



Informationstechnik

Hochschule Esslingen
University of Applied Sciences

IT-Innovationen

Band 28
Januar 2022

Grußwort des Dekans

Liebe Leserinnen und Leser,

In seiner Abschlussarbeit soll ein Student mit dem im Studium erworbenen Wissen und den erlangten Fertigkeiten eine klar umrissene Aufgabenstellung in einem vorgegebenen Zeitrahmen und mit einem vordefinierten Zeitbudget selbstständig lösen. Für das praxisnahe Studium an einer Fachhochschule liegt es nahe, Fragestellungen direkt aus der Industrie aufzugreifen und in enger Kooperation mit der Industrie, oftmals vor Ort im Industrieunternehmen, zu bearbeiten. So arbeitet er, noch als Student, genau in demselben Umfeld, das ihn nach seinem erfolgreichen Abschluss beim Eintritt in den Beruf erwartet.



Er arbeitet mit denselben Werkzeugen, zusammen mit potentiellen zukünftigen Kollegen. Die Fragestellungen sind dabei real, spiegeln die wahre Komplexität wider, ganz ohne die gewohnten Vereinfachungen aus dem bisherigen Hochschulalltag. Der Weg zum Ziel ist unscharf, Hindernisse müssen beseitigt werden, der Ausgang ist mitunter ungewiss. Häufig wird Neuland beschritten, neue Technologien sollen ausprobiert und evaluiert werden. Rasch ist der Student einziger Know-how-Träger und seinem Betreuer im Wissen voraus. Hier zeigt sich, was für den Erfolg entscheidend ist: vorhandenes Wissen in einen neuen Kontext einzupassen und zielgerichtet zu erweitern.

Bei der Betrachtung des vorliegenden Bandes der IT-Innovationen fällt die Vielfalt der zu lösenden Fragestellungen auf. Sie kommen aus sehr unterschiedlichen Kontexten. Ein Schwerpunkt lässt sich nicht ausmachen. So zeigen die Absolventen, dass sie mit den erlernten Methoden und Fertigkeiten, gepaart mit ihrem Wissen, vielfältige komplexe Aufgaben selbstständig auf hohem Niveau zu lösen wissen.

Machen Sie sich selbst ein Bild mit den vorliegenden Kurzbeschreibungen der Abschlussarbeiten des Abschlussjahrgangs Wintersemester 2021/22.

Viel Spaß beim Lesen wünscht Ihnen



Prof. Jürgen Nonnast

Dekan der Fakultät Informatik und Informationstechnik

IMPRESSUM

ERSCHEINUNGSORT

73732 Esslingen am Neckar

HERAUSGEBER

Prof. Jürgen Nonnast
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)711.397-4211
Telefax +49(0)711.397-4214
E-Mail it@hs-esslingen.de
Website www.hs-esslingen.de/it

REDAKTION, LAYOUT UND DESIGN

Prof. Dr.-Ing. Reinhard Schmidt
Hochschule Esslingen - University of Applied Sciences
Fakultät Informatik und Informationstechnik
Flandernstraße 101
73732 Esslingen am Neckar

SATZ, ANZEIGEN und VERLAG

Dipl.-Inf.(FH) Rolf Gassner
Tristan Löffler

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

ERSCHEINUNGSWEISE

Einmal pro Semester, jeweils Januar und Juni

DRUCK

Pixelgurus
Werbung – Werbetechnik – Digitaldruck.
Horbstraße 8
73760 Ostfildern

AUFLAGE

500 Exemplare

ISSN 1869-6457

Emre Can Altunkaya	Semantische Segmentierung von Verkehrsszenen	6
Marcel Autenrieth	Evaluierung für automatisierte Deployment-Methoden und Technologien für Kubernetes	8
Sinem Aylar	Stand des Influencer Marketing in Deutschland - Aufkommen, Bedeutung und empirische Untersuchung	10
Marcel Bauer	Verbesserung des klinischen Arbeitsablaufs in der Geburtshilfe- und Gynäkologie-Abteilung durch den Einsatz eines mobilen Ultraschallgeräts	13
Andreas Baulig	Improving Cartesian Plot Generation in a combined Multilateration/Passive Radar Data Fusion Application	16
Miguel Angel Beltra Mayoral	Analyse und Bewertung einer Teil-Integration (Hybride-Bereitstellung) einer bestehenden App in die Cloud	18
Ali Benli	Implementierung eines Algorithmus zur Stixel-Abstahierung auf Basis einer eigens konstruierten Stereokamera	20
Wolfgang Bradfish	Objekterkennung mit tiefen Neuronalen Netzen in der Logistikautomatisierung der Industrie 4.0	23
Tim Brecht	Konzeptionierung der Serialisierung von verpackten Kombinationsprodukten für einen Show Case	25
Senay Cakir	Konzeption und Entwicklung einer flexiblen Ablaufsteuerung zur Planung von Transport- und Arbeitsvorgängen in modularen Automationszellen	28
Ali Copurkuyu	Virtuelle Leadership Meetings als Element einer Kulturentwicklung	31
Alexander Dausel	Konzeption und Implementierung eines Konfigurationsprotokolls der EUCHNER Augmented Reality Submodule	33
Daniel Dippold	Secure Boot auf dem PicoCoreMX8MM	35
Anna Marie Dobler	Erstellung und Umsetzung eines Konzeptes zum Vergleich von Elektrofahrzeugen und Fahrzeugen mit Verbrennungsmotoren für Mercedes me connect mit Hilfe von Azure Databricks	39
Ebru Dogan	Verfahren zur Ermittlung der Support-Ticket Eskalationsstufe durch den Einsatz von Social Listening	42
Yannick Ege	Automatisierung der Regressionstests für eine Komplexe Applikationswelt	44
Jonathan Knuplesch	Gamification - Beurteilung der Einsatzmöglichkeiten im deutschen Finanzdienstleistungssektor	46
Marcel Gauss	Analyse, Entwurf und Implementierung einer Deep-Learning-Lösung zur visuellen Erkennung von Hand-Objekt-Interaktionen mit Werkzeugen	49

Jonathan Grau	Entwicklung eines Verkehrsschildsystems mit Maschinencode für eine autonome Buslinie zur Informationsgewinnung	53
Cengizhan Guenduezeri	Erhöhung der Prozessautomatisierung durch den Einsatz von RPA-Technologie an einem Beispielprozess bei GS/PE-SC	55
Umutcan Guerbuez	Prinzipien der IT-Sicherheit in der Automotive Cybersecurity	57
Kai Haeppler	Kamerabasierte Detektion von Fahrgassen und Pfadberechnung für landwirtschaftliche Fahrzeuge	60
Wolfgang Hamberger	Konzeption und Implementierung eines RFID-Firmwareupdate-Mechanismus für das EUCHNER Electronic-Key-System	63
Andreas Handel	Entwurf und Implementierung eines Damespiels unter Verwendung von Flutter mit der Auswahl eines geeigneten Backends und Kommunikationsprotokolls	66
Matthias Hauber	Konzept und prototypische Umsetzung der Portierung einer Ladeplattform zu einer Public-Cloud-Kubernetes-Umgebung unter dem Aspekt der Störungserkennung	69
Frank Holzmüller	Anwendung von Computer Vision zur Bewertung von hochlegierten Materialzuständen	72
Alexander Huebener	Analyse der Eignung der Programmiersprache Rust zur Entwicklung eines Kommunikationsstacks für eingebettete Systeme	75
Tom Jaschinski	Design and Implementation of Cryptographic-Key-Management-as-a-Service for Mission-critical Applications	79
Nicolas Kahle	Programmierung von Plugins zur Audiotbearbeitung und deren Einbindung in gängige DAWs	81
Kevin Kaiser	Performance und Speicheroptimierung einer bestehenden Netzwerkanalysesoftware	83
Radika Kapour	Strategie für die Testautomatisierung am Beispiel von C@P Marketing Porsche	85
Tunahan Kaya	Provisioning von Embedded Devices in der Azure Cloud	87
Jannick Keller	Konzeption und Implementierung eines Samplingalgorithmus für den Szenariobasierten XiL-Test für Fahrerassistenzsysteme	90
Jan Robin Kellner	Reengineering der Plugin-Verwaltung in der Desktop-Anwendung Elektra	92
Daniel Ketterer	Fluchtboote mit Satellitenbildern finden	95
Marei Kopf	Bewertung von IT-Unterstützungstools sowie von aktuellen Trends für Geschäftsprozesse in der Telekommunikation	97

Valdrin Krasniqi	Digitalisierung im Controlling - Stand und Potentiale	99
Eva Kuffler	Untersuchung der WLAN-Konfigurationsschnittstelle eines Steuerungssystems für Brenner unter dem Aspekt der Informationssicherheit	101
Florian Lauterbach	Reduktion der Nacharbeitskosten einer Ventilinsel	103
Marvin Louis	Automated reverse engineering of automotive bare metal firmware	106
Nico Maag	Evaluierung und Entwicklung eines Chatbots für die Gästekommunikation im Gastgewerbe	109
Moadh Mahmoud	Modeling and Simulation of reliability and availability for autonomous driving vehicles	112
Jo Maier	Single Pair Ethernet - An implementation of a 10Base-T1L physical layer for evaluation purposes	114
Daniel Pascal Mair	Herausforderungen und Potentiale der Migration auf Web-Technologien am Beispiel eines PLM-Systems	117
Amirhossein Marzani	Markendetektion mit maschinellem Lernen	119
Sueleyman Omurca	Requirements Engineering im Bereich Service Management zur Entwicklung einer neuen Datenquelle in einem Business Intelligence Analytics Tool	123
Yassine Ounajjar	Lidar-based localization and mapping for autonomous vehicle	126
Lara Rehbani	Empirische Untersuchung einer Scrum-basierten Arbeitsweise am Beispiel eines IT-Infrastrukturprojektes	128
Julian Schallmueller	Cloud Service Provider Evaluierung auf Basis von Infrastructure as Code Unterstützung	130
Matthias Schoelpfle	Evaluierung von Ähnlichkeitsmaßen für Zeichensequenzen im Machine Learning	133
Valentin Schuele	Batterieladesimulation zur Evaluierung von Lastverteilungsalgorithmen in einem Ladepark	135
Sven Schuler	Entwicklung eines Kalkulationsmodells für Softwaresysteme exemplarische Umsetzung für Siemens anhand von intralog WMS	139
Nicolas Schulz	Kinematic Calibration of a 6-DOF Robotic Manipulator	142
Timo Schwind	Untersuchung verschiedener Modelle des maschinellen Lernens hinsichtlich ihrer Eignung zur Unterscheidung von Anforderungstexten zu anderen Texten	145

Simon Seefeldt	Roboterassistiertes Container-Packing in der Kommissionierung	148
Imer Spatolaj	Transportverwaltung von ABAP-Coding mit GIT	151
Steffen Spieth	Optimierung des AET-Prozesses bei Mercedes-Benz Cars, Van, AMG & G-Klasse	153
Patrizia Spinola	Passwortlose Web-Authentisierungs-Konzepte in Industrieszenarien	155
Dominik Stach	Entwicklung eines generischen Beratungskonzeptes zur Einführung einer Predictive Analytics Lösung auf Basis der SAP Analytics Cloud	158
Jan Trillinger	Möglichkeiten und digitale Umsetzung eines Handels mit Emissionszertifikaten für ein Tiny House	161
Bahar Uzun	Design and implementation of a Data Quality Management Framework	164
Lora Venkova	Software-Geschäftsmodellentwicklung im Kontext der Bauteileoptimierung durch künstliche Intelligenz	166
Simon Weber	Konzeption und Implementierung einer Website zur Evaluation von 3D-Interaktionstechniken	169
Tobias Winkler	Analyse für den Einsatz von Machine Learning für die Security Log Korrelation und Alarmierung	171
Muhammed Yildirim	Entwicklung eines Vision-Objekterkennungs-Stacks für Autonomes Fahren	174

Semantische Segmentierung von Verkehrsszenen

Emre Can Altunkaya

Markus Enzweiler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Zielsetzung

In der vorliegenden Bachelorarbeit geht es um die Entwicklung eines Neuronalen Netzes zur pixelweisen Semantischen Segmentierung. Eine essenzielle Disziplin für das Autonome Fahren ist das Szenenverständnis. Diesbezüglich wird jedes Pixel, dass von der Kamera erfasst wird, einem Objekt zugeordnet und klassifiziert. Das Ziel dieser Arbeit ist es, ein neuronales Netz zu entwickeln, das eine Semantische Segmentierung auf Basis von Bilddaten durchführt. Das Neuronale Netz soll hierbei mit Hilfe eines Datensatzes trainiert werden. Dabei werden State-of-the-Art-Architekturen von Convolutional Neural Networks in Betracht gezogen.



Abb. 1: Beispielergebnis der semantischen Segmentierung [1]

Neuronale Netze

Künstliche Neuronale Netze beziehen sich auf das Neuronennetz des menschlichen Gehirns. Dies kann als Analogie und Inspiration für künstliche neuronale Netze verwendet werden, die in Computern simuliert werden. Durch Verwendung dieser komplexen Methode lassen sich Anwendungsprobleme aus den Bereichen Statistik, Wirtschaftswissenschaften und Technik effizienter lösen. Darüber hinaus sind künstliche neuronale Netze eine große Hilfe, um die Funktionsweise des menschlichen Gehirns wirksamer zu verstehen. Es ist durchaus ein attraktives Forschungsgebiet, das in der Industrie auf positive Resonanz stößt.

Funktionsweise eines neuronalen Netzes

Ein künstliches neuronales Netz ist eine Architektur, die üblicherweise aus mehreren Schichten besteht: einer

Input-Schicht, einer oder mehreren Zwischenschichten und einer Output-Schicht. Über die Input-Schicht werden Daten wie Bilder, Audiodaten oder Text in das neuronale Netz aufgenommen. Danach werden die eingespeisten Daten in den Zwischenschichten verarbeitet. Zum Schluss gibt das Netz die verarbeiteten Daten aus der Output-Schicht aus. In jeder Schicht befinden sich künstliche Neuronen, die über Gewichtungen miteinander vernetzt sind. Durch das Training des Netzes, wird versucht, die Gewichtungen des neuronalen Netzes durchgehend zu verbessern, um eine wirksamere Architektur zu erhalten. Wenn die Zwischenschicht aus mehreren Schichten besteht, wird von Deep Learning gesprochen.

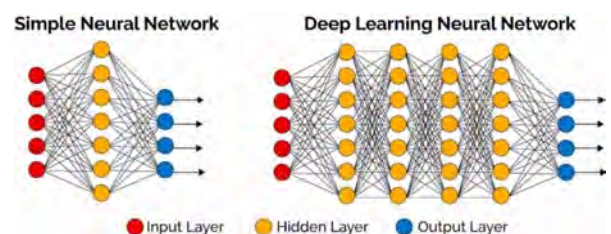


Abb. 2: Aufbau eines Neuronalen Netzes [6]

Convolutional Neural Networks

Convolutional Neural Network ist eine spezielle Deep Learning Architektur, das häufig in dem Bereich der Bilderkennung zum Einsatz kommt. Der Grund hierfür ist, dass die Architektur besonders auf die Verarbeitung von Pixeldaten anspricht. Hauptmerkmal der Architektur sind die sogenannten Filter (genannt Convolutional Layer). Mit den Filter lassen sich Faltungsoperationen über die Eingabedaten durchführen, um Merkmale wie Kanten zu extrahieren. Aufgrund dessen ist eine hohe Erkennungsleistung gewährleistet. Daraufhin folgt die Pooling-Schicht, über die man die Auflösung der erkannten Merkmale reduziert. Durch diese Operation lassen sich überflüssige Informationen verringern. Anschließend werden die Daten aus der vollständig verknüpften Schicht ausgegeben. [5]

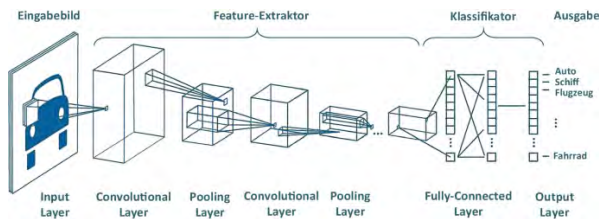


Abb. 3: Convolutional Neural Networks Architektur [4]

Semanische Segmentierung

Die semantische Segmentierung ist heute eines der Schlüsselthemen in der Computer Vision. Aus der Tatsache, dass immer mehr Anwendungen darauf angewiesen sind, Wissen aus Bildern zu erlangen,

steigt die Bedeutung des Verständnisses der Szene. Semantische Segmentierung ist der Prozess der Klassifizierung jedes Pixels, das zu einer bestimmten Objektklasse gehört. Eine allgemeine semantische Segmentierungsarchitektur kann als ein Encoder und Decoder-Netzwerk angesehen werden: Als Encoder wird ein Klassifizierungsnetzwerk bezeichnet, das die Aufgabe hat, Merkmale aus dem Bild zu extrahieren. Der Decoder versucht, die gelernten Merkmale, die eine niedrige Auflösung haben, semantisch auf eine höhere Auflösung zu übertragen. Schließlich erhält das System ein semantisch segmentiertes Bild, womit eine Maschine erhöhte Einsicht in eine Szene erhält. Im Gegensatz zur einfachen Erkennung eines einzelnen Objekts werden bei der semantischen Segmentierung auch Infrastrukturelemente wie Straßen und Gehwege abgebildet. [2]

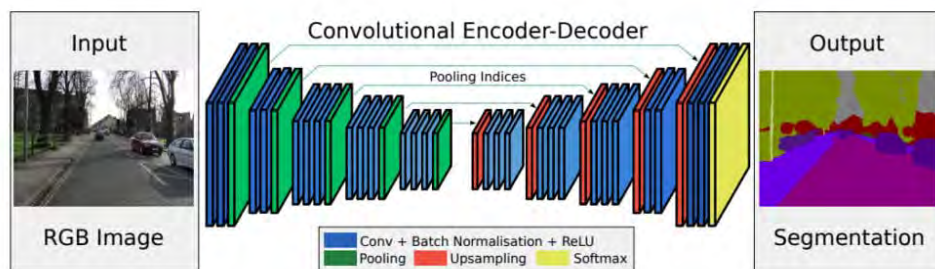


Abb. 4: Generelle Netzwerk-Architektur der pixelweisen Segmentierung [3]

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Anil Chandra Naidu Matcha. A 2021 guide to Semantic Segmentation. <https://nanonets.com/blog/semantic-image-segmentation-2020/>, 2021.
- [3] Derrick Mwit. The Definitive Guide to Semantic Segmentation for Deep Learning in Python. <https://cnvrg.io/semantic-segmentation/>, 2021.
- [4] Raghav Prabhu. Understanding of Convolutional Neural Network (CNN) — Deep Learning. <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>, 2018.
- [5] Christopher Thomas. An introduction to Convolutional Neural Networks. <https://towardsdatascience.com/an-introduction-to-convolutional-neural-networks-eb0b60b58fd7>, 2019.
- [6] Laurenz Wuttke. Deep Learning: Definition, Beispiele & Frameworks. <https://datasolut.com/was-ist-deep-learning/>, 2021.

Evaluierung für automatisierte Deployment-Methoden und Technologien für Kubernetes

Marcel Autenrieth

Kai Warendorf

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Novatec Consulting GmbH, Leinfelden-Echterdingen

Einleitung

Eine der großen Herausforderungen in der Softwareentwicklung ist der Prozess von der Idee einer Software oder einer neuen Softwareversion bis zur Auslieferung dieser an den Kunden oder Benutzer. Die Frage, mit der sich ein Softwareentwickler hier beschäftigen muss, lautet: Wie bekomme ich möglichst gute Software schnell zum Kunden? Für Unternehmen kann es hierbei von entscheidender Bedeutung sein, wie lange der Prozess von der Idee bis zur Auslieferung dauert. Ist ein Kunde unzufrieden, haben sich Fehler eingeschlichen oder benötigt er neue Features, so muss meist schnell reagiert werden, um die Kunden und Benutzer nicht an die Konkurrenz zu verlieren. Da es sich sowohl beim Bauen der Software sowie bei den automatisierten Tests und dem Ausliefern dieser auf die Produktivumgebung meist um sich wiederholende Schritte handelt, wurden für diese Schritte sogenannte Continuous Integration und Continuous Delivery bzw. Continuous Deployment Tools entwickelt. Sie automatisieren diese Prozesse und machen sie so reproduzierbar und testbar. Mit der Evolution von der Entwicklung und Auslieferung direkt auf der Maschine hin zur Entwicklung und Auslieferung von Software in virtuellen Maschinen und mittlerweile in Containern konnte der Auslieferungsprozess weiter vereinfacht werden. Der Trend weg von einer traditionellen monolithischen Architektur hin zur Microservice-Architektur beschleunigte diesen Prozess weiter. Außerdem entstand für eine bessere Zusammenarbeit zwischen der Entwicklung und dem IT-Betrieb sowie der Qualitätssicherung der DevOps Ansatz. Der Erfolg von Firmen, die die Migration in die Cloud erfolgreich geschafft haben, darunter beispielsweise Netflix, führte dazu, dass immer mehr Organisationen ihre Software für die Cloud entwickeln und in dieser betreiben. Die Vorteile, die daraus resultieren sind unlimitierte Skalierbarkeit, flexible Kosten durch Pay-per-Use, keine Ausfallzeit, Agilität und nicht mehr die eigene Infrastruktur betreiben zu müssen und sich so auf das eigentliche Kerngeschäft,

die Entwicklung von Software konzentrieren zu können. Kubernetes hat sich dabei als quasi Industrie Standard zur Orchestrierung von in Containern verpackten Microservices durchgesetzt.

Ziel der Arbeit

Das Ziel dieser Arbeit ist es, unterschiedliche Kubernetes native GitOps Tools anhand zuvor festgelegter Kriterien zu vergleichen, zu bewerten und die Notwendigkeit dieser Tools zu betrachten und daraus zu erkennen, welches Tool für welchen Einsatzzweck geeignet ist. Kubernetes nativ meint dabei Tools, die speziell für Kubernetes entwickelt wurden, also auch die Sprache von Kubernetes sprechen und auch selbst in einem Kubernetes Cluster betrieben werden. Das Hauptaugenmerk soll bei dieser Arbeit auf dem Deployment liegen und nicht auf dem gesamten CI/CD Prozess.

DevOps

Damit eine Software zur Wertschöpfung für ein Unternehmen beitragen kann, ist es wichtig, diese möglichst schnell in die Produktivumgebung auszuliefern. In der Entwicklung der Software wird aber meist kaum darauf geachtet, wie diese in der Produktivumgebung in einen einsatzfähigen Zustand gebracht werden kann. Dieses Problem zwischen dem IT-Betrieb und der Entwicklung und der durch die Trennung unzureichende Kommunikation wird auch Wall of Confusion genannt. Um eine bessere Kommunikation zu schaffen und auch kürzere Release-Zyklen zu ermöglichen, wurde der DevOps Ansatz entwickelt. Dabei handelt es sich nicht nur um eine Organisationsstruktur, sondern um eine Änderung der Denkweise mit einem Schwerpunkt auf Automation. Im Optimalfall gibt es dabei keine Trennung zwischen der Entwicklung (Dev) und dem IT-Betrieb (Ops) mehr. Beides geschieht im selben Team, damit Fehler und Missverständnisse bei der Übergabe zwischen den Teams nicht mehr entstehen.

Kubernetes

„Kubernetes ist ein Softwaresystem, das die einfache Bereitstellung und Verwaltung von Anwendungen in Containern ermöglicht“ [2]. Das Open-Source-Projekt wurde im Juni 2014 in der ersten Version released. Mit Kubernetes lassen sich Anwendungen auf Tausenden von Computerknoten ausführen, als wären diese ein einziger großer Computer. Es übernimmt die Bereitstellung der Anwendung und bietet weitere Funktionen wie die Skalierung, Lastenausgleich oder Selbstheilung. Kubernetes ist deklarativ. Das bedeutet, der Benutzer definiert den gewünschten Zustand des Systems und Kubernetes sorgt dafür, dass dieser hergestellt wird.

GitOps

Der Begriff GitOps wurde von Alex Richardson von Weaveworks eingeführt und basiert ebenfalls auf dem deklarativen Ansatz. Bei GitOps handelt es sich um ein neues Konzept, bei dem der IT-Betrieb weiter automatisiert werden soll [1]. Der Fokus dieses Ansatzes liegt dabei mehr auf dem Betrieb im Vergleich zum klassischen DevOps. Die Grundidee des GitOps Ansatzes ist, dass ein Git-Repository als die einzige verlässliche Datenquelle im Mittelpunkt steht. Es enthält die Konfigurationsdateien, welche das Systemabbild deklarativ beschreiben. Beim klassischen imperativen Continuous Delivery mit CI-Server, welches in Abbildung 1 dargestellt ist werden die Builds und Tests vom CI-Server imperativ durchgeführt und die im Build erstellten Applikationen auf der Registry abgelegt und anschließend nach dem Push-Prinzip ausgeliefert.

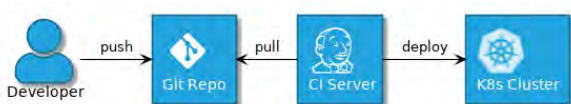


Abb. 1: Klassisches Continuous Delivery mit CI-Server [1]

GitOps trennt kontinuierliche Integration und den Deployment Prozess einer Anwendung. Dies ist in Abbildung 2 dargestellt. Der Prozess der Auslieferung wird nicht mehr direkt im Git-Repository mit einer

Pipeline ausgeführt. Ein Watcher in der Betriebsumgebung synchronisiert den Zustand direkt aus dem Git-Repository nach dem Pull-Prinzip. Der Sollzustand, der in den Konfigurationsdateien im Git-Repository definiert ist, wird mit dem Istzustand des Systems verglichen. Weicht der Istzustand vom Sollzustand ab, so werden die nötigen Maßnahmen unternommen, um diese Divergenz zu beheben.

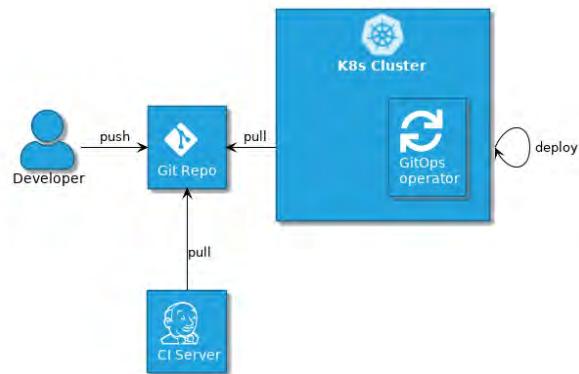


Abb. 2: Continuous Deployment mit GitOps [1]

Mit diesem Ansatz müssen nun keine Zugangsdaten zur Produktivumgebung in einer CI/CD Pipeline gespeichert werden. Dies führt zu einer höheren Sicherheit. Auch die Entwicklernähe ist ein Vorteil des Ansatzes, da Git mittlerweile als Basiswissen eines Entwicklers angesehen werden kann.

Ausblick

Anhand einer kleinen Microservice-Anwendung sollen die Open-Source GitOps Tools Argo CD und Flux miteinander verglichen werden. Die Kriterien für den Vergleich sind die Einfachheit der Installation, die Bedienbarkeit, die Darstellung bzw. UI falls vorhanden, die Konfigurierbarkeit, der Umfang der Tools und die zugehörige Dokumentation. Um die Tools objektiv miteinander zu vergleichen, werden diese jeweils auf demselben Kubernetes-Cluster unter den gleichen Bedingungen betrieben und die Anwendung jeweils unter den gleichen Voraussetzungen ausgeliefert. Die Beschreibungen der Kubernetes Ressourcen werden in einem privaten GitHub Repository abgelegt.

Literatur und Abbildungen

- [1] Cloudogu GmbH. GitOps in der Softwareentwicklung. <https://cloudogu.com/de/glossar/gitops/>, 2021.
- [2] Marko Lukša. *Kubernetes in Action: Anwendungen in Kubernetes-Clustern bereitstellen und verwalten*. Carl Hanser Verlag, 2018.

Stand des Influencer Marketing in Deutschland - Aufkommen, Bedeutung und empirische Untersuchung

Sinem Aylar

Anke Bez

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Die mit dem digitalen Wandel korrelierende vermehrte Nutzung des Internets brachte Unternehmen dazu, sich mit einer neuen Form des Marketings auseinander zu setzen, dem sogenannten Online-Marketing. Laut einer Studie der Faktenkontor GmbH zufolge halten sich 76 Prozent der Deutschen ab 16 Jahren in den sozialen Medien auf. [2] Dies führt dazu, dass sich vermehrt Unternehmen in den diversen Social-Media-Kanälen engagieren, um sich mit ihren Kunden zu vernetzen, denn die Devise lautet: Den Kunden dorthin zu folgen, wo dieser sich aufhält. Daraus resultierend entstand ein neuer Ansatz des Marketings, die sich dem Online-Marketing zuordnen lässt und darauf abzielt, Vermarktungsziele durch den zielorientierten Einsatz von sozialen Medien zu erreichen: das Social-Media-Marketing [3] Das Grundprinzip des Push-Marketings verliert durch seinen hohen werblichen Charakter an Bedeutung. Folglich nehmen das Vertrauen und Interesse an klassischer Werbung ab. [4] Die Konsumenten legen mehr Wert auf die Meinung und Empfehlung von Familie, Freunden und Bekannten. Tatsache ist, dass sich die Menschen schon immer untereinander und gegenseitig beeinflussen. Auch in der digitalen Welt hat sich daran nichts verändert. Das Internet und besonders die Social-Media-Kanäle stellen einen (virtuellen) Ort dar, an dem Meinungen frei geäußert und mit anderen geteilt werden können.

Aufkommen

Zunächst startete das Format des Blogs, wo sogenannte Blogger eine Art Online-Tagebuch führten. Dadurch, dass die Blogger Artikel veröffentlichten, in denen sie auf authentische Weise über ihre Erfahrungen berichteten, wurde dies nicht als Werbung wahrgenommen. Vielmehr fand ein Austausch zwischen Gleichgesinnten statt, ohne dass eine Werbebotschaft aufgezwungen wurde. Unternehmen erkannten, welches Potenzial sich dahinter verbirgt, und nach kurzer Zeit etablierte sich das Phänomen der Meinungsführer auch in den

Social-Media-Plattformen unter dem Begriff Influencer (englisch to influence). Influencer sind Personen mit einer großen Reichweite in den sozialen Medien, die zu einem bestimmten Themenbereich Content (Inhalte) kreieren und veröffentlichen. Mittlerweile haben sich Influencer als Markenbotschafter in der Unternehmenskommunikation fest etabliert und übermitteln in deren Auftrag Werbebotschaften. Diese Form des Marketings, für die der Begriff des Influencer Marketing geprägt wurde, erlangte im Jahr 2016 hohe mediale Aufmerksamkeit in Deutschland [5] und lässt sich dem Social-Media-Marketing zuordnen. Die Auswahl der passenden Influencer für Unternehmen spielt eine entscheidende Rolle, denn nicht jeder Meinungsführer eignet sich für jeden Zweck. Demnach gibt es verschiedene Kriterien (vgl. Abb. 1), nach denen man Influencer segmentieren kann:

Social-Media-Kanal	Themenbereich	Soziodemografische Merkmale	Reichweite (Abonnenten)
Blog	Fitness	Geschlecht	Nano (1K – 10)
YouTube	Fashion	Alter	Micro (10K – 50K)
Instagram	Food	Sprache	Macro (50K – 1Mio)
TikTok	Travel		Mega (1Mio +)
	Beauty		

Abb. 1: Typisierung von Influencern [1]

Bedeutung

Das Internet ist häufig die erste Anlaufstelle, wenn man auf der Suche nach Informationen ist und dient oftmals als Entscheidungsgrundlage. Die durch die Nutzer generierten Inhalte in den sozialen Medien stellen eine persönliche Meinung und somit eine Informationsquelle dar. Dies machen sich Unternehmen zunutze und setzen kommunikationsfreudige Anhänger der eigenen Marke als Markenbotschafter ein, um ihre Marketingziele zu erreichen. Sie verbreiten Inhalte über Produkte, Marken und Werbebotschaften und beeinflussen dadurch andere Personen sowie deren Kaufverhalten. Durch die Einbringung der Influencer resultiert eine

größere Aufmerksamkeit für die Unternehmen, was zur Erreichung von Vertriebszielen beiträgt. Weiterhin können Unternehmen ihre Kundenbasis um jüngere Zielgruppen erweitern. Besonders jüngere Generationen stellen den Fernseher auf stumm, sobald Werbung läuft, und nehmen das Smartphone zur Hand. Dadurch, dass Influencer in einer Beziehung zu ihrer Community stehen und eine gewisse Nahbarkeit entsteht, nehmen die Empfänger der Werbebotschaft diese als authentisch wahr. Hinzu kommt, dass der wesentliche Unterschied zwischen dem Influencer Marketing und herkömmlichen Marketingformen die Vermittlung der Botschaft ist. Diese erfolgt nicht wie gewohnt durch das Unternehmen selbst, sondern durch eine dritte individuelle Person, zu der die Konsumenten eine emotionale Bindung haben und dadurch die Werbebotschaft als viel glaubwürdiger aufnehmen.

Implementierung

Der gekonnte Einsatz des Influencer Marketing ist nicht nur effektiv für Unternehmen, sondern bringt auch einen Wettbewerbsvorteil gegenüber Konkurrenten, die nicht in den sozialen Medien vertreten sind. Unternehmen sind eng mit ihren Kunden vernetzt und es besteht eine ständige Kommunikation. Diese Form der Interaktion hilft dabei, eine Beziehung zum Kunden aufzubauen. Außerdem erhalten Unternehmen öffentlich geteiltes Feedback (zum Beispiel über die Kommentare), was als Möglichkeit genutzt werden kann, die Kundenzufriedenheit zu steigern. Die Umsetzung des Influencer Marketing umfasst einen zeitaufwendigen Prozess, der wie in Abbildung 2 dargestellt die Beachtung einiger Schritte bedingt.

