualität eingesetzt werden. Daten werden als qualitativ hochwertig angesehen, wenn sie den Erwartungen und Bedürfnissen der Datennutzer entsprechen und für die vorgesehenen Zwecke geeignet sind. Umgekehrt gelten Daten als geringe Qualität, wenn sie diese Anforderungen nicht erfüllen. Die Qualität der Daten ist daher kontextabhängig und orientiert sich an den spezifischen Bedürfnissen der Datennutzer. Die Kontextabhängigkeit bedeutet, dass Datenqualität für einen bestimmten Geschäftsvorfall ausreichend sein kann, während sie für einen anderen ungenügend ist. Da Daten nur ein Abbild der Realität sind und sich diese Realität ständig verändert, ändert sich auch die Datenqualität über die Zeit. Es gibt kein einziges Merkmal, das die Datenqualität vollständig beschreibt. Stattdessen gibt es verschiedene Datenqualitätsdimensionen, die zusammen die Qualität der Daten ausmachen. Der Begriff Dimension dient als Analogie zu den Dimensionen physischer Objekte (wie Länge, Breite, Höhe). Datenqualitätsdimensionen (siehe Abbildung 2) schaffen ein Vokabular zur Definition von Datenqualitätsanforderungen. Dadurch werden sowohl die Ergebnisse der anfänglichen Datenqualitätsbewertung als auch die kontinuierlichen Messungen festgelegt. Um die Qualität von Daten effektiv zu messen, muss eine Organisation Eigenschaften identifizieren, die sowohl für die Geschäftsprozesse relevant als auch messbar sind [5].

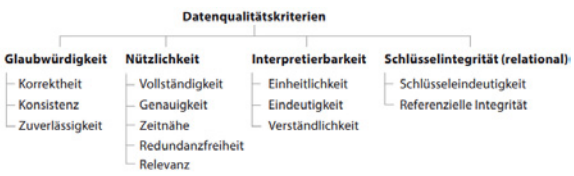

Abb. 2: Übersicht von Datenqualitätsdimensionen [1]

Datenqualitätsmanagement

Das Datenqualitätsmanagement (DQM) umfasst die Analyse, Verbesserung und Sicherung der Datenqualität. Gemäß der Definition der Data Management Association (DAMA) bezieht sich DQM auf sämtliche Aktivitäten, Verfahren und Systeme, die unter Verwendung von Methoden des Qualitätsmanagements die Eignung der Daten zur Nutzung messen, verbessern und sichern. Zur Analyse und Verbesserung der Datenqualität stehen viele Methoden und Techniken zur Verfügung. Man unterscheidet grundsätzlich zwischen zwei Ansätzen: der reaktiven, nachträglichen Datenbereinigung, wird Data Cleansing genannt, und der proaktiven, präventiven Verbesserung der Datenqualität [4].

Bedeutung von IT Capabilities im DQM-Kontext

Eine IT Capability ist eine organisierte Sammlung spezifischer Personen, Prozesse, Informationen, Technologien und anderer Ressourcen, die zusammenarbeiten, um eine IT-Fähigkeit zu realisieren [3].

1. Unterstützung bei der Datenqualitätsstrategie IT Capabilities ermöglichen die Entwicklung und Umsetzung von Datenqualitätsstrategien, die die Identifizierung, Bewertung und Korrektur von Datenqualitätsproblemen systematisieren. Technologien wie Datenmanagement-Plattformen und spezialisierte Software unterstützen dabei, Datenqualitätsziele zu definieren, zu überwachen und zu erreichen.

2. Automatisierung von Datenqualitätsprozessen Moderne IT-Tools und -Systeme bieten Möglichkeiten zur Automatisierung von Prozessen, die die Datenqualität sicherstellen. Dies umfasst die automatische Erkennung und Bereinigung von Fehlern, die Validierung und Standardisierung von Daten sowie die Implementierung von Geschäftsregeln, die sicherstellen, dass eingehende Daten den festgelegten Qualitätsstandards entsprechen. Zu den Datenqualitätsmanagement-Tools gehören unter anderem Data Profiling, Data Cleansing und Data Monitoring [7].

3. Echtzeit-Datenüberwachung und -analyse IT Capabilities erlauben die kontinuierliche Überwachung und Analyse von Daten in Echtzeit. Dies ist entscheidend, um die Integrität und Genauigkeit der Daten laufend zu gewährleisten und schnell auf potenzielle Probleme reagieren zu können.

4. Integration heterogener Datenquellen In der heutigen datengetriebenen Welt, wo Organisationen Daten aus einer Vielzahl von Quellen sammeln, ermöglichen IT Capabilities die Integration und Konsolidierung dieser Daten. Dies stellt sicher, dass die Daten

konsistent, vollständig und korrekt sind, unabhängig von ihrem Ursprung.

5. Einhaltung von Datenschutz- und Sicherheitsstandards IT Capabilities unterstützen die Einhaltung von gesetzlichen und internen Datenschutz- und Sicherheitsvorschriften durch Bereitstellung von Werkzeugen und Technologien, die den sicheren Umgang mit Daten gewährleisten. Sie helfen, sensible Daten zu schützen und Vertrauen bei den Nutzern und Stakeholdern zu schaffen.

6. Verbesserung der Datenzugänglichkeit und -nutzbarkeit Durch den Einsatz von IT Capabilities wird die Zugänglichkeit und Nutzbarkeit der Daten verbessert, indem Daten in benutzerfreundlichen Formaten zur Verfügung gestellt und durch leistungsfähige Such- und Abfragefunktionen ergänzt werden. Dies steigern die Effizienz und Produktivität der Nutzer. Insgesamt spielen IT Capabilities eine zentrale Rolle im DQM-Kontext, indem sie technologische Lösungen und Systeme bereitstellen, die eine hohe Datenqualität über den gesamten Datenlebenszyklus sicherstellen. Dies führt zu verbesserten Geschäftsentscheidungen, erhöhter Kundenzufriedenheit und letztendlich zu einem Wettbewerbsvorteil [2].

Literatur und Abbildungen

[1] Detlef Apel et al. *Datenqualität erfolgreich steuern: Praxislösungen für Business Intelligence-Projekte*. dpunkt.verlag, 3 edition, 2015.

[2] Susan Early et al. *DAMA-DMBOK: data management body of knowledge*. Technics Publication, 2 edition, 2017.

[3] Marc Gewertz. *DEFINING ENTERPRISE; A Systems View of Capability Management*. Marc H. Gewertz, 1 edition, 2016.

[4] Boris Otto and Hubert Österle. *Corporate Data Quality*. Springer Berlin Heidelberg, 2016.

[5] Laura Sebastian-Coleman. *Measuring data quality for ongoing improvement: a data quality assessment framework*. Morgan Kaufmann, 2013.

[6] Kilian Semmelmann. Die Bedeutung von Datenqualität für Unternehmen. https://datadrivencompany.de/die-bedeutung-von-datenqualitaet-fuer-unternehmen/, 07 2020.

[7] Kristin Weber and Christiana Klingenberg. *Data Governance: der Leitfaden für die Praxis*. Hanser, 2021.

Navigieren im Wandel: Ein Modell für effektives Change Management unter Berücksichtigung der kulturellen Etwicklung und Erfolgsfaktoren in Unternehmen

Maik Tobias **Catharina Kriegbaum-Kling**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Adesso SE, Stuttgart Vaihingen

Einleitung

Die Stärke, Komplexität und Unberechenbarkeit von Veränderungen haben seit der Zeit vor der Industriellen Revolution kontinuierlich zugenommen. Dieser Trend hat sich mit dem Übergang vom Industrie- zum Informationszeitalter weiter beschleunigt. Beispiele hierfür sind die rapide Zunahme der erteilten Patente oder die schnelle Verbreitung neuer Technologien wie Smartphones oder das Internet. In der heutigen Zeit führen Digitalisierung, Elektroautos, Energiewende und Industrie 4.0 sowie soziale und politische Bewegungen zu stetigen Veränderungen, die in die Organisationsabläufe integriert werden müssen. Damit einhergehend verändert sich das Anforderungsprofil sowohl an Unternehmen als auch an die Mitarbeiter. [4] [2]

Problemstellung

Unternehmen müssen in der Lage sein, sich schnell und effektiv anzupassen, um wettbewerbsfähig zu bleiben. Dies erfordert ein effizientes Change Management, das die spezifischen kulturellen Gegebenheiten des Unternehmens berücksichtigt. Ein Grund für das Scheitern vieler Change Projekte könnte in der unzureichenden Berücksichtigung kultureller Faktoren liegen. Dazu gehören beispielsweise die Führungskultur, Kommunikationswege und die allgemeine Bereitschaft der Mitarbeiter, Veränderungen zu akzeptieren und zu unterstützen. Darüber hinaus stellt sich die Frage, wie Maßnahmen und die nächsten Schritte im Change Prozess sinnvoll priorisiert werden können, um maximale Effizienz und Effektivität zu gewährleisten.

Change Management

Das Change Management arbeitet darauf hin, eine agile Organisation zu entwickeln, in der eine zielorientierte und wertschätzende Kultur zu nachhaltiger Wertschöpfung führt. Dabei müssen sowohl der Übergang von einem unbefriedigenden Zustand zu gemeinsam getragenen Ergebnissen als auch die individuellen Lernprozesse unterstützt und gefördert werden. [2] Ein klassisches Modell des Change Management ist das 8 Phasen Modell nach Kotter, in dem beschrieben wird, wie in 8 Schritten ein erfolgreicher Change realisiert werden kann.