Im Allgemeinen beschreiben Ziele einen angestrebten Zustand und sind notwendig für die Gestaltung einer Strategie. Doch bevor die Ziele definiert werden können, muss die Zielgruppe festgelegt und deren Onlinenutzungsverhalten untersucht werden, um zu sehen, wo und ob überhaupt sich die Zielgruppe in den sozialen Medien aufhält. Die Analyse des Nutzungsverhalten der anvisierten Zielgruppe dient nun auch als Vorlage, um die richtigen Kanäle und die passenden Content-Formate zu bestimmen. Mit der fortschreitenden Professionalisierung des Influencer Marketing besteht ein Übermaß an Influencern, worunter je nach Unternehmen und dem zu werbenden Produkten auch die passenden Influencer ausgewählt werden müssen. Die Suche und Auswahl wird durch Software-Tools unterstützt. Bevor die Kooperation nun starten kann, muss die Zusammenarbeit und Kampagne ausgestaltet werden. Dies beinhalten unter anderem die Budgetierung, Vergütung, Laufzeit und Inhalt der Kampagne.

Ausblick

Das Influencer Marketing hat sich den letzten Jahren weitestgehend professionalisiert und sich in den Marketing-Mix vieler Unternehmen etabliert. Aufgrund der dynamischen und schnelllebigen Eigenschaften der digitalen Welt sollte die Fortentwicklung des Influencer Marketing jedoch kritisch beobachtet werden. Mit wachsender Reichweite und zunehmender Anzahl an Kooperationen kann die Beliebtheit eines Influencers von heute auf morgen nachlassen und das wiederum geht mit dem Verlust der Glaubwürdigkeit einher.



Abb. 2: Umsetzung des Influencer Marketing [1]

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Faktenkontor GmbH. Social-Media-Nutzung 2021: Saarländer haben die Nase vorn. <https://www.faktenkontor.de/presse-meldungen/social-media-nutzung-2021-saarlaender-haben-die-nase-vorn/>, 2021.
- [3] Uwe Hettler. *Social Media Marketing: Marketing mit Blogs, Sozialen Netzwerken und weiteren Anwendungen des Web 2.0*. Oldenburg Wissenschaftsverlag, 2010.
- [4] Thomas Klein. *Der neue Corporate Influencer: Effizientes Social-Media-Marketing mit einem internen Content Creator*. Wiesbaden, Springer Fachmedien, 2020.
- [5] Erwin Lammenett. *Praxiswissen Online-Marketing: Affiliate-, Influencer-, Content- und E-Mail-Marketing, Google Ads, SEO, Social Media, Online- inklusive Facebook-Werbung*. Wiesbaden, Springer Fachmedien, 2019.

Verbesserung des klinischen Arbeitsablaufs in der Geburtshilfe- und Gynäkologie-Abteilung durch den Einsatz eines mobilen Ultraschallgeräts

Marcel Bauer

Harald Melcher

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Philips Medizin-Systeme Böblingen GmbH, Böblingen

Einleitung

Die typischen Vorsorgeuntersuchungen während einer Schwangerschaft sind: regelmäßige Blutdruck- und Gewichtsmessung, Analyse der Urinwerte, Stand der Gebärmutter ertasten, sowie Kontrolle der Lage und der Herztöne des Kindes. Außerdem werden in einem normalen Schwangerschaftsverlauf in den 12., 24. und 48. Schwangerschaftswochen Ultraschalluntersuchungen durchgeführt [2]. Für die Dokumentation der Vorsorgeuntersuchung werden Kreißsaal-Datenmanagementsysteme für die klinische Dokumentation medizinischer Versorgung von Mutter und Kind in der Perinatalen Versorgung eingesetzt. Philips bietet in diesem medizinischen Bereich die Systemsoftware IntelliSpace Perinatal (im weiteren Verlauf ISP genannt) an. Da ISP kein Interface zu einem Ultraschallgerät anbietet um die Bilder automatisch dokumentieren zu können werden öfters die Bilder in ISP wegen dem manuellen Aufwand nicht erfasst, somit ist die vollständige Kette der Dokumentation unterbrochen. Diese Arbeit beschäftigt sich mit der Ausarbeitung eines Ablaufs im Bereich der Geburtshilfe Versorgung, dessen Verbesserung durch ein mobiles Ultraschallgerät und einer dazugehörigen Programmiererweiterung für ISP.

Analyse und Ziel

Das Ziel dieser wissenschaftlichen Arbeit ist einen klinischen Arbeitsablauf in der Geburtshilfe zu identifizieren, bei dem ein stationäres Ultraschallgerät eingesetzt wird, und diesen mit Hilfe moderner, mobiler Ultraschallgeräte zu verbessern. Dabei wird der heutige Ablauf detailliert analysiert und seine Probleme/Schwierigkeiten aufgelistet. Die Identifizierung und Bewertung aus der klinischen Perspektive resultiert sich aus Expertengesprächen mit Ärzten und Hebammen im Universitäts-Frauenklinikum Tübingen. Das Klinikum verwendet bereits ISP als System auf der Geburts-

hilfestation, daher eignet sich ihre Expertise sehr gut für eine Auswertung und Validierung einer Umsetzung in ISP. Dies soll sicherstellen, dass eine zukünftige Verwendung der in die ISP App integrierten Features einen Mehrwert für den klinischen Arbeitsalltag mit sich bringt.

Die Software IntelliSpace Perinatal

IntelliSpace Perinatal (ISP) ist ein Kreißsaal Datenmanagementsystem für die klinische Dokumentation der medizinischen Versorgung von Mutter und Kind in der perinatalen Versorgung. In Abbildung 1 wird eine mütterliche und 3 fetale Herzfrequenzen, Kindeslage, Blutdruck, Muttermundöffnung und Temperatur dargestellt. Die Funktionen in ISP umfassen sowohl die Dokumentation von Parametern wie Sauerstoffsättigung, fetale Herzfrequenz (FHF) und mütterliche Herzfrequenz (MHF), Kindsbewegungen und Wehentätigkeit, als auch die Überwachung und Alarmfunktionen. Die Überwachung und Begleitung der Schwangeren kann von der ersten Vorsorgeuntersuchung über die Entbindung, bis hin zur Entlassung vollständig erfolgen und in einer digitalen Krankenakte dokumentiert und gespeichert werden.

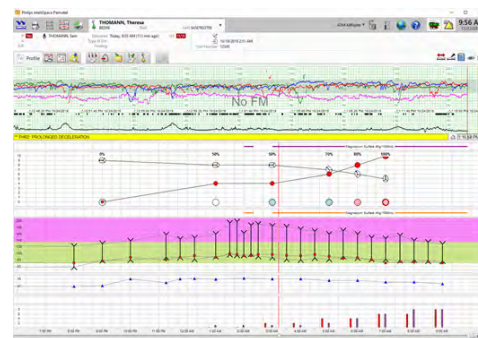


Abb. 1: Komplette Übersicht der Hauptfunktionen von ISP [1]

Mobiles Ultraschallgerät Philips Lumify

Im Jahr 2017 führt Philips das mobile Ultraschallgerät Lumify in den Markt ein. Lumify dient zum Erfassen von Ultraschallbilddaten, die von Ärzten und Kliniken für Screening-, Diagnose- und Verfahrenszwecke eingesetzt werden können. Mit dem Produkt soll das Erfassen klinisch akzeptabler Bilder und Ultraschalldaten für klinische Anwendungen und Anatomien ermöglicht werden. In der mobil Applikation von Lumify können Untersuchungen durchgeführt und exportiert werden [4].



Abb. 2: Mobiles Ultraschallgerät Philips Lumify [1]

DICOM-Server

Eine der Export Möglichkeiten ist ein DICOM-Server. Ein DICOM-Server ist ein digitales System zur Verarbeitung, Verwaltung und Archivierung von medizinischen Bildern und Daten. Hier können Bilddaten aller Medizingeräte, wie zum Beispiel Röntgen, Sonografie, CT, MRT oder Endoskopie gespeichert werden. Von dem DICOM-Server aus können dann die Bilder an Befundungs-, Betrachtungs- und Nachverarbeitungsplätze zu Verfügung gestellt werden. Ein DICOM-Server kann für eine bestimmte Abteilung oder ein gesamtes Krankenhaus bezogen werden [3].

Ultraschalldatenaustausch

Der Ultraschall Datenaustausch zwischen ISP und dem DICOM-Server findet über den DICOM-Web Standard statt. In Abbildung 3 wird der gesamte Datenaustausch grafisch dargestellt. Als erstes wird mit dem Mobilen Ultraschallgerät Philips Lumify ein Ultraschallbild aufgenommen und exportiert. Diese Daten werden dann an den DICOM-Server geschickt, dieser speichert und verwaltet die Ultraschallbilder in einer Datenbank. Wenn nun die Hebamme die Ultraschallbilder in ISP dokumentieren und kategorisieren möchte, öffnet sie die Krankenakte der Patientin in ISP. Dann klickt sie auf einen Button und die Daten werden spezifisch für diese Patientin vom DICOM-Server angefordert. Der DICOM-Server sendet zu der Anfrage die spezifischen Ultraschalldaten und die Hebamme kann diese dann in ISP dokumentieren und kategorisieren.

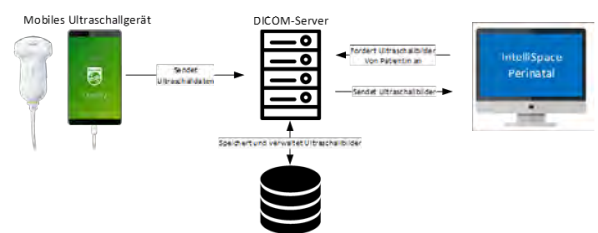


Abb. 3: Interaktion zwischen Lumify und ISP [1]

Verbesserung des Workflows

Zum Testen der Verbindung zwischen dem DICOM Server und ISP soll eine C# basierte Konsolenanwendung implementiert werden, die über die REST-API vom DICOM-Server und dem DICOMweb Standard mit diesem kommuniziert, um so die Ultraschalldaten einer speziellen Patientin zu bekommen. Dabei soll in der Konsolenanwendung für die Demonstration einer echten Patientin eine Patienten-ID und das Datum der Untersuchung, die man dokumentieren möchte, eingegeben werden. Zurück soll man dann die dazugehörige DICOM Datei komprimiert als PNG-Datei bekommen, dieses Bild kann dann angezeigt werden. Um den Workflow endgültig zu verbessern soll nach der Bachelorarbeit dieser Prototyp in ISP implementiert werden.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Bundeszentrale für gesundheitliche Aufklärung. Ultraschall-Untersuchung in der Schwangerschaft. <https://www.familienplanung.de/schwangerschaft/schwangerschaftsvorsorge-und-begleitung/vorsorge-zur-sicherheit/>, 03 2019.
- [3] Telemis GmbH. Was ist ein PACS? <https://www.itz-medi.com/was-ist-ein-pacs/>, 07 2017.
- [4] INC Philips Ultrasound. Lumify-Ultraschallsystem. https://www.philips.com/c-dam/b2bhc/master/feature-details/ultrasound-lumify-global-resources/manual_german.pdf, 04 2017.

Improving Cartesian Plot Generation in a combined Multilateration/Passive Radar Data Fusion Application

Andreas Baulig

Clemens Klöck

Department of Computer Science and Engineering, Esslingen University

Work carried out at Hensoldt Sensors GmbH, Ulm

Introduction

A Multilateration (MLAT) system is a type of cooperative independent surveillance system, that uses electromagnetic emissions of a target to determine its position. Employed as Wide-Area Multilateration (WAM) systems, this technique is used to determine an aircraft's location based on its Automatic Dependent Surveillance-Broadcast (ADS-B) transmissions.

Most systems use a three step process. First each station notes the Time of Arrival (TOA) of the arriving squitter message and forwards this information to a central processing node. There, one TOA from the set of all received TOAs is selected as reference. Subsequently, all other TOA values are subtracted by the chosen reference, producing a set of Time Differences of Arrival (TDOA). The third step involves finding intersections of the hyperboloids (or hyperbolas in 2D) defined by these TDOAs. Figure 1 illustrates the underlying principle in 2D. Depicted are three receiving stations titled Rx1 to Rx3. Rx1 is chosen as reference arbitrarily. TOAs from the other stations are compared against this reference to obtain the set of TDOAs. Each TDOA defines a hyperbola of possible target locations. For a system of $N + 1$ stations, the general hyperbola equation is given by

$$\Delta_i = \|\mathbf{p} - \mathbf{h}_i\|_2 - \|\mathbf{p} - \mathbf{h}_0\|_2,$$

where $\Delta_i = c(t_i - t_0)$, t_i being the TOA of the i -th node, t_0 being the TOAs of the reference station, and c being the speed of light. The target location in 3D-space is defined as $\mathbf{p} = [x, y, z]^T$. Similarly, $\mathbf{h}_i = [x_i, y_i, z_i]^T$, and $\mathbf{h}_0 = [x_0, y_0, z_0]^T$ define the position of the i -th node and reference station respectively. The $\|\cdot\|_2$ operator is defined as the L^2 vector norm.

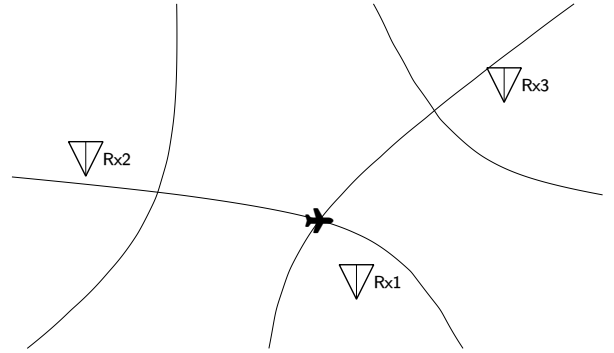


Fig. 1: Principle of the TDOA-based hyperbola intersection method. [5]

Various algorithms have been proposed to achieve position estimation based on TDOAs. A direct solution under the assumption of noise-free measurement known as Kleusberg's algorithm is proposed in [3]. Chan & Ho [1] use a two-step weighted least squares approach, wherein a quadratic equation has to be solved. Fränken [2] expands on this approach by deriving the parameter vector as a 6-th degree polynomial, the real roots of which lead to possible position estimates.

A common theme in these algorithms is, that they produce ambiguous estimates. Kleusberg and Chan & Ho both solve quadratic formulas, and thus, produce up to two real valued solutions. The application of position constraints, such as $\hat{z} > 0$ or $\arg \max_{\hat{z}} (\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2)$, where \hat{z} represents the height component of the position estimate $\hat{\mathbf{p}} = [\hat{x}, \hat{y}, \hat{z}]^T$, may already produce acceptable results. In case the discriminant of the quadratic formula becomes < 0 , such as through the introduction of measurement error, Kleusberg and Chan & Ho may produce two conjugated complex solutions. In such a case [4] proposes to take only the real part of either estimate, without further elaborating on this decision. [6] on the other hand, derives a decision tree, based on the sign of the quadratic term. However, so far no generic approach for an arbitrary sensor network has been proposed. Fränken's method

may produce up to six real solutions, where at least one solution is guaranteed to yield a positive range estimate. Position constraints, such as those mentioned earlier, may further refine the selection.

Motivation

Combining the measurement principles of Passive Radar (PR) and MLAT is expected to produce improved estimation results. While MLAT tries to find the target's location through intersecting hyperboloids, in PR the position is determined through intersecting ellipsoids. The idea is to co-locate MLAT and PR receiving stations. This typically results in higher intersection angles between hyperboloids and ellipsoids at the target location, thus decreasing Geometric Dilution of Precision (GDOP).

Data fusion of MLAT and PR can be done on different levels. 1) on a plot level. By intersecting hyperboloids and ellipsoids, this method would allow plot generation in situations where either measurement principle alone does not deliver enough data for plot generation. 2) on a track level. Assuming Cartesian plot generation is successful, plots from MLAT could allow improved track initialization and updating.

This thesis focuses on the latter method, by trying to improve Cartesian plot generation for a four and six sensor MLAT system. To this end, multiple estimation algorithms mentioned above are compared and rolled out in Hensoldt's existing tracking system. Great relevance is given to statistical analysis of GDOP and radial displacement in target estimation. The last phase of the thesis investigates the impact of expanding the existing four sensor network by an additional two.

This includes comparison of simulated performance using the well established Cramer-Rao Lower Bound (CRLB) against actual measurements.

State of Progress

Two algorithms, namely Kleusberg's [3] and Fränken's [2], were selected to be subject of investigation. Both algorithms were implemented in a simulated MATLAB environment. This allows different noise sources to be applied onto the artificial TOA measurements, the impact of which is analyzed through Monte Carlo simulations. Through these, the amplification effect of geometric factors on estimation error can be demonstrated. Predictions made by simulations of Kleusberg's algorithm have been validated by real measurements.

In Hensoldt's tracking system, the existing plot estimation routines for MLAT were extended by Fränken's method, to allow processing of real data. Additionally, the existing plot estimator has been readied for live operation.

Further Work

Though changes in the sensor network have been investigated through CRLB simulations, the true impact of adding two additional sensors remains to be validated by real measurements. However, said sensors are yet to be installed on location.

While simulations of GDOP have been successfully validated by real data, work continues to produce generalizable results based on the obtained insights.

References and figures

- [1] Y.T. Chan and K.C. Ho. A simple and efficient estimator for hyperbolic location. *IEEE Transactions on Signal Processing*, 42:1905–1915, 1994.
- [2] Dietrich Fränken. A robust algorithm with close to CRLB performance for localisation using TDOA measurements. In *Proceedings of the DGON International Symposium on Enhanced Solutions for Aircraft and Vehicle Surveillance Applications (ESAVS)*, pages 1–6. EVAS Enhanced Solutions for Aircraft and Vehicle Surveillance, 2013.
- [3] Alfred Kleusberg. Analytical GPS navigation solution, Quo vadis geodesia. *Festschrift for EW Grafarend on the occasion of his 60th birthday*, Eds. F. Krumm and VS Schwarze, Report, 1999.
- [4] Mateusz Malanowski and Krzysztof Kulpa. Two Methods for Target Localization in Multistatic Passive Radar. *IEEE Transactions on Aerospace and Electronic Systems*, 48:572–580, 2012.
- [5] Own representation.
- [6] Pei Xinxin, Huang Zhigang, Zhang Jun, Zhu Yanbo, and Liu Wei. An approach for close-form solution of multilateration equations. In *2010 8th World Congress on Intelligent Control and Automation*, pages 5464–5468. ICA Intelligent Control and Automation, 2010.

Analyse und Bewertung einer Teil-Integration, hybride Bereitstellung einer bestehenden App in die Cloud

Miguel Angel Beltra Mayoral

Rainer Keller

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma IT.TEM, Stuttgart-Vaihingen

Heutzutage sprechen alle Entwickler über Cloud und Cloud-Services, nun die Wahrnehmung dieses Begriffs scheint nicht bei jedem gleich zu sein. Dennoch sind sich alle einig, dass Cloud-Services (oder Cloud-Computing) das Bereitstellen von, unter anderen, Computingressourcen, Software, Datenspeicher und Infrastruktur [6]. All diese Services werden täglich von jedem benutzt, sei es E-Mails lesen, jemandem eine Nachricht schreiben oder online die Zeitung lesen. Die Cloud ist aktuell omnipräsent und wurde vom Trend zu Paradigma innerhalb weniger Jahre. Um eine Cloud-Migration durchführen zu können, müssen wir erstmal wissen, was Cloud-Computing ist, und was die Cloud alles anbietet. Denn es gibt nicht eine konkrete Lösung, die man einsetzt und dies war es, sondern es gibt eine große und vielfältige Auswahl an Services. Die Cloud, auch Cloud-Computing genannt, ist, wie bereits gesagt, die Bereitstellung von IT-Ressourcen über das Internet zu nutzungsabhängigen Preisen [1]. Die Cloud bietet viele Vorteile, die sich folgender beiden Hauptbereiche unterstellen lassen: Effizient und Flexibilität. Laut Rob Sauerwalt [7] sind die Vorteile folgende:

- **Effizienz:**
- Einfacher Zugriff: Cloud-basierte Anwendungen sind über das Internet zuzugreifen.
- Schnelle Markteinführung: Unternehmen können ihre Anwendungen schnell auf den Markt bringen.
- Datensicherheit: Vernetzte Sicherungen lassen Datenverluste nicht entstehen.
- Einsparung bei Hardware: Es werden Remote-Ressourcen genutzt, dies heißt, das Unternehmen braucht keinen eigenen Server besitzen.
- Zahlungsstruktur: Das Unternehmen zahlt nur für die tatsächlich benutzten Ressourcen.

Flexibilität:

- Skalierbarkeit: Die Infrastruktur passt sich nach der Arbeitslast an, die Ressourcen profitieren sowohl von scale-out, als auch von scale-up/down.

- **Speicheroptionen:** Abhängig von Überlegungen und Interessen kann ein Unternehmen zwischen Public-, Private- oder Hybrid-Speicher auswählen.
- **Auswahl des Tools:** Das Unternehmen hat die Wahl zwischen verschiedenen Tools, um zu den speziellen Anforderungen eine passende Lösung finden zu können.

Abgesehen von den vielen Benefits der Cloud muss man die Cloud in Cloud-Services und Cloud-Bereitstellungsmodelle unterteilen. Die Cloud-Services sind jedes einzelne Programm oder Service, die nicht physisch in unserem Rechner installiert ist. Die Cloud-Bereitstellungsmodelle beziehen sich darauf, wo die Cloud-Services gehostet oder deployed werden. Dabei wird eine Unterscheidung zwischen hauptsächlich public, hybrid und private gemacht. Die Cloud-Services unterscheiden sich voneinander bei der Abstraktionsebene, dies heißt, es hängt davon ab, wie viele Ressourcen der Endbenutzer im Endeffekt managen kann. Folgendes Bild beschreibt die Services:



Abb. 1: Bild mit Cloud-Services-Typen [3]

Ob diese Services On-Premise, auf einer Public-Cloud, Private-Cloud oder Hybride-Cloud gehostet werden, wird dem Unternehmen überlassen. Gründe wie Datenschutz, IT-Sicherheit, Performance oder Kosten sind hier entscheidend für unsere spätere

Wahl der finalen Lösung. Nun wissen wir alles was uns die Cloud anbietet, aber wenn wir von all diesen Vorteilen profitieren möchten, werden wir einen umfangreichen Ansatz auf unser System anwenden müssen. Dadurch klärt sich, ob sich es überhaupt lohnt, ein System in die Cloud zu migrieren. Die Schritten, den man folgen muss, sind folgende: Technische Analyse, Risiko -Analyse, Business-Case, Planung und Validation, Data- und App-Migration und Optimierung. Bei der technischen Analyse muss die Architektur der App mithilfe von UML-Modellen beschrieben werden, sodass mögliche/potenzielle Lösungen später betrachtet werden können [2]. Jede Cloud-Lösung hat genuine Eigenschaften, deswegen wird in den nächsten Schritten eine Erwägung erfolgen müssen. Zunächst folgt die Risiko Analyse. Dies ist dafür notwendig, um Risiken zu identifizieren und um keine Überraschungen zu bekommen. Alle Risiken bringen Kosten mit, die bei der Business Case zu berücksichtigen sind. Diese müssen aufgelistet werden, eine Priorität zugewiesen bekommen und im Anschluss eine Entscheidung zugunsten des Unternehmens. Diese variieren zwischen Risiko behalten, die Wahrscheinlichkeit reduzieren, die Auswirkung reduzieren oder das Risiko vermeiden [4]. Bei der Business Case, abgesehen der möglichen Kosten der durch Risiken, müssen die potenziellen Kosten der Migration aufgelistet werden. Ein Business-Case bewertet Nutzen, Kosten und Risiken eines neuen Projekts [5]. Die Kosten werden im nächsten Schritt zur Erwägung gebracht, wo die Endlösungen vorgestellt werden. Dann

wird sich die Hauptfrage des Business-Case beantwortet werden: Lohnt es sich, ein System zu migrieren? Die nächste und wichtigste Phase ist die Planung und Validation. Diese besteht darin, mit der gewonnenen Information, mögliche Strategien zur Migration zu überlegen und anschließend eine Entscheidung treffen. Die bekanntesten Strategien sind die 6R-Strategien. Die „R“s stehen für Rehost, Replatform, Repurchase, Refactor, Retain und Retire [8]. Jede Strategie wird einzeln analysiert und die Ergebnisse von dieser werden zur Erwägung genommen. Mindestens eine der Strategien muss nach dem Vergleich genommen und durchgeführt werden. Die Endlösung wird dann mithilfe des ausgewählten Verfahrens erst nach dem Proof-of-Concept oder Pilot. Der Pilot muss erstellt werden um den künftigen Stand/Architektur unseres Systems validieren zu können. Zunächst kommt die tatsächliche Data- und App-Migration. In dieser Phase werden die notwendigen Teile unseres Systems endgültig migriert. Dafür eignen sich Tools, die vom Cloud-Anbieter entwickelt wurden, um das Migrationsverfahren zu vereinfachen. Nachdem die Migration erfolgreich stattgefunden hat, wird unsere neue Umgebung optimiert. Dafür ist es empfehlenswert, Zeit und Ressourcen zu investieren. Die Optimierung muss in den Bereichen: Skalierbarkeit, Elastizität, Sicherheit, Überwachung (Monitoring) und Wartung (Maintenance) erfolgen. Folgt man all diesen Schritten, wird man ein neues System in der Cloud-Umgebung haben, das von allen möglichen Vorteilen profitiert.

Literatur und Abbildungen

- [1] Inc Amazon Web Services. Was ist Cloud Computing? <https://aws.amazon.com/de/what-is-cloud-computing/>, 2021.
- [2] Piotr Filipowicz. 6 Cloud Migration Strategies: How To Choose the Proper One? <https://inspeerity.com/blog/cloud-migration-strategies/>, 2020.
- [3] Michelangiolo Mazzeschi. What Are Cloud IAAS, PAAS, SAAS, FAAS, And Why We Use Them. <https://pub.towardsai.net/what-are-cloud-iaas-paas-saas-faas-and-why-we-use-them-8af979dad141>, 2021.
- [4] Sun Mengjie. Risk assessment-based decision support for the migration of applications to the Cloud. <https://d-nb.info/1049931173/34>, 2014.
- [5] Tool Micro. Was ist ein Business Case? <https://www.microtool.de/wissen-online/was-ist-ein-business-case/>, 2021.
- [6] Sales Cloud Salesforce. What is Cloud Computing? Types and Examples. <https://www.salesforce.com/products/platform/best-practices/cloud-computing/>, 2021.
- [7] Rob Sauerwalt. Vorteile von Cloud-Computing. <https://www.ibm.com/de-de/cloud/learn/benefits-of-cloud-computing>, 2019.
- [8] Florian Wirthensohn. THE 6 R'S OF CLOUD MIGRATION AND WHY YOU NEED TO KNOW THEM. <https://txture.io/en/blog/6-Rs-cloud-migration-strategies>, 2019.

Implementierung eines Algorithmus zur Stixel-Abstahierung auf Basis einer eigens konstruierten Stereokamera

Ali Benli

Markus Enzweiler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Zielsetzung

Das Ziel dieser Bachelorarbeit liegt darin, einen geeigneten Algorithmus zur Stixel-Abstahierung zu finden und in ein Versuchsfahrzeug zu implementieren, der dann in Echtzeit funktionsfähig ist. Die dafür notwendige Stereokamera besteht aus zwei Monokameras, für die eine entsprechende Halterung konstruiert wurde. Am Schluss soll die Pipeline zwischen Hardware und Software in Python aufgestellt werden. Zusätzlich muss der Stixel-Algorithmus an die Pipeline angepasst werden.

Stixel World

Eine schnelle und ressourcenarme Bildverarbeitung in Fahrzeugen wird immer relevanter. Dabei ist die Stixel World eine ausgezeichnete Alternative, 3D Objekte kompakt in Form von Stixeln darzustellen. Die einzelnen Stixel werden durch eine Disparitätskarte extrahiert. Jeder einzelne Stixel ist in der Breite vordefiniert, hat eine Position und Höhe in 3D und repräsentiert einen bestimmten Teil eines Objektes. So lässt sich die zum automatisierten Fahren relevante Informationen kompakt in Form von Stixel darstellen. [6]

Es gibt 3 Klassifikationen für den Stixel-Algorithmus

- Objekt
- Himmel
- Boden

Vorteile

- Millionen von Pixeln werden auf Hunderte von Stixeln reduziert
- Kompakt, signifikante Reduzierung des Datenvolumens
- Vollständigkeit, die Tiefeninformation geht nicht verloren
- Stabil, kleine Änderung der Daten führen nicht direkt zu einer schnellen Veränderung
- Robust, Ausreißer verfälschen das Ergebnis nicht
- Unterscheidung zwischen beliebigen Objekten und Freiraum

Stereo-Vision

Eine gewöhnliche Monokamera erzeugt ein Bild, in dem es eine dreidimensionale Szene auf eine zweidimensionale Ebene projiziert. Dadurch geht die Tiefeninformation verloren. Diese wichtige Information kann durch zwei einzelne Monokameras mit einem definierten Abstand aufrechterhalten werden. Dieser Vorgang wird als Stereo-Vision bezeichnet. Es entstehen zwei Bilder aus zwei verschiedenen Perspektiven. Als nächstes müssen korrespondierende Punkte in beiden Bildern gefunden werden. Für jeden dieser korrespondierenden Punkte wird eine Triangulation durchgeführt, um dann den Abstand zu dem gewünschten Punkt zu erhalten. [3]

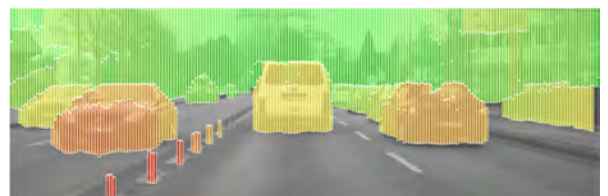


Abb. 1: Stixel World [5]

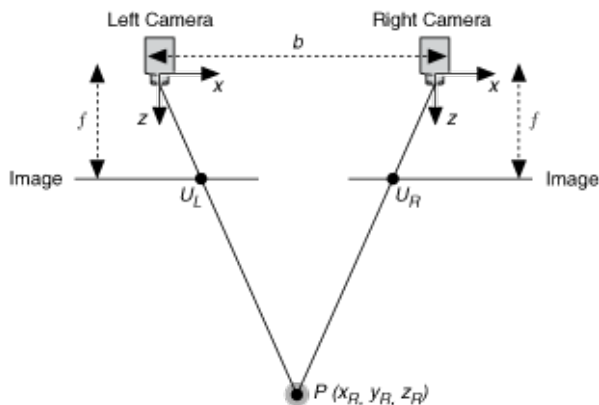


Abb. 2: Stereo-Vision [4]

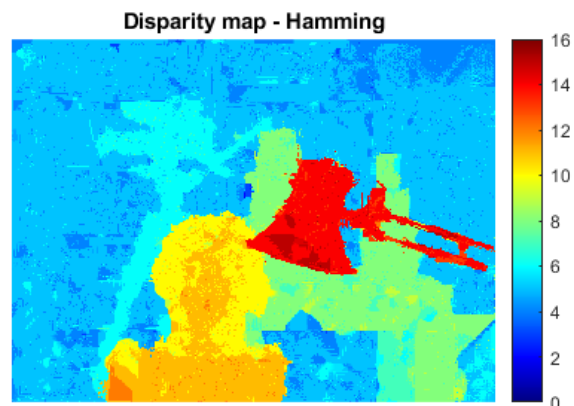


Abb. 3: Disparitätskarte [2]

Semi Global Matching (SGM)

Um korrespondierende Punkte in zwei Bildern zu finden, gibt es verschiedene Verfahren, von denen Semi Global Matching (SGM) häufig verwendet wird. Das SGM Verfahren versucht für jeden Pixel eine Korrespondenz im anderen Bild zu finden. Dies geschieht durch eine globale Kostenfunktion, welche entlang von acht Pfadrichtungen über das Bild optimiert wird. Das Ergebnis dieses Prozesses ist eine dichte Disparitätskarte. [2]

Ergebnisse und Ausblick

Der aktuelle Stand ist, dass der Algorithmus implementiert wurde und einwandfrei funktionsfähig ist. Zusätzlich wurde eine Halterung, die eigens für die beiden Monokameras konstruiert wurde, mit einem 3D Drucker erstellt. Die Pipeline zwischen den Algorithmus und den beiden Kameras ist integriert. All diese Komponenten müssen miteinander harmonisieren, sonst entstehen am Schluss Fehler in der Stixel-Erzeugung.

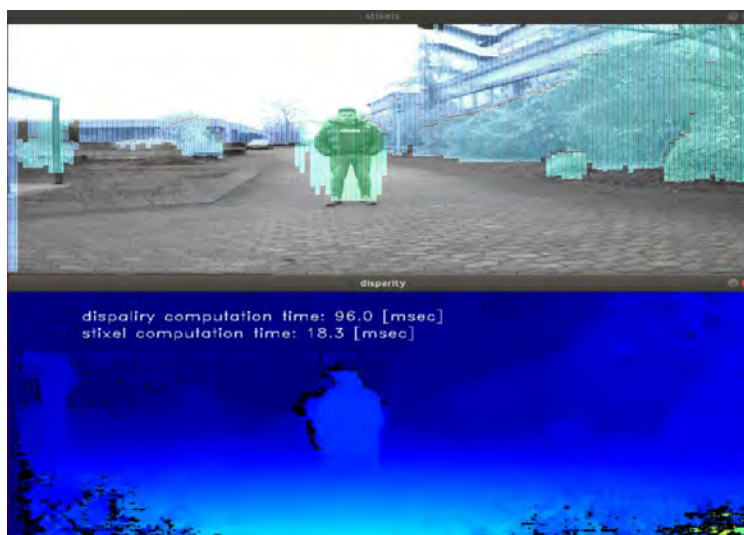


Abb. 4: Eigene Disparitätskarte und Stixel [1]

Die nächsten Schritte liegen darin, dass der Aufbau im Straßenverkehr getestet werden muss, um die Ergebnisse genauer zu analysieren. Wenn diese Schritte erfolgreich durchgeführt sind, wird zusätzlich ein Algorithmus integriert, der eine semantische-Stixel

Erzeugung durchführt. Das bedeutet, dass eine Disparitätskarte und eine semantische Szenenbeschriftung als Input eingefügt werden, um semantische Stixel zu erhalten.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Heiko Hirschmüller. Stereo Processing by Semiglobal Matching and Mutual Information. <https://ieeexplore.ieee.org/abstract/document/4359315>, 2007.
- [3] Jörg Langwald. Stereo Vision. <https://www.dlr.de/rm/desktopdefault.aspx/tabid-9389/>, 2015.
- [4] Dinesh Nair. A Guide to Stereovision and 3D Imaging. <https://www.techbriefs.com/component/content/article/tb/pub/features/articles/14925>, 2012.
- [5] David Pfeiffer. The Stixel World. <https://d-nb.info/1025886399/34>, 2011.
- [6] David Pfeiffer. The Stixel World. <https://edoc.hu-berlin.de/handle/18452/17228?show=full>, 2017.