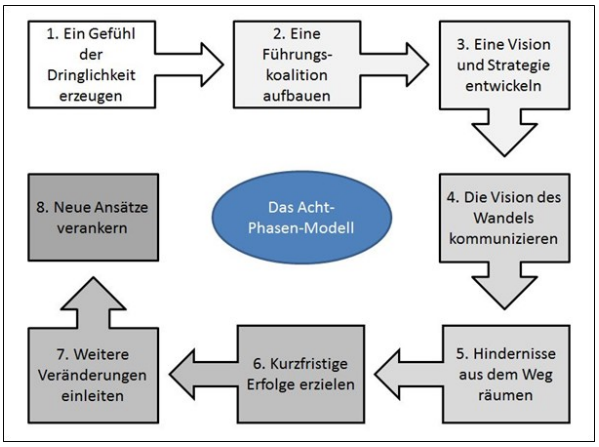

Abb. 1: Das Acht Phasen Modell [3]

Dringlichkeit erzeugen: In dieser Stufe wird ein starkes Dringlichkeitsgefühl bei den Mitarbeitern erzeugt, um eine große Chance der Organisation wahrzunehmen. Führungskoalition aufbauen: Das Dringlichkeitsgefühl wird genutzt, um ein Netzwerk aus motivierten Mitarbeitern aus allen Organisationsebenen zu bilden, die den Wandel vorantreiben wollen. Strategische Vision formulieren: Die Vision und strategische Initiativen richten die Organisation schnell und flexibel auf eine große Chance aus. Freiwillige mobilisieren: Mitarbeiter werden über die Vision und Initiativen informiert, um möglichst viele zur Unterstützung zu bewegen. Barrieren abbauen: Hindernisse, die strategisch wich-

tige Aktivitäten verlangsamen oder stoppen, werden identifiziert und beseitigt. Kurzfristige Erfolge erzielen: Sichtbare Erfolge werden geschaffen und gefeiert, um die Motivation und Glaubwürdigkeit zu fördern. Kontinuierlich weitermachen: Das System wird in Bewegung gehalten, indem ständig neue Gelegenheiten und Herausforderungen angegangen werden. Wandel fest etablieren: Die Erfolge werden dauerhaft in die Prozesse und Kultur der Organisation integriert, um die Veränderungen langfristig zu verankern. [3]

Unternehmenskultur

Jeder Mensch hat Denk-, Fühl- und Handlungsmuster, die er im Laufe seines Lebens erlernt hat, vor allem in der frühen Kindheit, da er in dieser Zeit am empfänglichsten für Lernprozesse ist. Diese Muster müssen erst abgelegt werden, bevor Neues gelernt werden kann, was schwieriger ist als das ursprüngliche Lernen. Die Quellen dieser Muster liegen im sozialen Umfeld, in dem wir aufgewachsen sind, wie Familie, Nachbarschaft, Schule, Jugendgruppen, Arbeitsplatz und Partnerschaft. In der Sozialanthropologie umfasst der Begriff 'Kultur' alle diese Denk-, Fühl- und Handlungsmuster, einschließlich alltäglicher und grundlegender Dinge wie Grüßen, Essen und das Zeigen von Gefühlen. Kultur ist ein kollektives Phänomen, da sie von Menschen geteilt wird, die im selben sozialen Umfeld leben. Sie besteht aus ungeschriebenen Regeln des sozialen Spiels und ist eine kollektive Programmierung des Geistes, die Mitglieder einer Gruppe von anderen unterscheidet. Kultur wird erlernt und leitet sich aus dem sozialen Umfeld ab, nicht aus den Genen. Organisationskultur ist folglich die kollektive Programmierung des Geistes, die die Mitglieder einer Organisation von anderen unterscheidet. Diese Kultur wird nicht nur im Kopf der Mitglieder, sondern auch bei anderen Beteiligten wie Kunden, Lieferanten und Behörden aufrechterhalten. [1] Die Organisationskultur zeichnet sich nach Hofstede durch sechs Kulturdimensionen aus.

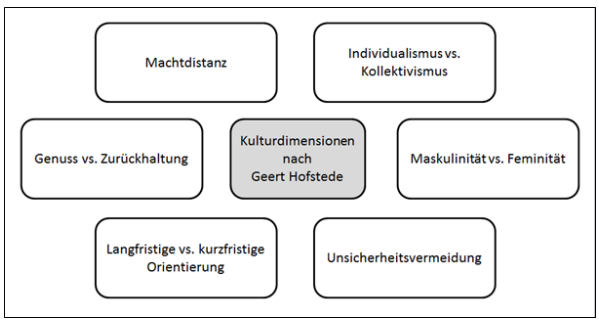

Abb. 2: Die Kulturdimensionen nach Hofstede [1]

Machtdistanz: Beschreibt die Akzeptanz von Hierarchie. Hohe Machtdistanz bedeutet Akzeptanz von Hierarchie, geringe Machtdistanz zeigt den Wunsch nach Gleichheit. Individualismus vs. Kollektivismus: Definiert, ob individuelle Selbstbestimmung oder kollektive Integration wichtiger ist. Hohe Werte stehen für individuelle Selbstbestimmung, niedrige Werte für ein starkes Wir-Gefühl. Maskulinität vs. Femininität: Charakterisiert das vorherrschende Wertesystem. Niedrige Werte stehen für feminine Werte wie Kooperation und Fürsorglichkeit, hohe Werte für maskuline Werte wie Dominanz und Leistung. Unsicherheitsvermeidung: Erklärt, wie stark eine unsichere Zukunft als negativ empfunden wird. Hohe Werte bedeuten den Wunsch nach mehr Regeln und Sicherheit, niedrige Werte zeigen höhere Risikobereitschaft. Lang- oder kurzfristige Ausrichtung: Beschreibt die zeitliche Planung. Hohe Werte stehen für langfristige Planung und Beharrlichkeit, niedrige Werte für kurzfristige Flexibilität. Genuss vs. Zurückhaltung: Beschreibt, ob man den eigenen Wünschen nachgeht oder sie kontrolliert. Hohe Werte stehen für Nachgiebigkeit, niedrige Werte für Kontrolle. [5]

Ausblick

Im Rahmen der Arbeit werden Experteninterviews durchgeführt, um den im theoretischen Teil erarbeiteten Zusammenhang zwischen Change Management und Unternehmenskultur zu bestätigen. Die Ergebnisse der Interviews sollen die enge Verknüpfung zwischen diesen beiden Faktoren bestätigen und Aufschluss über mögliche Erfolgsfaktoren geben. Auf Basis der Erkenntnisse soll ein Modell für effektives Change Management entwickelt werden, das die Aspekte der Unternehmenskultur berücksichtigt. Durch die Berücksichtigung kultureller Aspekte und die Identifikation von Erfolgsfaktoren soll das Change Management optimiert und an die spezifischen Bedürfnisse und Herausforderungen der jeweiligen Ausgangssituation angepasst werden. Das Modell soll neue Perspektiven für die Praxis eröffnen und Unternehmen die Möglichkeit bieten, Veränderungsprozesse erfolgreicher zu gestalten. Zukünftige Forschungen könnten sich darauf konzentrieren, das Modell in verschiedenen organisatorischen Kontexten zu testen und um weitere Dimensionen zu erweitern.

Literatur und Abbildungen

[1] Geert Hofstede et al. *Lokales Denken, globales Handeln – interkulturelle Zusammenarbeit und globales Management*. dtv, 2017.

[2] Claudia Kostka. *Change Management – Wandel gestalten und durch Veränderungen führen*. Carl Hanser Verlag, 2017.

[3] John P Kotter. *Accelerate – Strategischen Herausforderungen schnell, agil und kreativ begegnen*. Franz Vahlen GmbH, 2015.

[4] John P Kotter, Vanessa Akhtar, and Gaurav Gupta. *Change – Wie Unternehmen in unbeständigen Zeiten herausragende Ergebnisse erzielen*. Wiley-VCH GmbH, 2022.

[5] Ian Towers and Alexander Peppler. *Geert Hofstede und die Dimensionen einer Kultur*. Springer Gabler, 2017.

Einsatz von Machine Learning zur automatisierten Anomalieerkennung in Systemlogs

Mustafa Salih Uenal **Gabriele Gühring**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Accenture Industry X, Böblingen

Einleitung

Die Digitalisierung ist für moderne Unternehmensstrategien unerlässlich. Unternehmen nutzen digitale Technologien, um wettbewerbsfähig zu bleiben und die Kundenzufriedenheit zu maximieren [4]. Auch die Automobilbranche transformiert sich digital durch mehr Elektrik, Elektronik und Software in Fahrzeugen [3]. Moderne Autos sind vernetzt, digital und haben zahlreiche Fahrerassistenzsysteme und Infotainmentsysteme mit Internetzugang.
Infotainmentsysteme sind dabei ein zunehmend beliebtes Feature in Fahrzeugen. Dabei handelt es sich um einen integrierten Computer in Fahrzeugen, das sowohl Informations- als auch Unterhaltungsfunktionen wie Navigation, Radio, Smartphone-Integration und Fahrerassistenzsysteme bietet.

Abb. 1: Videospiele auf Tesla-Infotainmentsystem (Shakir, 2024) [2]

Moderne, softwareintensive Fahrzeugsysteme bieten viele Möglichkeiten, erfordern jedoch eine sorgfältige Planung, Umsetzung und Überprüfung. Methoden und Werkzeuge der Softwaretechnik müssen spezifisch für die Automobilbranche angewandt werden, um Sicherheit- und Zuverlässigkeitsanforderungen zu erfüllen [3].
Accenture ist ein globales Dienstleistungsunternehmen, das eine breite Palette von Dienstleistungen und Lösungen in den Bereichen Strategie, Beratung, Digitalisierung und Technologie anbietet. Der Standort in Böblingen konzentriert sich dabei hauptsächlich auf die Prüfung von Steuergeräten und Fahrzeugfunktionen. Hier werden Tests durchgeführt, um sicherzustellen, dass die elektronischen Systeme und Softwarekomponenten in Fahrzeugen höchsten Qualitäts- und Sicherheitsstandards entsprechen.

Problembeschreibung und Zielsetzung

Ein wesentlicher Aspekt der modernen Fahrzeugtechnologie ist die Fähigkeit zur Diagnose und Fehlersuche. Dabei spielen die AUTOSAR Diagnostic Logs und Traces (DLT) eine zentrale Rolle. Diese Logs protokollieren umfassend die Abläufe und Ereignisse in den verschiedenen Systemen eines Fahrzeugs und sind daher ein wertvolles Werkzeug zur Fehleranalyse. Durch die steigende Komplexität der Softwaresysteme steigt aber auch die Anzahl der Logs stark an, was die manuelle Analyse deutlich zeitaufwändiger und fehleranfälliger macht. Traditionelle regelbasierte Algorithmen sind oft nicht in der Lage, die Vielfalt und Komplexität der Datenmuster effektiv zu bewältigen.
Machine Learning-Modelle hingegen können sich an neue Daten anpassen, sowie komplexe und subtile Muster erkennen, die traditionelle Algorithmen übersehen könnten. Die Anomalieerkennung ist dabei ein Teilbereich des Machine Learning, das dabei helfen kann, indem sie Muster und Anomalien in Logs identifiziert. Dies könnte bei der Erkennung potenzieller Fehler helfen und somit den Zeitaufwand reduzieren, sowie die Wartung vereinfachen.
Das Ziel der Arbeit ist es, geeignete Machine Learning-Verfahren zu identifizieren und zu implementieren, welche es ermöglichen, Fehlverhalten in Fahrzeugsystemen anhand der Logs effizient und zuverlässig zu erkennen.

Ansatz

Die Grundidee der Anomalie-Erkennung in Log-Daten besteht darin, das Modell so zu trainieren, dass es normales Verhalten von anormalem unterscheiden kann. In Abbildung 2 sind die hierfür notwendigen Schritte illustriert.

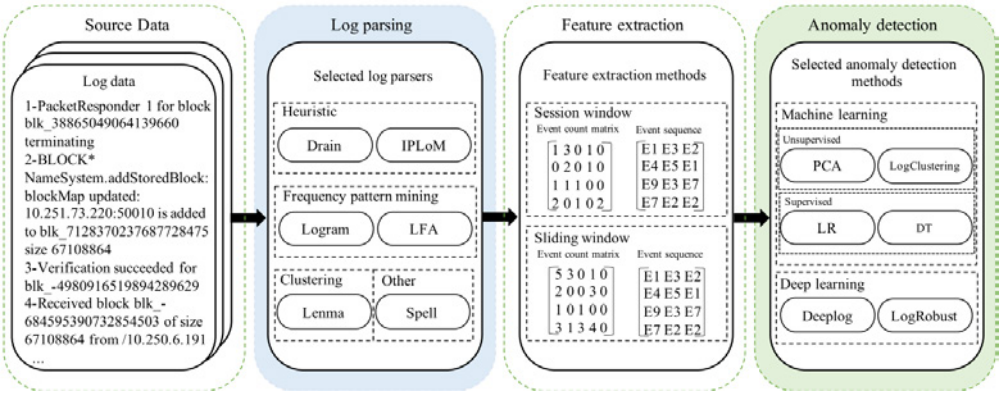

Abb. 2: Schritte zur Erkennung von Anomalien in Log-Daten [1]

1. *Log Parsing*: Log-Einträge bestehen aus halbstrukturierten Textdaten mit einem festen und einem variablen Teil. Der feste Teil wird im Quellcode definiert und bleibt unverändert, während der variable Teil dynamische Informationen wie Zeitstempel und Fehlercodes enthält. Das Ziel des Log Parsings ist es, den variablen Teil vom festen zu trennen. So wird aus dem in Abbildung 2 dargestellten Log-Eintrag »PacketResponder 1 for blk_388 terminating« das Template »PacketResponder <*> for <*> terminating« erstellt [1].
2. *Feature Extraction*: Dieser Prozess extrahiert relevante Daten aus den Logs und führt sie in ein Format über, das für Machine Learning-Modelle geeignet ist. Dafür werden Log-Sequenzen in »Windows« eingeteilt und eine Matrix erstellt, welche zählt, wie oft die Templates in den jeweiligen Log-Sequenzen auftreten [1].
3. *Anomaly Detection*: Nach der Extraktion der relevanten Features aus den Logs werden diese Daten genutzt, um Machine-Learning-Modelle zu trainieren. Während des Trainingsprozesses lernt das Modell, normale Muster und Verhaltensweisen in den Log-Daten zu erkennen. Sobald das Modell ausreichend trainiert ist, kann es auf neue, unbekannte Log-Daten angewendet werden, um zu bestimmen, ob diese eine Anomalie darstellen [1].

Ausblick

Die Anomalieerkennung in Log-Daten bietet vielversprechende Möglichkeiten, die Verfügbarkeit und Zuverlässigkeit von Software-Systemen zu verbessern. Im Kontext von AUTOSAR DLT wird die Anomalieerkennung verwendet, um ungewöhnliche Muster in den Log-Daten von Fahrzeugsteuergeräten zu identifizieren. Die bisherigen Schritte legen den Grundstein für ein robustes Überwachungssystem, das in der Lage ist, kontinuierlich die Systemperformance zu überwachen und Anomalien zu identifizieren.
Insgesamt zeigt die Anomalieerkennung ein großes Potenzial, die Fahrzeugdiagnose und -wartung zu revolutionieren und somit sicherere und zuverlässigere Fahrzeuge zu gewährleisten.

Literatur und Abbildungen

[1] Ying Fu et al. An empirical study of the impact of log parsers on the performance of log-based anomaly detection. *Empirical Software Engineering*, 2022.

[2] Umar Shakir. New Teslas might lose Steam. https://www.theverge.com/2024/5/17/24158929/tesla-steam-discontinue-new-model-x-delivery-s-cybertruck, 05 2024.

[3] Miroslaw Staron. *Automotive Software Architectures: An Introduction*. Springer, 2021.

[4] Uwe Winkelhake. *Die digitale Transformation der Automobilindustrie*. Springer-Verlag, 2024.

Growth Hacking - Implementierung und Anwendung innerhalb eines Start-ups

Curtis Walch **Anke Bez**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma pipappl UG, Esslingen

Growth Hacking: Ein innovativer Ansatz zur Förderung von Start-up-Wachstum

Einleitung In der modernen Geschäftswelt stehen Start-ups vor der Herausforderung, schnell und effizient zu wachsen, um in einem wettbewerbsintensiven Umfeld zu überleben. Growth Hacking hat sich als ein kostengünstiger und wirksamer Ansatz zur Förderung dieses Wachstums etabliert. [2] Diese Methodik kombiniert technisches Wissen, wie Programmierung und Automatisierung, kreative Marketingstrategien und eine datengetriebene Testkultur, siehe Abbildung 1, um schnell Erfolge zu erzielen und langfristiges Wachstum zu sichern. [4]

Abb. 1: Bereiche des Growth Hackings [3]

Theoretischer Hintergrund

Definition und Einordnung Growth Hacking, erstmals von Sean Ellis im Jahr 2010 geprägt, beschreibt eine Methodik, die auf schnellen, iterativen Experimenten basiert, um die gesamte Kundenreise zu optimieren. Diese Experimente zielen darauf ab, schnelle Erfolge (Quick Wins) zu erzielen, die durch kontinuierliche Verbesserung in signifikante Fortschritte (Big Steps) überführt werden können. [4]

Voraussetzungen und Herausforderungen Die Grundvoraussetzung für Growth Hacking ist die Erreichung des Product-Market-Fit (PMF). Ohne PMF können selbst die innovativsten Wachstumsstrategien langfristig scheitern. [2], [4] Eine der größten Herausforderungen im Growth Hacking besteht darin, relevante Daten zu identifizieren und zu interpretieren. Fehlinterpretationen können zum Scheitern von Growth Hacking-Initiativen führen. [2]

Umsetzung und Methoden Growth Hacking folgt einem kontinuierlichen Analyse-Ideen-Priorisierung-Testen-Zyklus (siehe Abbildung 2). Dieser Zyklus ermöglicht es, agil auf Marktveränderungen zu reagieren und innovative Lösungen zu entwickeln. [4]

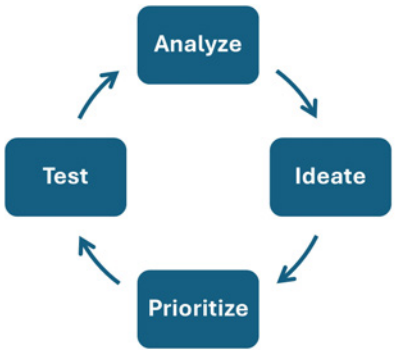

Abb. 2: Growth-Hacking-Zyklus [3]

Bekannte Beispiele erfolgreicher Growth-Hacking-Strategien sind Hotmail, Dropbox und Airbnb. Diese Unternehmen nutzten kreative Taktiken wie virales Marketing und Empfehlungsprogramme, um exponentielles Wachstum zu erzielen.

Praxisbeispiel: pipappl UG

Unternehmensvorstellung Die pipappl UG ist ein junges Start-up, das ein innovatives Zimmerpflanzenbewässerungssystem anbietet. Aufgrund begrenzter Ressourcen und eines dynamischen Marktumfelds

entschied sich das Unternehmen, Growth Hacking als Wachstumsstrategie zu implementieren.
Implementierung des Growth Hacking-Zyklus Der Growth Hacking-Zyklus wurde in der pipappl UG in mehreren Phasen umgesetzt:

1. Analysephase: Identifikation von Kundenbedürfnissen und Marktanforderungen.
2. Ideenfindung: Entwicklung kreativer Lösungen zur Nutzerakquise und -bindung.
3. Priorisierung: Bewertung und Priorisierung der entwickelten Ideen (siehe Abbildung 3).
4. Testphase: Umsetzung der priorisierten Idee und Implementierung von A/B-Tests.
5. Lernphase: Auswertung der Testergebnisse und Ableitung von Handlungsempfehlungen.