Objekterkennung mit tiefen Neuronalen Netzen in der Logistikautomatisierung der Industrie 4.0

Wolfgang Bradfisch

Thao Dang

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma SICK AG, Waldkirch

Abstract

In der vorliegenden Arbeit geht es um die Anwendungen von, Neuronalen Netzen zur Objekterkennung. Der in dieser Arbeit betrachtete Anwendungsfall in der Logistikautomatisierung betrifft folgende Problemstellung, es soll die Position und Anzahl der Pakete erkannt werden, die sich auf einem Förderband befinden. Dies wird durchgeführt, um eine Fehlbestimmung der Zustelladresse durch Überlagerung von Paketen zu verhindern. Um dieses Problem zu beheben, macht eine Kamera Aufnahmen dieses Förderbandes, auf diesen werden dann Objekterkennungsalgorithmen angewandt. In dieser Arbeit wird der YOLOv3 Objektdetektionalgorithmus für diesen Anwendungsfall untersucht. Der vorliegende Artikel umfasst drei Abschnitte. Im ersten wird eine Einführung in das Thema Objekterkennung gegeben, im zweiten wird der Algorithmus YOLO erläutert werden und der letzte Abschnitt beschäftigt sich mit den Ergebnissen mit YOLOv3 im Anwendungsfall in der Logistikautomatisierung.

Objekterkennung

Unter dem Begriff Objekterkennung wird in diesem Text sowohl die Objektklassifikation als auch die Objektdetektion verstanden [7]. Bei der Objektklassifikation besteht das Ziel darin, die Klasse eines Objektes in einem Bild aus einer Anzahl gegebener Klassen vorherzusagen. Die Objektdetektion verfolgt das zusätzliche Ziel, die Position der Objekte, mit z.B. Begrenzungsrahmen (engl. *Bounding box*) um die zu detektierenden Objekten vorherzusagen. Trotz der unterschiedlichen Ziele, die die einzelnen Objekterkennungsdisziplinen verfolgen, geht es im Allgemeinen darum, die bestimmten Merkmale in einem Bild zu entdecken, die Aufschluss über das Objekt geben. Durch die stetige Steigerung der Rechengeschwindigkeit haben sich in den letzten Jahren in vielen Bereichen Algorithmen, die auf maschinelles Lernen beruhen, durchgesetzt. Dies ist auch in der Ob-

jekterkennung der Fall, hier kommen immer mehr Algorithmen mit tiefen neuronalen Netzen zum Einsatz. So wurden seit dem Durchbruch des sogenannten AlexNet [2] fast ausschließlich Algorithmen verwendet, deren Bildmerkmalserkennung durch tiefe neuronale Netze, vorwiegend Faltungsnetze, geschieht.

You Only Look Once (YOLO)

Der You Only Look Once oder kurz YOLO Algorithmus gehört dem Bereich der Objektdetektion an. Vorteil von YOLO ist vor allem, dass es im Vergleich zu anderen Objektdetektionsalgorithmen wie z.B. R-CNN einen geringeren Rechenaufwand benötigt [4]. Dies liegt daran, wie YOLO seine Vorhersagen der Position gibt. YOLO filtert nicht erst wie R-CNN interessante Regionen heraus, sondern es bestimmt seine Vorhersagen aus den Merkmalskarten des gesamten Bildes. Dazu wird das Bild auf ein $S \times S$ großes Gitter transformiert, wo jede Zelle die Wahrscheinlichkeit für C verschiedenen Klassen vorhersagt und B verschiedene Begrenzungsrahmen. Die Begrenzungsrahmen sind vordefinierte und auf den jeweiligen Anwendungsfall abgestimmt, sogenannte Ankerbegrenzungsrahmen. YOLO hat sich nach seinem ersten Auftreten 2016 [5] in der Art Merkmalskarten zu erstellen ständig weiterentwickelt, in dieser Arbeit wurde die Version YOLOv3 [6] verwendet. Hier wird das tiefe neuronale Netz Darknet53 verwendet, was aus 53 Faltungsschichten besteht.

Anwendungsfall

Der Datensatz, der in dieser Arbeit zu Verfügung stand, besteht aus Schwarzweißbildern von Förderbändern, auf denen sich Verpackungs-Kartons und Säcke befinden. Die Labels der Daten waren die Art der Verpackung sprich Karton oder Sack sowie die Begrenzungsrahmen (x und y sowie Höhe und Breite), also die Position des Objektes. Des Weiteren gab es eine zusätzliche Information, ob es sich um ein Bild handelt, auf dem mehrere Objekte oder nur ein einzel-

nes Objekt vorhanden ist. Das Faltungsnetz Darknet53, was von YOLOv3 für die Merkmalskartenbestimmung verwendet wird, wurde vor dem Training auf dem MS COCO Datensatz [3] vortrainiert. Dies führte zu einem schnelleren Training der Architektur, vor allem in den Bereichen der Vorhersage der Begrenzungsrahmen und der Sicherheit, dass sich ein Objekt in dem Begrenzungsrahmen befindet.

Die Ergebnisse des so trainierten YOLOv3 sind in Abbildung 1 zu erkennen. Die Ergebnisse werden in dieser Anwendung so gemessen, dass die Prozent Anzahl der richtig erkannten Multi und Single Bilder gemessen wird. Hierbei wurde mit YOLOv3 auf einem unabhängigen Testdatensatz eine Single-Genauigkeit von ungefähr 93% und eine Multi-Genauigkeit von ungefähr 91% erzielt, was zu einer gesamten Genauigkeit von 92% führt. Die Genauigkeit lässt sich gut in der in einer Konfusionsmatrix darstellen welche in Abbildung 2 abgebildet ist.

Zusammenfassend lässt sich feststellen, dass mit einem Standard Objektdetektion Algorithmus ein gutes Ergebnis für diesen Anwendungsfall erzielen werden kann. Dennoch kann ein Algorithmus, der auf die Anwendung zugeschnitten ist, ein besseres Ergebnis erzielen.

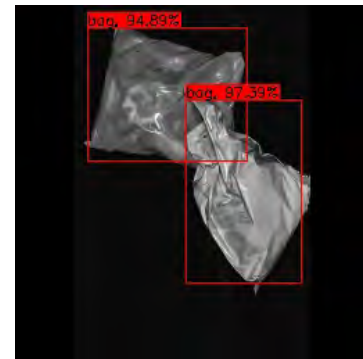


Abb. 1: Vorhergesagte Position der Objekte auf einem unabhängigen Datensatz. Die Prozentangabe, neben der Klasse, gibt an, wie sicher sich die Architektur bei ihrer Vorhersage ist. [1]

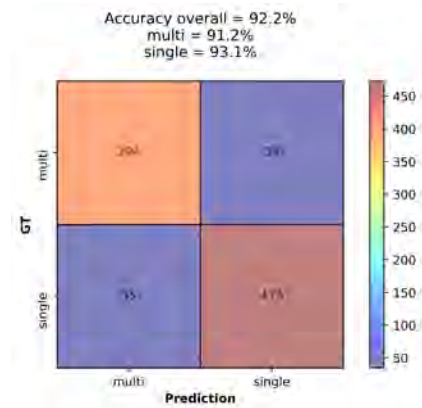


Abb. 2: Ground truth (GT) ist der Sollwert und Prediction die Vorhersage von Multi oder Single Bildern. [1]

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
- [3] Tsung-Yi Lin et al. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [4] Li Liu et al. Deep learning for generic object detection: A survey. *International journal of computer vision*, 128:261–318, 2020.
- [5] Joseph Redmon et al. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788. CVPR, 2016.
- [6] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- [7] Olga Russakovsky et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015.

Konzeptionierung der Serialisierung von verpackten Kombinationsprodukten für einen Show Case

Tim Brecht

Catharina Kriegbaum-Kling

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Haselmeier GmbH, Stuttgart-Möhringen

Einleitung

Der illegale Handel mit gefälschten Arzneimitteln ist ein großes Problem und schadet der Weltwirtschaft und der öffentlichen Gesundheit. Patienten auf der ganzen Welt gefährden ihre Gesundheit durch die unwissentliche Einnahme dieser Arzneimittel. Das Geschäft mit den Fälschungen kann für Kriminelle profitabler sein als der Handel mit Drogen wie Kokain oder Heroin, da sie damit bei minimalem Risiko große Gewinne erzielen können. Um dieser Kriminalität entgegenzuwirken und die Patienten zu schützen, wurde in Europa ein Schutzsystem eingeführt, welches das Eindringen gefälschter Arzneimittel in die legale Lieferkette verhindern soll. Das Ziel dieser Arbeit ist es, ein Konzept zu entwickeln, welches zur Umsetzung der rechtlichen Vorgaben und der damit verbundenen Serialisierung genutzt werden kann. [6]

Motivation

Die Firma Haselmeier GmbH ist ein Hersteller von patentgeschützten Injektionssystemen, wie beispielsweise Pen-Injektoren für die Applikation von Medikamenten. Bei den Pens handelt es sich so lange um ein Medizinprodukt, bis diese mit der Karpule, welche das Medikament enthält, zusammengesetzt werden. Ist das der Fall, spricht man von einem Kombinationsprodukt, welches nicht mehr nur den Vorgaben für Medizinprodukte, sondern auch denen für Arzneimittel genügen muss. Ein Kombinationsprodukt ist also ein Produkt, das aus einer Kombination eines Arzneimittels mit einem Medizinprodukt besteht. [4]

Ziel der Arbeit

In dieser Arbeit werden sowohl aktuelle Herausforderungen der Pharmaindustrie als auch die theoretischen Grundlagen des Themas Serialisierung dargestellt. In der EU ist die Serialisierung für den Handel mit Arzneimitteln seit 2019 verpflichtend. Es werden alle

regulatorischen und technischen Anforderungen in Bezug auf die Serialisierung von Kombinationsprodukten betrachtet. In der Arbeit soll zum einen die praktische Umsetzung der rechtlichen Vorgaben durch Unternehmen und Organisationen und zum anderen das konkrete Konzept als eine Art praktischer Umsetzungsplan für das Unternehmen aufgezeigt werden. Der Fokus der Arbeit liegt auf der Entwicklung eines Konzeptes, welches dem Unternehmen zur Umsetzung des Vorhabens der Serialisierung von Kombinationsprodukten dienen soll. Dieses Konzept enthält relevante Themen, wie die Anpassung der Arbeitsprozesse und die Integration der Serialisierung in die Produktion. Ebenso soll jeweils ein Zeit-, Maßnahmen- und Investitionsplan ausgearbeitet werden.

Rechtlicher Rahmen

Die gesetzlichen Grundlagen wurden am 06. November 2001 durch die EU-Richtlinie 2001/83/EG gelegt. Diese ist als Richtlinie nicht unmittelbar verpflichtend und musste durch die Mitgliedsstaaten der EU in nationales Recht umgesetzt werden, um rechtlich bindend zu sein. Das Hauptziel dieser Richtlinie ist die Schaffung eines Gemeinschaftskodex für Humanarzneimittel in der EU. [7] Dieser gilt für alle in der EU verkauften verschreibungspflichtigen Humanarzneimittel (Rx), inklusive der in der „Blacklist“ (Anhang 2 EU 2016/161) genannten nicht verschreibungspflichtigen (OTC - Over-The-Counter) Arzneimittel und jene, die importiert werden. Ausnahme stellen die in der „Whitelist“ (Anhang 1 EU 2016/161) aufgeführten Arzneimittel dar. [5] Die Bestimmungen aus der Richtlinie 2001/83/EG wurden durch die Vorgaben der EU-Richtlinie 2011/62/EU (Fälschungsschutzrichtlinie – FMD) erweitert. Hauptziel der FMD ist es, das Eindringen von gefälschten Arzneimitteln in die legale Lieferkette zu verhindern. Wesentliche Punkte der FMD sind die Einführung eines individuellen Erkennungsmerkmals auf jeder Packung, eines Manipulationsschutzes sowie eines zentralen Datenspeicher und -abrufsystems zur Verifikation der

Produkte. [1] Die Vorgaben der Verordnung (EU) 2017/745 vom 05. April 2017 (gültig ab 09.02.2019) änderten die Richtlinie 2001/83/EG. So wurde zum Beispiel in Kapitel 3 Artikel 27 das System zur eindeutigen Produktidentifikation vorgeschrieben. Dieses System, auch UDI (Unique Device Identifikation) genannt, hat den Zweck der eindeutigen Produktidentifikation und ist somit für die praktische Umsetzung der Serialisierung relevant. Das Hauptziel dieser Verordnung ist die Rückverfolgbarkeit von Produkten bis hin zur Ausgabe an den Patienten (in der Apotheke). [8]

Serialisierung

Um die eindeutige Identifikation einer Packung zu gewährleisten, wird, wie in Abbildung 1 skizziert, ein individuelles Erkennungsmerkmal (unique identifier) aufgedruckt. Dieses besteht aus dem Produktcode, der Seriennummer, der Chargenbezeichnung und dem Verfallsdatum. Diese Daten stehen zusätzlich in Form eines 2D-Data Matrix Codes zur Verfügung.



Abb. 1: Musterpackung mit Data Matrix Code [2]

Manipulationsschutz

Um die Originalität der Packung sicherzustellen und somit das unbemerkte Austauschen des Inhaltes vorzubeugen, muss zusätzlich zur Serialisierung ein Manipulationsschutz (Anti-tampering device) angebracht werden.

End-to-End-Verifizierung

Beim Verpacken der Arzneimittel werden die oben genannten Daten in klarer Schrift und als Data Matrix Code aufgedruckt. Nach Verifikation der Lesbarkeit und Qualität des Codes und der Klarschrift wird die

Seriennummer in ein europaweites Datenbanksystem hochgeladen. Die Verifikation des Arzneimittels erfolgt, wie in Abbildung 2 verdeutlicht, bei Ausgabe an den Patienten durch den Apotheker. Dieser prüft durch Scannen des Data Matrix Codes die Echtheit des Arzneimittels durch Abrufen der Informationen aus der Datenbank. Ist der Status der Packung in Ordnung, wird diese an den Patienten ausgegeben und im System ausgebucht. [1] Ab diesem Schritt würde jedes wiederholte Scannen des Produktes einen entsprechenden NIO Status (nicht in Ordnung) anzeigen, da die Seriennummer bereits ausgegeben und ausgebucht wurde. Um den Datenschutz beim Abgleich mit der Datenbank zu gewährleisten, sind die Datenbanken, auf welche die Apotheker zugreifen, physisch und logisch von denen der pharmazeutischen Industrie getrennt und kommunizieren wie in der Abbildung skizziert über einzelne Anfragen und Antworten. [6]

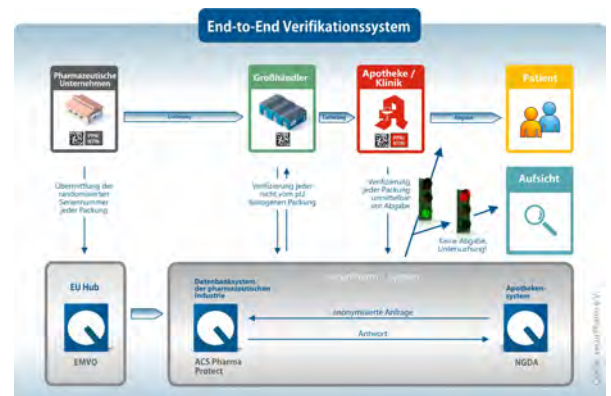


Abb. 2: Datenspeicher und -abrufsystem [3]

Ausblick

Das kurzfristige Ziel in der EU ist es, die Behörden an das System abzubinden, damit diese ihren Aufgaben nachkommen können. Außerdem muss die Anzahl der Fehlermeldungen des Systems deutlich reduziert werden, um einen tatsächlichen Nutzen daraus ziehen zu können. Langfristig könnte ein Ziel sein, das System mit der Blockchain Technologie umzusetzen, um deren Vorteile nutzen zu können. Des Weiteren wäre die Umsetzung einer vollständigen „Track & Trace“ Lösung denkbar, bei der jederzeit nicht nur der Status einer Packung, sondern auch der Standort bestimmbar wäre.

Literatur und Abbildungen

- [1] Arzneimittelkommission der deutschen Ärzteschaft. Europäische Fälschungsschutzrichtlinie und ihre Umsetzung in Deutschland. <https://www.akdae.de/Arzneimitteltherapie/AVP/Artikel/201901-2/087h/index.php>, 11 2018.
- [2] securPharm eV. Pressematerial. <https://www.securpharm.de/pressematerial/>, 12 2016.
- [3] securPharm eV. Pressematerial. <https://www.securpharm.de/pressematerial/>, 12 2020.
- [4] Concept Heidelberg GmbH. Was ist ein Kombinationsprodukt? <https://www.gmp-navigator.com/gmp-news/was-ist-ein-kombinationsprodukt>, 09 2019.
- [5] Europäische Kommission. Delegierte Verordnung (EU) 2016/161 der Kommission vom 2. Oktober 2015. <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0161&from=LT>, 10 2015.
- [6] Pharmazeutische Zeitung online. Gewinnspanne höher als im Drogenhandel. <https://www.pharmazeutische-zeitung.de/gewinnspanne-hoeher-als-im-drogenhandel/>, 02 2019.
- [7] Europäisches Parlament and und der Rat der Europäischen Union. Richtlinie 2001/83/EG des Europäischen Parlaments und des Rates vom 6. November 2001. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:2001L0083:20070126:de:PDF>, 11 2001.
- [8] Europäisches Parlament and und der Rat der Europäischen Union. Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates vom 5. April 2017. <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:02017R0745-20200424&from=DE>, 04 2017.

Konzeption und Entwicklung einer flexiblen Ablaufsteuerung zur Planung von Transport- und Arbeitsvorgängen in modularen Automationszellen

Senay Cakir

Thao Dang

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma HEITEC PTS GmbH, Kuchen

Einleitung

In der Produktion werden Bearbeitungsmaschinen häufig mit einer Automation (siehe Abbildung 1) ergänzt, um diese nahtlos in den Fertigungsablauf zu integrieren. Diese Automation beinhaltet verschiedene wiederkehrende Stationen wie Werkstückzuführungen, Speichersysteme, Waschmaschinen, Entgratstationen usw. welche sich auch an der Fertigung des Werkstücks beteiligen. Diese werden immer kundenspezifisch in die Ablaufsteuerung integriert und verkettet. Die Werkstückzuführung solcher Automationszellen wird oft von einem industriellen Roboter übernommen. Folglich kann die Ablaufsteuerung mitunter sehr komplex werden, da mehrere Stationen sich an der Fertigung beteiligen, Transporte geplant oder auch Ressourcen geteilt werden müssen. Die Programmierung solcher komplexen Konstellationen gestaltet sich in der Automatisierung häufig schwierig und ist in der Umsetzung und in den Tests zeitaufwändig.

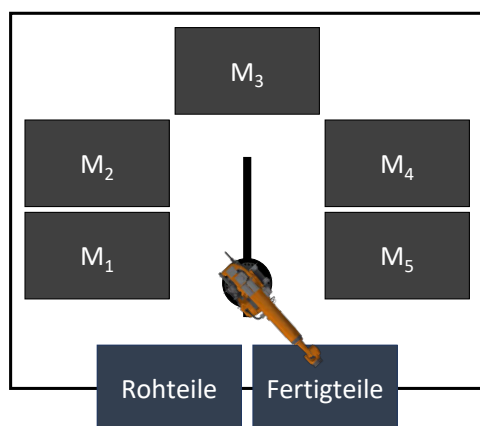


Abb. 1: Eine beispielhafte Darstellung der Automationszelle [1]

Was ist eine Ablaufsteuerung?

Eine Ablaufsteuerung plant und steuert den Ablauf von Fertigung und Transport aller Werkstücke. Die erforderlichen Vorgänge für die Fertigung eines Werkstücks werden dabei vom Entwurf abgeleitet. Diese Vorgänge, die auch als Arbeitsvorgänge bezeichnet werden, bilden den Arbeitsplan eines Werkstücks. In den einzelnen Arbeitsvorgängen eines Arbeitsplans wird das Werkstück, von dem einen Zustand zu dem anderen transformiert. Auf Basis des Arbeitsplans wird ein Transportplan für den Durchlauf des Werkstücks in der Automationszelle erstellt. Die zeitliche Ausführung der Arbeits- und Transportvorgänge erfolgt dabei über die Ablaufsteuerung. Darüber hinaus können auch die Rüstvorgänge der Stationen in der Ablaufsteuerung mitbetrachtet werden.

Ziel der Arbeit

Das Ziel dieser Abschlussarbeit ist die Konzeption und Entwicklung einer flexiblen Ablaufsteuerung zur Planung von Transport- und Arbeitsvorgängen und somit das Aufzeigen der Möglichkeiten in modularen Automationszellen. Hierfür soll ein anpassbares und übergeordnetes System entstehen, welches den Zustand aller Werkstücke und Stationen kennt und den Ablauf von Transport und Fertigung plant. Es sollen Möglichkeiten geschaffen werden, diese möglichst einfach in der Ablaufsteuerung konfigurieren zu können, damit diese zukünftig möglichst wenig kundenspezifisch ausprogrammiert werden müssen. Hierbei muss die Komplexität der zusammenhängenden Daten reduziert werden und diese in leichter verständliche Teilbereiche aufgeteilt werden. Darüber hinaus soll eine Testapplikation für die Validierung und Auswertung erstellt und die Verkettung der Modulfunktionen realisiert werden.

Flexibilität

Um einen höheren Grad an Flexibilität in einer Automationszelle zu erreichen, werden in den Planungen alternative Möglichkeiten berücksichtigt. Dadurch kann die Ablaufsteuerung bei Veränderung der Verfügbarkeit von Ressourcen ohne großen Aufwand auf eine alternative Möglichkeit zugreifen. Im Folgenden werden einige relevante Flexibilitätstypen beschrieben [4]:

- **Operation-Flexibilität:** Die Möglichkeit eine Operation bzw. einen Arbeitsvorgang auf alternativen Bearbeitungsstationen auszuführen.
- **Sequenz-Flexibilität:** Die Möglichkeit die auszuführende Sequenz an Operationen zu verändern.
- **Prozess-Flexibilität:** Die Möglichkeit ein Merkmal eines Werkstücks mithilfe von alternativen Operationen oder Operationssequenzen zu fertigen.

In der folgenden Abbildung werden die oben genannten Flexibilitätstypen anhand eines flexiblen Arbeitsplans eines Werkstücks dargestellt. Dabei wird in einem Arbeitsvorgang die Nummer des Arbeitsplans und Arbeitsvorgangs, die alternativen Maschinen, die das Merkmal des Werkstücks fertigen können, und die entsprechenden Bearbeitungszeiten aufgezeigt.

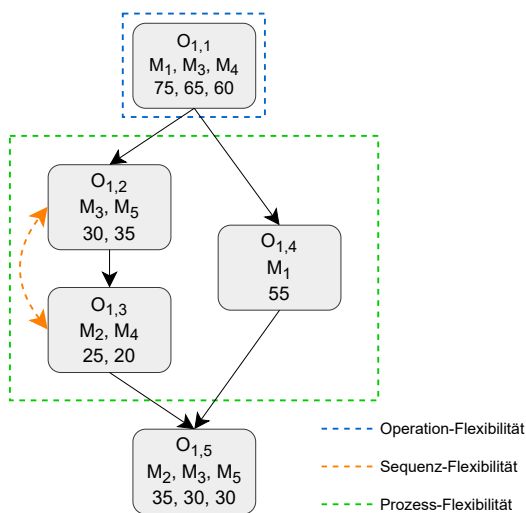


Abb. 2: Die Darstellung der Flexibilitätstypen anhand eines flexiblen Arbeitsplans eines Werkstücks. In Anlehnung an [6]

Integrated Process Planning and Scheduling

Bei einem konventionellen Ansatz sind die Arbeits- und Ablaufplanung zwei separate Phasen, die auf-

einanderfolgend durchgeführt werden. Folglich führt dies aufgrund der fehlenden Flexibilität zu einer geringeren Reaktionsfähigkeit, so dass die Planungen nicht rechtzeitig auf die Änderungen in der Produktion reagieren können. Aus diesem Grund ist eine Integration der beiden Phasen, die Integrated Process Planning und Scheduling (IPPS) genannt wird, unvermeidlich. Im Folgenden werden einige Kategorien näher erläutert, die Ansätze zur Lösung des IPPS Problems anbieten [5]:

- Bei den metaheuristischen Algorithmen werden akzeptable Lösungen für komplexe Probleme gesucht. Hierbei wird keine optimale Lösung angestrebt, sondern eine gute Lösung, die in einer angemessenen Zeit gefunden werden kann. Mit der Verwendung eines Ausgleichs zwischen lokaler Suchen und globaler Explorationen soll dieses erreicht werden [3]. Das IPPS Problem führt die Suche über mehrere Iterationen aus, um einen guten Ablaufplan zwischen den alternativen Plänen auszuwählen. Einige Ansätze hierfür ist der Genetic Algorithm (GA) oder die Ant Colony Optimization (ACO)
- Bei den regelbasierenden Ansätzen kann das IPPS Problem mittels Prioritätsregeln bzw. deren Kombination in kurzer Zeit optimiert oder gelöst werden. Ein Ansatz hierfür ist der Priority Scheduling Approach (PSA) mit Regeln, wie z.B. First In First Out (FIFO) [2].

Realisierung und Ausblick

Zur Realisierung des übergeordneten Systems der Ablaufsteuerung wird eine C# Testapplikation erstellt, die eine Verbindung zu einer Datenbank aufbauen wird. Die Datenstruktur, die hierfür konzipiert wurde, wird bei der Implementierung als Grundbasis verwendet. In der Ablaufsteuerung sollen die Zustandsübergänge und die hierfür verfügbaren Stationen in dem Arbeitsplan vordefiniert werden. Der Transportplan soll auf Basis dieses Arbeitsplans automatisch generiert werden. Die automatische Generierung von alternativen Transportplänen soll dabei die Flexibilität eines Materialhandhabungssystems gewährleisten. Darüber hinaus soll einer der Ansätze von IPPS zur Auswahl der alternativen Möglichkeiten und Optimierung der Ablaufsteuerung verwendet werden. Schlussendlich sollen verschiedene Fertigungsabläufe mit der Testapplikation getestet und die zu optimierende Größen ausgewertet werden.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Hyoung-Ho Doh, Jae-Min Yu, et al. A priority scheduling approach for flexible job shops with multiple process plans. *International Journal of Production Research*, pages 3748–3764, 2013.
- [3] Amir Hossein Gandomi. *Metaheuristic Applications in Structures and Infrastructures*. Elsevier, 2013.
- [4] Xinyu Li and Liang Gao. *Effective Methods for Integrated Process Planning and Scheduling*, volume 2. Springer, Berlin, Heidelberg, 2020.
- [5] R K Phanden, A Jain, and R Verma. *Review on Integration of Process Planning and Scheduling*. Katalinic Branko, DAAAM International Scientific Book, 2011.
- [6] S Zhang and T N Wong. Integrated process planning and scheduling: an enhanced ant colony optimization heuristic with parameter tuning. *Journal of Intelligent Manufacturing*, pages 585–601, 2018.

Virtuelle Leadership Meetings als Element einer Kulturentwicklung

Ali Copurkuyu

Clemens Klöck

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Robert Bosch GmbH, Stuttgart-Feuerbach

Einleitung

Es gibt kein Unternehmen, das keine Kultur hat. Alle sozialen Systeme prägen automatisch eine bestimmte Kultur aus. So sollte eine Unternehmenskultur etabliert werden, die den größten Nutzen für das Unternehmen hat [2]. Die Unternehmenskultur entsteht nicht von heute auf morgen. Es handelt sich hierbei um einen langfristigen Lernprozess. Die Mitarbeiter eines Unternehmens sind hierbei für die Ausgestaltung der Kultur verantwortlich. Starke Führungspersönlichkeiten, wie die Generation der Unternehmensgründer haben eine besonders starke Auswirkung auf die spätere Kultur, da in den Anfangsphasen eines Unternehmens noch keine genügend überprüften Normen, Werte und Annahmen vorhanden sind. Die Unternehmenskultur entsteht also durch Überzeugung, soziale Interaktion und die soziale Validierung erfolgreicher Konzepte und Ideen. Der Unternehmenserfolg ist von mehreren Faktoren abhängig. Das Handeln der Mitarbeiter beeinflusst diese Faktoren. Die Unternehmenskultur wirkt sich auf das Verhalten von Individuen und Organisationen aus. Somit hat die Unternehmenskultur auch Auswirkungen auf den Erfolg des Unternehmens und spielt deshalb eine wichtige Rolle [3]. Die Unternehmenskultur ist ein Phänomen, welches es in jedem Unternehmen existiert. Sie hat einen großen Einfluss auf das Mindset und Handeln der Mitarbeiter. Jedoch ist es schwierig diese Kultur zu verändern oder neu zu gestalten. Das liegt hauptsächlich daran dass die Kultur nicht ohne weiteres gemessen werden kann und es ein langwieriger Prozess ist. Mit verschiedenen Methoden kann man dennoch an der Kulturentwicklung arbeiten.

Zielsetzung

Diese Bachelorarbeit beschreibt zu Beginn die Ebenen, Funktionen und den Einfluss der Unternehmenskultur auf den Erfolg, um ein besseres Verständnis für den Begriff und die Notwendigkeit der Unternehmenskultur zu geben. Des Weiteren werden Methoden zur Kulturentwicklung vorgestellt. Hierbei spielen die Führungskräfte eine große Rolle. Zudem werden Leadership Meetings genauer betrachtet. Sie werden als ein wichtiges Element gesehen, um eine Kulturentwicklung voran zu treiben. Mit dieser Bachelorarbeit wird ein Einblick in ein Leadership Meeting in der Robert Bosch GmbH gewährt. Thematisiert werden die genutzten Methoden sowie deren Auswirkungen auf die Unternehmenskultur und auf die Führungskraft. Unterstützt werden die Aussagen mit den Ergebnissen einer Umfrage, welcher im Anschluss eines Leadership Meetings durchgeführt wurde. Das Ziel ist mögliche Verbesserungspotenziale zu erkennen, damit die Auswirkungen eines Leadership Meetings auf die Unternehmenskultur gesteigert werden kann.

Problemstellung

Der Bereich Global Business Services (GS) der Robert Bosch GmbH ist ein relativer neuer Bereich, welcher aus mehreren Funktionsbereichen zusammengestellt wurde (vgl. Abb1). Der Geschäftsbereich steckt mitten in einer Transformation. Durch diese Neugestaltung, wird versucht die Unternehmenskultur weiter zu entwickeln, damit der Bereich als eine Einheit erfolgreich fungieren kann.

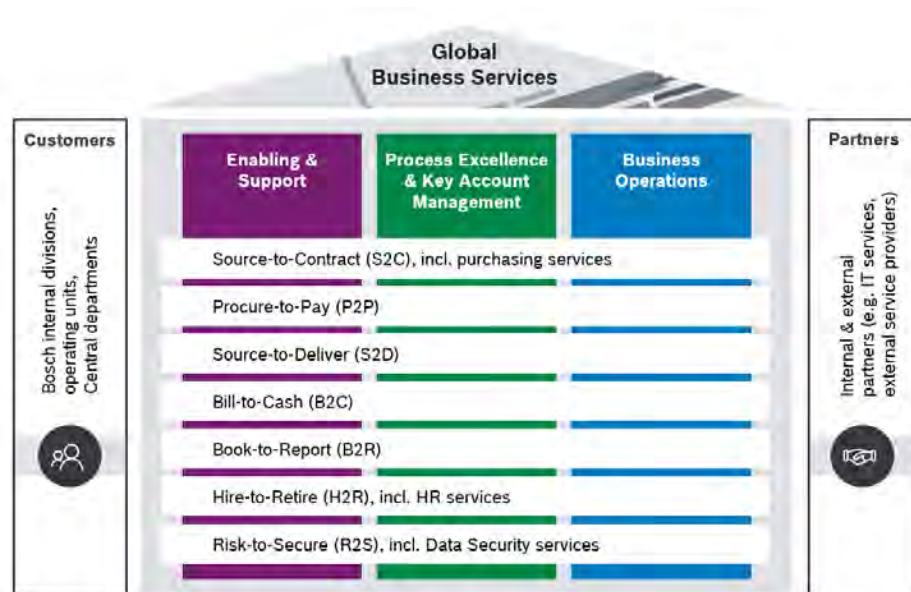


Abb. 1: Organizational Setup [1]

Leadership Meetings im Bereich GS

Im Bereich Global Business Services fand das Leadership Meeting dieses Jahr virtuell über MS Teams statt. Die Anzahl der Leadership Meetings in einem Jahr sind unterschiedlich. Meistens finden ungefähr 2-4 Leadership Meetings pro Jahr im GS statt. Bei dieser Bachelorarbeit geht es hauptsächlich um das Leadership Meeting im Oktober 2021. Zu diesem

Termin wurden ca. 200 Führungskräfte aus der ganzen Welt eingeladen. Das Ziel war es die Führungskräfte untereinander besser zu vernetzen, um somit eine einheitliche Führungsmannschaft und ein gemeinsames Führungsverständnis zu kreieren. Hierbei wurden verschiedene Methoden angewendet. Außerdem haben die Führungskräfte Informationen über diverse Themen und Projekte erhalten.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Josef Herget. *Unternehmenskultur gestalten*. Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature, 2020.
- [3] Norbert Homma, Rafael Bauschke, and Laila Maija Hofmann. *Einführung Unternehmenskultur*. Springer Fachmedien Wiesbaden, 2014.