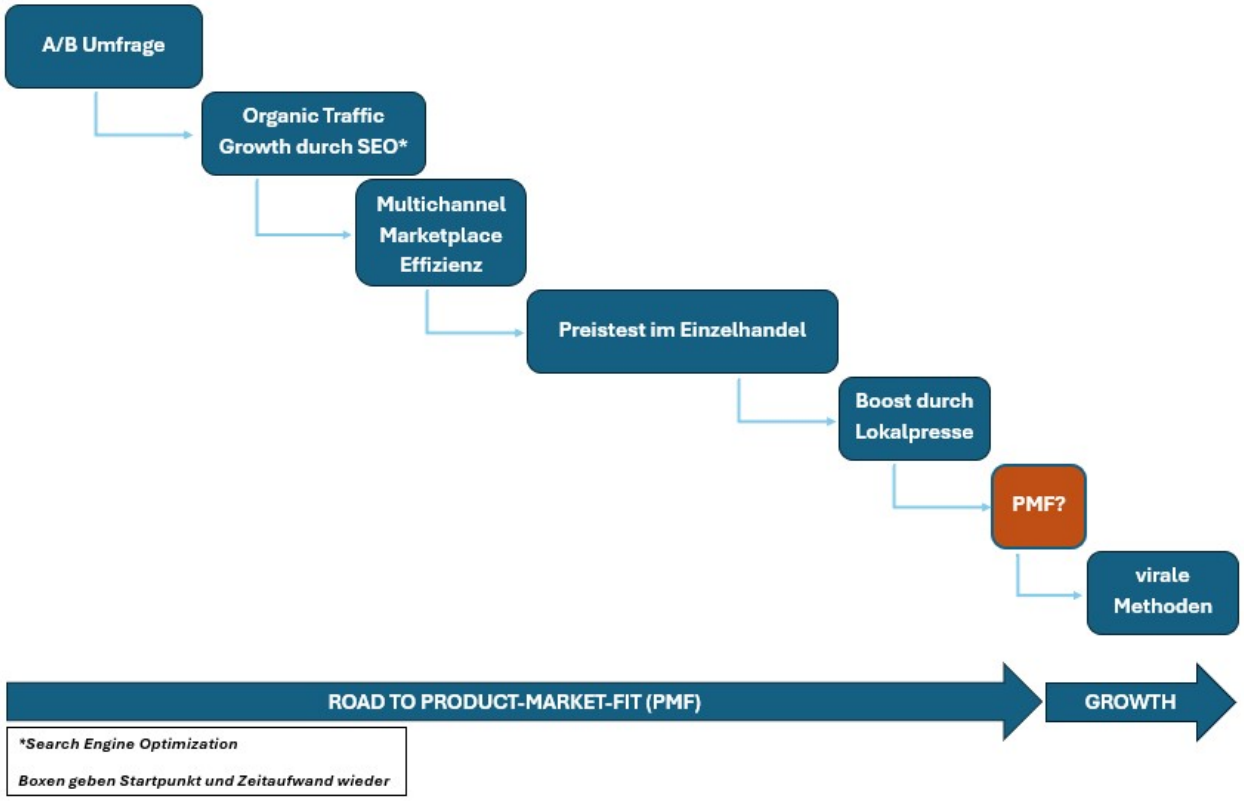

Abb. 3: Ideenpipeline für pipappl UG nach Phase 3 [3]

Ergebnisse und Diskussion Die Growth-Hacking-Umfrage, die A/B-Testing beinhaltet, hat das Start-up pipappl UG einen Schritt näher an den PMF gebracht, indem sie wertvolle Einblicke in die Zielgruppe und deren Präferenzen lieferte. Die folgenden Punkte fassen die wichtigsten Beiträge zusammen:

- Verbesserung der Produktwahrnehmung: Die Daten zur Verpackung und Anleitung geben konkrete Hinweise darauf, wie das Produkt besser auf die Bedürfnisse der Zielgruppe abgestimmt werden kann.
- Preisgestaltung: Die Vorliebe für separate Versandkosten und die Akzeptanz des Produktpreises liefern wichtige Informationen für die Preisstrategie.
- Zielgruppenverständnis: Die detaillierten Informationen über die Zielgruppe und deren Präferenzen helfen, zukünftige Marketing- und Produktentwicklungsmaßnahmen besser zu fokussieren.
- SEO und Vertrieb: Die identifizierten Schlüsselwörter und bevorzugten Vertriebskanäle bieten eine Grundlage für zukünftige SEO- und Marketplacement-Strategien.

Empfehlungen für weitere Maßnahmen:

- Langfristige SEO-Strategien: Implementierung der identifizierten Schlüsselwörter in die SEO-Strategie, um den organischen Traffic nachhaltig zu erhöhen.

- Erweiterung der Umfrage: Weitere Umfragen oder Tests, um spezifischere Daten zu sammeln und die Zielgruppe noch genauer zu segmentieren.
- Optimierung der Konversionsrate: Maßnahmen zur Steigerung des Vertrauens der Kunden und zur Erhöhung der Konversionsrate, wie z.B. die Nutzung von Marktplätzen mit höherem Kundenvertrauen.

Durch diese Maßnahmen kann pipappl UG die gewonnenen Erkenntnisse nutzen, um den PMF zu erreichen und das Wachstum des Unternehmens nachhaltig zu fördern.

Grünes, nachhaltiges Wachstum

Studien zeigen, dass die Digitalisierung erheblich zur Erreichung der Klimaziele beitragen kann. [1] Initiativen wie der Europäische Green Deal und die ESG-Kriterien unterstützen Unternehmen dabei, nachhaltige Praktiken zu implementieren. [5] Durch die Nutzung digitaler Technologien können Effizienzsteigerungen erzielt werden, die sowohl ökologisch als auch ökonomisch vorteilhaft sind. Dies kann mit Bemühungen der Implementation von Growth Hacking vereinbart werden und zu kongruenten Zielen führen.

Fazit und Ausblick

Growth Hacking bietet Start-ups eine flexible und kosteneffiziente Methode zur Förderung des Wachstums. Durch die Etablierung von Growth Hacking können Unternehmen schnell auf Marktveränderungen reagieren und langfristigen Erfolg sichern. Zukünftige Forschung sollte sich auf die Weiterentwicklung und Anpassung von Growth-Hacking-Methoden an verschiedene Branchen und Unternehmensgrößen konzentrieren, um deren Effektivität weiter zu untersuchen und steigern.

Literatur und Abbildungen

[1] Redaktion accenture. Klimaeffekte der Digitalisierung 2.0 - Studie zur Abschätzung des Beitrags digitaler Technologien zum Klimaschutz in Deutschland. *Bitkom e.V.*, 2024.

[2] Augusto Bargoni et al. *Highway to hell or paradise city? Exploring the role of growth hacking in learning from innovation failure*. Elsevier Ltd. | Technovation, 2024.

[3] Eigene Darstellung.

[4] Sean Ellis and Morgan Brown. *Hacking growth: how today's fastest-growing companies drive breakout success*. Virgin Books, 2017.

[5] Redaktion IHK. Von Green Deal bis ESG. *Magazin Wirtschaft*, 2024.

Probabilistische Erkennung und Analyse schlafbasierter Arousals und Schlafkrankheiten mittels Bayesian Deep Learning

Philipp Walter **Gabriele Gühring**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma IT-Designers GmbH, Esslingen am Neckar

Motivation und Problemstellung

Erholsamer Schlaf ist essenziell für die Gesundheit des Menschen. Gut ein Drittel der Menschen in Deutschland leiden unter Schlafproblemen [9]. Ungenügender Schlaf jedoch kann zu gesundheitlichen Problemen wie Depressionen, Übergewicht, Herz-Kreislauf-Erkrankungen oder Demenz führen [7] [6]. In diesem Zusammenhang spielen auch sogenannte Schlaf-Arousals eine wichtige Rolle. Dabei handelt es sich um kurzzeitige Aktivierungen des Gehirns, die zu einer plötzlichen Erhöhung der Wachheit führen, wodurch es zu einer Unterbrechung des Schlafprozesses oder einer Änderung der Schlafphase kommen kann, was die Erholungsfunktion des Schlafes beeinträchtigt [8]. Folge dieser Schlaffragmentierungen können beispielsweise Tagesmüdigkeit oder -schläfrigkeit sowie eingeschränkte Leistungsfähigkeit sein [8].

Die American Academy of Sleep Medicine (AASM) definiert Arousals als abrupte Frequenzänderungen im Schlaf-Elektroenzephalogramm (EEG) während der Schlafstadien N1, N2, N3 und REM, denen mindestens zehn Sekunden Schlaf vorausgeht und die mindestens drei Sekunden andauern [1] [7], exemplarisch dargestellt in Abb. 1. Arousals können spontan auftreten oder durch Krankheiten verursacht werden, wie z. B. häufig durch eine sogenannte obstruktive Schlafapnoe, die im Schlaflabor zu diagnostizieren sind. Um die Ursachen der Arousals zu differenzieren, sind zusätzliche Sensordaten notwendig [10].

Die Erkennung und Quantifizierung von Schlaf-Arousals und das Stellen einer einschlägigen Diagnose sind entscheidend, da sie direkte Auswirkungen auf die Schlafqualität und somit auf die Gesundheit haben. Die manuelle Analyse von Schlafdaten ist jedoch zeitaufwändig und erfordert spezielle Kenntnisse. Daher bieten automatisierte Methoden zur Erkennung von Arousals einen wichtigen Beitrag, um die Diagnose und Behandlung von Schlafstörungen effizienter zu gestalten. Hierfür gibt es in der Literatur bereits verschiedene auf Deep Learning basierende Ansätze zur Vorhersage von Arousals oder möglicher Arousal-Intervalle für die im Rahmen einer Polysomnographie aufgezeichneten Daten [6] [11]. Ebenfalls ist denkbar, Machine Learning für die Diagnose von Schlafkrankheiten auf Basis dieser Daten einzusetzen.

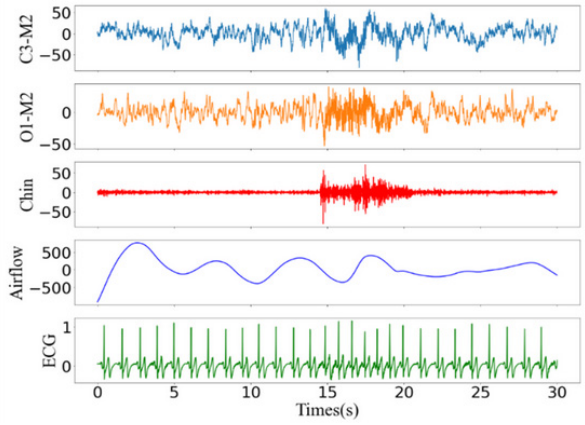

Abb. 1: EEG-Kanal-Diagramm während eines Arousal-Events (lizenziert unter CC BY 4.0 - https://creativecommons.org/licenses/by/4.0/) [7]

Ein Problem von Machine Learning-Modellen ist jedoch die oft falsche Sicherheit ihrer Vorhersagen, insbesondere bei ungesehenen Trainingsdaten. Dies kann besonders in sensiblen Bereichen wie der Medizin, wo diese zur Entscheidungsunterstützung eingesetzt werden, problematisch sein, da eine fehlerhafte Diagnose, insbesondere im falsch-negativen Fall, schwerwiegende Folgen haben kann. Bayes'sche Neuronale Netze bieten hier eine Lösung, indem sie Unsicherheiten in ihren Vorhersagen schätzen und somit eine robustere und vertrauenswürdigere Analyse ermöglichen. Dies ist im vorliegenden Fall besonders nützlich, um die Verlässlichkeit der Erkennung von Arousals oder Dia-

gnosen zu erhöhen und potenziell falsche Schlüsse oder Entscheidungen zu reduzieren, was letztlich zu einer besseren individuellen Schlafbehandlung führen dürfte.

Bayes'sche neuronale Netze als Lösungsansatz

Bayes'sche neuronale Netze (BNNs) stellen eine Erweiterung traditioneller, frequentistischer neuronaler Netze dar und verwenden probabilistische Modelle, um Unsicherheiten in den Vorhersagen zu berücksichtigen [2] [4]. Im Gegensatz zum frequentistischen Ansatz, der die Gewichte eines Modells als fest, unbekannte und durch Training zu optimierende Größen betrachtet, erlaubt der Bayes'sche Ansatz eine probabilistische Interpretation der Modellparameter. Während ersterer Punktvorhersagen liefert und oft die Unsicherheiten in den Schätzungen ignoriert, was zu übermäßig selbstsicheren Modellen führen kann, berücksichtigen BNNs hingegen diese Unsicherheiten, indem sie Verteilungen über die Gewichte lernen. [2] [4]. In Abb. 2 ist dies schematisch dargestellt. Das ist aus genannten Gründen besonders in sensiblen Anwendungsbereichen wie der Medizin von großem Nutzen.
Die Unsicherheiten werden durch den Posterior (die Nach-Wahrscheinlichkeit) der Gewichte nach der Beobachtung der Daten beschrieben. Mathematisch lässt sich dies als $p(w|D)$ darstellen, wobei w die Gewichte des neuronalen Netzes und D die Eingabedaten sind. Der Posterior wird mithilfe des Bayes'schen Theorems berechnet:

$$p(w|D) = \frac{p(D|w)p(w)}{p(D)} \tag{1}$$

Hierbei ist $p(D|w)$ die Likelihood, $p(w)$ die Prior-Verteilung der Gewichte und $p(D)$ die sogenannte marginale Likelihood oder auch Evidenz, die sich auch als Integral über alle möglichen Gewichte berechnen lässt:

$$p(D) = \int p(D|w)p(w)d \tag{2}$$

So kann die sogenannte Vorhersage-Verteilung bestimmt werden:

$$p(\hat{y}|D) = \int p(\hat{y}|w)p(w|D)dw \tag{3}$$

Diese Verteilung integriert über alle möglichen Gewichte und liefert damit eine Verteilung der möglichen Vorhersagen. Dies ermöglicht nicht nur eine Unsicherheitsabschätzung, sondern auch die Ableitung einer Punktvorhersage, z.B. in Form des Erwartungswerts der Verteilung. Diese Integration wird als Marginalization bezeichnet und ermöglicht es, die Unsicherheit in den Vorhersagen zu berücksichtigen und dadurch robustere und verlässlichere Modelle zu erstellen.
Die Schätzung des Posteriors ist in der Praxis analytisch jedoch aufgrund der hohen Dimensionalität in der Regel nicht durchführbar. Daher werden Ansätze zur Approximation wie sogenannte Sampling-Methoden und Variational Inference eingesetzt [5]. Sampling-Methoden wie z.~B. Markov Chain Monte Carlo (MCMC) generieren Beispiele aus dem Posterior, was allerdings rechenintensiv sein kann. Bei der Variational Inference hingegen nähert man einen Posterior durch eine einfachere Verteilung an und optimiert die Parameter dieser Verteilung, um sie möglichst nah an den echten Posterior heranzuführen. Mittels dieser Näherungen ist es möglich, Bayes'sche neuronale Netze effizient zu trainieren und die Unsicherheiten in den Vorhersagen realistisch zu modellieren.

Ausblick

Im Rahmen dieser Arbeit soll untersucht werden, wie Bayes'sche neuronale Netze zur Arousal-Erkennung und Diagnostik von Schlafkrankheiten eingesetzt werden können. Hierfür soll ein entsprechendes Modell zur Diagnose entwickelt und vorhandene Ansätze zur Arousal-Erkennung erweitert und unter verschiedenen Posterior-Approximationen evaluiert werden. Ein wesentlicher Schritt dabei ist auch die Feature Selection, d. h. die Auswahl der wichtigsten Input-Variablen aus den vorliegenden Schlafaufzeichnungen sowie allgemeinen medizinischen Informationen. Zum Training und zur Evaluation der Modelle steht ein Datensatz zur Verfügung, der 113 diagnostische polysomnografische Schlafaufzeichnungen von Patient:innen umfasst. Die Aufzeichnungen bestehen aus bis zu 36 Rohdatenkanälen sowie 23 abgeleitete Datenkanäle und enthalten 81 Typen annotierter Ereignisse für alle Teilnehmenden, die automatisiert und manuell durch erfahrende Schlafmediziner:innen vorgenommen wurden. Ergänzt wird der Datensatz durch Informationen aus verschiedenen Fragebögen und den jeweiligen Diagnosen. Der Datensatz wurde im Rahmen einer medizinischen Studie in den Jahren 2021-2022 in Zusammenarbeit zwischen dem Klinikum Esslingen, der IT-Designers Gruppe und der NRI Medizintechnik GmbH, alle ansässig in Deutschland, gesammelt.
Ziel ist es, die Erkennung von Arousals und die Diagnose von Schlafkrankheiten zu verbessern, indem Unsicherheiten in den Vorhersagen berücksichtigt werden. Dies soll zu robusteren und vertrauenswürdigeren Modellen führen, die eine bessere patientenspezifische Behandlung ermöglichen.

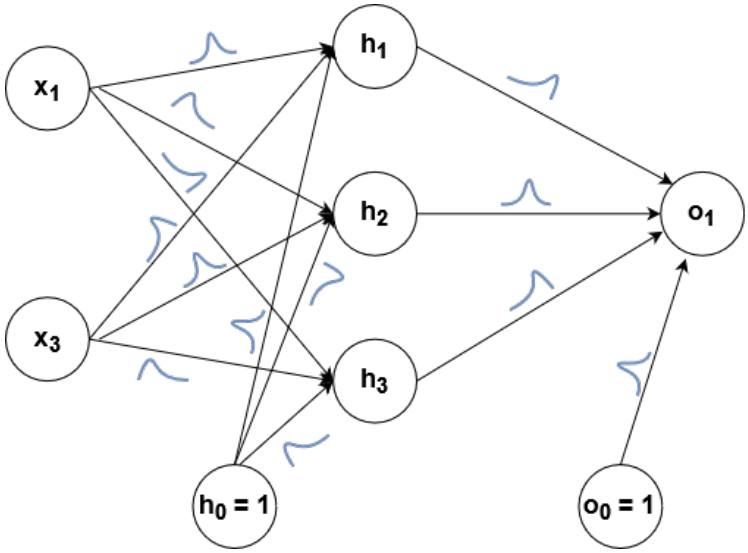

Abb. 2: Schematische Darstellung eines Bayes'schen Neuronalen Netzes (BNN) mit Verteilungen für die einzelnen Gewichte anstelle von fixen Gewichten [3]

Literatur und Abbildungen

[1] Richard B. Berry, Rohit Budhiraja, Daniel J. Gottlieb, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *Journal of Clinical Sleep Medicine*, 8:597–619, 2012.

[2] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural networks. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, volume 37, pages 1613–1622. JMLR.org, 2015.

[3] Eigene Darstellung.

[4] Zoubin Ghahramani. Bayesian non-parametrics and the probabilistic approach to modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371, 2013.

[5] Ethan Goan and Clinton Fookes. Bayesian Neural Networks: An Introduction and Survey. In *Case Studies in Applied Bayesian Data Science. Lecture Notes in Mathematics*, volume 2259, pages 45–87. Springer International Publishing, 2020.

[6] Hongyang Li and Yuanfang Guan. DeepSleep convolutional neural network allows accurate and fast detection of sleep arousal. *Communications Biology*, 4, 2021.

[7] Xiangyu Qian, Ye Qui, Qingzu He, et al. A Review of Methods for Sleep Arousal Detection Using Polysomnographic Signals. *Brain Sciences*, 11, 2021.

[8] Friedhart Raschke. Arten von Arousal. *Somnologie - Schlafforschung und Schlafmedizin*, 19:6–11, 2015.

[9] Robert Schlack et al. Häufigkeit und Verteilung von Schlafproblemen und Insomnie in der deutschen Erwachsenenbevölkerung. *Bundesgesundheitsblatt 2013*, 56:740–748, 2013.

[10] Boris A. Stuck, Christoph Schöbel, Alfred Wiater, and Dora Triché. Klug entscheiden: Obstruktive Schlafapnoe. *Deutsches Ärzteblatt*, 19:996–1000, 2021.

[11] Hasan Zan and Abdulnasır Yildiz. Multi-task learning for arousal and sleep stage detection using fully convolutional networks. *Journal of Neural Engineering*, 20, 2023.

State of the art in automated API fuzzing

Henning Weise **Dominik Schoop**

Department of Computer Science and Engineering, Esslingen University

Work carried out at Mercedes-Benz Tech Innovation GmbH, Ulm

Introduction

As software systems evolve, the role of Application Programming Interfaces (APIs) becomes increasingly pivotal. The API serves as the fundamental conduit for communication between different software components, enabling seamless data exchange and interaction [5]. Over the past decade, the use of APIs has become increasingly popular among developers, enabling them to decouple backend and frontend components. Statistics from Cloudflare show that in the first week of December 2021, API calls represented 54 % of their total network traffic. [6] In a survey conducted by RapidAPI, 850 developers from over 100 countries were queried regarding their reliance on APIs [3]. Over 62.6 % of respondents indicated that they relied on APIs to a greater extent in 2022 than in 2021. The complexity of APIs coupled with the rise in cyber threats necessitates effective testing techniques to ensure the security and reliability of APIs [9]. Checking APIs for vulnerabilities can be a time-consuming and labor-intensive process when done manually, and is often prone to human error [1] [4] [8]. This is especially challenging in large-scale systems.

Efforts to identify and mitigate vulnerabilities encompass a spectrum of techniques, primarily categorized into Static Application Security Testing (SAST) and Dynamic Application Security Testing (DAST). Static analysis involves scrutinizing source code or compiled binaries without execution, while dynamic analysis entails runtime examination, often facilitated by debuggers. Within dynamic analysis, fuzz testing, or fuzzing, stands out as a prevalent automated technique for vulnerability discovery. [2] API fuzzing tests the API with random or semi-random input. This approach can uncover quality or security problems that SAST and other DAST methods may not detect. Furthermore, automating an API fuzzer offers the possibility to apply it to multiple applications simultaneously without additional human effort.