Konzeption und Implementierung eines Konfigurationsprotokolls der EUCHNER Augmented Reality Submodule

Alexander Dausel

Mirko Sonntag

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma EUCHNER GmbH + Co. KG, Leinfelden-Echterdingen

Einleitung

Die Augmented Reality (AR) ist für die meisten Menschen noch immer ein relativ unbekanntes Gebiet der Technik. Sie beschäftigt sich hauptsächlich mit der Erweiterung der Realität um visuelle Informationen. Somit grenzt sie sich deutlich von der Virtual Reality (VR) ab, welche den Benutzer in eine vollständig künstlich erschaffene Umgebung eintauchen lässt. Heutige AR-Anwendungen legen ihren Fokus auf die interaktive Verarbeitung von Informationen. Benutzer können sich beispielsweise über eine spezielle Brille aufbereitete Messwerte direkt auf dem Display anzeigen lassen.

Die Firma Euchner ist seit vielen Jahren erfolgreich in der industriellen Sicherheitstechnik vertreten. Gerade die Konfiguration von Sicherheitsmodulen ist ein ideales Anwendungsgebiet der AR. Wo bisher eher kleine und simple Bedienfelder auf den einzelnen Steuerungen zu finden waren, soll in Zukunft mithilfe der AR deutlich einfacher und flexibler gearbeitet werden können. Die AR ermöglicht eine intuitive und direkte Einrichtung des Systems. Anwender erhalten ein unmittelbares visuelles Feedback. Ein AR-Modul wird hierbei auf ein Zuhaltmodul, welches beispielsweise eine Türzuhaltung steuert, aufgesteckt und verbleibt im Gerät. Um das System zu konfigurieren, verbindet sich ein Benutzer wahlweise per Rechner oder Handy-App über Bluetooth mit dem AR-Modul und kann Konfigurationsdateien im JSON Format hoch bzw. herunterladen. Diese Konfigurationsdatei legt fest, welche Funktionalitäten auf dem Zuhaltmodul vorhanden sind und angezeigt werden sollen. So kann beispielsweise festgelegt werden welche Taster zur Verfügung stehen und welche Einstellungen zugänglich sein sollen. Dies schafft eine zusätzliche Sicherheit. Ist die Konfiguration erfolgreich, kann mit einer AR-Brille oder Handy-App über Displayeingaben das Modul gesteuert werden.

Aufgabenstellung

Der Schwerpunkt der Bachelorarbeit lag bei der hardwarenahen Programmierung. Aufgabe war es, eine stabile Bluetooth Low Energy Verbindung von einem Bluetooth Device zu dem AR-Modul herzustellen und die Konfigurationsdatei zu senden oder zu empfangen. Um den Datenaustausch zwischen AR-Modul und Bluetooth Device realisieren zu können, bestand ein weiterer Teil der Arbeit darin, eine simple Handy-App zu programmieren die alle notwendigen Funktionalitäten bereitstellt. Auf dieser Grundlage sollte daraufhin ein Konfigurationsprotokoll erstellt werden, welches genau festlegt wie die einzelnen Schritte im Detail ablaufen müssen.

Konzept und Implementierung

Zunächst war eine gründliche Einarbeitung in den Bluetooth LE Standard notwendig. Auf dem AR-Modul befindet sich ein FreeRTOS welches die einzelnen Tasks steuert. Zu diesen Tasks gehört unter anderem der Bluetooth LE Dienst. Im ersten Schritt musste die verwendete Hardware initialisiert werden. Nach der Initialisierung der Hardware, musste ein sogenannter Bluetooth Service definiert werden. Services dienen den Bluetooth Devices als Verbindungspunkte. Jeder Service besitzt eine spezielle 16-bit UUID. Diese UUID dient dazu, den Service eindeutig zu kennzeichnen. Allerdings können auch eigene UUIDs definiert werden. Diese sind 128-bit lang und können frei bestimmt werden.

Sobald der Service definiert ist und von anderen Devices erkannt wird, kann eine Verbindung stattfinden. Die Übertragung der Konfigurationsdatei erfolgt in Paketen. Hierbei muss empfangenseitig eine Prüfung erfolgen. Die Pakete werden zunächst im RAM abgespeichert und können daraufhin entweder mithilfe einer CRC-Checksumme oder eines FreeRTOS JSON-parsers

überprüft werden. Sind alle Prüfungen erfolgreich, wird die Konfigurationsdatei in den Flash geschrieben.

Ausblick

Mögliche interessante zukünftige Aspekte sind ausge-reifte Zugangsbeschränkungen. So könnte mithilfe von Anmeldedaten zwischen Technikern und Facharbeitern

unterschieden werden und so nur notwendige Informa-tionen angezeigt werden. Diese zusätzliche Sicherheit schützt vor versehentlicher, aber auch böartiger Manipulation. Neben der Übertragung über Bluetooth ist auch eine zentrale Verteilung über Server denkbar und praktikabel. Ein möglicher, interessanter Ansatz wäre der Azure IoT-Hub Dienst von Microsoft.



Abb. 1: Komplettsset MGB2-L.HB-PN mit Submodul-Slots [1]

Literatur und Abbildungen

- [1] Vanessa Lagnach. KOMPLETTSETS MGB2-L.HB-PN. <https://assets2.euchner.de/files/89/17/728917.jpg>, 2021.

Secure Boot auf dem PicoCoreMX8MM

Daniel Dippold

Dominik Schoop

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma F&S Elektronik Systeme GmbH, Stuttgart

Ziel und Motivation

Technische Systeme, welche in der Industrie genutzt werden, benötigen Schutz vor Angriffen. Angreifer können viele verschiedene Intentionen haben, ein System zu infiltrieren. Ob es sich um Spionage handelt oder der Angriff nur Schaden anrichten soll, ist hierbei gleichermaßen zu betrachten. Ziel ist es, zu jeder Zeit einen möglichen Angriff präventiv zu verhindern, um Schäden jeder Art grundsätzlich zu vermeiden. Wenn man den Begriff "Security" hört, denkt man meist an einen Schutz, während das System läuft. Dieser greift jedoch nicht, während das System hochfährt. Somit ist es auch nötig den Bootprozess abzusichern, da ansonsten ein Angreifer seinen manipulierten Code in das System laden könnte. Im schlimmsten Fall wird nicht bemerkt, wenn im Hintergrund Schadsoftware läuft. Genau aus diesem Grund muss gewährleistet werden, dass jeder geladene Code nachweislich von einer Person stammt, die berechtigt war diesen Code zu erstellen und in das System zu laden. Ansonsten muss das System in der Lage sein, den Bootvorgang abubrechen, um möglichen Schaden zu vermeiden. Hierzu wurde Secure Boot entworfen. Durch diese Technologie soll gewährleistet werden, dass jede einzelne Bootebene authentisch ist und geladen werden darf. Um dieses Ziel zu erreichen, ist jedoch verschiedene Soft- und Hardware notwendig, in Abhängigkeit von der verwendeten Plattform. Die Arbeit wird bei der Firma F&S Elektronik Systeme durchgeführt. F&S entwickelt und verkauft embedded Devices welche überwiegend NXP Prozessoren verwenden. Secure Boot wird bereits für Systeme, welche auf NXPs i.MX6 Architektur basieren angeboten. Das Ziel der Arbeit ist es, Secure Boot auf F&S Systeme zu integrieren, welche auf der neueren i.MX8 Architektur basieren. Dies soll durch eine Chain of Trust [1] implementiert werden. Die Chain of Trust soll über die Signatur der Images nachweisen, ob einem Image vertraut werden kann. Zusätzlich könnten auch Features wie Verschlüsselung integriert werden.

Secure Boot

Bei i.MX8 Prozessoren von NXP gibt es die HAB Library im ROM-Code der CPU. HAB ist kurz für High Assurance Boot. Diese Library wurde speziell für die Verifizierung von Software erstellt. Die Library wird vom ROM-Bootloader, der ersten Boot-Ebene, für das kommende Image aufgerufen. Da sowohl der ROM-Bootloader, als auch die HAB Library unveränderlich sind, kann dieser Prozess als authentisch angesehen werden. Von jetzt an muss jede Ebene in der Lage sein, das nächste Image sowie andere wichtige Daten, welche geladen werden, zu verifizieren. Wenn zum Beispiel die Daten zur Initialisierung des D-RAM manipuliert wurden, kann dieser durch falsche Verwendung beschädigt werden. Natürlich muss es für die Verifikation einen zuverlässigen Weg geben, die Herkunft eines Codes nachzuweisen. Dazu muss dieser signiert werden. Wenn ein Code mit seiner Signatur nicht verifiziert werden kann, oder gar keine besitzt, muss das System reagieren. Eine Reaktion wäre zum Beispiel, den Bootvorgang anzuhalten. [3] Der ROM-Bootloader wendet die notwendigen HAB Funktionen bereits an und bildet somit die Grundlage für die "Chain of Trust". [1]

NXP stellt zur Signierung ein Programm namens "Code Signing Tool" bereit. Durch dieses ist es möglich Zertifikate zu erstellen und, basierend auf diesen, gültige Signaturen für einen Code. Zusätzlich kann dieses Tool auch Images verschlüsseln. Die Zertifikate, welche erstellt und verwendet werden, folgen dem X.509 Standard. Die verwendeten Schlüssel sind asymmetrisch. Systeme bekommen den öffentlichen Schlüssel zur Verfügung gestellt um Signaturen zu verifizieren, welche mit dem privaten Schlüssel signiert werden. Somit kann niemand, der nur den öffentlichen Schlüssel kennt, selber gefälschte Signaturen erstellen. Images können auch verschlüsselt werden. Hierzu erstellt das Tool einen symmetrischen Schlüssel. Dieser wird an das Board gesendet, welches ihn durch einen internen sicheren Mechanismus verschlüsselt. Damit kann nur das Board selber den Schlüssel finden und mit diesem das Image entschlüsseln. In Abbildung

1 ist beschrieben welche Komponenten das Code Signing Tool benötigt, um ein Image zu signieren. Die Signaturbeschreibung beinhaltet Informationen über den auszuführenden Vorgang. Sie ist der einzige Input, den ein Nutzer angibt. Welches Image signiert und welches Zertifikat verwendet werden soll, ist in der Beschreibung angegeben. Zudem werden hier alle weiteren relevanten Informationen aufgeführt. Dazu gehört zum Beispiel, welche Bereiche des Images signiert werden sollen. [3]

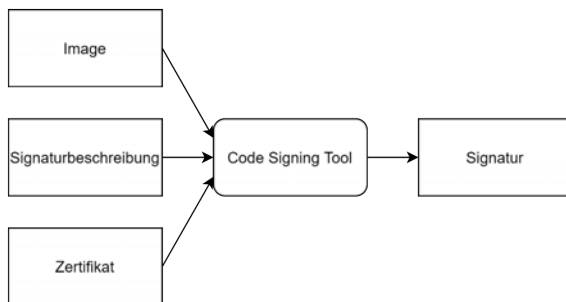


Abb. 1: Code Signing Tool [2]

Das verwendete Zertifikat muss dem System, welches den Code laden soll, bekannt gemacht werden. Danach sollte niemand das Zertifikat ändern können. Ansonsten könnte ein Angreifer einfach den öffentlichen Schlüssel verändern und so seine eigenen Zertifikate verwenden. Außerdem sollte auch die Einstellung von Secure Boot an sich permanent sein, da ansonsten ein Angreifer einfach Secure Boot ausschalten könnte. Den Schlüssel in den ROM (Read-Only Memory) zu schreiben, wäre eine unveränderliche Implementierung.

Das Problem hierbei ist, dass ein Nutzer am ROM nichts verändern kann. Demnach müsste NXP bei der Fertigung der Prozessoren die Zertifikate kennen. Das wäre für NXP und Partnerunternehmen ein unmöglicher Aufwand. Dazu soll nicht jeder Nutzer seine Zertifikate preisgeben müssen. Deshalb verwendet man hierfür eine Hardware-Einheit namens OCOTP oder eFuses. OCOTP steht für On-Chip-One-Time-Programmable. Diese Hardwareeinheit besteht aus Registern, welche ein einziges Mal beschrieben werden können. Spezielle Adressen in den "Fuse Banks" sollen den öffentlichen Schlüssel eingebrannt bekommen. Zuletzt gibt es noch eine Fuse, welche für das Aktivieren von Secure Boot zuständig ist. Dadurch ist die Aktivierung von Secure Boot permanent. Der Angreifer kann somit weder den öffentlichen Schlüssel verändern, noch Secure Boot ausschalten. [1]

Da F&S Elektronik Systeme ein anderes Bootkonzept verwendet als NXP, muss die Implementierung im Voraus geplant werden. Zudem soll die Verwendung des "Code Signing Tools" für den Nutzer möglichst simpel gestaltet werden. Durch den unterschiedlichen Aufbau der Bootcontainer muss hier evaluiert werden, welche Bereiche der Bootcontainer signiert und zu welchem Zeitpunkt sie überprüft werden. Das User Tool von F&S, das den Kunden zur Verfügung gestellt wird, muss hierauf angepasst werden. Teilimages dürfen während des Prozesses nicht beschädigt und Informationen, wie z.B. die Größe der Images, müssen richtig gesetzt werden. In Abbildung 2 kann man sehen, dass ein Nutzer dem Tool nur noch ein Image übergeben muss und das Tool die Vorarbeit übernimmt, welche nötig ist, um das Code Signing Tool zu verwenden. Die Signaturbeschreibung wird automatisch erstellt und das Code Signing Tool ausgeführt.

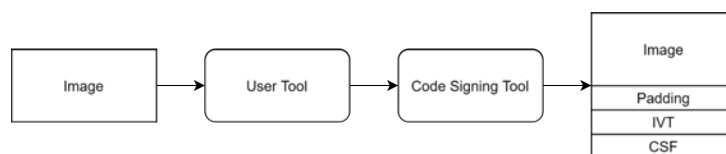


Abb. 2: User Tool [2]

Die Integration von Secure Boot ist sehr flexibel. Wichtig sind hierbei jedoch ein paar entscheidende Dinge. Zuerst einmal muss jedes signierte Image auch eine sogenannte IVT haben. IVT bedeutet Image Vector Table und ist mit einem struct in C zu vergleichen. Jede vier Byte, gibt es ein fest definiertes Feld, welches eine Hardware-Adresse vorgibt. Eines dieser Felder ist die Adresse auf das CSF (Command Sequence File). Mit CSF ist die Signatur gemeint. Das ist deshalb wichtig, weil die HAB Funktion zum Authentifizieren eines Images ein Argument hat, welches die Adresse

der IVT ist. Wenn die IVT gültig ist, wird die HAB Funktion die CSF finden und das Image überprüfen. Die HAB Funktionen selber haben feste Adressen im ROM. Dadurch kann ein Programmierer diese, nachdem er sie in Erfahrung gebracht hat, nutzen. Dazu muss er die Funktionen über einen Funktionspointer initialisieren und nutzen. Während der Entwicklung muss auch auf sogenannte HAB Events geachtet werden. Diese Events werden im HAB Event Log gespeichert. Jedes dieser Events zeigt einen Fehler an, der bei einer HAB Funktion aufgetreten ist. Da die Funktionen auch bei

einem Board verwendet werden können, auf dem Secure Boot noch nicht aktiviert wurde, lassen sich potentielle Fehler debuggen. Ein "geschlossenes" Board würde wiederum nicht mehr booten.

Umsetzung

Da das Bootkonzept bei F&S anders funktioniert als bei NXP, muss die "Chain of Trust" an das Bootkonzept von F&S angepasst werden. Wichtig ist zu verstehen, wo sich welche Komponente befindet und wann sie in das System eingespielt wird. Im Grunde ist die Reihenfolge selber aber gleich. Abbildung 4 zeigt die Reihenfolge der Images und begleitet deren Erklärung im folgenden Absatz. Zuerst lädt der ROM-Bootloader den SPL (Secondary Program Loader). Der SPL ist ein kleines Programm, das in den internen RAM des Prozessors passt. Der DRAM ist zu diesem Zeitpunkt noch nicht verfügbar. Die Images der anderen Bootebenen sind daher zu groß für den verfügbaren Speicher. Deshalb benötigt man den SPL, um den DRAM und die boardspezifische Hardware zu initialisieren. Danach lädt er sowohl die Arm Trusted Firmware (ATF) und den U-Boot in den RAM. Die ATF ist ein Secure Monitor, der später sicherheitsrelevante Aufgaben übernehmen soll, anstatt dass das Betriebssystem direkt auf Komponenten wie Clocks, Cores oder aber auch die HAB Library zugreifen kann. Der U-Boot ist ein Bootloader, welcher die restlichen Konfigurationen übernimmt und das Linux System startet. Zudem kann ein Nutzer auch eine Kommandozeile verwenden und selber bei Bedarf Einstellungen vornehmen. Sobald beide Images im RAM liegen, startet der SPL die ATF. Sobald diese sich selber initialisiert hat, startet diese den U-Boot und läuft im Hintergrund weiter. Der U-Boot startet nun Linux.



Abb. 4: Boot Order [2]

Der Aufruf von Prüfungen innerhalb des SPL, kann unter Umständen problematisch sein. Da die Hardware noch nicht voll initialisiert wurde, kann es dazu kommen, dass ein sonst erfolgreicher Aufruf das Board zum Neustart zwingt. Da keine Informationen über den für den Neustart verantwortlichen Fehler ausgegeben werden, wird der Debugging Prozess erschwert. Es muss also ein Punkt gefunden werden, an welchem die nötige Hardware vollständig initialisiert wurde. Zudem ist es wichtig zu wissen, ab wann ein Image im RAM liegt und an welcher Adresse.

Da die meisten Images feste Ladeadressen haben, kann

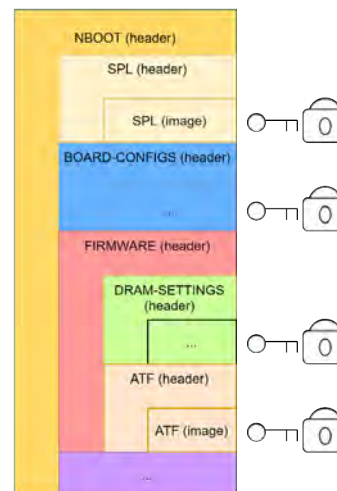


Abb. 3: NBoot [2]

Der Unterschied in den Bootstrategien liegt darin, dass NXP viele der Images in einen Boot Container legt, in welchem einige Images gemeinsam signiert werden. F&S verwendet einen eigenen Bootcontainer mit dem Ziel, low-level Code vor dem Nutzer zu verbergen. Zudem hat es den Vorteil, dass über den sogenannten N-Boot Container sich für Kunden beim Aufspielen der Images nichts zu älteren Boards verändert, da es dort bereits einen N-Boot gab, welcher jedoch andere Aufgaben hatte. Somit kann ein Nutzer gewohnten Vorgängen nachgehen, ohne sich selbst um Images wie SPL oder ATF kümmern zu müssen. Abbildung 3 zeigt, wie der Container aufgebaut ist. Vor jedem Image befindet sich ein Header, welcher unter anderem die Größe des Bereiches angibt, welchen er einschließt. Dadurch lässt sich eine Art "Baum" erstellen. Innerhalb dieses Baumes kann ein System jetzt Images nach bekannten Mustern finden.

man bei diesen die IVT nicht an den Anfang des Images legen, wie man es bei dem SPL machen würde. Eine Möglichkeit wäre, die Images auf eine gewisse Adresse zu paden und dort an eine feste Adresse die IVT zu legen. Dadurch kann das Image selber in Zukunft auch unter Umständen größer sein. Der Test könnte immer noch diese Adresse ansprechen, da das Image an derselben Stelle seine IVT und Signatur hat. Der N-Boot wird nicht vollständig gespeichert. Nur die für das System relevanten Images werden behalten. In Abbildung 3 kann man die Hierarchie erkennen, nach welcher der N-Boot aufgebaut ist. Um die relevanten

Images zu signieren, müssen diese also separiert werden. Der N-Boot wird deshalb bei der Signierung durch das F&S-Tool auseinandergenommen. Die einzelnen Images werden signiert. Jetzt müssen alle Einzelteile des zerlegten N-Boot wieder zusammengefügt werden. Für die Images verwendet man hier die signierten Varianten. Da sich jetzt wieder der reguläre Aufbau des N-Boot ergibt, kann ein F&S Board diesen erkennen und abspeichern.

Ausblick

Für die Zukunft ist das Ziel, die Signierung für jede Bootebene zu integrieren. Dazu gehören die

Anpassung der Tools sowie die Integration der Prüfungen. Das System muss in der Lage sein, sowohl richtige Signaturen, als auch fehlende oder falsche Signaturen zu erkennen und entsprechend zu reagieren. Ein wichtiger Punkt ist auch die Authentifizierung von Images dynamisch zu gestalten. Images werden durch Updates über die Zeit größer. Deshalb sollte ein Weg implementiert werden, bei dem kein fester Offset verwendet wird, wie zuvor beschrieben. Zusätzlich kann dieses System auf weitere F&S Boards mit i.MX8 Prozessoren implementiert werden. Außerdem wäre es möglich, eine Verschlüsselung der Images zu integrieren.

Literatur und Abbildungen

- [1] Akshay Bhat. Secure Boot,ChainofTrustand DataProtection. https://www.timesys.com/pdf/Session_4.3_II_Bhat.pdf, 02 2019.
- [2] Eigene Darstellung.
- [3] Shyam Saini. i.MX6 HABv4. https://wiki.amarulasolutions.com/uboot/secure_boot/imx6_habv4.html#high-assurance-boot-hab, 04 2019.

Erstellung und Umsetzung eines Konzeptes zum Vergleich von Elektrofahrzeugen und Fahrzeugen mit Verbrennungsmotoren für Mercedes me connect mit Hilfe von Azure Databricks

Anna Marie Dobler

Catharina Kriegbaum-Kling

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Mercedes-Benz AG, Stuttgart

Der durch Menschenhand erzeugte Klimawandel ist eines der wichtigsten Themen der aktuellen Zeit. Ein großer Anteil des Ausstoßes klimaschädlicher Treibhausgase ist auf die Mobilität mit Verbrennungsmotoren zurückzuführen. Eben hier setzen weltweit viele Regierungen an und forcieren einen Wandel hin zur Elektromobilität mit dem Ziel den Ausstoß von Treibhausgasen zu verringern. Auch Automobilhersteller sehen sich in der Pflicht zu handeln und erweitern ihre klimaneutrale Angebotspalette stetig. In Kooperation mit Mercedes me im Hause der Mercedes-Benz AG soll im Rahmen dieser Bachelorarbeit das Thema aufgegriffen werden. Das Hauptaugenmerk liegt auf dem Vergleich der Elektromobilität und Fahrzeugen mit Verbrennungsmotoren. Hierbei sollen die unterschiedlichen Kennzahlen von Mercedes me im Hinblick auf die unterschiedlichen Antriebsarten analysiert, aufbereitet und ausgewertet werden. Dies erfordert eine ausführliche Datenanalyse via Azure Databricks, welche mithilfe von Python ausgeführt wird. Das zweite Hauptaugenmerk dieser Arbeit, das Aufsetzen einer Strategie zur nachhaltigen Datenanalyse, soll an eben diesem Punkt ansetzen und einen Lösungsvorschlag für nachhaltige Datenanalysen auf Basis der zuvor durchgeführten Analyse bereitstellen.

Mercedes me

Bei Mercedes me handelt es sich um einen eigenständigen Konzernteil innerhalb der Mercedes-Benz AG. Der Fokus von Mercedes me liegt auf dem Angebot von digitalen Produkten und Diensten rund um die Fahrzeuge von Mercedes-Benz. Am bekanntesten hiervon ist die Mercedes me App. Hiermit können unterschiedlichste Themen rund um das Fahrzeug gesteuert und erfragt werden. Darüber hinaus liegt auch die Sprachsteuerung mit „Hey Mercedes“ in der Verantwortung von Mercedes me. Die Voraussetzung

zur Nutzung der Dienstleistungen von Mercedes me ist die Verknüpfung des Mercedes mit der App. Nach der Verknüpfung können die unterschiedlichen Services und Produkte von Mercedes me aktiviert und benutzt werden. Diese Verknüpfungen und Aktivierungen bilden das Herzstück von Mercedes me und stellen die Basis für den durch die Arbeit generierten Mehrwert für die Mercedes-Benz AG dar. Die Datenanalyse soll hier ansetzen und die unterschiedlichen Kennzahlen von Mercedes me vergleichen.

Zielsetzung

Die Arbeit verfolgt zwei Hauptziele. Das erste Hauptziel ist die Analyse des Vergleiches von Fahrzeugen mit Elektroantrieb und Fahrzeugen mit Verbrennungsmotor. Hier sollen, im Hinblick auf die für Mercedes me relevanten Verknüpfungen und Aktivierungen, wertvolle Erkenntnisse gewonnen werden sowie eine Handlungsempfehlung für die Ausrichtung zukünftiger Produkte und Vorgehensweisen getroffen werden. Der Fokus liegt dabei auf einem vermuteten unterschiedlichen Nachfrageverhalten der Dienste in Bezug auf Fahrzeuge mit Elektroantrieb und Fahrzeugen mit Verbrennungsmotor. Das zweite Hauptziel ist das Erarbeiten und Aufstellen einer Strategie für eine nachhaltige Datenanalyse bei Mercedes me. Diese soll ermöglichen, dass Datenanalysen, welche von externen Mitarbeitern oder zeitlich befristeten Mitarbeitern aufgestellt werden, auch beim Austritt dieser Mitarbeiter weiterhin im Unternehmen verwendet werden können und das dabei entstandene Knowhow nicht verloren geht. Diese Strategie soll anhand der Datenanalyse für das erste Hauptziel erstellt werden.

Datenanalyse

Im ersten Schritt soll der theoretische Hintergrund der Datenanalyse beleuchtet werden. Neben der Ausarbeitung relevanter Begriffe und Themen rund um die Datenanalyse, wie Big Data, Datenspeicher und unterschiedlichster Analysearten, liegt das Augenmerk auf dem Cross-Industry Standard Process for Data Mining CRISP-DM Modell, welches das am häufigsten eingesetzte Prozessmodell für Datenanalysen darstellt. [2]

Ziel dieses Modells ist, durch das iterative Verfahren relevante Erkenntnisse aus den Daten zu erlangen. Die einzelnen Phasen dienen hierzu als Step-by-Step Anleitung und können je nach Verlangen wiederholt und iterativ durchlaufen werden. Die erste Phase, das Unternehmensverständnis, zielt auf ein generelles Verständnis des Unternehmensumfeldes ab. Hierzu gehört bspw. die Erarbeitung der Zielsetzung oder der Organisationsstruktur. Im Hinblick auf diese Arbeit versteht sich darunter die Ausarbeitung der Hypothesen sowie das Verständnis der vorliegenden IT- und Dateninfrastruktur. Dies wird durch die zweite Phase, dem Datenverständnis, verfeinert, indem ein Augenmerk auf die vorhandenen Daten und Datenquelle gelegt wird. Im Rahmen der dritten Phase, der Datenaufbereitung, werden die Daten gruppiert und bei Bedarf mit Metadaten ergänzt. Diese gruppierten Daten dienen als Eingabe für die zu verwendenden Analysemethoden (Modellierungsphase). Die Ergebnisse hieraus werden mit den in der ersten Phase festgelegten Zielen verglichen. Sofern dies zufriedenstellend ist, wird in der letzten Phase die Umsetzung der Analyse durchgeführt. [2]

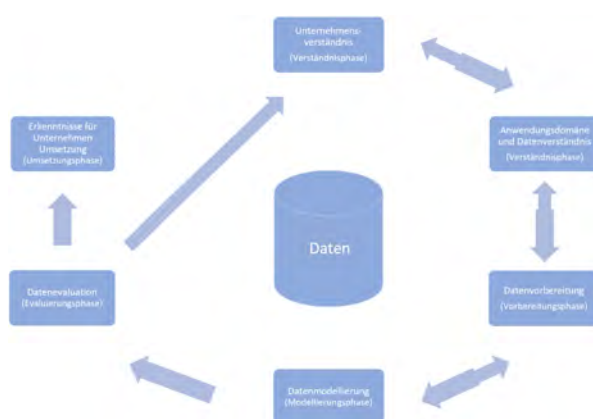


Abb. 1: Eigene Darstellung CRISP-DM Modell [1]

Mit Hilfe des zuvor erarbeiteten theoretischen Hintergrunds wird die Datenanalyse für Mercedes me erstellt. Die hierfür benötigten Daten liegen bei Mercedes me in einem Data Lakehouse vor, auf welches über Azure Databricks zugegriffen werden kann. Die Datenanalyse

wird mithilfe von SQL-Abfragen in Azure Databricks erstellt und durch Python sowie Power BI für die Auswertung und Visualisierung der Abfragen erweitert. Für die Datenanalyse müssen folgende Grundvoraussetzung erfüllt sein: Die Fahrzeuge von Mercedes müssen in die Kategorien Verbrenner und E-Mobilität einsortiert werden können. Nur dadurch können die unterschiedlichen Kennzahlen verglichen werden. Im Hinblick auf die Kennzahlen sollen im ersten Schritt die Verknüpfungs- und Aktivierungsraten verglichen werden. Danach wird die Analyse auf den Vergleich der von Mercedes me unterschiedlich angebotenen Services gerichtet. Neben diesen Mercedes me spezifischen Kennzahlen soll auch noch auf käuferspezifische Kennzahlen eingegangen werden. Zur Erweiterung der Analyse und Vergrößerung des Erkenntnisgewinns sollen im Hinblick auf die bereits zuvor genannten Kennzahlen noch zusätzlich zwei Märkte miteinander verglichen werden. Hierfür wird ein europäischer Markt gewählt, welcher ähnlich zu Deutschland ist, und ein asiatischer Markt mit starken Wachstumschancen für Mercedes. Die für das erste Hauptziel angelegten Analysen sollen im nächsten Schritt als Basis für das zweite Hauptziel verwendet werden. Mit Hilfe einer guten Dokumentation, leicht verständlichen Abfragen in Azure Databricks, welche durch Kommentare und den Einsatz von Variablen die Benutzerfreundlichkeit erhöhen, soll die nachhaltige Datenanalyse ermöglicht werden. Hierfür wird ein Konzept sowie eine Handlungsempfehlung für Mercedes me aufgestellt. [2]

Hypothesen

Im Rahmen der Datenanalyse sollen die folgenden Hypothesen überprüft werden:

1. Bei Fahrzeugen mit Elektromotoren ist die Verknüpfungs- sowie Aktivierungsrate höher als bei Fahrzeugen mit Verbrennungsmotoren. Dies ist auf spezielle Services für Elektrofahrzeuge zurückzuführen, die von entscheidender Relevanz für die Nutzung des Fahrzeuges sind.
2. Die Verknüpfungs- und Aktivierungsrate ist im asiatischen Markt im Vergleich zum europäischen Markt generell höher.
 - a. Dies ist auf ein grundsätzlich jüngeres Alter bei den Fahrzeughaltern zurückzuführen, wodurch diese in der Regel technikaffiner sind.
 - b. Die Quote von Elektrofahrzeugen ist im asiatischen Markt deutlich höher als im europäischen Markt.
3. Bei Elektrofahrzeugen ist der Nachkauf von Services und Produkten deutlich höher als bei Fahrzeugen mit Verbrennungsmotor. Der Grund hierfür sind die speziell für Elektrofahrzeuge zugeschnittenen Services, welches es für Verbrennern nicht gibt.

Abschluss

Am Ende der Arbeit sollen die anfangs gesetzten Ziele erfüllt sowie die Hypothesen bestätigt oder entkräftet sein. Zum anderen soll auf Basis der Analyse ein konkretes Ergebnis zur Erhöhung der Verknüpfungs-

und Aktivierungsrate präsentiert werden. Hierdurch ist es möglich, auf Basis der gewonnenen Erkenntnisse aus dem Vergleich der beiden Fahrzeuggruppen, eine Handlungsempfehlung zum Beispiel für die Implementierung neuer Services bei Mercedes me abzuleiten.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Sara D'Onofrio and Andreas Meier. *Big Data Analytics Grundlagen, Fallbeispiele und Nutzungspotenziale*. Springer Vieweg, 2021.

Verfahren zur Ermittlung der Support-Ticket Eskalationsstufe durch den Einsatz von Social Listening

Ebru Dogan

Anke Bez

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Kunden kaufen oft aus den unterschiedlichsten Gründen ein: Um Bedürfnisse zu befriedigen, wegen des Ansehens, aus reiner Neugier oder einfach um eigene Wünsche zu erfüllen. [1] Der Support der Kunden ist auch nach dem Kauf sehr wichtig, um erstens die Kunden langfristig an sich zu binden und zweitens, um die Kundenzufriedenheit zu erhöhen. So werden Marke und Produkt zukünftig von Kunden weiter empfohlen. ‚Nach dem Kauf ist vor dem Kauf‘ beschreibt dieses Ziel am besten.

Motivation

Häufig, wenn ein Kunde ein kleines oder großes Problem mit einem gekauften Produkt hat, ist die erste Reaktion meistens folgende: Im Internet nach einer Lösung suchen. Das Internet bietet die Möglichkeit mit einem Klick auf viele Informationen schnell und einfach zuzugreifen. Dadurch können die unterschiedlichsten Plattformen, wie z.B. Firmenwebsites, Social Media, Foren usw. nach einer Lösung abgesucht werden. Hierbei fällt dann auf, dass bereits frühere Kunden evtl. dasselbe Problem hatten. Daher ist dies auch meistens die erste Option des Kunden, um nach einer möglichen Lösung zu suchen. Die zweite Option, die für die Kunden danach meistens in Frage kommt, ist der Support selbst, der von den meisten Unternehmen zusätzlich zu den verkauften Produkten angeboten wird. Es gibt viele Varianten, wie der Support von Unternehmen kontaktiert werden kann, z.B. FAQs, Telefon, E-Mail, Chat-Bot usw.. Unabhängig davon, welche Möglichkeiten zur Kontaktaufnahme das Unternehmen anbietet, entstehen hierfür im Hintergrund interne Kosten. Diese Kosten können je nach Aufkommen der Supportanfragen hoch oder niedrig sein.