Problem Statement and Objectives

Our work is focused on REST APIs due to their widespread usage. Designing test cases to effectively fuzz REST APIs' behaviors is challenging. APIs comprise a number of operations that may be executed in various orders, thereby changing the state of the database. The order in which the different paths are tested is of critical importance for the efficiency of fuzzing the API. If an attempt is made to delete a resource without first creating it, the behavior of that path cannot be tested in detail. Errors could also occur in the application if a resource is first deleted and then an attempt is made to retrieve it. Fuzzers must address this problem by intelligently determining the order in which to test the paths.

It is common for APIs to have paths secured with authentication. This raises the question of how well fuzzers can handle different authentication methods. Additionally, it is important to consider the extent to which API fuzzers can deal with applications that manage sessions. In such applications, a user's interaction with the application is not limited to a single request and response cycle. Instead, it involves a sequence of interactions that depend on the information stored from previous interactions. To illustrate, a user may log in to an application, thereby initiating a session. They may then perform various actions, all of which are tied to that session. Eventually, the user may log out, thereby ending the session. Each step in this sequence relies on the context provided by the session. API fuzzers must therefore be able to handle these sessions correctly in order to effectively test the application. The fuzzer should be able to manage session timeouts and renew sessions as needed to continue testing.

Rate limiting, unreachable servers, and different HTTP versions (HTTP/2 and HTTP/3) introduce further challenges in API fuzzing. Rate limiting can throttle the number of requests a fuzzer can make, potentially hindering comprehensive testing. Unreachable servers can disrupt the fuzzing process, causing incomplete testing cycles or producing false

positive results. The differing characteristics of HTTP versions may influence how fuzzers interpret and generate requests, potentially affecting their capacity to identify vulnerabilities.
In order to gain a comprehensive understanding of the capabilities and limitations of existing fuzzers, we employ a systematic experimental approach. This methodical examination involves multiple steps designed to rigorously test the fuzzers' functionality.

- Review of existing non-commercial API fuzzers, including those published in scientific work and available in public repositories.
- Testing the basic functionality of the identified fuzzers using applications configured with specific scenarios and known vulnerabilities.
- Evaluating how fuzzers handle different authentication methods, including basic authentication, JWT, and OAuth2.
- Evaluating the performance of fuzzers with rate limiting, unreachable servers, HTTP/2, and HTTP/3.
- Evaluating how good Fuzzers can correlate errors and the sequence of requests responsible for them.
- Analyzing the experimental results to identify the strengths and weaknesses of current fuzzers.
- Proposing a model for an optimal API fuzzer, addressing the limitations identified in the experiments.

Selecting of Fuzzers

In our initial experiment, we check, whether the publicly available fuzzers can be run and operated on our system. We use a virtual machine with Kali 2024.1, which is a Debian-based Linux distribution. The system uses a Linux kernel version 6.6.9-amd64, Python 3.11.8, and Docker version 20.10.25. Despite a comprehensive search, we have been unable to locate any installable software or source code for the MOREST and bBOXRT fuzzers. Consequently, we are unable to run them on our system. When installing Hsuan-Fuzz, RestTest, openapi3-fuzzer, TnT-Fuzzer, and RestCT, errors occur that we cannot solve. We have created issues on GitHub, but they have not yet been answered. Additionally, the ZAP software is used together with the fuzzing add-on. It became evident that fuzzing with ZAP requires a significant amount of manual effort, for instance, the necessity of fuzzing each path manually and identifying the fields to be fuzzed. Consequently, we exclude ZAP in subsequent investigations.

First results and limitations of the fuzzers

Figure 1 presents the fuzzers running on our system, the version we use, and their limitations.
We note that RestTestGen and OFFAT are the only two fuzzers that do not strictly adhere to the OpenAPI Specification (OAS) when generating test cases. If it is specified that certain parameters are required for a path, RestTestGen tests which results are produced by omitting the parameters. OFFAT is the only fuzzer that checks the API for undocumented HTTP verbs. This means that the fuzzer checks for each path documented in the OAS whether a result can be produced if an undocumented verb is used.
It is important to note that, except for openapi-fuzzer and RestTestGen, all fuzzers recognize only 5xx status codes as errors. This means that if a fuzzer manages to obtain an error code other than 5xx that is not documented, only the two will recognize it. This missing feature in the other fuzzers represents a significant problem. The 4xx range encompasses a variety of client errors, including *404 Not Found*, *400 Bad Request*, and *401 Unauthorized*. These errors can indicate potential issues or unexpected behavior within the API.

Fuzzer	Version	Test generation not strictly adhere to OAS	Detects other codes than 5xx	Basic auth	Token auth	Request sequence
APIFuzzer	0.9.13	no	no	no	no	no
OFFAT	0.17.5	yes	no	yes	no	no
openapi-fuzzer	0.2.0	no	yes	yes	no	no
RESTler	9.2.4	no	no	yes	yes	no
RestTestGen	24.03	yes	yes	yes	yes	no

Fig. 1: Working fuzzers with results on our first experiments [7]

Additionally, some undocumented error codes may indicate security vulnerabilities. For example, repeated undocumented *401 Unauthorized* errors could suggest authentication issues. Identifying these errors can assist in the detection of security flaws that require attention.
The utilization of basic and JWT authentication renders APIFuzzer inoperable. For OFFAT, openapi-fuzzer, RESTler, and RestTestGen, the use of basic authentication with username and password is not problematic. Nevertheless, authentication via tokens presents a significant challenge, particularly when the token is only valid for a short period of time. In RESTler and RestTestGen, the fuzzer is capable of obtaining a new token when it expires through the use of a script that must have been created.
In addition, we can already say at this stage that, with all fuzzers, the tracing of requests that have led to errors is only possible to a very limited extent. RESTler has attempted to implement this feature, but the results it produces in our experiments are not convincing.

Outlook

Since this is work in progress, in the next steps, we will inspect the behavior of the fuzzers with regard to denial-of-service vulnerabilities, unreachable servers, and rate limiting. Based on the results and the identified limitations of the current fuzzers, we propose a theoretical model for a better fuzzer. We address the current problems and describe how they can be solved.

References and figures

[1] Andrea Arcuri. An Experience Report On Applying Software Testing Academic Results In Industry: We Need Usable Automated Test Generation. *Empirical Software Engineering*, 23:1959–1981, 2018.

[2] James Fell. A Review of Fuzzing Tools and Methods. *PenTest Magazine*, 2017.

[3] RapidAPI Inc. State of APIs 2022 | Rapid Developer Survey Results. https://stateofapis.com/, 2022.

[4] Isha, Abhinav Sharma, and M. Revathi. Automated API Testing. In *Proceedings of the International Conference on Inventive Computation Technologies (ICICT-2018)*, pages 788–791. Institute of Electrical and Electronics Engineers (IEEE), 2018.

[5] Alberto Martin-Lopez, Sergio Segura, and Antonio Ruiz-Cortés. Test coverage criteria for RESTful web APIs. In *Proceedings of the 10th ACM SIGSOFT International Workshop on Automating TEST Case Design, Selection, and Evaluation (A-TEST '19)*. Association for Computing Machinery, New York, NY, United States, 2019.

[6] Daniele Molteni. Landscape of API Traffic. https://blog.cloudflare.com/landscape-of-api-traffic, 01 2022.

[7] Own representation.

[8] Ari Takanen, Jared DeMott, Charles Miller, and Atte Kettunen. *Fuzzing for Software Security Testing and Quality Assurance*. Artech House, 2 edition, 2018.

[9] Xiaogang Zhu, Sheng Wen, Seyit Camtepe, and Yang Xiang. Fuzzing: A Survey for Roadmap. *ACM Computing Surveys*, 2022.

Sichere Codeausführung in Docker Containern

Willy Matthew Xamountry **Dominik Schoop**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma IT-Designers GmbH, Entennest 2, 73730 Esslingen

Einleitung

In einem Zeitalter der progressiven Digitalisierung spielt die Sicherheit eine große Rolle. Das Ziel von Sicherheit ist es, die Funktions- und die Informationssicherheit eines Systems zu gewährleisten. Grob beschrieben schützt die Informationssicherheit das System vor Angreifern, beispielsweise durch das Verhindern von unautorisierten Zugriffen auf sensible Daten. Die Funktionssicherheit hingegen schützt den Nutzer vor dem System, beispielsweise durch das verhindern von illegalen Zuständen eines Systems. Eine besondere Herausforderung für die Sicherheit stellen dabei Systeme dar, die den Code von Nutzern ausführen sollen. Um solch ein System zu konzipieren, muss man sich überlegen, wie man Code sicher ausführen kann, ohne das System zu gefährden.
Die Lernumgebung der Kata-Plattform bietet Nutzern die Möglichkeit, sich Programmierkenntnisse anzueignen, indem sie Aufgaben mit wachsender Schwierigkeit überwinden. Um diese Aufgaben zu lösen, müssen Nutzer eine Lösung programmieren und auf der Kata-Plattform hochladen. Die Abgabe wird dann in Docker Containern kompiliert und ausgeführt. Es stellt sich die Frage, wie sicher eine Codeausführung in Docker Containern ist. Zu denkbaren Schwachstellen werden mögliche Lösungsansätze entworfen [3].

Kata-Plattform

Die Kata-Plattform wird von der IT-Designers Gruppe entwickelt. Die Anwendung nutzt verschiedene Docker Container für die Funktionen der Anwendung. Da jedoch alle Container auf einem Host laufen, handelt es sich hierbei um ein verteiltes System auf einem Monolithen. Als Lernumgebung für die Softwareentwicklung soll die Plattform die Fähigkeiten der Nutzer fördern. Dies erfolgt durch das Lösen von Aufgaben, wie zum Beispiel *Programmieren Sie die Fibonacci-Folge*. Dabei kann eine Aufgabe in verschiedenen Programmiersprachen gelöst werden. Gibt ein Nutzer eine Lösung ab, wird ein Docker Container für die entsprechende Programmiersprache gestartet. Die Abgabe wird in diesem Container kompiliert und es werden Paare von Eingabe- und Ausgabeparametern getestet. Entspricht die Ist-Ausgabe der Soll-Ausgabe, ist die Lösung korrekt.

Docker

Docker ist ein Open-Source-Framework für die Anwendungsentwicklung und verwendet eine Client-Server-Architektur für die Kommunikation zwischen Docker-Daemon und Docker Container. Möchte der Nutzer Operationen ausführen, sendet er als Docker-Client seine Anfragen über eine REST-API unter Unix-Sockets an den Docker-Daemon. Container entstehen aus Images. Diese sind damit die Blaupausen und die Container die Instanzen dieser Blaupausen. Docker Container sind leichtgewichtig und enthalten alles Relevante für die Lauffähigkeit, sodass man nicht von den Bibliotheken des bereitstellenden Computers abhängig ist [2].