Zielsetzung

Ein mittelständisches Maschinenbauunternehmen aus Baden-Württemberg möchte mithilfe einer Social Listening-Lösung herausfinden, ob ein Zusammenhang zwischen den Supportanfragen und den Social Media-Diskussionen hergestellt werden kann. Auf welchen Social Media-Plattformen sprechen Kunden aktiv über die Produkte bzw. produktbezogenen Probleme und was sind die Themen? Dadurch soll dann bestimmt werden, welche Probleme ein Einzelfall sind und welche Probleme mehrmals auftreten, um dann dem Ganzen, z.B. durch erhöhte Social Media-Aktivität oder Suchmaschinenoptimierung, entgegenzuwirken. Am Ende soll letztendlich die Kundenzufriedenheit durch die verbesserte Reaktion des Unternehmens erhöht werden und die internen Kosten sollen durch die Abnahme der Supportanfragen gesenkt werden.

Einsatz von Social Listening im Support

Social Listening durchsucht die sozialen Medien nach den für ein Unternehmen relevanten Informationen. Social Listening-Tools beobachten Plattformen wie YouTube, Facebook, Twitter und Instagram, aber auch kleinere Online-Auftritte wie Bewertungsportale, Blogs oder Foren. Es hilft Unternehmen zu erkennen, was über sie in den sozialen Medien gesprochen wird. Aber auch wo sich die Konkurrenz bewegt oder unternehmensrelevante Themen können auf Beobachtung gesetzt werden. Auf Basis der gewonnenen Daten können Trends analysiert werden. Im Laufe dieser Bachelorarbeit wird das Social Listening-Tool Talkwalker verwendet. [3]



Abb. 1: Prozess zum Social Listening [2]

Vorgehensweise

Der Prozess vom Social-Listening sieht folgendermaßen aus (vgl. Abb. 1): Mithilfe von Talkwalker soll erkannt werden, auf welchen Plattformen die Kunden unterwegs sind und was die (häufigsten) Gesprächsthemen sind. Jedoch muss hierfür eine Query erstellt werden, die mit sinnvollen Suchbegriffen gefüllt ist. Für diesen Schritt sind folgende Mittel eine Hilfestellung: Dashboards von Power BI, in denen die bekannten Fehler weitgehend kategorisiert werden, und das CRM-Tool, mit dem die Support-Anfragen bearbeitet werden. Da im Rahmen dieser Bachelorarbeit nur ein bestimmtes Produkt des Unternehmens angeschaut wird, wurden nur die fünf häufigsten produktbezogenen Fehler aus Power BI ermittelt, um anschließend diese dann im CRM zu suchen. So ist die Auflistung von fast allen wichtigen (Haupt-)Suchbegriffen abgedeckt. Wenn die Query steht, beginnt Talkwalker mit der Suche, indem es verschiedene Plattformen nach den festgelegten Suchbegriffen absucht. Mithilfe der ersten Ergebnisse

wird die Query dann nochmal erweitert, indem fehlende Suchbegriffe mit erzielten Treffern miteinbezogen werden. Die (End-)Ergebnisse werden analysiert und anschließend werden Verbesserungsvorschläge bestimmt, die das Unternehmen bei voraussichtlichen Maßnahmen unterstützen sollen. Ziel ist es, wie bereits am Anfang erwähnt, durch diese Maßnahmen die internen Kosten zu senken und die Zufriedenheit der Kunden zu erhöhen.

Aktueller Stand und Ausblick

Zum Zeitpunkt der Verfassung dieses Artikels wurden bereits die ersten Suchbegriffe identifiziert und zu einer Liste erarbeitet. Im nächsten Schritt wurden die Queries mithilfe dieser Suchbegriffe aufgestellt, die im Hintergrund bei Talkwalker laufen und Ergebnisse generieren. Im nächsten Schritt werden dann diese Ergebnisse analysiert, um Verbesserungen und Handlungsempfehlungen auszusprechen.

Literatur und Abbildungen

- [1] Ralf Armbrüster. Warum Kunden kaufen – 8 Gründe für eine Kaufentscheidung. <https://www.souveraen-verkaufen.de/warum-kunden-kaufen-gruende-fuer-eine-kaufentscheidung/>, 04 2016.
- [2] Eigene Darstellung.
- [3] bvik eV. Was ist Social Listening? Eine Erklärung. <https://bvik.org/b2b-glossar/social-listening/>, 2019.

Automatisierung der Regressionstests für eine komplexe Applikationswelt

Yannick Ege

Astrid Beck

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Daimler AG, Sindelfingen

Motivation

Bei der Softwareentwicklung zielt man darauf ab vordefinierte Anforderungen umzusetzen, ein gewisses Maß an Qualität bereitzustellen. Ist die Software qualitativ nicht ausreichend, kann es zu großen Problemen führen. Es kann zu Verlust von Geld und Zeit kommen, der Geschäftsruf kann geschädigt werden, bis hin zu Verletzungen von Personen. Beispielsweise der Autopilot in Fahrzeugen verdeutlicht das Sicherheitsrisiko von Software. Daher ist es wichtig die Qualität und Funktionalität der Software zu prüfen, um Risiken eines Softwareausfalles oder Softwarefehler zu minimieren und sicherzustellen, dass die Software den Anforderungen entsprechen. Durch Softwaretests erhält man eine Einschätzung der Softwarequalität und kann auf erkannte Fehler reagieren. [3] Mit Software-Testing ist Aufwand verbunden, sei es die Testplanung, der Testentwurf, die Testrealisierung, die Bewertung oder der Abschluss der Testaktivität. Unter anderem durch Regressionstests versucht den Aufwand zu minimieren und die Testausführungszeit/Testzyklen zu verkürzen. Regressionstests haben zum Ziel, dass durch Veränderungen oder Modifikationen am System keine Seiteneffekte oder neuen Fehlerzustände entstehen (Regressionen). Vorhandene Tests werden nach Veränderungen am System durchgeführt, diese prüfen unveränderte Features, dass diese weiter lauffähig sind und unverändert funktionieren. Eine Art das Verfahren anzuwenden ist, vorhandenen Tests an der neuen Produktversion zu wiederholen. Es müssen Tests ausgewählt werden, die alte Funktionalitäten zum Gegenstand haben. Diese Tests müssen wiederholbar sein, damit das der Fall ist, ist es wichtig, dass diese manuellen Tests gut dokumentiert sind. Regressionstests müssen oft und regelmäßig ausgeführt werden, somit eignen sich diese gut für die Testautomatisierung. [2]

Zielsetzung

Ziel der Arbeit ist für ein bestehendes System eine Toolchain zu entwickeln und zu implementieren, so dass die UI Tests automatisiert durchgeführt werden können. Die Toolchain soll in einen bestehenden Entwicklungsprozess implementiert werden, nach einem bestimmten Ereignis, sollen automatisierte UI Tests das bestehende System testen. Vordefinierte und gleichbleibende Testfälle sollen in der Testumgebung das vorhandene System automatisch testen, damit aufgrund der Veränderungen keine Fehler bei den schon lauffähigen Funktionen entstehen (Regressionen). Fragen wie Beispielsweise:

- Erscheint die gewünschte Tabelle beim Klick des entsprechenden Buttons?
- Wird ein eingefügter Datensatz eindeutig integriert? (der aktuelle Eintrag soll auch bei passender Filtereinstellung als erstes angezeigt werden)
- Bleibt das Layout konsistent?
- Sind die Metriken (Ladezeit etc.) performant/benutzerfreundlich für die GUI?

Diese und weitere Themen sollen auf der UI getestet werden. Die Tests sollen für verschiedenen Applikationen anwendbar sein, auf Web-Applikationen, auf Android und auf IOS. Die ausgewählten Tools und die darauf aufbauende Toolchain soll möglichst so gewählt werden, dass eine Erweiterung zu API Tests möglich ist. Um einen gezielten Mehrwert aus dem Test zu generieren, sollen die Daten gesammelt und in einem Monitoring dargestellt werden.

Vorgehensweise

Zu Beginn wird auf die Grundlagen des Testens von Software eingegangen, dabei wird der Testbegriff dargestellt, Bestandteile eines Tests aufgezeigt und

Testverfahren vorgestellt. Die jeweiligen Teststufen werden aufgezeigt. Speziell auf UI Tests und Regressionstests wird eingegangen. Weiterführend wird eine allgemeine Definition der Testautomatisierung vorgestellt, Ziele und Vorteile der Testautomatisierung erarbeitet und die Risiken der Testautomatisierung dargelegt. Zudem werden Teststrategien verglichen. Es wird ein Vorgehen, wie Testing automatisiert werden kann, definiert. Die Rahmenbedingungen der Arbeit und somit unveränderliche Faktoren werden beschrieben und ein Bild über das bestehende System gebildet. Somit werden die relevanten Komponenten und ihre Funktionen untersucht, der technische Hintergrund und ihre Zusammenhänge werden durchleuchtet. Das Altsystem wird mit dem aktuell entwickelten System verglichen, speziell die Verbesserungen der UI hinsichtlich der Benutzerfreundlichkeit werden festgehalten. Die Nutzergruppen des Systems werden aufgezeigt. Schließlich werden die Anforderung an die Testautomatisierung definiert. Wurden die Rahmenbedingungen definiert, wird der Bewegungsspielraum und damit veränderliche Komponenten durchleuchtet. Der Entwicklungsprozess wird analysiert, mögliche Prozess, in der die UI-Tests integriert werden können, identifiziert. Um ein Konzept der Toolchain zu erstellen, wird

eine Benchmark der zur Verfügung stehenden Tools durchgeführt und anhand der erarbeiteten Kriterien das passende Werkzeug zur Automatisierung ausgewählt. Letztendlich wird ein Konzept der Toolchain erarbeitet. Prüfpläne für exemplarische Tests werden konzipiert und aufgebaut. Sind die Konzepte erstellt, beginnt die Implementierung. Die ausgewählten Tools werden in den bestehenden Entwicklungsprozess implementiert. Es werden exemplarische Tests durchgeführt, die entstandenen Daten werden gesammelt und ein Testmonitoring aufgesetzt. Schließlich wird ein Fazit der Arbeit festgehalten.

Ausblick

Im Rahmen der Arbeit werden exemplarische Tests automatisch durchgeführt, die im Projekt aktuell manuell durchgeführten Tests müssen zunächst automatisiert werden. Als nächster Schritt wird angestrebt, dass die Tests auf Integrationsebene automatisiert werden. Schließlich wird im Projekt ein Test-Driven-Development angestrebt, wonach anstatt erst den Quellcode zu schreiben und diesen dann zu testen, zuerst die Tests geschrieben werden. [1]

Literatur und Abbildungen

- [1] Ionos. Test Driven Development: So funktioniert die Methode. <https://www.ionos.de/digitalguide/websites/webentwicklung/was-ist-test-driven-development/>, 2020.
- [2] Richard Seid, Manfred Baumgartner, Thomas Bucsics, and Stefan Gwihs. *Basiswissen Testautomatisierung: Konzepte, Methoden und Techniken*. dpunkt.verlag, 2015.
- [3] Andreas Spillner and Tilo Linz. *Basiswissen Softwaretest. Aus- und Weiterbildung zum Certified Tester Foundation Level nach ISTQB-Standard. 6., überarbeitete und aktualisierte Auflage*. dpunkt.verlag, 2019.

Gamification - Beurteilung der Einsatzmöglichkeiten im deutschen Finanzdienstleistungssektor

Jonathan Knuplesch

Anke Bez

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Gamification

Gamification ist ein Begriff, der besonders im letzten Jahrzehnt immer mehr an Bedeutung gewonnen hat. Diese Entwicklung hängt zum einen mit der voranschreitenden Digitalisierung als auch mit dem breiten Einsatzspektrum zusammen. Im Allgemeinen wird Gamification als der Einsatz von Spieldesign-Elementen in nicht Spielkontexten, definiert. [4] Spieldesign-Elemente sind die grundlegenden Bausteine einer Gamification und können bspw. Abzeichen, Bestenlisten, Leistungsgraphen oder Punkte sein. Einsatzfelder findet Gamification vor allem innerhalb des Marketings, in der Verhaltensänderung und im Lernumfeld. [11] So ist der Kerngedanke von Gamification Menschen auf emotionaler Ebene einzubeziehen und bei diesen Motivation auszulösen. [2] In diesem Bezug bietet es sich an Gamification mit dem Finanzdienstleistungssektor, in einem digitalen Kontext, zu verknüpfen.

Problemstellung

Kaum ein Bereich ist einem so stetigen Wandel ausgesetzt wie der Finanzdienstleistungssektor. [8], [10] In der aktuellen Zeit können daher für Finanzdienstleister vor allem Herausforderungen, in Verbindung mit technologischem Fortschritt, zunehmendem Wettbewerb und verändertem Kundenverhalten, identifiziert werden. [1] [5] Der Treiber dieser Herausforderungen ist dabei vor allem die Digitalisierung. So treten infolge dessen, immer mehr branchenfremde Wettbewerber in den Sektor ein und erhöhen den Druck auf die Finanzdienstleister. [9], [6] Dazu sorgen das Internet und die damit einhergehenden Vergleichsmöglichkeiten für die Kunden, zu einer Markttransparenz, die die Wettbewerbssituation zusätzlich verschärft. [3] Weiterführend sind in Folge der Digitalisierung Veränderungen im Kundenverhalten und in den Anforderungen an Finanzprodukte erkennbar, die dazu führen, dass auch hier eine Anpassung notwendig ist (vgl. Abbildung 1). [7] Letztlich wird auch eine Kundengruppe immer relevanter, die unter dem Namen 'Digital Natives'

zusammengefasst wird. Diese charakterisiert sich durch die intensive Nutzung digitaler Kanäle und stellt die zukünftige Generation an Kunden dar. [1] Folglich müssen Finanzdienstleister sich mit den veränderten Kundenanforderungen auseinandersetzen und Wege finden, um zukünftige Generationen (bspw. "Digital Natives") als Kunden zu gewinnen. Zusätzlich müssen Möglichkeiten zur Differenzierung identifiziert werden, um sich, mit Blick auf die Zukunft, auf dem Markt durchzusetzen.

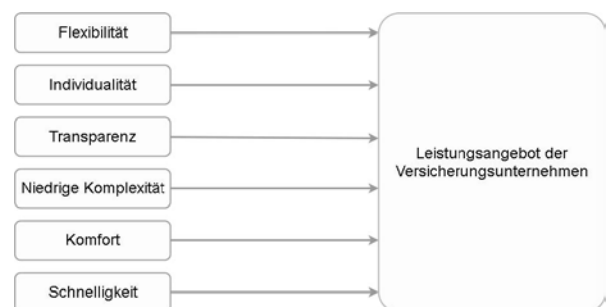


Abb. 1: Aktuelle Kundenanforderungen am Beispiel Versicherung in Anlehnung an Altunas und Uhl [1]

Zielsetzung

Mithilfe dieser Arbeit soll aufgezeigt werden, in welchem Umfang und in welcher Form Gamification im deutschen Finanzdienstleistungssektor bereits eingesetzt wird und welche Rolle Gamification, mit Blick auf die Zukunft, einnehmen wird. In diesem Kontext findet eine Verknüpfung mit den Herausforderungen des Finanzdienstleistungssektors statt und es soll aufgezeigt werden, inwiefern der Einsatz von Gamification, diesbezüglich unterstützen kann.

Struktur und Methodik

Für eine Verknüpfung von Gamification mit dem Finanzdienstleistungssektor wurde sowohl eine intensive Literaturrecherche als auch Interviews mit Gamification Experten aus dem Banken- und dem Versicherungssektor, durchgeführt. Auf Grundlage dessen sollten Einsatzgebiete für Gamification identifiziert werden und die Bereiche Bank und Versicherung, in diesem Bezug, verglichen werden.

Ergebnis

Die Verknüpfung und die Auswertung der Literaturrecherche und der Experteninterviews, zeigt auf, dass eine Anwendung von Gamification im Finanzdienstleistungssektor sinnvoll ist. Für eine erfolgreiche Implementierung allerdings sowohl ein zielgruppenorientiertes Konzept als auch eine geschickte Kombination von Spieldesign-Elementen, notwendig ist. Gamification kann in diesem Bezug dazu eingesetzt werden, um:

- die Kundenbindung zu verbessern.
- Digital Natives als Kundengruppe zu gewinnen.
- eine Differenzierung gegenüber dem Wettbewerb zu erreichen.

Im Hinblick auf die Unterschiede zwischen dem Banken- und dem Versicherungssektor lassen sich besonders Unterschiede in der Form identifizieren, wie Gamification eingesetzt werden sollte. Dies hängt unter anderem damit zusammen, dass Banken deutlich mehr

Kontaktpunkte mit Kunden aufweisen als Versicherer. Innerhalb des Bankensektors könnte Gamification eingesetzt werden, um die digitale Customer Journey zu beeinflussen und diese zu steuern. Auf diese Weise könnten Operationen innerhalb der Bankingapp vereinfacht werden und in einer Form gestaltet werden, die angenehmer für den Kunden ist. Aufgrund des Charakters der Kundenbeziehung ist eine Gamification Umsetzung innerhalb, des Bankensektors, in Kombination mit bestehenden Prozessen und Operationen umsetzbar und unterscheidet sich in der Form, grundlegend von einer Umsetzung, innerhalb des Versicherungssektors. So macht es im Versicherungskontext wenig Sinn, die komplette Customer Journey zu gamifizieren. Gamification sollte hierbei eher als eine Ergänzung gesehen werden, die durch eine sinnvolle Verknüpfung mit anderen Lebensbereichen, wie Mobilität, Vitality, Nachhaltigkeit und Fitness, einen Mehrwert für den Kunden schaffen kann.

Zusammenfassend kann durch einen sinnvollen Einsatz von Gamification, Vorteile sowohl aufseiten der Finanzdienstleister als auch aufseiten der Kunden, generiert werden. In der aktuellen Zeit wird Gamification noch kaum innerhalb des deutschen Finanzdienstleistungssektors eingesetzt. Mit Blick auf die Zukunft lässt sich allerdings sagen, dass Gamification sicherlich auch im deutschen Finanzdienstleistungsbereich, Anwendungsfelder finden wird. So wird Gamification zukünftig, besonders in Bezug auf die Steuerung und die Beeinflussung der Kundenbeziehung eingesetzt werden, und zwar eine wichtige Rolle im digitalen Kontext einnehmen, allerdings den Sektor nicht grundlegend revolutionieren.

Literatur und Abbildungen

- [1] Muhammed Altuntas and Pacal Uhl. *Herausforderungen an die Versicherungswirtschaft*. Springer Gabler, 2016.
- [2] Brian Burke. *Gamify*. Taylor and Francis, 2014.
- [3] Pascal Bühler. Industrialisierung der Assekuranz in einer digitalen Welt. <https://www.ivw.unisg.ch/~media/internet/content/dateien/instituteundcenters/ivw/studien/industrialisierung-digital2015.pdf>, 2015.
- [4] Sebastian Deterding. Gamification: Toward a definition. https://www.researchgate.net/publication/273947177_Gamification_Toward_a_definition, 2011.
- [5] Andreas Dombret. Die Herausforderungen der Digitalisierung für Banken und Bankenaufsicht. <https://www.bundesbank.de/de/presse/reden/die-herausforderungen-der-digitalisierung-fuer-banken-und-bankenaufsicht-710900>, 2016.
- [6] Lars Fend and Jürgen Hofmann. *Digitalisierung in Industrie-, Handels- und Dienstleistungsunternehmen*. Springer Gabler, 2018.
- [7] Ute Lohse and Annemarie Will. Rahmenbedingungen und Strategische Herausforderungen für die Versicherungsbranche. In *Handbuch Versicherungsmarketing*, volume 2019, pages 3, 14. Zerres Reich, 2019.
- [8] Detlef Rahmsdorf. Trends im finanzdienstleistungssektor und die Folgen für die Strategie von Banken. *Konsequenzen wirtschaftsrechtlicher Normen*, 2002:119, 2002.
- [9] Bernd Rolfes and Holger Wessling. *Innovative Strategien für bekannte Herausforderungen - Optimistischer Ausblick in die Zukunft?* Springer Fachmedien Wiesbaden, 2020.
- [10] Georg Walter. *Banking Innovation 2016*. Springer Gabler, 2020.
- [11] Gabe Zichermann and Christopher Cunningham. *Gamification by design*. O'Reilly, 1. edition, 2011.

Analyse, Entwurf und Implementierung einer Deep-Learning-Lösung zur visuellen Erkennung von Hand-Objekt-Interaktionen mit Werkzeugen

Marcel Gauss

Reiner Marchthaler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma AKKA Industry Consulting GmbH, Stuttgart

Motivation

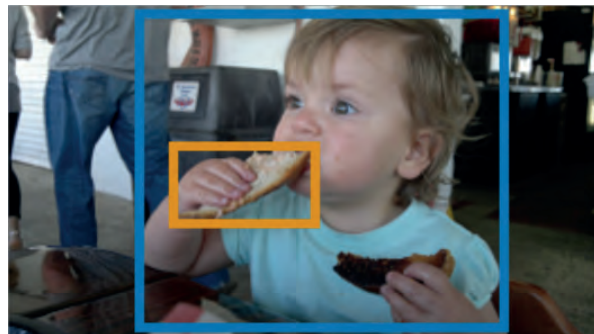
Manuelle Montagetätigkeiten sind ein wesentlicher Bestandteil der industriellen Wertschöpfungskette, wobei sie je nach Produkt bis zu 70% der Gesamtfertigung in Anspruch nehmen können. Hierbei sind die Fehler die durch Menschen bei der Montage eines Produktes auftreten, das Ergebnis einer mangelhaften Arbeitsausführung. Durch sogenannte Assistenzsysteme können menschliche Fehler in der manuellen Montage vermieden werden [5]. Die Assistenzsysteme kombinieren die Vorteile der computergestützten Datenverarbeitung mit den Fähigkeiten eines Menschen. Mithilfe von Augmented Reality (AR) können die Monteure dazu angeleitet werden, Montageschritte in der richtigen Reihenfolge, an der korrekten Stelle mit den richtigen Komponenten und Werkzeugen auszuführen. Dadurch kann sichergestellt werden, dass kein fehlerhaft montiertes Produkt das Unternehmen verlässt [3]. Aufgrund der sich durch die Assistenzsysteme ergebenden Vorteile, hat die AKKA Industry Consulting GmbH die Vision eines Assistenzsystems, welches Monteure während des Montageprozesses anleitet und eine fehlerfreie Fertigung sicherstellt.

Ziel der Arbeit

Das Ziel dieser Arbeit ist es, prototypisch ein System zu entwerfen und zu implementieren, welches visuell die Interaktionen mit Werkzeugen erkennt, damit zukünftig in einem Assistenzsystem geprüft werden kann, ob die erforderlichen Interaktionen tatsächlich stattgefunden haben. Im genaueren soll im Bereich des Deep Learnings eine effektive und effiziente Methode ausfindig gemacht werden, sodass das System in Echtzeit mit einer möglichst hohen Genauigkeit auf einer HoloLens V2 lauffähig ist. Hierfür muss das System in der Lage sein, aus egozentrischer Sicht die Interaktionen mit Werkzeugen als auch dessen Position zu bestimmen.

Human Object Interaction Detection

Die sogenannte Human Object Interaction Detection (HOI-Detection) hat das Ziel, Menschen und Objekte zu lokalisieren und zudem die Interaktion zwischen ihnen zu erkennen. Hierfür erfordert die Aufgabe die Lokalisierung und Beschreibung von Paaren, interagierender Menschen und Objekten [6]. Abbildung 1 zeigt hierbei an einem Beispiel einer Prädiktion einer HOI-Detection auf.



<Mensch isst Brot>

Abb. 1: Darstellung einer HOI-Detection Prädiktion. Prädiziert werden die enthaltenen Subjekte (Mensch) und Objekte (Brot) sowie das Prädikat (isst). Die Position von Subjekt und Objekt werden mittels einer Bounding Box prädiziert. [1]

Eine Human-Object-Interaction (HOI) wird dabei als Triplet (Subject, Prädikat, Objekt) definiert. Bei HOI-Triplets ist das Subjekt immer ein Mensch, sodass sich die interessierenden Interaktionen auf Paare von Prädikaten und Objekte beschränken. Typischerweise werden Subjekt sowie Objekt als beschriftete Bounding Box dargestellt [6].

Datenbeschaffung

Damit das System auf der Zielhardware die bestmöglichen Ergebnisse bezüglich der Genauigkeit erzielt, wird der Trainingsdatensatz mittels der HoloLens V2 aufgenommen. Hierbei werden die Interaktionen mit den Werkzeugen in einer simplifizierten Umgebung aufgenommen. Abbildung 2 zeigt hierbei die Umgebung auf. Durch die gewählte Umgebung ist es möglich, eine Aussage treffen zu können, mit welcher Genauigkeit die Interaktionen aus unterschiedlichen Perspektiven erkannt werden.



Abb. 2: Dargestellt ist die simplifizierte Umgebung für die Erkennung der Interaktionen mit Werkzeugen (links). Zudem ist in Farben (rechts) aufgezeigt, zwischen welchen Perspektiven unterschieden wird. [2]

Data Augmentation

Die sogenannte Data Augmentation ist eine Technik, womit die Größe des Trainingsdatensatzes künstlich vergrößert werden kann, indem modifizierte Daten aus einem vorhandenen Datensatz erstellt werden. Dadurch kann dem Overfitting entgegengewirkt werden und eine bessere Leistung eines Modells erreicht werden. Die Aufgabe der Data Augmentation besteht darin, Translationsinvarianzen wie beispielsweise Blickwinkel, Maßstäbe, Verdeckungen, Bewegungsunschärfe und Beleuchtungen in den Datensatz so einzubauen, sodass das resultierende Modell trotz dieser Herausforderungen gut funktioniert. Bei den Techniken der Data Augmentation für Bilder wird zwischen grundlegenden Bildmanipulationen und synthetischen Daten unterschieden [4]. In dieser Arbeit finden ausschließlich grundlegende Bildmanipulationen statt. Die im folgenden genannten Bildmanipulationstechniken werden hierbei auf den Trainingsdatensatz angewandt:

- Spiegeln
- Translation
- Zuschneiden
- Farbraumtransformationen
- Noise Injection
- Random Erasing

Abbildung 3 zeigt die in dieser Arbeit verwendeten Techniken auf. Die Transformationen werden so implementiert, dass diese labelbewahrend sind, sodass keine weitere manuelle Annotationen notwendig sind.



Abb. 3: Darstellung der in dieser Arbeit verwendeten Techniken der Data Augmentation [2]

Realisierung und Ausblick

Zur Realisierung des Systems der visuellen Erkennung von Interaktionen mit Werkzeugen wird ein HOI-Detection Modell der Aufgabe entsprechend trainiert. Abbildung 4 zeigt dabei eine vereinfachte Darstellung des Systemaufbaus auf. Zu Beginn werden die Bilddaten mittels einer HoloLens V2 aufgenommen. Anschließend werden diese manuell annotiert. Die annotierten Daten werden aufgeteilt in einen Datensatz für das Training und die Validation. Der Datensatz für das Training wird mittels der Data Augmentation künstlich vergrößert, sodass das Modell robust bei

Translationsinvarianzen ist. Mittels dem Trainingsdatensatz wird ein HOI-Detection Modell der Aufgabe entsprechend trainiert. Das HOI-Detection Modell wird anhand einer Evaluation bezüglich der Echtzeitfähigkeit als auch Genauigkeit ausgewählt. Das resultierende Modell aus dem Training wird infolgedessen bezüglich der Genauigkeit validiert. Bei Optimierungspotenzialen wie Over- oder Underfitting fließen diese Erkenntnisse erneut in die Datenbeschaffung und Datenaufbereitung ein. Sobald das Modell einen zufriedenstellenden Stand erreicht, kann es für eine zukünftige Inferenz auf der HoloLens V2 verwendet werden.

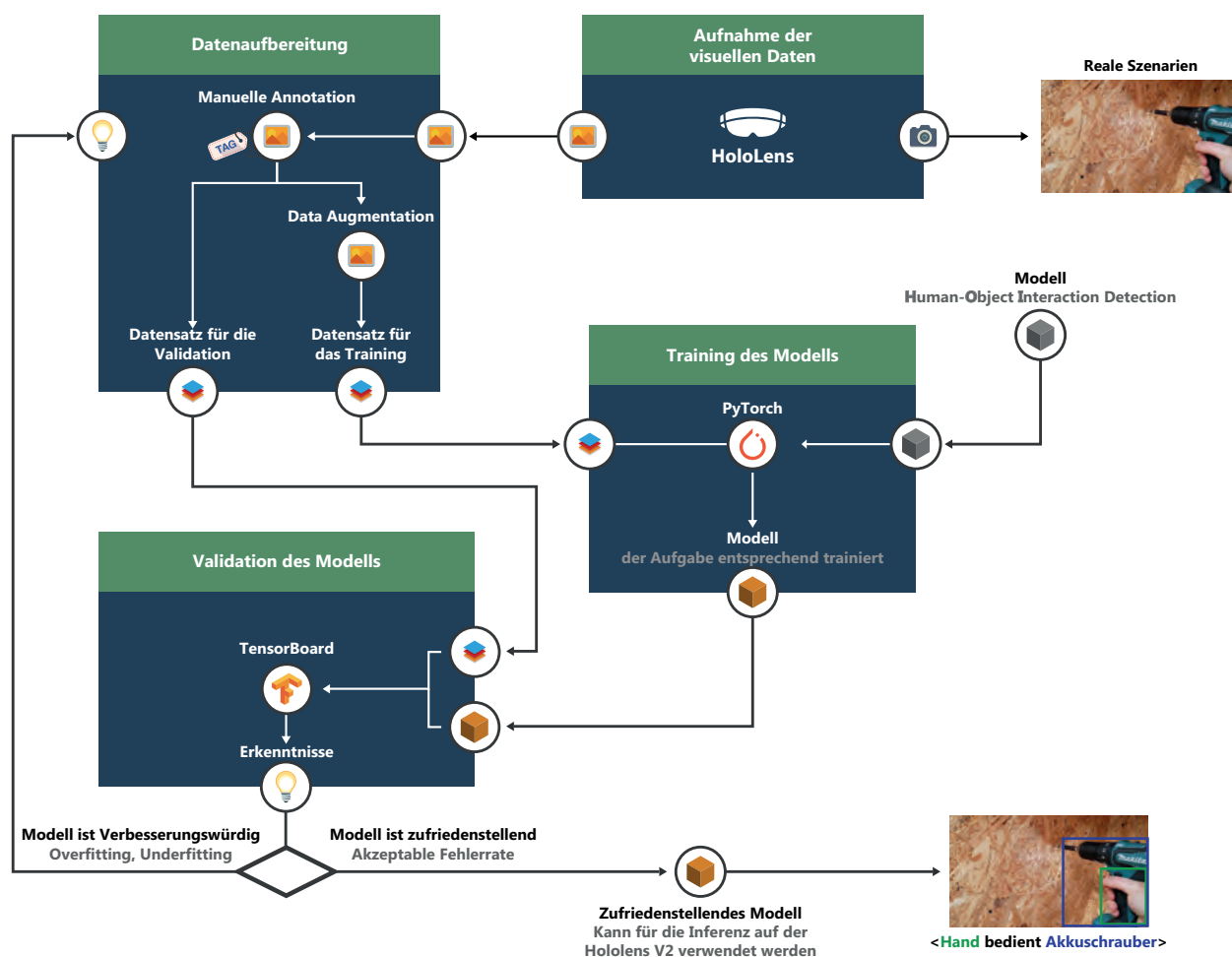


Abb. 4: Darstellung des Systemaufbaus [2]

Literatur und Abbildungen

- [1] Kim Bumsoo, Lee Junhyun, Kang Jaewoo, Kim Eun-Sol, and J. Kim Hyunwoo. HOTR: End-to-End Human-Object Interaction Detection with Transformers. *arXiv*, 2021.
- [2] Eigene Darstellung.
- [3] Susanne Nördinger. So funktioniert Montage 4.0. <https://www.produktion.de/technik/so-funktioniert-montage-4-0-351.html>, 2020.
- [4] Connor Shorten and Taghi M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, pages 4–14, 2019.
- [5] Oliver Sträter and Karsten Loer. *Risikofaktor Mensch? Zuverlässiges Handeln gestalten*. Beuth Verlag, 1 edition, 2019.
- [6] Cheng Zou, Bohan Wang, Yue Hu, Junqi Liu, Qian Wu, Yu Zhao, Boxun Li, Chenguang Zhang, Chi Zhang, Yichen Wei, and Jian Sun. End-to-End Human Object Interaction Detection with HOI Transformer. *arXiv*, page 1, 2021.

Entwicklung eines Verkehrsschildsystems mit Maschinencode für eine autonome Buslinie zur Informationsgewinnung

Jonathan Grau

Reiner Marchthaler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Ziel

Das Ziel der Bachelorarbeit ist die Entwicklung eines Verkehrsschildsystems zur Informationsgewinnung für eine autonome Buslinie. Ein Maschinen 2D-Code, welcher beim Lesen auf einen Server linkt, liefert mit Hilfe dieses Servers verschiedenste Informationen. Dazu gibt der Bus Informationen an den Server weiter, z. B. zur Überwachen des Verkehrsaufkommens. Dieser Informationsaustausch ermöglicht nicht nur dem Bus einfacher zu fahren, sondern auch die anderen Verkehrsteilnehmer.

Ein genereller Überblick über die Bachelorarbeit zeigt das Flussdiagramm. Ein Modell aus dem maschinellen Lernen wird durch einen erstellten Datensatz auf das Erkennen von den 2D-Codes trainiert. Hierfür wird zuvor eine Menge an Codes generiert und markiert. Die angeschlossene Kamera nimmt dann ein Bild auf und das trainierte Modell überprüft, ob sich ein Code darauf befindet. Wenn ein Code erkannt wird schneidet ein drittes Programm das restliche des Bildes weg, so dass nur noch der erkannte Code übrig bleibt. Mithilfe des einzelnen Codes wird wie oben schon beschrieben zu einem Server verbunden und der Informationsaustausch durchgeführt.

Problembehandlung

Ein Maschinencodesystem entwickeln, welches auch über 60-80 Meter durch eine Kamera für einen fahrenden Bus erkennbar ist. Hierfür die vorhandenen 2D-Codesysteme verstehen und ein eigenes verbessertes System entwickeln. Informationen variabel verfügbar zu machen, zum Beispiel durch einen Server auf dem sich der Bus verbindet, wenn er den Code gelesen hat. Dazu soll der Bus Informationen zur Verfügung stellen, welche der Server für andere Verkehrsteilnehmer nutzen kann.