Ziel der Arbeit

Da es sich bei der Arbeit um eine Machbarkeitsstudie handelt, wird evaluiert, wie sicher die Ausführung von Code in Docker Containern ist. Die Forschung widmet sich dabei der Aufgabe, mögliche Schwachstellen bei dieser Ausführung zu finden und Maßnahmen für eine erhöhte Sicherheit zu entwickeln. Im Fokus liegt vor allem die willkürliche Ausführung von jeglichem Programmcode. Es wird untersucht, welche impliziten Angriffe über diesen Angriffsvektor möglich sind und welche Auswirkungen diese haben. Als mögliche Gegenmaßnahmen werden statische Code-Analysen implementiert und andere Methoden recherchiert und evaluiert. Die Docker Container sind damit der Fokus dieser Arbeit, jedoch werden weitere Sicherheitsrisiken der Kata-Plattform ebenfalls betrachtet und theoretisch evaluiert.

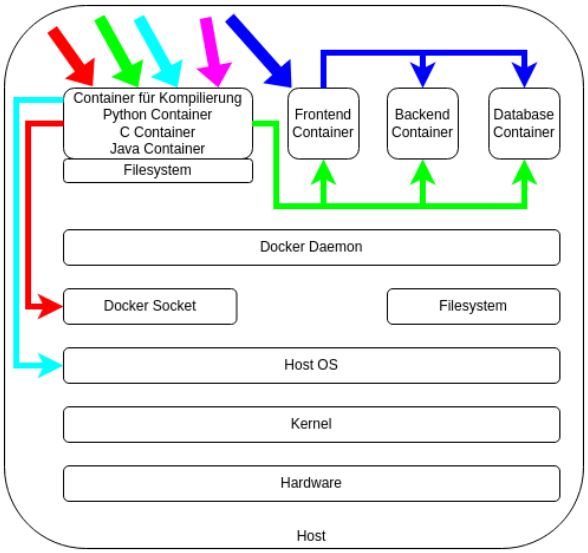

Abb. 1: Schematische Darstellung der Angriffsvektoren [1]

Angriffsvektoren

Abbildung 1 zeigt mögliche Angriffsvektoren. Dabei stellt jede Farbe einen eigenen Angriffsvektor dar. Mehrere Angriffe starten mit dem Container für die Kompilierung. Dieser Container entspricht dem Container, der gestartet wird, um die Abgaben der Nutzer auszuführen.

- **Roter Angriffsvektor**: Ein Angriff auf den Docker Socket. Erlangt ein Angreifer Zugriff auf den Socket, erhält er auch Zugriff auf den Docker-Daemon und kann damit bestehende Container beeinflussen oder neue Container starten.
- **Grüner Angriffsvektor**: Ein Angriff auf Container im selben Netzwerk.
- **Hellblauer Angriffsvektor**: Ein Ausbruch aus dem Container, um sich Zugriff zum Host-System zu verschaffen.
- **Lila Angriffsvektor**: Ein Angriff auf den Container selber.
- **Blauer Angriffsvektor**: Ein Angriff auf die Web-Applikation, wie Code-Injektion oder den Upload einer Archivbombe, da das Hochladen von Archiven valide ist.

Code Analyse

Durch das Hochladen von Abgaben ist sämtlicher Code lesbar. Deswegen erscheint die Maßnahme einer statischen Codeanalyse vielversprechend. Mit Fokus auf die Programmiersprache Python wird ein Analyseprogramm entwickelt, das bösartige Abgaben herausfiltern soll. Wichtig dabei ist, sowohl die Effektivität als auch die Effizienz dieser Methode zu evaluieren. Dazu gehört es auch zu untersuchen, wie mit falschen Positiven und falschen Negativen umgegangen wird. Zusätzlich werden andere Methoden der Code-Analyse in Betracht gezogen, wie die Nutzung von Large Language Models oder Control Flow Integrity.
Bei Large Language Models handelt es sich um sogenannte Modelle künstlicher Intelligenz, die darauf trainiert werden, natürliche und logische Sprachen zu erkennen und zu generieren. Besonders fortschrittliche und wahrscheinlich bekannte Modelle sind ChatGPT und LLaMa [5].
Control Flow Integrity ist eine Überwachung, die auf der statischen Codeanalyse basiert. Bei der Analyse liegt der Fokus auf dem Arbeitsfluss des Codes. Es werden die validen Zustände eines Codes bestimmt und mit den Ist-Zuständen verglichen. Dies verhindert einen Ausbruch aus dem natürlichen Arbeitsfluss des Codes [4].

Ausblick

Zum Zeitpunkt der Erstellung dieses Artikels sind die Arbeiten an der Analyse noch nicht abgeschlossen. Fortlaufend soll festgestellt werden, wie sicher die Ausführung von Code in Docker Containern ist. Es besteht auch die Möglichkeit, dass dies eine sichere Umgebung ist und diese Methode keine oder nur unbedenkliche Risiken mit sich bringt. Ein Angriff auf den Container, um diesen lahmzulegen, ist eine definitive Schwachstelle. Ist die Ausführung von Code im Hinblick auf das Hostsystem unbedenklich, wird zumindest diese Schwachstelle anhand der Python Programmiersprache evaluiert und es werden Gegenmaßnahmen dagegen entwickelt. Weitere Sicherheitsmaßnahmen werden für die Kata-Plattform vorgeschlagen, welche eine Grundlage für die Verbesserung der Anwendung bilden. Dazu gehören Sicherheitsrisiken in der Web-Applikation. Im Hinblick auf falsche Ergebnisse werden vor allem die falschen Negativen betrachtet. Diese stellen im Vergleich zu falschen Positiven eine größere Gefahr dar.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Inc. Docker. Docker overview. https://docs.docker.com/get-started/overview/, 2024.

[3] Claudia Eckert. *IT-Sicherheit : Konzepte, Verfahren, Protokolle*. De Gruyter Oldenbourg, Verlag, 11 edition, 2023.

[4] J. Li, L. Chen, G. Shi, K. Chen, and D. Meng. ABCFI: Fast and Lightweight Fine-Grained Hardware-Assisted Control-Flow Integrity. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on. 39(11):3165-3176 Nov, 2020*, 2020.

[5] J. Wang, Y. Huang, C. Chen, Z. Liu, S. Wang, and Q. Wang. Software Testing With Large Language Models: Survey, Landscape, and Vision. *IEEE Transactions on Software Engineering IIEEE Trans. Software Eng. Software Engineering, IEEE Transactions on. 50(4):911-936 Apr, 2024*, 2024.

Evaluation der Performance von Softwarebasierten Verschlüsselungen im Kontext von Hochbandbreitigen ADAS-Logging Anwendungen

Max von Berg **Tobias Heer**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Vector Informatik GmbH, Stuttgart

Motivation und Problemstellung

Die Automobilindustrie hat stetig steigende Anforderungen an Sicherheit und Komfort. Um diese zu erfüllen, werden immer mehr Sensoren von Fahrerassistenzsystemen verwendet. Diese Systeme werden Advanced Driver Assistance Systems (ADAS) genannt und generieren durch den Einsatz von Kameras, Radar und anderen Sensoren enorme Datenmengen von bis zu 5 GB/s. Um alle gängigen Sensordaten und Protokolle transparent im Steuergeräte-Verbund während Erprobungsfahrten zu loggen, bietet die Firma Vector Informatik GmbH selbstentwickelte ADAS-Datenlogger als universelles Tool an. Die Datensicherheit der auf Erprobungsfahrten erfassten Messdaten spielt eine wichtige Rolle, da sie vertrauliche Informationen über Fahrzeugdetails und Kamerabilder enthalten, die vor Datendiebstahl und Missbrauch geschützt werden müssen. Die Messdaten müssen in Echtzeit, selbst bei hohem Datendurchsatz, zuverlässig verschlüsselt und auf Festplatten gesichert werden, um vor unberechtigtem Zugriff geschützt zu sein. Zur Datenverschlüsselung in Vectors Datenlogger gibt es zwei grundlegende Ansätze: Hardware- und softwarebasierte Verschlüsselung. Hardwarebasierte Verschlüsselung, wie sie bei Self-Encrypting Drives (SEDs) zum Einsatz kommt, bietet eine effiziente Verschlüsselungslösung. Um jedoch auf Dateiebene verschlüsseln zu können, ist eine Software-Lösung erforderlich. Softwarebasierte Verschlüsselungsmethoden sind in der Lage, einen kontinuierlichen Datenstrom bei sehr hohen Datenraten zu verschlüsseln. Dabei gilt es folgende Kriterien zu beachten:

- **Datenschutz:** Unbefugten Zugriff auf Daten verhindern
- **Datenrate:** Um eine hohe Datenrate von bis zu 5 GB/s verschlüsselt loggen zu können wird ein sehr effizienter Algorithmus benötigt.
- **CPU-Last:** Die Messleistung der Datenlogger darf nicht beeinträchtigt werden.
- **Datenverlust:** Bei der Echtzeit-Verschlüsselung darf kein Datenverlust entstehen.
- **Integration:** Die Verschlüsselungslösung muss nahtlos und transparent in den bestehenden Vector Software-Stack integriert werden können
- **Schlüsselaustausch/Rechte-Management:** Der Schlüsselaustausch muss sicher und praktikabel sein
- **Kryptoagilität:** Wenn Algorithmen in Zukunft möglicherweise als unsicher gelten oder effizientere Lösungen entwickelt werden

Ziel ist es, unter Berücksichtigung der aufgeführten Kriterien eine robuste und effiziente softwarebasierte Verschlüsselungsmethode im Kontext eines ADAS-Logging-Systems zu erstellen und deren Leistungsgrenzen zu evaluieren.