Forschung

Durch dieses Projekt soll geforscht werden, auf welcher Distanz ein Maschinencode lesbar ist. Dazu kommt, dass aktuelle automatisierte Fahrzeuge sich darauf fokussieren, aktuelle Verkehrsschilder zu erkennen. Mithilfe eines speziellen Verkehrsschildsystems nur für Autos ist es möglich mehr Informationen auszutauschen und auch das Auto Informationen weitergeben lassen. Das neue Verkehrsschildsystem soll eine verbesserte Version der vorhandenen 2D-Codes verwenden.

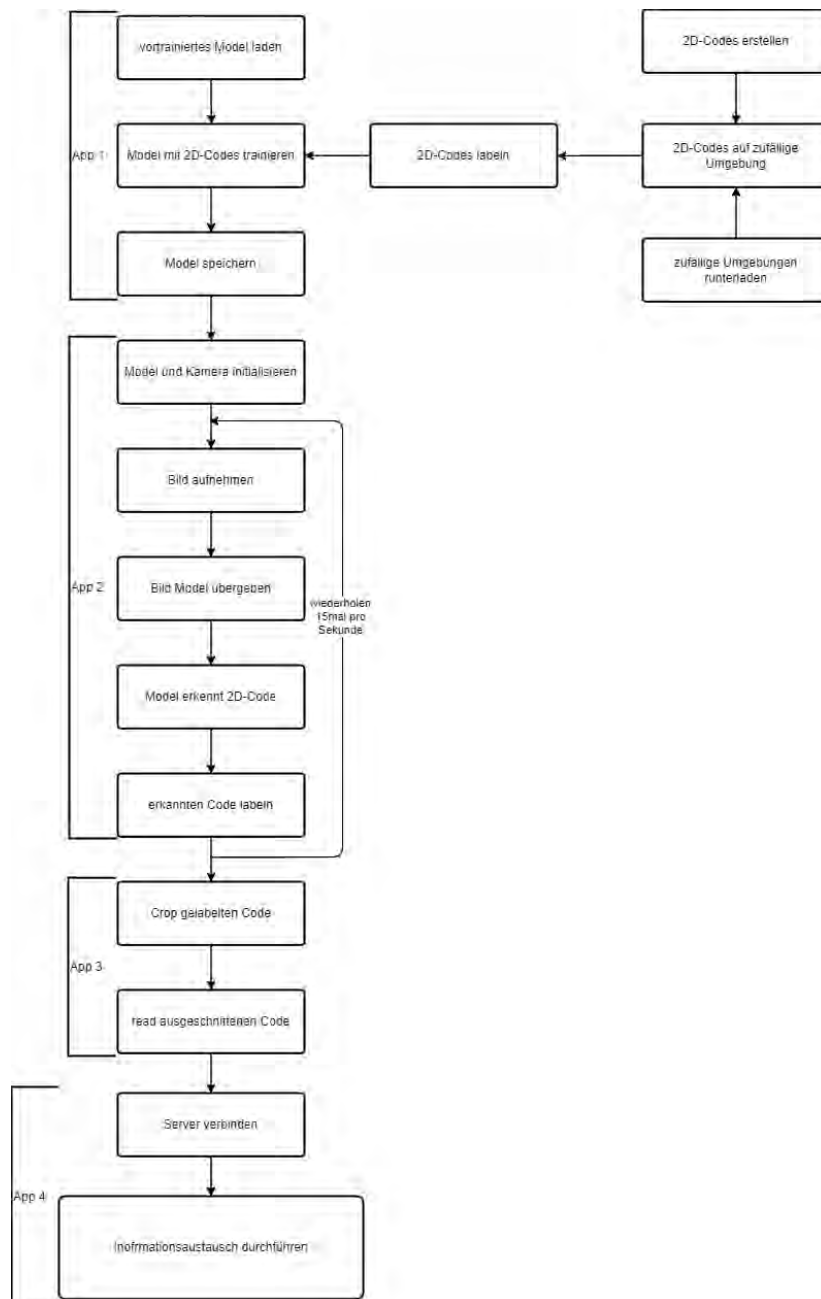


Abb. 1: Flussdiagramm als Überblick über die Bachelorarbeit [1]

Literatur und Abbildungen

[1] Eigene Darstellung.

Erhöhung der Prozessautomatisierung durch den Einsatz von RPA-Technologie an einem Beispielprozess bei GS/PE-SC

Cengizhan Guenduezeri

Clemens Klöck

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Robert Bosch GmbH, Stuttgart

Einleitung

In wirtschaftlich schwierigen Zeiten überprüfen Unternehmen permanent ihre Position im Wettbewerb und es wird immer wichtiger, Innovationen und Wettbewerbsvorteile zu schaffen. Ihre Prozesse effizient und profitabel zu gestalten, ist für viele Unternehmen zu einer der wichtigsten Aufgaben geworden. Die Digitalisierung und Automatisierung von einzelnen Geschäftsprozessen kann Unternehmen einen enormen Vorteil verschaffen. Allerdings ist dies ebenfalls mit Kosten verbunden. Hierbei kann der Einsatz von Softwaresystemen eine wichtige Rolle spielen. Die Digitalisierung erfordert Innovationskraft und steht oft den Einsparungszielen entgegen. Die Einführung von Robotic Process Automation (RPA) ist eine Methode, die heutzutage oft zur Anwendung gebracht wird. Diese wird genutzt, um mittels Softwareroboter, Geschäftsprozesse zu automatisieren.

Problemstellung

Die Problemstellung bei dieser Arbeit ist der erfolgreiche Einsatz von RPA-Technologie im Prozess der E-Mail-Kommunikation und Reporting für den Bereich GS/PE-SC. Es stellt sich die Frage, wie der Einsatz dieser Technologie die manuelle Arbeitskraft ersetzen und somit effizientere Arbeit fördern, die Fehleranfälligkeit reduzieren und den manuellen Aufwand verringern kann.

Zielsetzung

Ziel und Zweck der Arbeit ist es, den Prozess der E-Mail-Kommunikation mithilfe von RPA-Technologie zu automatisieren, einen Mehrwert zu bieten, den Kundennutzen zu stiften, eine Erhöhung der Qualität zu erreichen, manuelle Prozessschritte zu reduzieren und diesen Prozess zu beschleunigen. Durch diese

Automatisierung wird einem bestimmten Mitarbeiterkreis automatisch eine Erinnerungsmail verschickt, mit der Option, diese interaktiv zu beantworten und dem Central Testmanagement Team die Planung der Tests zu erleichtern, die Qualität der Antworten zu steigern und die Antwortgeschwindigkeit zu erhöhen. Dieser Einsatz soll die Abteilung entlasten, um den bisher manuell getätigten Prozess viel schneller auszuführen, die Qualität der Ergebnisse zu erhöhen, die Kundenzufriedenheit durch einen standardisierten Prozessablauf anzuregen und die gesparte Zeit anderweitig für andere Tätigkeiten zu nutzen. Des weiteren soll durch die Automatisierung die Planung der Tests erleichtert werden und mögliche Fehl- oder Falscheinaben vermieden werden. Dies ist ein wichtiger Schritt zur Automatisierung des Fachbereiches und eine Motivation, auch weitere Tätigkeiten im Bereich GS/PE-SC zu automatisieren.

Robotic Process Automation (RPA)

Durch die zunehmende Automatisierung in der heutigen Zeit haben sich eine Vielzahl von Technologien entwickelt. Robotic Process Automation (RPA) ist eine dieser Technologien. Der Einsatz von RPA ermöglicht es, Software-Roboter zu erstellen, die menschliche Handlungen imitieren. Diese Roboter sind in der Lage, ganze Geschäftsprozesse oder einzelne Prozessschritte selbstständig und automatisch auszuführen. Der Einsatz von künstlicher Intelligenz (KI) und Machine-Learning Fähigkeiten kommen dabei zum Einsatz. Menschliche Handlungen oder Benutzerinteraktionen im Prozess werden vom Roboter nachgeahmt und die Interaktion mit den am Prozess beteiligten Systemen oder Anwendungen werden somit durchgeführt. [3] Mit dem Einsatz von RPA-Bots werden folgende Verbesserungen beabsichtigt: Qualität, Produktivität, Flexibilität, Sicherheit. [2]

Prozessablauf

Der Prozess startet an bestimmten Daten im Jahr, indem E-Mails an die jeweiligen Personen der Entwickler-Teams verschickt wird. Die Daten sind in einer Excel Datei gespeichert und müssen von dem Bot gelesen werden. Die gesendete E-Mail leitet die Personen zu einem Formular weiter, dieses kann

ausgefüllt werden, sofern Test-Support benötigt wird. Nach erfolgreicher Eingabe der relevanten Daten, wird das Formular abgeschickt und die Daten werden in einer Datei festgehalten. Die Übersicht ist zugänglich für das Central Testmanagement Team. Anschließend wird aus dieser Datei ein Report hergestellt. Die erste Version des genaueren Prozessablaufes ist hier abgebildet. (siehe Abbildung 1)

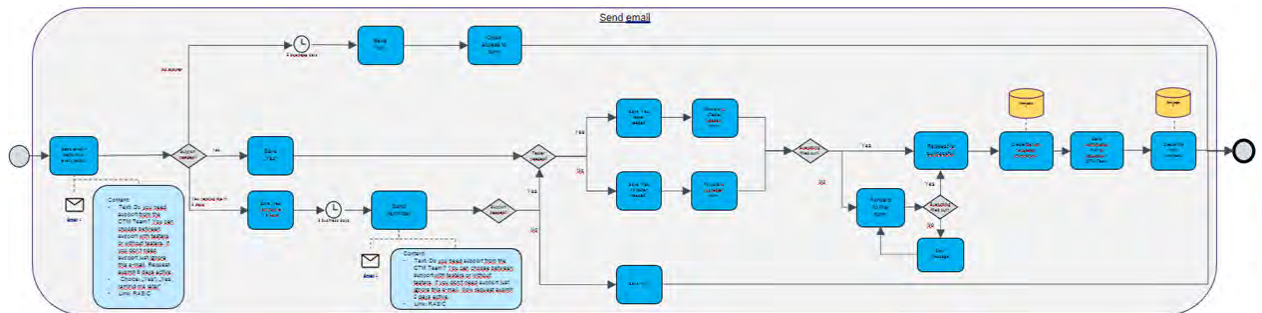


Abb. 1: Prozessablauf [1]

Realisierung

Die Realisierung des Projektes erfolgt über die Plattform Catalytic. Catalytic bietet eine Cloud-Plattform und dient einmal als Schnittstelle zum Erstellen von No-Code-Workflows und als Anwendung zum Ausführen und Verwalten der Workflows. Die Oberfläche der Plattform ist benutzerfreundlich für sofortige Ein-

richtung, schnelles Erstellen und agile Änderungen. Unternehmen verwenden die maßgeschneiderten, digitalisierten Workflows von Catalytic, um Aufgaben im Hintergrund zu koordinieren und zu automatisieren, damit Mitarbeiter in ihrem Betrieb in kurzer Zeit mehr zu erreichen können. [4] Die Bausteine von Catalytic ermöglichen es, den Prozessablauf zu automatisieren und den manuellen Prozess automatisch zu tätigen.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Olaf Eisele. Robotic Process Automation (RPA). https://www.arbeitswissenschaft.net/fileadmin/Downloads/Angebote_und_Produkte/Zahlen_Daten_Fakten/ifaa_Zahlen_Daten_Fakten_RPA_2019-12-16.pdf, 2019.
- [3] Alok Mani Tripathi. *Learning Robotic Process Automation*. Packt Publishing, 2018.
- [4] Catalytic Webseite. Frequently asked questions. <https://www.catalytic.com/platform/faqs>, 2015.

Prinzipien der IT-Sicherheit in der Automotive Cybersecurity

Umutcan Guerbuez

Clemens Klöck

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma ALTEN GmbH, Stuttgart

Einleitung

Cyber-Angriffe auf Autos: „Der Fluch der Vernetzung“ [1], „Der Feind fährt mit“ [3], „Bundesamt warnt vor Cyberangriffen auf vernetzte Autos“ [5]. Dies sind nur ein paar der wenigen Schlagzeilen der Medien über die Thematik von Cyberangriffen auf vernetzte Autos. Obwohl diese Thematik für viele noch vor wenigen Jahren kein wirkliches Gesprächsthema war, gewinnt Cybersecurity im Automotive Bereich, durch die fortgeschrittene Automatisierung von Fahrzeugen immer mehr an Bedeutung. Die Zukunft der Automobilwelt entwickelt sich immer weiter in Richtung autonomes Fahren. Vollständig autonom fahrende Fahrzeuge ohne Fahrer klingen zunächst interessant und verlockend, doch ist der Trend von autonomen Fahrzeugen und dem autonomen Fahren auch ohne Cybersecurity möglich? Welche Bedrohungen bringen vollständig autonom fahrende Fahrzeuge mit sich? Wie sieht die aktuelle Infrastruktur im Bereich Cybersecurity für Fahrzeuge aus? Dies sind vermutlich Fragen, die in den nächsten Jahren häufiger auftreten werden. Große Automobilhersteller, investieren immer mehr Kapital für die Entwicklung autonomer Fahrzeuge. Der steigende Automatisierungsgrad der Fahrzeuge bringt auch immer weiter einen größeren Bedarf an Cybersecurity. Die Implementierung von einer Cybersecurity Infrastruktur spielt dabei eine wichtige Rolle. Unternehmen müssen sich zukünftig immer weiter mit Thematiken bezüglich Cybersecurity im Automotive-Bereich beschäftigen. Zu diesen Thematiken zählt auch unter anderem eine erfolgreiche und sichere Vernetzung der Fahrzeuge miteinander. Jedoch ist die Vernetzung der Fahrzeuge allein nicht für eine erfolgreiche Cybersecurity Infrastruktur ausreichend. Die Vernachlässigung von Cybersecurity kann die Ursache für ein großes Problemfeld werden. Ein Cyberangriff auf ein autonomes Fahrzeug kann das Leben von Menschen kosten. Die Aufgabe gilt deshalb darin, diese Angriffe durch die Implementierung von Cybersecurity zu hindern. Diese Arbeit beschäftigt sich

mit den Prinzipien von IT-Sicherheit im Automotive Bereich und klärt grundlegenden Fragen in diesem Themengebiet auf. Hierfür werden Standards wie die ISO 26262-Norm und die ISO 21434 Norm aufgegriffen. Ebenfalls soll diese Arbeit zeigen, in welchen Bereichen Potential vorhanden ist. Hierbei werden die Potentiale analysiert und dem Leser nähergebracht.

Problemstellung

Autonome Fahrzeuge bilden die Zukunft der Automobilwelt. Der damit zusammenhängende Cybersecurity Bedarf, welcher gedeckt werden muss, sorgt für die Entwicklung eines Kreislaufes. Dieser Kreislauf muss ständig aktuell gehalten werden. Mehr Autonomie bedeutet auch mehr Angriffsfläche für ein Fahrzeug. Diese Angriffsfläche kann nur durch eine durchdachte Cybersecurity-Infrastruktur beseitigt werden. Die Aufgabe liegt darin entscheiden zu können, welche Risiken, welchen Einfluss mit sich bringen. Diese Entscheidung zu treffen bedingt die Durchführung einer effektiven Risikoanalyse und Bewertung. Diese Arbeit ermöglicht den Studenten das Potenzial, welches noch im Automotive Cybersecurity-Bereich verfügbar ist, zu analysieren. Die ständigen Fortschritte von Cybersecurity im Automotive Bereich, bringen auch immer Aktualisierungen und Anpassungen der Normen mit sich.

Zielsetzung

Die Zielsetzung dieser Abschlussarbeit ist es die Rolle von Cybersecurity im Automotive-Bereich zu analysieren und diese im Anschluss dem Leser näherzubringen. Diese Arbeit soll grundlegendes Wissen über die Infrastruktur von Cybersecurity im Automotive-Bereich vermitteln und die Antwort, für die am häufigsten gestellten Fragen in diesem Themengebiet bieten. Ein weiteres Ziel dieser Arbeit ist Themen wie Notfallmanagement, Hochverfügbarkeit und Business Continuity auf die IT-Sicherheit im Automotive-Bereich

zu übertragen. Der Leser soll im Anschluss der Arbeit grundlegendes Wissen über die Standards von Cybersecurity im Automotive-Bereich erhalten und die Wichtigkeit und das Potenzial von Cybersecurity in diesem Bereich erkennen und kritisch bewerten können.

Was ist IT-Sicherheit?

Unter dem Begriff IT-Sicherheit ist die Sicherheit von Informationen und Daten, welche durch elektronische Systeme verarbeitet werden, zu verstehen. IT-Sicherheit ermöglicht das Nutzen von IT-Services, obwohl diese mit Risiken verbunden sind. [2] IT-Sicherheit lässt sich wiederum in weitere Sicherheitsbereiche unterteilen. Diese sind die Anwendungssicherheit, Informationssicherheit, Netzwerksicherheit und die Systemsicherheit. Unter der Anwendungssicherheit lässt sich die Sicherheit von verschiedenen Anwendungen beschreiben, die Dienste, Benutzer- und Anwendungsschnittstellen anbieten, welche zusammen verknüpft werden können. Bei der Informationssicherheit geht es wiederum, um die Sicherheit von Informationen und Daten, welche von einer Anwendung, erzeugt, gemessen, gespeichert und übertragen werden. Die Netzwerksicherheit beschreibt die Sicherheit des Netzes, welches auf Hardware- und Softwarekomponenten basiert. Diese Komponenten werden von verteilten Anwendungen für die Übertragung von Daten und Informationen, über verschiedene Standorte verwendet. Die Systemsicherheit beschreibt hingegen die Sicherheit von Computersystemen, wie Handys, Laptops oder Tablets, welche von verteilten Anwendungen verwendet werden. [4]

Warum braucht die Automobilwelt Cybersecurity?

Eine erfolgreiche Cybersecurity-Infrastruktur ist für die Automobilwelt unumgänglich. Ein unberechtigter

Zugriff auf sicherheitsrelevante elektronische Steuerungssysteme von Fahrzeugen kann hierbei eine ernsthafte Bedrohung darstellen. Dies hat sich im Jahr 2015 durch ein Forschungsprojekt bestätigt. [6] Im Rahmen dieses Forschungsprojekts haben die Automobilsicherheitsforscher Charlie Miller und Chris Valesek ein Fernangriff auf den Jeep Cherokee durchgeführt. Durch diesen Fernangriff haben die Forscher die Kontrolle über die Bremsen und über das Gaspedal vom Fahrzeug übernommen. Es stellte sich heraus, dass für diesen Angriff eine Sicherheitslücke verwendet wurde. [7] Durch diesen Angriff hat sich die Bedeutung von Cybersecurity verstärkt. Um Angriffe solcher Art zu verhindern ist die Implementierung einer Cybersecurity Infrastruktur für die Automobilwelt wichtig.

Ausblick

Durch die steigende Nachfrage für autonome Fahrzeuge wird Cybersecurity im Automotive-Bereich weiterhin an Bedeutung gewinnen. Die steigende Autonomie der Fahrzeuge wird auch in der Zukunft den Cybersecurity-Bedarf in der Automobilbranche steigern. Große Automobilhersteller werden weiterhin auf Cybersecurity setzen müssen, um die Sicherheit für autonom Fahrzeuge im Straßenverkehr gewährleisten zu können. Normen wie die ISO 21434 können der Automobilindustrie bei der Implementierung einer Cybersecurity-Infrastruktur helfen. Jedoch ist die ISO 21434 Norm derzeit in der Entwicklungsphase, deshalb lassen sich hierbei durchaus Potentiale ermitteln. Der laufende Wandel in Richtung autonomes Fahren sorgt dafür, dass diese Normen sich auch in der Zukunft weiterhin anpassen und fortschreiten müssen, um die Automobilwelt während dieses Wandels zu unterstützen.

Literatur und Abbildungen

- [1] Marie-madeleine Aust. Cyber-Angriffe auf Autos: Der Fluch der Vernetzung. <https://www.next-mobility.de/cyber-angriffe-auf-autos-der-fluch-der-vernetzung-a-1066429/>, 2021.
- [2] Eberhard von Faber. *IT und IT-Sicherheit in Begriffen und Zusammenhängen*. Eberhard von Faber, 2021.
- [3] Tom Grünweg. Cyberattacken auf Autos: Der Feind fährt mit. <https://www.spiegel.de/auto/aktuell/cyberattacken-auf-autos-der-feind-faehrt-mit-a-1084059.html>, 2016.
- [4] M. Kathiresh and R. Neelaveni. *Automotive Embedded Systems*. M. Kathiresh ■ R. Neelaveni, 2021.
- [5] Deutschland Redaktionsnetzwerk. Bundesamt warnt vor Cyberangriffen auf vernetzte Autos. <https://www.rnd.de/mobilitaet/bundesamt-warnt-vor-cyber-angriffen-auf-vernetzte-autos-5LYSTM2O5ZBXRHP2EENUKQGDIA.html>, 2021.
- [6] Lars Schnieder and René Sebastian Hosse. *Leitfaden Automotive Cybersecurity Engineering*. Schnieder, Lars; Hosse, René Sebastian, 2018.
- [7] Madhusudan Singh. *Information Security of Intelligent Vehicles Communication*. Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland, 2021.

Kamerabasierte Detektion von Fahrgassen und Pfadberechnung für landwirtschaftliche Fahrzeuge

Kai Haeppler

Thao Dang

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma AGCO GmbH, Marktoberdorf

Einleitung

Die Landwirtschaft und ihre Erzeugnisse sind ein wichtiger Bestandteil für die Gesellschaft. Einerseits trägt sie dazu bei, Nahrungsmittel für die Bevölkerung zu erzeugen. Andererseits werden auch Energieträger wie Strom, Kraftstoff und Wärme aus nachhaltig angebauten Rohstoffen gewonnen. Da aktuelle landwirtschaftliche Betriebe eine immer größere Fläche bewirtschaften, benötigen Landwirte Landmaschinen mit einer hohen Flächenleistung [1]. Um das Arbeiten mit diesen Landmaschinen zu erleichtern und den Automatisierungsgrad zu steigern sind verschiedene Fahrerassistenzsysteme nötig. Beispielsweise halten GNSS-gestützte (Global Navigation Satellite System) Spurführungssysteme die landwirtschaftlichen Fahrzeuge exakt auf dem im Vorfeld geplanten Pfad bzw. der Spurlinie. Dadurch wird die Überlappung bei der Feldbearbeitung minimiert und zusätzlich der Fahrer entlastet. Somit kann sich der Fahrer verstärkt auf den eigentlichen Arbeitsprozess konzentrieren, um in weiterer Folge eine bessere Arbeitsqualität zu erreichen. Da eine durchgängige Verfügbarkeit der vorgeplanten Spurlinien bzw. des relativ genauen GNSS mit RTK-Korrektur (Real Time Kinematic) nicht über alle Bearbeitungsprozesse sichergestellt werden kann, soll im Rahmen dieser Arbeit eine kamerabasierte Spurführung entlang der bestehenden Fahrgassen untersucht werden.

Problemstellung

Damit ein landwirtschaftliches Fahrzeug mittels Kamerasystem automatisiert entlang einer Fahrgasse geführt werden kann, muss die Fahrgasse innerhalb des Kamerabildes erkannt und der zugehörige Pfad daraus abgeleitet werden. Abbildung 1 zeigt exemplarisch das Bild einer Fahrgasse, das im März beim Düngen aufgenommen wurde. Dabei liegt die Herausforderung der Fahrgassenerkennung in der Diversität der landwirtschaftlichen Umgebung. Diese beinhaltet viele verschiedene Kulturen, wie Weizen, Gerste oder

Mais, die sich je nach Jahreszeit in unterschiedlichen Wachstumsstadien befinden. Außerdem erhöhen inhomogene Wetter- und Lichtverhältnisse die Diversität der Umgebung. Neben den Umwelteinflüssen stellt die Eigenbewegung der Kamera hervorgerufen durch die Nick- und Wankbewegungen des Fahrzeugs eine weitere Herausforderung dar.

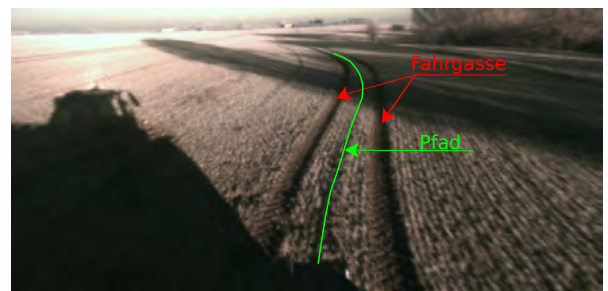


Abb. 1: Dargestellt ist eine Fahrgasse (rote Markierung) eines Weizenfelds im März. Dabei sind die inhomogenen Lichtverhältnisse, verursacht durch Sonne und Schatten deutlich zu erkennen. Die grüne Linie zeigt den Pfad des landwirtschaftlichen Fahrzeugs. [4]

Ziel der Arbeit

Das Ziel dieser Arbeit ist die Konzeption und prototypische Entwicklung eines Systems, das die Fahrgassen innerhalb des Kamerabildes erkennt und anschließend daraus den Pfad für das landwirtschaftliche Fahrzeug berechnet. Dabei soll die verwendete Algorithmik echtzeitfähig sein, sodass im späteren Betrieb Fahrgeschwindigkeiten von bis zu 25 km/h erreicht werden können. Zusätzlich sollte das entwickelte System möglichst robust gegenüber den beschriebenen Umwelteinflüssen sein. Daher soll in dieser Arbeit untersucht werden, welche Bildsegmentierungsverfahren geeignet sind, um die Fahrgassen zu erkennen. Außerdem wird

analysiert, wie der Pfad anhand der segmentierten Bilder bestimmt werden kann.

Konzeption

Die in Abbildung 2 dargestellte Konzeption des Systems zeigt eine Übersicht der einzelnen Prozessschritte. Dabei wird zunächst ein Bild der landwirtschaftlichen Umgebung durch eine am Fahrzeug montierte Kamera aufgenommen. Anschließend folgt eine Bildsegmentierung, welche die Fahrgasse innerhalb des Bildes erkennt. Im nächsten Schritt findet die Pfadberechnung statt. Der berechnete Pfad wird infolgedessen durch eine Koordinatentransformation von Bild- in Fahrzeugkoordinaten transformiert. Somit kann dieser an den Pfadregler übergeben werden, welcher den Lenkwinkel des Fahrzeugs einstellt.

Bildsegmentierung

Das Ziel einer Bildsegmentierung ist die pixelweise Klassifizierung des Eingabebildes. Dabei zeigt Abbildung 3 einen Überblick verschiedener Segmentierungstechniken, wie beispielsweise Thresholding, Region Based Segmentation oder Deep Neural Network Based Segmentation. Einige kamerabasierte Reihen-erkennungsverfahren aus dem landwirtschaftlichen Bereich sowie Linienerkennungsverfahren aus dem Bereich des autonomen Fahrens arbeiten mit einer schwellwertbasierten Segmentierung (Thresholding). Da sich hierbei das Farbspektrum der zu erkennenden Merkmale gut vom Hintergrund unterscheidet und annähernd konstant ist, kann dieses durch eine Vorverarbeitung verstärkt werden. Beispiele hierfür sind grüne Farbtöne von Pflanzen auf braunem Ackerboden oder weiße Farbtöne von Spurlinien auf dunklem Asphalt. Anschließend können die entsprechenden Pixel durch Thresholding Methoden wie beispielsweise der Otsu-Methode detektiert werden [3], [6], [7]. Unterscheiden sich die zu erkennenden Merkmale kaum vom Hintergrund, so sind schwellwertbasierte Segmentierungen eher ungeeignet, da die Häufigkeitsverteilung der Pixelwerte nicht mehr bimodal ist. Daher werden bei relativ diversen Umgebungen häufig CNNs (Convolutional Neural Networks) eingesetzt, um mit Faltungsoperationen die Strukturen in den Bilddaten zu erkennen. Hierbei bewegt sich ein Faltungskernel durch das Eingabebild und extrahiert die Merkmale. Ein komplettes CNN besteht dabei aus mehreren sequenziell angeordneten Schichten, welche Faltungsoperationen durchführen, um somit schrittweise die relevanten Merkmale zu extrahieren. Damit ein CNN für eine spezifische Anwendung verwendet werden kann, müssen die Faltungskernel und gegebenenfalls weitere Parameter der Aufgabe entsprechend trainiert werden. Daher ist ein aussagekräftiger Datensatz nötig, der die

Beschaffenheit der Umgebung möglichst gut beschreibt [8].

Pfadberechnung

Für die Berechnung des Pfades wird das segmentierte Bild häufig durch Anwendung einer Homographie in die Draufsicht transformiert. Somit kann dem perspektivischen Effekt der Kameraperspektive (siehe Abb. 1) entgegengewirkt werden, bei dem sich parallel verlaufenden Spurlinien im hinteren Bildbereich verjüngen. Das vereinfacht die Annäherung des Pfades an die segmentierten Pixel durch nachfolgende Algorithmen. Die beiden Fahrspuren können beispielsweise durch eine Polynomapproximation zweiten Grades angenähert werden. Dabei wird für jede Fahrspur ein adaptives, verschiebbares Rechteck definiert, das den Bereich der Polynomapproximation begrenzt. Somit können fehlerhafte Pixel der Segmentierung ausgeschlossen werden. Die Rechtecke werden schrittweise entlang der beiden Fahrspuren verschoben. Zusätzlich sorgt ein integrierter Kalmanfilter dafür, dass die Polynomapproximation robust funktioniert, auch wenn eine Fahrspur unterbrochen ist [3]. Neben der Polynomapproximation mit adaptivem Rechteck zur Begrenzung ist die Modellierung der Fahrspur als B-Spline dritter Ordnung ein weiterer Ansatz den Pfad zu bestimmen. Dabei wird der RANSAC (Random Sample Consensus) Algorithmus verwendet, um die Kontrollpunkte des B-Splines zu bestimmen. Durch diesen iterativen Prozess können fehlerhafte Detektionen aussortiert werden. Somit wird ein B-Spline ermittelt, welcher die Fahrspur möglichst gut approximiert [2].

Ausblick

Für die prototypische Entwicklung des Systems wird zunächst die Diversität des landwirtschaftlichen Umfelds beschränkt. Beispielsweise genügt es vorerst einen Datensatz zu generieren, der eine Kultur zu einer bestimmten Jahreszeit repräsentiert, jedoch aber einige unterschiedliche Lichtverhältnisse zu unterschiedlichen Tageszeiten enthält. Basierend auf diesem Datensatz werden im Rahmen der Arbeit verschiedene Bildsegmentierungsalgorithmen untersucht und evaluiert, ob diese für die Fahrgassenerkennung geeignet sind. Dabei ist die Robustheit gegenüber inhomogener Lichtverhältnisse ein zentraler Bestandteil der Evaluation. Des Weiteren werden die Ergebnisse der Bildsegmentierung für die Entwicklung der Pfadberechnung verwendet. Es wird untersucht, welche Algorithmen für die Pfadberechnung geeignet sind. Hierbei muss zusätzlich berücksichtigt werden, dass die Fahrgassen auch Abzweigungen, Einmündungen oder Kreuzungen enthalten können.



Abb. 2: Konzeption des Systems bestehend aus den einzelnen Prozessschritten. Der rote Rahmen markiert dabei den Fokus der Arbeit. [4]

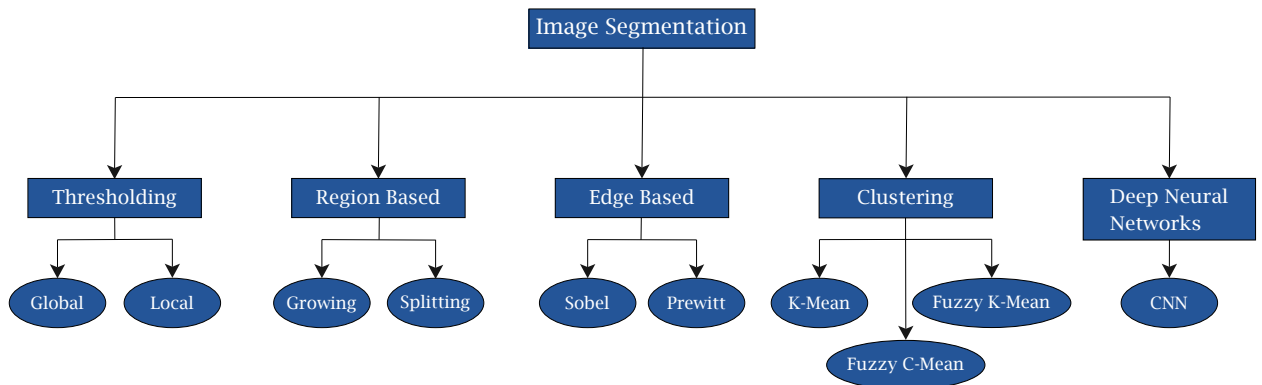


Abb. 3: Übersicht verschiedener Bildsegmentierungsmethoden. Eigendarstellung angelehnt an [5]

Literatur und Abbildungen

- [1] Statistisches Bundesamt. Strukturwandel in der Landwirtschaft hält an. https://www.destatis.de/DE/Presse/Pressemitteilungen/2021/01/PD21_028_412.html, 01 2021.
- [2] J. Cao, C. Song, S. Song, F. Xiao, and S. Peng. Lane Detection Algorithm for Intelligent Vehicles in Complex Road Conditions and Dynamic Environments. *MDPI sensors*, 2019.
- [3] C. Chen, L. Tang, Y. Wang, and Q. Qian. Study of Lane Recognition in Haze Based on Kalman Filter. *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, 2019.
- [4] Eigene Darstellung.
- [5] S. K. Dubey and S. Vijay. A Review of Image Segmentation using Clustering Methods. *Number 5 International Journal of Applied Engineering Research*, pages 2484–2489, 2018.
- [6] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Pearson Education, 4 edition, 2018.
- [7] C. Tu, B. J. van Wyk, K. Djouani, and Y. Hamam. An Efficient Crop Row Detection Method for Agriculture Robots. *2014 7th International Congress on Image and Signal Processing*, 2014.
- [8] A. Zhang et al. Dive into Deep Learning. <https://d2l.ai/d2l-en.pdf>, 2021.

Konzeption und Implementierung eines RFID-Firmwareupdate-Mechanismus für das EUCHNER Electronic-Key-System

Wolfgang Hamberger

Mirko Sonntag

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Euchner GmbH + Co. KG, Leinfelden-Echterdingen

Motivation

EUCHNER entwickelt und fertigt Produkte für die Industrie im Bereich der Sicherheitstechnik. Das EUCHNER Electronic-Key-System (EKS) ist ein System zur Bestimmung von Zugangsberechtigungen und der Festlegung von Betriebsmodi an Maschinen und Anlagen, die sogenannte sichere Betriebsartenwahl. [2] Mitarbeiter erhalten Electronic-Keys in Form von Schlüsselanhängern, die sie am Schreib-/Lesegerät mittels RFID-Technik authentifizieren.

Euchner entwickelt ein kundenspezifisches EKS, bei welchem die sichere Betriebsartenwahl mit RFID-Karten im Kreditkartenformat erfolgt. Die Firmware des EKS wird via USB aufgespielt und aktualisiert; im eingebauten Zustand ist über diese Schnittstelle jedoch der Rechner der Anlage verbunden. Somit ist nach bestimmungsgemäßem Einbau des EKS ein Zugriff von außen nicht mehr möglich, weswegen die RFID-Technik auch bei der Übertragung eines Firmwareupdates Verwendung finden soll. Die neu entwickelte EUCHNER Active Configuration Card (ACC) ist eine Platine mit den Abmessungen einer RFID-Karte, welche mittels USB-Kabel mit einem Rechner verbunden wird. Sie soll die Kommunikation mit dem EKS ermöglichen.

Aufgabenstellung

Ziel dieser Arbeit ist die Softwareentwicklung aller an der kabellosen Aktualisierung der EKS-Firmware beteiligten Komponenten. Dieses Programm spricht die ACC an, welche Befehle und Nutzdaten drahtlos via RFID an das EKS überträgt. Das EKS kann die Daten verarbeiten und sich anschließend selbst aktualisieren. Der Vorgang weicht aus Sicht des Benutzers nicht vom bisherigen Updateprozess ab. Die Firmware der ACC wird aufgrund der Hardwarenähe in C geschrieben und implementiert das Modbus-Protokoll zur Kommunikation mit dem PC. Auf der ACC findet ein STM32F302K8 als Mikroprozessor

Verwendung, der über I²C mit dem Dynamic Tag IC M24SR64 verbunden ist, welcher die Kommunikation von und zur RFID-Antenne ermöglicht.

Konzept

Die grundsätzliche Überlegung ist, die im Vergleich hohe Rechenleistung und den Speicherplatz vom PC soweit zu beanspruchen, dass ACC und EKS entsprechend ressourcenschonend betrieben werden können. Somit werden bereits im PC werden die jeweiligen Datenpakete erstellt. Aufgrund der zwei unterschiedlichen Übertragungswege – von PC zu ACC über USB-Kabel und drahtlos von ACC zu EKS – sind zwei verschiedene Protokolle im Einsatz.

Die Pakete zwischen ACC und EKS stellen sich laut Standard des Near Field Communication Forums (NFC-Forum) als maximal 8192 Bytes große NDEF-Dateien dar (*NFC Data Exchange Format*-Dateien). Im ersten Schritt wird die Firmware als Bytestring eingelesen und in Abschnitten von 4096 Bytes aufgeteilt, der Größe eines Segments im Speicher des EKS. Vor jeden Abschnitt wird ein Header vorangestellt, der die für das EKS relevanten Befehle enthält. Die ersten beiden Bytes einer NDEF-Datei sind für die zu berechnende NDEF Message Length reserviert, der Größe des Payloads inklusive des erzeugten Headers.

Im zweiten Schritt des Python-Skripts wird die NDEF-Datei für die Kommunikation mit der ACC über das Modbus-Protokoll vorbereitet. Das Datenfeld eines Modbus-Frames kann bis zu 252 Bytes betragen, weswegen die NDEF-Datei in Abschnitten von 128 Bytes zerteilt wird. Jedes Modbus-Datenfeld erhält einen Header, der im Zustandsautomat der ACC ausgewertet wird. Wird das Kommando *Firmwareupdate* erkannt, speichert die ACC den Payload der Datenfelder in einem Buffer, die zusammengesetzt wieder eine NDEF-Datei ergeben. Nach vollständiger Übertragung einer NDEF-Datei empfängt die ACC den Befehl *Write* und

schickt sie an das EKS über die RFID-Schnittstelle. Der Ablauf ist in Abbildung 1 kompakt dargestellt. Die Firmware des EKS wird um die NDEF-Funktionalität erweitert, sodass dieses sowohl ein Update via Kabel, als auch drahtlos verarbeiten kann. Hierbei kann auf Teile der ACC-Firmware zurückgegriffen werden, um empfangene Pakete, die als Daten vom Typ NDEF erkannt wurden, von der Antenne einzulesen und deren Inhalte zu verarbeiten. Das EKS kann nach der Implementierung auf ähnliche Weise Daten wie z. B. Logs oder seinen Status als NDEF-Dateien an die ACC senden.

Das Python-Skript, welches die NDEF-Dateien erzeugt, dient als Schnittstelle, wodurch sich für den Anwender nur der Funktionsaufruf ändert.

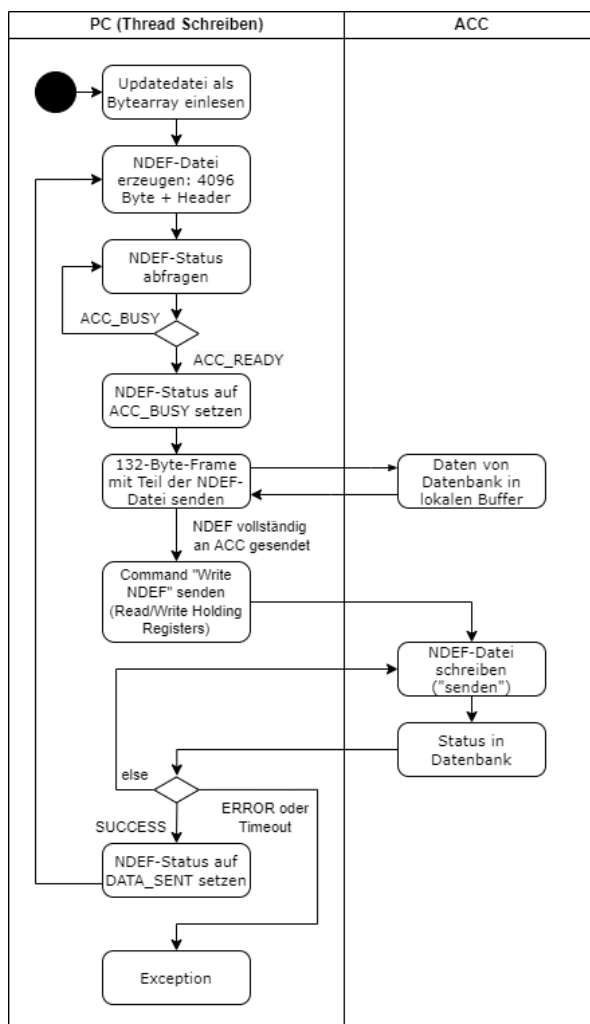


Abb. 1: Ablauf Firmwareupdate PC→ACC [1]

Implementierung

Da die Kommunikation mit der ACC über Modbus erfolgt, wie es bislang mit dem EKS der Fall ist, können bestehende Python-Module hierzu verwendet werden. Das Python-Skript erhält beim Aufruf das auszuführende Kommando, ggf. die Firmwareupdate-Datei und für das EKS relevante Informationen als Argumente. Im Falle eines Firmwareupdates wird im ersten Schritt ein bis zu 4096 Bytes großer Abschnitt der Firmware zwischengespeichert und dafür ein Header erstellt, der u. a. den Befehl für das EKS enthält. Davor wird nach NDEF-Standard die 2 Bytes lange *NDEF Message Length* gesetzt, also die Größe der NDEF-Datei. Dieses Bytearray wird an eine Funktion übergeben, die es für das Versenden via Modbus in gleicher Vorgehensweise noch weiter zyklisch zerteilt und mit je einem Header versieht.

Der Zustandsautomat zur Verarbeitung von Befehlen des PCs wird im Task *DynamicTagTask* realisiert. Tasks werden vom Scheduler verwaltet und zyklisch aufgerufen. Nachdem ein Modbus-Frame empfangen wurde, lädt der Task die enthaltenen Daten in einen lokalen Buffer und wertet den Befehl aus, der in den ersten Bytes jedes Frames enthalten ist. Wird eine Firmwaredatei übertragen, speichert der Task den Payload des Frames nacheinander in ein Array, das die NDEF-Datei darstellt. Empfängt der Task das Kommando *Update*, ruft er die Funktion *WriteNDEF* mit dem Array als Argument auf, wodurch die Daten an den Tag-IC M24SR64 gesendet werden.

Um den Aufbau bereits in einem frühen Stadium testen zu können und um eine potenzielle Fehlerquelle auszuschließen, wird zu Beginn ohne das EKS gearbeitet. Die NDEF-Dateien werden somit direkt nach dem Schreiben auf die Antenne wieder eingelesen und zurück an den PC gesendet, ohne, dass eine drahtlose Datenübertragung stattfindet. Das Python-Skript besteht in dieser Phase also aus einer schreibenden und einer lesenden Funktion, die in einer Endlosschleife mittels Token – ähnlich einem Mutex nebenläufiger Prozesse – abwechselnd aufgerufen werden, sobald eine Funktion den Token frei gibt.

Ausblick

Im weiteren Verlauf dieser Arbeit wird die Firmware im EKS soweit angepasst, dass die vollständige Kommunikation realitätsnah im Versuchsaufbau validiert werden kann. Das Prinzip der kabellosen Datenübertragung ließe sich auch auf das bereits etablierte EKS anwenden, da diese sich, bezogen auf die hier beschriebene Vorgehensweise, nicht unterscheiden.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] EUCHNER GmbH und Co KG. Electronic-Key-System EKS. *Dok. Nr. 086808, Version 15-07/19*, 2019.

Entwurf und Implementierung eines Damespiels unter Verwendung von Flutter mit der Auswahl eines geeigneten Backends und Kommunikationsprotokolls

Andreas Handel

Mirko Sonntag

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma pep.digital, Stuttgart

Einleitung

Gesellschaftsspiele erfreuen sich einer immer größer werdenden Beliebtheit. Ungeachtet des Booms der Videospielbranche werden auch Umsatzsteigerungen im Bereich Gesellschaftsspiele in Deutschland verzeichnet. Eine noch größere Aufmerksamkeit wird den Gesellschaftsspielen während der Corona-Pandemie geschenkt. Da jegliche Veranstaltungen eingeschränkt oder abgesagt werden, finden immer mehr Menschen Spaß an Gesellschaftsspielen. Das Spielen wird in der Pandemie auch als Therapiemittel gesehen [1]. Neben Videospielen etablieren sich immer mehr Gesellschaftsspiele im virtuellen Kontext. Das Gesellschaftsspiel Dame ist sehr bekannt und eignet sich gut für eine Untersuchung von Technologien.

Die Bachelorarbeit befasst sich mit dem Framework Flutter im Bereich der mobilen App Entwicklung. Ziel der Arbeit ist die Implementierung des Gesellschaftsspiels Dame. Eine Partie Dame soll dabei auf zwei unterschiedlichen Smartphones gespielt werden, jeweils eines pro Spieler. Die App für die Darstellung des Damespiels wird als Frontend im Framework Flutter implementiert. Für die Realisierung einer Vernetzung von zwei Smartphones gibt es zwei Möglichkeiten. Eine davon ist die direkte Verbindung mittels Peer-to-Peer zum Beispiel über Bluetooth. Die zweite Möglichkeit benötigt eine zentrale Instanz, einem sogenannten Backend oder Server. Der Server dient der Erstellung einer Spielpartie und zur Überwachung der Spielzüge. Vorteilhaft bei einer serverbasierten Architektur ist, dass die Spieler geografisch unabhängig sind. In der Bachelorarbeit wird deshalb ein Server genutzt. Der Datenaustausch zwischen den Apps und dem Server erfordert ein Kommunikationsprotokoll, das von beiden Seiten implementiert wird. Die Wahl des Backends und Kommunikationsprotokolls wird im Laufe der Bachelorarbeit durch Vergleiche entschieden.

Dame

Das Brettspiel Dame ist ein bekanntes Strategiespiel für zwei Spieler. Dame wird in Großbritannien Draughts und in den USA Checkers genannt. Das Spielbrett besteht aus quadratisch angeordneten Feldern, welche im Wechsel die Farbe Schwarz und Weiß haben. Die schwarzen oder weißen Spielsteine haben die Form einer Scheibe. Jeder Spieler hat eine Farbe und platziert die Spielsteine auf die eigenen Reihen aller schwarzen Felder. Die Spielsteine dürfen nur auf den schwarzen Feldern bewegt werden. Abwechselnd darf jeder Spieler einen Stein diagonal auf ein freies Feld bewegen. Gegnerische Steine werden übersprungen und dann vom Spielfeld entfernt. Wenn ein Stein die gegnerische Seite erreicht, wird dieser zu einer Dame, indem ein weiterer Stein daraufgelegt wird. Die Dame hat mehr Bewegungsfreiheiten und darf sowohl vorwärts als auch rückwärts ziehen. Ziel des Spiels ist es, alle gegnerischen Steine zu entfernen. Die Implementierung des Damespiels ist eines der ältesten Computerspiele überhaupt und hat den Grundstein für künstliche Intelligenz gelegt [3]. Es gibt eine Vielzahl von Varianten und kann auf unterschiedlichen Spielbrettern gespielt werden. Zu den regionalen Varianten gehören zum Beispiel die internationale, kanadische, deutsche, italienische, brasilianische, russische, spanische, türkische und friesische Dame.

Wahl des Kommunikationsprotokolls

Für die Kommunikation zwischen Server und Client muss ein Kommunikationsprotokoll verwendet werden. Eine Auswahl wird zwischen den Protokollen GraphQL, REST, gRPC und SOAP getroffen, da sie weit verbreitet und etabliert sind. Zu den unterschiedlichen Arten von Kommunikationsprotokollen gehören RPC, message und streaming basierte Protokolle. SOAP, REST und GraphQL sind message und RPC basiert.

gRPC implementiert RPC und unterstützt messages sowie streaming.

SOAP ist ein XML basiertes Datenformat und wurde im Jahre 2000 entwickelt. Das Nachrichtenformat und Processing Model löste mit anderen Technologien wie UDDI und WDSL einen Webservice Hype aus. Meistens wird SOAP mit HTTP übertragen, kann jedoch auch mit anderen Protokollen verwendet werden.

SOAP hat sich aufgrund der Standardisierung sehr stark verbreitet. Nachteilig war jedoch die Schwergewichtigkeit der Nachrichten, weshalb eine andere Technologie entwickelt wurde. REST ist eine einfachere und leichtgewichtigere Alternative zu SOAP. Jede Anfrage enthält ein HTTP Verb und eine URL. Die wichtigsten Verben sind GET, POST, PUT und DELETE mit denen die CRUD Operationen durchgeführt werden.

GraphQL ist eine Abfragesprache für APIs, die Anfragen von vordefinierten Typen verarbeitet und je nach Bedarf dem Client zur Verfügung stellt. Das Prinzip von GraphQL ist dem von REST ähnlich und kann als Alternative angesehen werden.

gRPC ist ein Cross-Plattform RPC Framework, dass ursprünglich unter dem Namen Stubby bei Google entwickelt wurde. Das darunterliegende Kommunikationsprotokoll ist standardmäßig Protocol Buffers, kann jedoch auch mit anderen Datenformaten wie zum Beispiel json verwendet werden. Protocol Buffers sind eine effiziente Form der Serialisierung strukturierter Daten. Die Schnittstellenbeschreibung von Client und Server ist in Services aufgeteilt, die ihrerseits durch RPC Funktionen definiert sind.

Für den Vergleich der oben genannten Kommunikationsprotokollen werden verschiedene Vergleichsaspekte herangezogen. Wichtige Kategorien sind die Aktualität, Performance, Popularität, Verbreitung und Komplexität der Protokolle. Als Sieger geht gRPC hervor, da es effizient ist und bidirektionales Streaming ermöglicht. Damit kann anders als bei Stateless Anwendungen der Server eine Nachricht versenden. Das ist hilfreich, wenn der nächste Spieler zum Zug aufgefordert wird.

Wahl des Backends

Bei der Wahl des Backends werden die Technologien Java, Go, Dart, C#, Python und Node.js auf Tauglichkeit untersucht, da diese von gRPC unterstützt werden. Es ist vor allem eine schnelle Entwicklungszeit und Nebenläufigkeit wichtig. Nebenläufigkeit wird für die Verarbeitung der Client Requests, Aktualisierung des Spielstands und Server Antworten benötigt. Diese Anforderungen erfüllt die Programmiersprache Go aufgrund der Einfachheit und der eingebauten Unterstützung von Parallelität. Go ist eine kompilierte, nebenläufige, und statisch typisierte Programmierspra-

che, die auch Garbage-Collection durchgeführt. Sie wurde ursprünglich von Google entwickelt.

Flutter

Flutter ist ein von Google entwickeltes UI Framework zur Erstellung von mobilen, desktop- und webbasierten Anwendungen. Mit nur einer Codebasis werden viele verschiedene Plattformen unterstützt. Das Damebrett wird aus einzelnen Widgets zusammengesetzt und reagiert auf Toucheingaben des Benutzers. Grafische Richtlinien, Werkzeuge und Komponenten bietet die Designsprache Material. Auch über Fehleingaben kann der Benutzer angemessen informiert werden.

Liste valider Spielzüge

Für die Implementierung der Spiellogik wird eine Liste valider Spielzüge benötigt. Da mehrfache Sprünge in einem Zug möglich sind, reicht eine einmalige Überprüfung der linken und rechten Diagonalen nicht aus. Eine Möglichkeit besteht darin, die Verknüpfungen in einem Baum darzustellen 1. Ausgehend von jedem einzelnen Spielstein werden alle Eltern-Kind-Beziehungen in eine Map geschrieben. Der Elternteil repräsentiert die Zielposition und der Kindteil die Startposition einer einzelnen Bewegung. Wenn die Map eines einzelnen Spielsteins vollständig ist, wird diese in einzelne Listen transformiert. Die Listen werden aus den Wurzel- nach Blattknoten Pfaden gebildet. Jede einzelne Liste steht für einen möglichen Spielzug. Die Liste wird an den Client gesendet, der dann eine Auswahl trifft und an den Server zurücksendet. Nachfolgend ist ein Beispiel für einen doppelten Sprung gezeigt.

$$\text{map} \left[\begin{array}{c} \text{Position} \\ \underbrace{\{x \ y\}}_{\text{Ziel (Schlüssel)}} : \underbrace{\{x \ y\}}_{\text{Start (Wert)}} \cdots \end{array} \right]$$

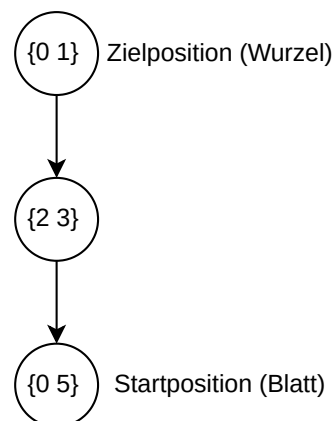


Abb. 1: hierarchische Baumstruktur [2]

$$\text{map}[\{2 \ 3\} : \{0 \ 5\} \ \{0 \ 1\} : \{2 \ 3\}]$$

$$\underbrace{\{0 \ 5\}}_{\text{Start}} \ \{2 \ 3\} \ \underbrace{\{0 \ 1\}}_{\text{Ziel}}$$

Ausblick

In der Bachelorarbeit werden die elementaren Funktionalitäten eines Damespiels implementiert. Weitere mögliche Aspekte sind ein Matchmaking System, Registration und Highscore-Tabelle. Der Einsatz von Containern und Containerorchestrierung mit Kubernetes erleichtern die fortlaufende Entwicklung und den Betrieb der Software.

Literatur und Abbildungen

- [1] Till Bücken. Gesellschaftsspiele statt Partys. <https://www.tagesschau.de/wirtschaft/spieleboom-corona-101.html>, 2021.
- [2] Eigene Darstellung.
- [3] John McCarthy. Arthur Samuel: Pioneer in Machine Learning. <http://infolab.stanford.edu/pub/voy/museum/samuel.html>, 2021.

Konzept und prototypische Umsetzung der Portierung einer Ladeplattform zu einer Public-Cloud-Kubernetes-Umgebung unter dem Aspekt der Störungserkennung

Matthias Hauber

Mirko Sonntag

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Die wachsende Anzahl verteilter und vernetzter IT-Systeme verzeichnen eine steigende Komplexität aufgrund der dynamischen IT-Komponenten als auch der Microservice-Landschaft. Aufwände zur Entdeckung eines fehlerhaften Zustand und der Ursache steigen daher exorbitant an [4]. Auch ein fehlendes Lifecycle-Management der mittels Container umgesetzten Microservices kann zu unerwünschten Ausfällen und nur aufwendig manuell zu beheben den Störungen führen. Dem kann durch Einsatz von Container-Orchestrierung entgegengewirkt werden, welche Deployment, Management, Skalierung und Vernetzung von Containern automatisieren soll. Kubernetes ist hierbei die Open Source Standardlösung für Container-Orchestrierung. Diese wird von allen Public Cloud Plattformen angeboten und ist auch On-Premises einsetzbar. Die mit Kubernetes einhergehende rudimentäre Fehlererkennung und -behebung auf Ebene der Container gilt es mit weiteren etablierten Methodiken der Störungserkennung zu ergänzen. Diese Methodiken, die unter anderem Logging, Tracing, Monitoring und Alerting umfassen, werden teilweise von Cloud-Anbietern angeboten und automatisch mit einem „managed“ Kubernetes Service offeriert. Hierbei gilt es zu untersuchen, inwieweit eine Störungserkennung bereits mit diesen Angeboten umgesetzt werden kann und wie diese in der Funktionalität mit direkt auf dem Kubernetes Cluster installierten Frameworks vergleichbar sind. Somit steht im Mittelpunkt der Masterthesis die Forschungsfrage „Wie können die Aspekte der Störungserkennung in einer durch Microservices realisierten verteilten IT Landschaft im Umfeld von Kubernetes umgesetzt werden?“

Zielsetzung

Die Arbeit wird im Rahmen eines Forschungsprojekts in Kooperation mit dem Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO verfasst. Das Projekt ist im Bereich der Elektromobilität angesiedelt. Hier gilt es den Sprung von Microservices bestehend aus einzelnen Docker Containern hin zu einer mit Kubernetes umgesetzten IT-Landschaft inklusive umfassender Störungserkennung zu schaffen. Es wird das Ziel verfolgt, sowohl eine Einführung in die Thematik zu geben als auch die Problematiken der Portierung zu der Lifecycle-Management-Plattform Kubernetes zu analysieren. Diese Problematiken wurden konkret dem Projektkontext entnommen und anhand dessen sollen geeignete Lösungswege aufgezeigt werden. Dabei soll ein Konzept für die Implementierung von Störungserkennung in einer Kubernetes-Umgebung entstehen, das sowohl im vorliegenden Anwendungsfall als auch als Referenz für ähnlich gestellte Problemstellungen dienen kann. Die Masterarbeit betrachtet die gesamte Thematik mit einem Hauptmerk auf der Störungserkennung. Dies beinhaltet, wie diese bei den zu evaluierenden Public-Cloud-Anbietern und auch allgemein im Kubernetes Umfeld zum Einsatz kommen kann. Das entstandene Konzept soll einer Validierung in Bezug auf die Forschungsfrage unterzogen werden.

Störungserkennung

Die Störungserkennung befasst sich sowohl mit der Erkennung als auch mit der darauffolgenden Behebung einer Störung. Darüber hinaus wird auch das Erkennen von Ursachen der Störungen angestrebt. Unter Störungserkennung können „Werkzeuge und Maßnahmen bezeichnet [werden], die dazu dienen, Fehlerursachen, Fehler und Störungen zu erkennen, die im Betrieb von Softwaresystemen auftreten“ [3].

Grundlegend können vier Methoden dazu beitragen Störungen zu erkennen. Diese sind Monitoring, Logging, Tracing und die Verwendung eines Service Meshs. Das Thema Monitoring steht stark im Fokus, um ein System oder einen Vorgang zu überwachen. **Monitoring** wird häufig für die Überwachung der Verfügbarkeit, der Performance und der Sicherheit eingesetzt. Dabei werden festgelegte Messgrößen (Metriken) kontinuierlich festgestellt, zumeist mit Referenzwerten verglichen und anschaulich über Dashboards visualisiert. Auch die Methodik des Alertings wird bei der Störungserkennung gerne im Zusammenspiel mit Monitoring eingesetzt. Diese bezeichnet das Warnen vor einer möglicherweise gefährlichen Situation. So kann zum Beispiel bei Erreichen eines bestimmten Zustands oder einer Metrik eines IT-Systems eine E-Mail an den Betrieb versandt werden.

Jede Komponente in einem verteilten System sollte die wichtigsten Ereignisse lokal auf dem der Komponente zur Verfügung stehenden Speicherplatz in Form sogenannter fortlaufender Log-Dateien protokollieren. Diese Logs können dann in einem verteilten System zur Störungserkennung genutzt werden, indem sie zentral durch eine separate Komponente in Form eines Tools wie Elasticsearch aggregiert und gespeichert werden. Hier wird von **Logging** in der Störungserkennung gesprochen. Die gespeicherten Daten können visualisiert, durchsucht und anhand von Mustern analysiert werden. Über das Logging hinaus gibt es das **Tracing**. Hier werden Systemereignisse und Nachrichten zumeist in Form von HTTP-Anfragen auf ihrem Weg durch das verteilte System aufgezeichnet. Damit kann nachvollzogen werden, welche Komponenten oder Microservices an der Beantwortung einer Anfrage beteiligt waren. Metriken zur Anfrage, wie die Latenzzeit, können berechnet werden und es kann erforscht werden, an welcher Stelle eventuell ein Fehler aufgetreten ist.

Des Weiteren gibt es das **Service Mesh** als Methode für die Störungserkennung. Diese ist eine Infrastrukturschicht, die grundsätzlich den Netzwerkverkehr einer Microservices Landschaft analysiert, kontrolliert und regelt [2]. Wichtigste Funktionalitäten sind das Traffic Management und die Transportverschlüsselung mittels mTLS. Ohne diese Verschlüsselung wäre der Netzwerkverkehr zwischen den auf verschiedenen Rechenzentren verteilten Nodes unverschlüsselt.

Konzeption

Um Störungserkennung für einen Kubernetes Cluster in der Cloud umzusetzen, können drei verschiedene Herangehensweisen betrachtet werden, aus denen jeweils Konzepte abgeleitet werden können. Das erste und flexibelste Konzept sieht vor, dass nur virtuelle Maschinen (VMs) zum Einsatz kommen und kein von Cloud-Anbietern verwalteter Kubernetes Service

verwendet wird. Die Frameworks und Tools zur Störungserkennung werden dabei auf den gleichen Kubernetes Nodes betrieben, auf denen auch die zu überwachende Applikation ausgeführt wird. Vorteil hierbei ist, dass flexibel zwischen Cloud-Anbietern gewechselt werden kann und auch ein On-Premises Betrieb möglich ist. Nachteil ist der erhöhte Verwaltungsaufwand für den Cluster.

Das zweite und am schnellsten zu implementierende Konzept sieht vor, dass ein vom Cloud-Anbieter verwalteter Kubernetes Service zum Tragen kommt, der automatisch eine Integration in die Services des Anbieters zur Störungserkennung ermöglicht. Vorteil ist der geringe Aufwand für das Aufsetzen und die Verwaltung des Clusters, Nachteil ist die volle Abhängigkeit vom Cloud-Anbieter.

Das dritte und favorisierte hybride Konzept, wie in Abb. 1 dargestellt, kombiniert beide Ansätze und bietet so alle Vorteile, die die Nutzung eines verwalteten Kubernetes Service mit sich bringt, ohne dabei die Nachteile überwiegen zu lassen. Alle Methodiken der Störungserkennung werden im Kubernetes Cluster mit der Anwendung zusammen ausgeführt und sind somit portabel im Falle eines Cloud-Anbieter Wechsels. Darüber hinaus werden zusätzlich die Cloud-Anbieter spezifischen Services zur Störungserkennung genutzt, die bereits automatisch bei der Erstellung eines Kubernetes Clusters zur Verfügung stehen. Vorteile und Nachteile halten sich bei diesem Konzept die Waage und es werden alle verfügbaren Möglichkeiten genutzt, ohne eine einseitige Abhängigkeit zu schaffen.

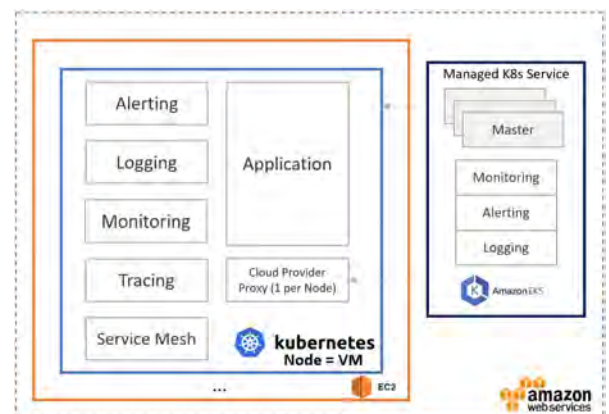


Abb. 1: Umsetzung der Störungserkennung innerhalb des Kubernetes Clusters und durch den Cloud-Anbieter (hier: AWS) [1]

Evaluation Cloud-Anbieter

Um die Vorteile der Cloud, wie Skalierung, Verfügbarkeit und Kosten nur für verbrauchte Ressourcen, vollumfänglich in Kombination mit Kubernetes aus-

nutzen, kann ein vom Cloud-Anbieter verwalteter Kubernetes Service zum Einsatz kommen. Hierbei kann zumeist binnen Minuten über die Weboberfläche des Anbieters ein Cluster erstellt werden, wobei viele Konfigurationsmöglichkeiten, wie Ressourcen der Nodes und Autoscaling, festgelegt werden können. Weitere Vorteile eines „managed“ Kubernetes Service sind die Verteilung der Nodes über verschiedene geographisch verteilte Rechenzentren hinweg, die Replizierung des Kubernetes Masters und die automatische Installation von Updates.

Anbieter von „managed“ Kubernetes Services in der Cloud häufen sich. Für die Evaluation gilt es eine Vorauswahl an Cloud-Anbietern anhand von Kriterien einzugrenzen. Diese Kriterien umfassen beispielsweise, dass der Anbieter eigene Rechenzentren in Deutschland betreibt, für Privatpersonen zugänglich ist (Public Cloud), Erfahrung vorzuweisen hat und Störungserkennung bereitstellt. Sechs Anbieter werden so in die Gruppen „Leaders“ (AWS, Azure, GCP), „Challengers“ (IBM, Oracle) und „Disruptor“ (DigitalOceans) eingeteilt und evaluierend miteinander verglichen (siehe Abb. 2).



Abb. 2: Zur Evaluation ausgewählte Cloud-Anbieter [1]

Der Kriterienkatalog zur Evaluation schließt folgende Kriterien mit ein:

- Initialisierung Cluster & Aufwand Portierung
- Skalierung (horizontal, vertikal, Nodes, Master)

- Kosten (Cluster Management Fee & Ressourcen)
- Benutzerfreundlichkeit & Support
- Sicherheit & Wiederherstellung
- Zuverlässigkeit & Verfügbarkeit
- Störungserkennung out-of-the-box
- Integration in Continuous Deployment Pipeline

Zusammenfassung und Ausblick

Störungserkennung ist in verteilten IT-Systemen zwingend nötig, um Ausfallsicherheit, Verfügbarkeit und Nachvollziehbarkeit über einen längeren Zeitraum hinweg sicherzustellen. Im Kontext des Forschungsprojekts wird ein verteiltes IT-System durch Microservices mittels Containern repräsentiert. Für das Lifecycle-Management dieser Container soll das Framework Kubernetes eingesetzt werden. Dieses wird als von Cloud-Anbietern verwalteter Service in der Cloud angeboten. Die Störungserkennung kann hier sowohl auf Kubernetes Ebene implementiert als auch vom Cloud-Anbieter bereitgestellt werden. Das favorisierte Konzept verortet die Störungserkennung sowohl innerhalb des Kubernetes Clusters als auch beim Cloud-Anbieter durch Nutzung des „managed“ Kubernetes Services. Hierdurch fallen die Nachteile nicht schwer ins Gewicht, währenddessen die Vorteile trotzdem gewahrt bleiben.

Im weiteren Verlauf der Arbeit werden die vorausgewählten Cloud-Anbieter verglichen und die Portierung im Rahmen des Forschungsprojekts prototypisch durchgeführt. Dies geschieht anhand des hergeleiteten Konzepts mit dem für den Anwendungsfall geeignetsten Anbieter. Störungserkennung wird in naher Zukunft einen immer höheren Stellenwert besitzen, da durch die voranschreitende Digitalisierung auch kritischer Bereiche, wie des Gesundheitssystems, ein ausfallsicherer und stabiler Betrieb der IT-Systeme gewährleistet sein muss.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Istio Istio. The Istio service mesh. <https://istio.io/latest/about/service-mesh/>, 2021.
- [3] Falko Kötter, Andreas Freymann, et al. *Überwachung und Störungserkennung in vernetzten IT-Systemen*. Fraunhofer Verlag, 2020.
- [4] Martin Lehmann and Renato Vinga-Martins. Microservices nur bei ausreichender Komplexität. <https://www.informatik-aktuell.de/entwicklung/methoden/microservices-nur-bei-ausreichender-komplexitaet.html>, 2018.