Design

Um geeignete Algorithmen für diese Arbeit auszuwählen, werden in einem ersten Schritt basierend auf verwandten Arbeiten verschiedene symmetrische Verschlüsselungsalgorithmen ausgewählt. In einer im Rahmen der Arbeit entwickelten Testumgebung, geschrieben in C++, werden die ausgewählten Algorithmen hardwareunabhängig untersucht. Diese Testumgebung ist entscheidend für das Sammeln von Daten, die Aufschluss über Leistungskriterien wie Durchsatz, CPU- und Speicherlast geben, um eine Analyse der Algorithmen zu ermöglichen. Es werden Benchmark-Szenarien implementiert, um Daten über die Verschlüsselunsgszeit, Datendurchsatz bei unterschiedlichen Puffergrößen sowie die CPU- und

Speicherauslastung zu sammeln. Zudem ist die Umgebung plattformunabhängig, um Unterschiede zwischen den von Vector verwedeten Betriebssystemen Windows und Linux zu identifizieren. Um sicherzustellen, dass die Verschlüsselungsalgorithmen auch in Zukunft für zeitkritische Anwendungen geeignet sind, werden sie sowohl mit Datenflüssen in der Größenordnung heutiger Systeme als auch mit darüber hinausgehenden Datenmengen getestet. Die maximal testbare Datenrate wird dabei durch die Leistung des aktuellen Logger-Systems begrenzt. Um Daten unter realen Messbedingungen zu sammeln, werden die vielversprechendsten Algorithmen aus den Benchmarks in den Vector-Stack (Hardware und Software) integriert. Bei der Implementierung wird auf Kryptoagilität geachtet, um in Zukunft bei geänderten Performance- oder Sicherheitsanforderungen flexibel agieren zu können. Für einen direkten Vergleich zwischen einer software- basierten Verschlüsselung und einer Hardwareverschlüsselung mit Self-Encrypting Drives, wird der Vector Datenlogger mit entsprechenden SEDs ausgestattet. Evaluierungsskripte und Tests werden implementiert, um Daten zu CPU- und RAM- Performancedaten und File I/O erheben und einen Vergleich der Leistungsgrenzen ermöglichen zu können. Da bei symmetrischen Verschlüsselungen der Schlüsselaustausch ein grundlegendes Problem ist, wird anhand eines Proof-of-Concept-Prototyps eine potentielle Lösung für das Schlüsselverteilungsproblem im Rahmen des Vector Datenloggers vorgestellt. Ein zentrales Ziel dabei ist, ein Rechte-Management zu implementieren, welches gewährleistet, dass unterschiedliche Benutzergruppen spezifische Daten nach Bedarf entschlüsseln können.

Evaluation

Mithilfe der entkoppelten Testumgebung werden gezielt Daten erhoben, die einen Vergleich verschiedener softwarebasierter Verschlüsselungsalgorithmen ermöglichen. Abbildung 1 zeigt exemplarisch einen Vergleich verschiedener Betriebsmodi von AES-256 und ChaCha20. Anhand der erhobenen Daten werden vielversprechende Algorithmen für den Logging-Kontext identifiziert, wobei die Faktoren Datensicherheit, Performance und Effizienz berücksichtigt werden. Im Rahmen der Untersuchung wird ein Vergleich zwischen den softwaregestützten Verschlüsselungsansätzen und der Verschlüsselung durch Self-Encrypting Drives durchgeführt. Ziel ist es, die Vor- und Nachteile jeder Methode herauszustellen. Algorithmen aus den verwandten Arbeiten werden dabei genauer untersucht und es werden Vor- und Nachteile gegeneinander abgewogen. Falls vorhanden, werden die verschiedenen Schlüssellängen und Betriebsmodi dieser Algorithmen analysiert, um Aufschluss über die Sicherheit und Performance der jeweiligen Konfiguration zu geben.

Abbildung 2 (Originalbild entnommen aus einem Demovideo des DYNA4-Simulationstools [6]) zeigt beispielsweise erkennbare Muster im Chiffretext, die potenziell wertvolle Informationen preisgeben können. Die zu untersuchenden Algorithmen werden unter Systemlasten, ähnlich einer Messfahrt, auf dem Datenlogger getestet. Dabei werden entweder kontinuierlich Daten von ADAS- Sensoren generiert oder aufgezeichneter Datenverkehr einer Messfahrt in Form von Pcap-Dateien abgespielt. In einer dedizierten Messkampagne werden verschiedenen Lastszenarien und Datendurchsatzraten simuliert, um Daten aus diesem praxisnahen Umfeld sammeln und analysieren zu können. Die Analyse der Daten soll Aufschluss über die Performance und speziell die Leistungsgrenzen des jeweiligen Algorithmus geben. Evaluiert werden dabei Kennzahlen, wie der Datendurchsatz, potentielle Verluste von Messdaten, die CPU- Auslastung pro Kern und insgesamt, sowie die Datenintegrität. Das Konzept für die Lösung des Schlüsselverteilungsproblems wird einer Schwachstellenanalyse unterzogen und orientiert sich dabei an Normen wie ISO 27001 und NIST SP 800-57. Dabei wird anhand einer Risikoanalyse jeder Angriffsvektor hinsichtlich seiner Wahrscheinlichkeit und potentiellen Auswirkungen bewertet, um die Funktionalität und Sicherheit des Konzepts in verschiedenen Szenarien zu überprüfen. Diese Analyse zielt darauf ab, ein Proof-of-Concept-Modell vorzulegen, das die Integrität und Vertraulichkeit der Daten gewährleistet.

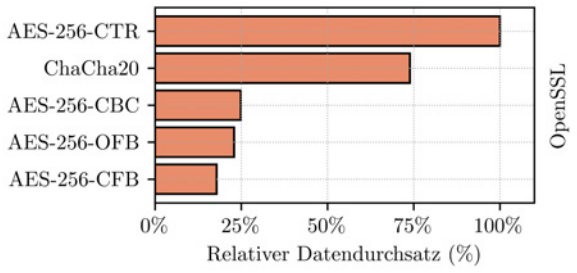

Abb. 1: Relativer Datendurchsatz von ausgewählten symmetrischen Verschlüsselungsalgorithmen implementiert in OpenSSL (AES und ChaCha20) mit einer Schlüssellänge von 256 Bit, relativ zum schnellsten Algorithmus (AES-256-CTR) gemessen [4]

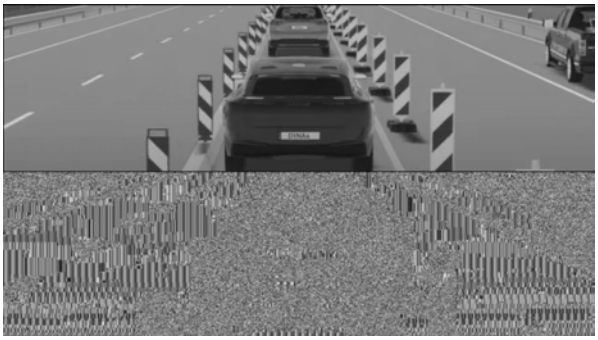

Abb. 2: Originalbild (oben, entnommen aus [6]), verschlüsselt mit AES-128 im Electronic-Codebook Modus (unten) [4]

Verwandte Arbeiten

In der Studie von HDibas und Khair Eddin Sabri [5] wird die Leistungsfähigkeit der symmetrischen Verschlässelungsalgorithmen wie AES, 3DES, Blowfish und Twofish auf Basis von Laufzeit, Speicherverbrauch und Ciphertextgröße im Verschlüsselungs- und Entschlüsselungsprozess verglichen. Die Tests erfolgen auf aktueller Hardware, welche gleich dem Datenlogger von der Firma Vector, einen AES-beschleunigenden Prozessor verwendet. Hierbei zeigt sich, dass der AES-Algorithmus bei Dateigrößen von 1KB bis zu 100MB schneller als die anderen Algorithmen verschlüsselt, was ihn zu einem vielversprechenden Algorithmus für dieser Arbeit macht. D. Bujari et al. führen in ihrer Studie [3] einen Vergleich verschiedener Betriebsmodi hinsichtlich Sicherheit, Effizienz und Leistung durch. Die vorliegende Arbeit greift Ergebnisse der Studie auf und evaluiert verschiedene Betriebsmodi im Kontext eines ADAS-Logging-System. Neben AES stellt Daniel J. Bernstein in seiner Studie [2] den Algorithmus ChaCha vor. Dieser Algorithmus zielt darauf ab, hohe Sicherheit bei gleichzeitig guter Leistung auf einer Vielzahl von Plattformen zu bieten. In dieser Studie zeigen die Autoren, dass ChaCha auch ohne Hardwarebeschleunigung eine konkurrenzfähige Leistung erbringen kann, wodurch sich der Algorithmus für die Analyse in dieser Arbeit eignet. Aufgrund der ständig steigenden Sicherheitsanforderungen benötigt man Kryptoagilität, um die Sicherheit eines Systems langfristig gewährleisten zu können. Die Studie von Nouri Alnahawi [1] zeigt die Probleme, die in einem Sicherheitssystem ohne Kryptoagilität entstehen können. Zudem stellt Alnahawi Methoden vor, die bei der Implementierung von Kryptoagilität beitragen können. Diese Methoden werden in der vorliegenden Arbeit angewandt.

Ergebnis

Diese Thesis soll Aufschluss darüber geben, welche symmetrischen Verschlüsselungsalgorithmen in einem zeitkritischen System mit sehr hoher Datenrate nutzbar sind. Anhand eines Testaufbaus sollen reale Szenarien simuliert werden. Nach der Analyse der gesammelten Daten lassen sich Empfehlungen ableiten, welche Algorithmen für den Einsatz in einem hochbandbreitigen ADAS-Logging-Systemen am besten geeignet sind.

Literatur und Abbildungen

[1] N. Alnahawi, N. Schmitt, A. Wiesmaier, A. Heinemann, and T. Graßmeyer. On the State of Crypto-Agility. In *18. Deutscher IT-Sicherheitskongress*. SecuMedia Verlags-GmbH, 2022.

[2] Daniel Julius Bernstein. ChaCha, a variant of Salsa20. In *Workshop record of SASC*, pages 3–5. European Network of Excellence in Cryptology, 2008.

[3] D. Bujari and E. Aribas. Comparative analysis of block cipher modes of operation. In *International Advanced Researches & Engineering Congress 2017 Proceeding Book*, pages 1–4. Recep Halicioğlu, 2017.

[4] Eigene Darstellung.

[5] H. Dibas and K. E. Sabri. A comprehensive performance empirical study of the symmetric algorithms: AES, 3DES, Blowfish and Twofish. In *Proceedings of the 2021 International Conference on Information Technology (ICIT)*, pages 344–349. Institute of Electrical and Electronics Engineers, 2021.

[6] Vector Informatik GmbH. Testing Camera-Based ADAS Functions with DYNA4. https://youtu.be/z5wdsccX-DUY, 2022.