Anwendung von Computer Vision zur Bewertung von hochlegierten Materialzuständen

Frank Holzmueller

Steffen Schober

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Robert Bosch GmbH, Renningen

Einführung

Durch die Anwendung von Convolutional Neural Networks konnten in den letzten Jahren komplexe Probleme in der Medizin, dem autonomen Fahren und vielen weiteren Szenarien gelöst werden. Auch der Bereich der Materialanalytik ist auf dieses Thema aufmerksam geworden, weshalb vermehrt Untersuchungen zur Nutzung von CNNs für analytische Problemstellungen durchgeführt werden. Die hier vorgestellte Arbeit beschäftigt sich mit der Bewertung von hochlegierten Materialzuständen unter Verwendung einer U-Net Architektur. Dabei wird der gesamte Prozess von der Erstellung der Labels bis zur Auswertung der Ergebnisse abgedeckt.

Motivation und Zielsetzung

Bei der metallographischen Untersuchung von Werkstoffen fallen enorme Datenmengen in Form von Bildern an, welche digital gespeichert und im Nachgang manuell analysiert und ausgewertet werden. Diese Bilder enthalten einen sehr großen Informationsgehalt, der allerdings bei der manuellen Auswertung mit dem menschlichen Auge nicht voll ausgeschöpft werden kann. Aus diesem Grund führen die Analysen von unterschiedlichen Personen zu abweichenden Ergebnissen. Zudem ist der Prozess der Auswertung sehr zeitintensiv. Kommerzielle Software unterstützt den Anwender zwar vermehrt mit Ansätzen aus dem Machine Learning, diese Anwendungen sind allerdings durch den kommerziellen Aspekt der breiten Nutzbarkeit gebremst und entfalten somit nicht ihr volles Potential. Um die Bewertung von lichtmikroskopischen Aufnahmen konstanter und effizienter zu gestalten, wird im Rahmen dieser Arbeit eine spezifische Implementierung für lichtmikroskopische Aufnahmen eines speziellen Bauteils entwickelt. Dabei liegt der Fokus auf der Generierung von numerischen Kennzahlen für die Abschätzung der Qualität eines untersuchten Objektes.

Methodik

Das vorhandene Szenario eignet sich hervorragend für die Anwendung einer semantischen Segmentierung. Bei dieser Art der Segmentierung wird jeder Pixel eines Eingangsbildes einer der möglichen Ausgangsklassen zugeordnet. Dadurch entsteht eine Segmentierungsmaske, welche das analysierte Bild als pixelweise Detektion der Bildinformation darstellt. Aufgrund der Ergebnisse vorhandener Literatur [4], [2] zu diesem Thema wird als Architektur des CNNs ein U-Net verwendet [3]. Die Verwendung des U-Nets ist in Abb. 1 dargestellt.

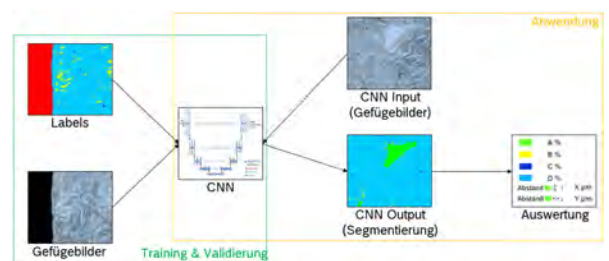


Abb. 1: Workflow für Training und Anwendung des U-Net [1]

Ein weiteres Verfahren, das in vorherigen Untersuchungen gute Ergebnisse liefern konnte, ist das Transfer Learning [2]. Dabei wird der Encoder, also der Teil des U-Nets der sich um die Faltung der Bilder zu Feature Vektoren kümmert, durch ein vortrainiertes Netz ersetzt. Die Architektur des verwendeten Encoders ist dabei austauschbar, wodurch mehrere Architekturen von Netzen (ResNet, VGG...) verwendet werden können.

Erstellen von Labels

Die größte Herausforderung im Rahmen der Arbeit besteht darin, Labels für das Training und die Validierung des U-Nets zu erzeugen. Es liegt zwar eine große Menge an rohen Bilddaten vor, doch ohne die

Information der darin enthaltenen Gefügebestandteile können diese Bilder für den Ansatz des Supervised Learnings nicht verwendet werden. Eine manuelle Annotation jedes einzelnen Pixels ist extrem aufwändig und zudem aufgrund der variierenden Interpretation des Gefüges durch verschiedene Personen nicht eindeutig umsetzbar. Deshalb wird für die Erstellung eines Labels zusätzlich zu den geätzten Bildern der Lichtmikroskopie ein weiteres Verfahren mit dem Namen Elektronenrückstreuung (EBSD) verwendet. Die Ergebnisse zweier unterschiedlicher Lichtmikroskopie-Aufnahmen werden mit dem Resultat aus EBSD überlagert, um ein Label aus der Kombination der Messergebnisse zu generieren (vgl. Abb. 2). Die korrespondierenden Ausschnitte der jeweiligen Bilder sind grün umrahmt. Für die Überlagerung ist eine Verzerrung und Skalierung der Bilddaten nötig, was halbautomatisiert über ein Python Programm umgesetzt wird.

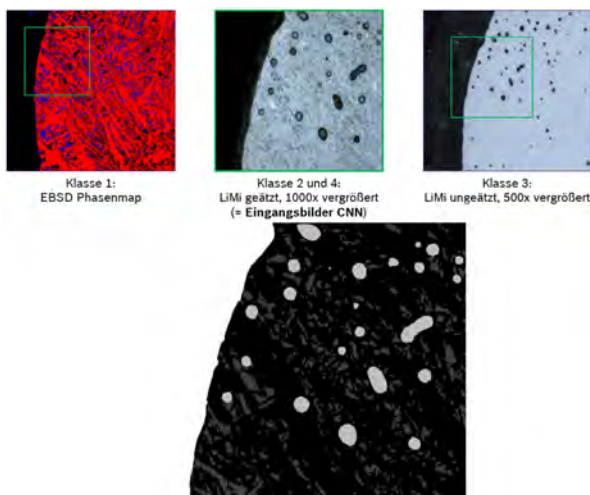


Abb. 2: Datengrundlage für das Erstellen von Labels (EBSD, LiMi geätzt, LiMi ungeätzt) und fertiges Label [1]

Die in Abb. 2 gezeigten Lichtmikroskopie-Bilder werden durch die zuvor angesprochene kommerzielle Software vorsegmentiert (vgl. Abb. 3), was den Prozess der Erstellung von Labels beschleunigt. Die halbautomatisch vorsegmentierten Bilder bieten Anhaltspunkte für die Labels, wodurch nur noch eine manuelle Bereinigung der Detektionen durchgeführt werden muss. Alle diese separat generierten Labels für die verschiedenen Klassen werden final zu einem One-Hot Label zusammengeführt. Dabei überschreiben die Klassen sich gegenseitig nach Priorität, wodurch doppelte Labels für einen einzelnen Pixel ausgeschlossen sind. Ein fertiges Label ist ebenfalls in Abb. 2 dargestellt.

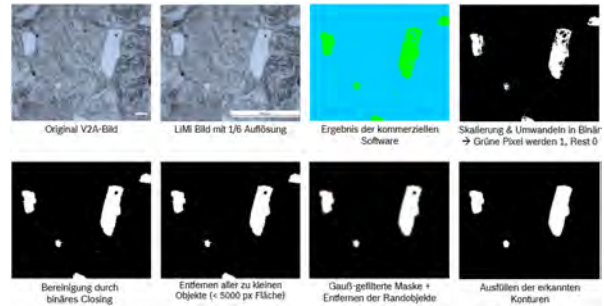


Abb. 3: Prozess der KI-unterstützten Generierung von Labels [1]

Trainieren des U-Nets

Für das Training des U-Nets werden die großen Originalbilder (2464x2056 Pixel) in kleinere Bilder mit 256x256 bzw. 480x480 Pixeln unterteilt. Diese Entscheidung ist der Limitierung der vorhandenen Hardware geschuldet. Für das Training selbst wird ein U-Net mit ResNet18 bzw. VGG-16 als Encoder verwendet. Die Gewichte der vortrainierten Encoder basieren auf dem ImageNet Datensatz. Als Optimizer kommt Adam mit einem exponentiellen Abfall der Learning Rate zum Einsatz, als Loss Funktion wird die Summe aus dem Dice Loss und der Binary Crossentropy angewendet. Um einen Vergleich der resultierenden Netze durchführen zu können, wird die Performance anhand des Jaccard-Index (= Intersection over Union) gemessen.

Auswertung der Segmentierung

Die resultierende Segmentierung wird im Rahmen einer automatisierten Python Toolchain ausgewertet. Dabei können sehr große, aus mehreren Einzelbildern bestehende Lichtmikroskopie-Aufnahmen als Eingangsdaten verwendet werden. Die Auswertung nutzt ein Sliding-Window, welches diese Bildgrößen von bis zu 40 GB in viele kleine, für das Netz behandelbare Teilbilder trennt. Die unterteilten Bilder werden im Anschluss wieder zusammengeführt und in Form einer Segmentierung dargestellt (vgl. Abb. 4).

Abb. 4 zeigt eine resultierende Segmentierung im direkten Vergleich mit dem Eingangsbild für das Netz. Alle numerischen Ergebnisse (Verteilung der Klassen, horizontaler und vertikaler Abstand der speziellen Klasse) werden außerdem am Ende der Auswertung angezeigt und gespeichert, damit diese Ergebnisse mit anderen Proben verglichen werden können.

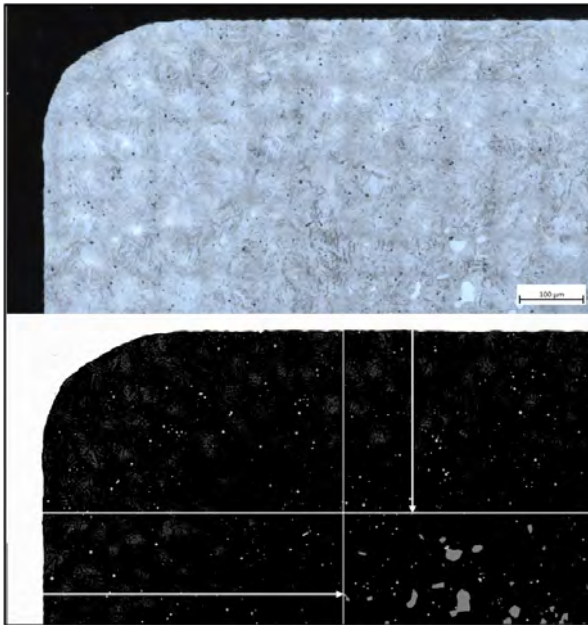


Abb. 4: Eingangsbild und Ergebnis der Auswertung [1]

Fazit und Ausblick

Das U-Net liefert sehr gute Ergebnisse für die dargestellte Aufgabe. Speziell mit der Unterstützung durch Transfer Learning können schon mit wenigen (<100) Labels ordentliche Ergebnisse generiert werden. Trotzdem sind die segmentierten Resultate nicht optimal, weshalb eine größere Menge an Labels noch bessere Ergebnisse hervorbringen könnte. Speziell für die Robustheit des Netzes gegen unscharfe Bilder oder sonstige Bildfehler werden noch deutlich mehr und deutlich diversere Labels benötigt.

Das erreichte Ergebnis kann bereits in dem aktuellen Zustand zur Unterstützung bei metallographischen Analysen von Gefügen eingesetzt werden. Mit etwas mehr Trainingsdaten und weiteren Verbesserungen der Hyperparameter könnten unter Umständen noch bessere Ergebnisse erreicht werden, wodurch diese Methodik die manuelle Analyse ersetzen könnte. Auch eine Anwendung in der Produktion ist möglich, da ein System dieser Art zur automatisierten Qualitätskontrolle eingesetzt werden kann.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Elizabeth A. Holm, Ryan Cohn, Nan Gao, et al. Overview: Computer Vision and Machine Learning for Microstructural Characterization and Analysis. *Metall Mater Trans A*, 51:5985–5999, 2020.
- [3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, pages 234–241, 2015.
- [4] Stylianos Tsopanidis, Raúl Herrero Moreno, and Shmuel Osovski. Toward quantitative fractography using convolutional neural networks. *Engineering Fracture Mechanics*, 231, 2020.

Analyse der Eignung der Programmiersprache Rust zur Entwicklung eines Kommunikationsstacks für eingebettete Systeme

Alexander Huebener

Markus Enzweiler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma IT-Designers GmbH, Esslingen

Einleitung

Pkws, Lkws oder fliegende Verkehrsmittel, seien diese bemannt oder unbemannt, benötigen Sensoren und Aktoren zur Steuerung, Navigation und Stabilisierung. Aufgrund steigender Anforderung an Verarbeitungsleistung z.B. durch autonomes Fahren, werden kleine abgeschlossene Systeme eingesetzt, welche untereinander mit Hilfe von Netzwerktechnik Daten austauschen. Ein System kann hierbei aus einem Mikrocontroller und angeschlossener Peripherie bestehen. Die eingesetzten Mikrocontroller haben meist nur eingeschränkte Ressourcen bezüglich Speicher und CPU-Leistung. Demnach ist es wichtig, dass die verwendete Software für diese Einschränkungen optimiert wird und somit effizienter ausgeführt werden kann.

Die verwendete Programmiersprache soll den Programmierer in diesem Umfeld vor Ausführungsfehlern bewahren, aber auch gleichzeitig eine speicher- und ausführungseffiziente Firmware gewährleisten. In dieser Arbeit wird hierfür die Programmiersprache Rust für eine Kommunikationsstack-Implementation auf einem eingebetteten System angewandt und evaluiert.

UAVCAN

Um eine einheitliche Kommunikation unter den Mikrocontrollern gewährleisten zu können, wird der öffentlich zugängliche UAVCAN Standard entwickelt. Die Spezifikation kann laut [6] in 1000 logischen Codezeilen umgesetzt oder mit Hilfe einer bereits entwickelten Bibliothek verwendet werden.

UAVCAN soll ein offenes und leichtgewichtiges Protokoll sein, welches zur zuverlässigen und deterministischen Kommunikation innerhalb Fahrzeuge wie Autos, Anwendungen der Luft- und Raumfahrt und Roboterapplikationen entwickelt wird [7].

Zur Zeit der Arbeit stehen Implementationen in C/C++, Python und zu Teilen in Rust zur Verfügung. Für diese Arbeit sind die in C/C++ und Rust

geschriebenen Bibliotheken von Bedeutung, da sich diese für den Einsatz auf eingebettete Systeme eignen.

Rust

Die Programmiersprache Rust wurde ab 2006 von Graydon Hoare als ein Hobbyprojekt entwickelt [2]. In [4] nannte er als Grund, wieso er mit dem Rust Projekt begonnen hat, dass C++ "[...] ziemlich fehlerträchtig" sei und verwies auf die Speichersicherheit, welche nicht gewährleistet ist.

Rust wird öffentlich durch die Community über ein Request For Comments (RFC) System weiterentwickelt. Durch den *rustc* Compiler wird der Quellcode direkt in Maschinen-Instruktionen kompiliert. Dies, und die Vermeidung eines Garbage Collectors, macht Rust zu einer schnellen und vergleichbaren Sprache zu C/C++ und lässt sie bei Projekten für eingebettete Systeme mit begrenzten Ressourcen einsetzen.

In Rust geschriebene Programme werden häufig modular aus Bibliotheken zusammengesetzt. Daher ist viel Funktionalität, welche die Sprache speziell für Betriebssysteme liefert, in der so genannten *std* Bibliothek vorhanden. Für Arbeiten auf einem eingebetteten System kann auf die *core* Bibliothek zurückgegriffen werden, welche alle Funktionalität enthält, die nicht das Betriebssystem involviert.

Rust bietet durch das Angeben eines Targets, eine schnelle Möglichkeit eine ausführbare Datei für verschiedene Architekturen zu kompilieren. Demnach kann in wenigen Schritten mit der Entwicklung auf anderen Plattformen begonnen werden.

In Rust werden Fehlerquellen, welche aus anderen Sprachen bekannt sind, durch Funktionen und Einschränkungen vermieden.

- Darunter fällt die Eigentümerschaft, welches zu jeder Zeit nur einen Besitzer für einen Speicherbereich oder eine Variable zulässt. Dieses soll Speichersicherheit garantieren [3].

- Variablen, welche verändert werden sollen, müssen speziell als veränderbar gekennzeichnet werden. Dies soll die Leserlichkeit des Programmes erhöhen [3] und verhindern, dass Variablen unbeabsichtigt verändert werden.
- Außerdem wird durch den Borrow Checker Data-Racing vermieden [3]. Damit wird überprüft, dass entweder an mehreren Stellen eine Variable gelesen werden kann oder nur an einer Stelle die Variable verändert wird.
- Hat der Compiler zu wenig Informationen über den Programmcode, sei es ein Fremdfunktionsaufruf zu C oder ein direkten Speicherzugriff, reagiert der Compiler pessimistisch und meldet einen Kompilierungsfehler [3]. Daher gibt es in Rust die Möglichkeit solche Bereiche mit *unsicher* zu markieren, um somit die Verantwortung der Ausführungssicherheit dem Programmierer zu übergeben.

Diese Funktionen werden zur Kompilierzeit überprüft und durch Kompilierfehler dem Programmierer angezeigt. Demnach wird kein zusätzlicher Mehraufwand während der Ausführung benötigt.

Schritte während der Arbeit

Zu Beginn der Arbeit ist die offizielle Rust-Umsetzung der UAVCAN Spezifikation, namens *uavcan.rs* welche auf GitHub entwickelt wird, nur für Linux Betriebssysteme einsetzbar. Daher sollen Anpassungen an der Bibliothek umgesetzt werden, welche es erlauben, diese auch in einem eingebetteten System auf einem Mikrocontroller verwenden zu können. Die Änderungen werden mit Merge-Requests in dem Repository angewandt.

Damit Anpassungen für eingebettete Systeme evaluiert werden können, muss ein Test-Programm erstellt werden, welches auf einem Mikrocontroller ausführbar ist. Dieses enthält einen HAL, welcher es ermöglicht die Peripherie des Mikrocontrollers zu konfigurieren und zu steuern.

Die Hardware besteht dabei aus zwei STM Nucleo-G431RB Boards welche über CAN-Tranceiver an einen CAN-Bus angeschlossen sind. Die Boards bauen auf einem Arm-Cortex-M4 auf und haben einen CPU-Takt von 170MHz. Außerdem ist ein Debugger auf den Boards inkludiert.

Das eine Board soll über die Rust UAVCAN Bibliothek Daten veröffentlichen und das andere diese empfangen. Um die UAVCAN Rust-Implementation auf dem Mikrocontroller zu verwenden, müssen Funktionalitäten, welche auf die Standard-Bibliothek von Rust zurückgreifen, umgeschrieben werden, da diese nicht für

eingebettete Systeme zur Verfügung steht. Eine solche Implementation wird in Rust mit *no_std* versehen.

Die Änderungen beziehen sich hauptsächlich auf Zeitstempel. Diese basieren zu Beginn auf der Standard-Bibliothek und werden auf einem Mikrocontroller durch einen Zähler umgesetzt, welcher beim Zurücksetzen des Controllers gestartet wird.

Des Weiteren muss die dynamische Speicherverwaltung angepasst werden. Es soll weiterhin dynamischer Speicher verwendet werden. Aus diesem Grund wird die Bibliothek *alloc* verwendet, welche die Collections der Standard-Bibliothek enthält. Durch einen global definierten Allokator, welcher den *tlsf* Algorithmus zur Speicherverwaltung eines statischen RAM-Segments verwendet, können die Strukturen dynamisch Speicher anfordern und wieder freigeben.

Außerdem werden Optimierungen während durchgeführten Zeitmessungen an der UAVCAN Rust-Implementation vorgenommen. Diese beinhalten die Verwendung des dynamischen Speichers zu optimieren, da dies sehr zeitintensiv ist. Demnach dürfen diese Allokationen nur sehr selten vorkommen.

Zur weiteren Zeitoptimierung gehört, dass Zugriffe in *Vec* (eine dynamische erweiterbare Liste in Rust) und *BinTree* (Baumstrukturen in Rust) möglichst wenig auftreten. Dies kann durch bessere Programmlogik und der Verwendung besserer Zugriffsfunktionen in Rust erzielt werden.

Für die Messungen wurde eine Bench-Suite erstellt. Diese soll die Zeitmessungen für den weiteren Verlauf der UAVCAN Rust-Implementations-Entwicklung automatisch durchführen. Dadurch kann jede neue Version der Bibliothek evaluiert werden.

Messungen

Die Rust-Implementation soll durch Aspekte der Softwarequalität beurteilt werden. Softwarequalität wird durch mehrere Merkmale, welche die Software erfüllen muss, definiert. Im ISO-25010 Softwareproduktqualitätsbaum, wird die Leistungsfähigkeit als eins dieser Merkmale definiert und als Ausführungsmerkmale kategorisiert [5].

Zur Evaluation der Rust Implementation sollen Zeitmessungen durchgeführt werden. Als Referenz dient eine Implementation der UAVCAN Spezifikation auf der Arduino Plattform. Diese ist in C umgesetzt mit einer API in C++. Im Folgenden wird diese Bibliothek als Arduino oder C++ Implementation bezeichnet.

Die Wahl ist auf diese Referenz gefallen, da die Arduino Umsetzung für eingebettete Systeme verwendbar ist. Außerdem ist die zu nutzende API in Punkten wie Typsicherheit der Rust-Implementation ähnlich.

Bei den Messungen werden die Aufrufzeiten der Bibliotheksfunktionen, jeweils für die Rust und die Arduino-Implementation, gemessen. Hierbei wird nicht

die Buskommunikation mit einbezogen. Die Messreihen werden in Senden und Empfangen von Übertragungen unterteilt.

Die Messungen werden in einem Testprojekt durchgeführt, welches zwischen einzelnen Messreihen neu kompiliert und auf das Board geladen wird. Das Testprojekt besteht dabei aus notwendigen Konfigurationsaufrufen für das Board und der zu messenden Funktionalität der Bibliothek.

Das Programm wird in Rust mit dem Marker `opt-level=3` und in C++ mit `-O3` kompiliert. In Rust wird außerdem `lto` verwendet, welches für eine bessere Optimierung sorgen soll.

Danach werden Messungen mit dem CAN-Bus durchgeführt. Dabei agiert ein Board als Mess-Knoten und das andere als Echo-Knoten. Hierbei wird die gesamte Umlaufzeit gemessen, sowie die Zeit, welche der Echo-Knoten benötigt, die gesendete Nachricht zu verarbeiten und wieder zurückzusenden.

Zum Schluss wird der Speicherverbrauch gemessen. Hierbei wird der statische Speicherbedarf und die Auslastung des Stacks und des Heaps betrachtet.

Messergebnisse

In 1 sind die Zeitmessungen der Arduino- und Rust-Implementation der Bibliotheksfunktionen im Vergleich zu sehen.

Es ist zu sehen, dass die Rust-Implantation (grüne Kurven) weniger Zeit für dieselbe Nachrichtengröße

benötigt. Deutlich ist dies beim Senden größerer Nachrichten erkennbar.

Für Übertragungen, welche in einen CAN-Datenrahmen passen, liegt der Unterschied beim Senden bei 7,6 und beim Empfangen bei 5,4 Mikrosekunden. Bei einer Übertragung, für welche mehrere Rahmen zum Übertragen benötigt werden, wird pro Übertragung bei der Arduino-Implementation initial 2,7 beim Senden und 3,3 Mikrosekunden beim Empfangen benötigt. Diese Zeit wird grundsätzlich, unabhängig der Datenrahmenanzahl, benötigt. Bei der Rust-Implementation ist diese Zeit vernachlässigbar. Pro Rahmen, welcher gesendet wird, unterscheiden sich die Implementationen beim Senden um 4,8 und beim Empfangen um 1,4 Mikrosekunden.

Aufgrund der Tatsache, dass die Arduino-Implementation initial und pro Datenrahmen mehr Zeit benötigt, ergeben die Messungen eine signifikante Differenz der Bibliotheken bei Übertragungen mit 8 Datenrahmen.

Beim Senden benötigt die Rust-Implementation 18 Mikrosekunden, wobei die Arduino Implementation um die 60 Mikrosekunden braucht.

Beim Empfangen benötigt die Rust-Implementation 35 Mikrosekunden. Dies sind 15 Mikrosekunden weniger als bei der C++-Implementation.

Als Referenz für die genannten Werte: Die CPU benötigt bei einer Taktfrequenz von 170MHz für einen CPU-Zyklus 0,006 Mikrosekunden. Demnach sind dies um die 166 Zyklen in einer Mikrosekunde.

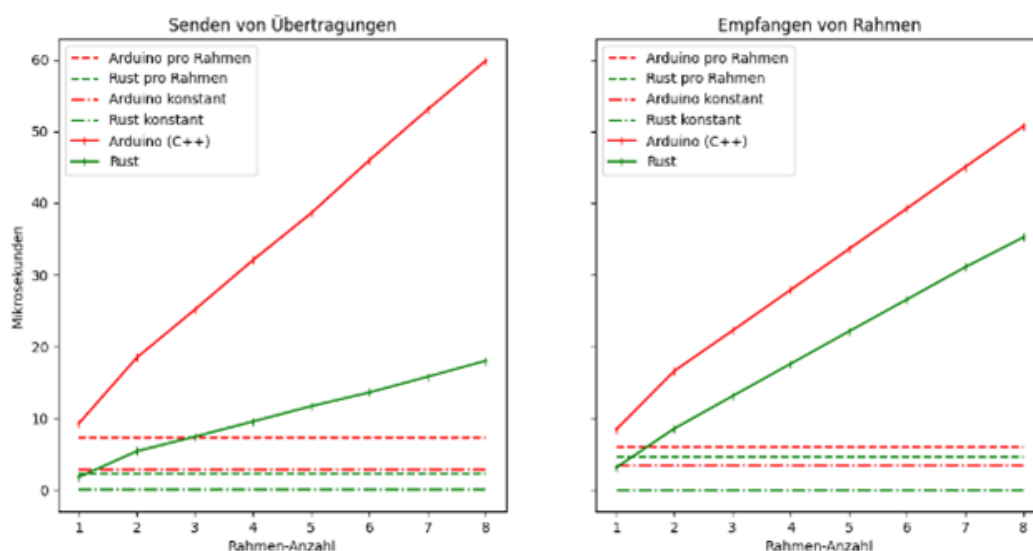


Abb. 1: Zeigt die Messergebnisse der Rust-/Arduino-Implementationen im Vergleich [1]

Bei Messungen von dem Echo-Knoten - zum einen in C++, zum anderen in Rust - benötigt der Rust-

Echo-Knoten weniger Zeit. Die Differenz bei einem Rahmen liegt hierbei bei 18 und bei 8 Rahmen bei 57 Mikrosekunden.

Auswertung und Ausblick

Durch das Anwenden der Programmiersprache Rust in dem UAVCAN Projekt hat sich gezeigt, dass diese einsetzbar für das Programmieren eingebetteter Systeme ist. Die verfügbaren Tools und ein guter Support für verschiedene Ausführtargets erleichtern es, Projekte für Mikrocontroller zu entwickeln und zu kompilieren.

Durch die Eigentümerschaft, Unveränderlichkeit zu Beginn, Überprüfen von Überläufen bei Rechenoperationen und weiteren Funktionen der Sprache wird der Programmierer gut unterstützt, sodass Fehler in Bezug auf Speichersicherheit und Logik vermieden werden können. Verstöße der Funktionen und Regeln werden während der Kompilierung überprüft und dem Entwickler in Form einer Fehlermeldung mitgeteilt. Dies macht das Programmieren für eingebettete Systeme sehr effizient und sicher, da viele mögliche Fehlerquellen bereits vor dem Übertragen auf den Mikrocontroller überprüft werden und das Programm nicht erst ausgeführt werden muss.

Aufgrund der großen Community von Rust, werden viele Bibliotheken, welche in eigenen Projekten verwendet werden können, öffentlich entwickelt. Speziell für eingebettete Systeme ist guter Support für einige Mikrocontroller vorhanden, jedoch gibt es dabei noch Lücken für manche Funktionalitäten. Da Support-Bibliotheken für Boards von den Herstellern meistens in C gestellt werden, muss auf die Community Arbeit gesetzt werden, oder durch eigene Initiative die benötigte Funktionalität implementiert werden. Der letzte Punkt wurde während der Arbeit festgestellt. Durch

fehlenden Support für CAN für die verwendeten Mikrocontroller werden zuerst automatisch generierte Bindungen zu dem C HAL des Herstellers verwendet. Nachdem während der Arbeit ein Rust HAL für CAN verfügbar wurde, wird dieser verwendet.

Die Messergebnisse der Aufrufzeiten ergeben, dass die Rust-Implementation eine gute Alternative zu der für Arduino gedachte Implementation ist. Die unterschiedlichen Messergebnisse entstehen durch Faktoren wie Implementationslogik und der Programmiersprache. Durch ein Angleichen der Implementationslogik sind die Differenzen der Ausführungsgeschwindigkeiten an den Programmiersprachen und den verwendeten Compilern festzumachen.

Durch die Einführung einer Bench-Suite, kann die Qualität der Bibliothek während der Entwicklung weiterhin überwacht werden und auch den Entwickler überzeugen die Rust-Implementation zu verwenden.

Zur weiteren Evaluation und Entwicklung der Rust-Implementation der UAVCAN Spezifikation können weitere Szenarien, welche bei UAVCAN Kommunikation auftreten können, gemessen und ausgewertet werden. Dazu können Messungen durchgeführt werden, wenn mehrere Knoten gleichzeitig kommunizieren. Dies ist interessant, wenn Übertragungen aus mehreren Rahmen bestehen und die Bibliothek diese zusammensetzen muss.

Außerdem kann die umgesetzte Bench-Suite in CI/CD eingebunden werden. Dabei können die Tests bei Merge-Requests automatisch auf der Hardware ausgeführt werden. Damit kann evaluiert werden, ob Änderungen die Ausführungszeiten beeinträchtigen.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Team Rust. Frequently Asked Questions. <https://prev.rust-lang.org/en-US/faq.html>, 2018.
- [3] Team Rust. The Rust Programming Language. <https://doc.rust-lang.org/stable/book/>, 2021.
- [4] Julia Schmidt. Graydon Hoare im Interview zur Programmiersprache Rust. <https://prev.rust-lang.org/en-US/faq.html>, 2013.
- [5] Gernot Starke. Question C-1-2: What are quality goals (aka quality attributes)? | arc42 FAQ. <https://faq.arc42.org/questions/C-1-2/>, 2019.
- [6] Consortium UAVCAN. UAVCAN. <https://uavcan.org/>, 2021.
- [7] Team UAVCAN Development. UAVCAN Specification v1.0-beta. https://uavcan.org/specification/UAVCAN_Specification_v1.0-beta.pdf, 2021.

Design and Implementation of Cryptographic-Key-Management-as-a-Service for Mission-critical Applications

Tom Jaschinski

Tobias Heer

Department of Computer Science and Engineering, Esslingen University

Work carried out at Thales Deutschland GmbH, Ditzingen

Introduction

The German Federal Office for Information Security (BSI) describes in its report about the state of IT security in Germany in 2021 that the threat landscape in cyber security is increasing. Attacks on critical infrastructures that provide mission-critical services such as oil distribution facilities or hospitals show that the impact of cyber-attacks can be very severe. [2] Therefore, there is an increasing need for cyber defence solutions that mitigate or prohibit such attacks. One countermeasure against rising security threats is the use of a Key Management System providing management capabilities for cryptographic operations within an organization.

According to ISO 7498-2:1989, Key Management is defined as the "generation, storage, distribution, deletion, archiving and application of keys in accordance with a security policy." [4] An important requirement for any Key Management System is the control of the complete lifecycle of cryptographic assets that is displayed in figure 1. The Key Management System secures the keys against unauthorized access, modification, replay, substitution and misuse. [1]

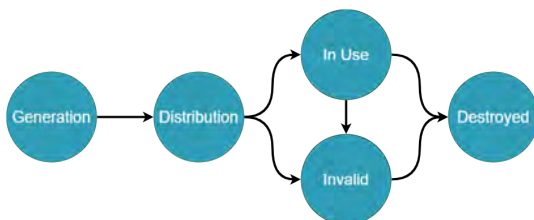


Fig. 1: Cryptographic Key Lifecycle [3]

Problem Statement and Motivation

In mission-critical environments (e.g. defence, space, aerospace, and ground transportation) current Key

Management Systems are deployed on-premise. Compared to cloud deployments that are usually operated by another organization and optimized to scale, on-premise systems introduce the following problems:

- Customers pay per system provisioning, not per system use.
- Systems do not scale as easily as as-a-service approaches and occupy more resources.
- Customer IT administrators are responsible for the system's operation and maintenance and must be trained and accordingly.
- The complete technology stack of a Key Management System is typically subject to a security accreditation resulting in an additional need for coordination on the customer's side when installed on-premise.
- New features and security patches are integrated based on the customer's patch policy instead of an instant installation after release.
- Customers must acquire and maintain the system's hosting infrastructure resulting in high capital and operational expenditures.

To meet the issues stated above, a Software-as-a-Service (SaaS) service model could be adopted for Key Management in mission-critical environments. The service providers are responsible for the system's operation while multiple customers utilize the service based on a pay per use billing policy.

Currently, many software providers already adopted an as-a-service approach which is not common in high-security environments. This is due to long-lasting legacy system architectures, the high impact of possible security incidents and due to prejudices against cloud computing and its implicated security challenges.

Advances in cloud computing and cyber security enable more secure cloud technologies that could meet the imposed challenges. Therefore, this thesis will investigate the idea of Cryptographic-Key-Management-as-a-Service (CKMaaS) for mission-critical environments.

Design

In the initial step, different as-a-service deployment scenarios (e.g. private cloud on-premise, public cloud in vendor managed infrastructure) are identified and compared.

The next step focuses on the analysis and definition of typical Key Management use cases. The analysis is performed in context of the mission-critical domains space, aerospace, ground transportation and defence. This could uncover similarities and differences and provide insight to the applicability of the previously defined as-a-service scenarios in each contemplated environment. Depending on the amount and severity of differences, the scope of the following steps could be narrowed to a selection of the environments to manage the thesis' complexity. By the time of this article, due to different and complex Key Management use cases in the contemplated mission-critical environments, the thesis focuses on Key Management for the European Rail Traffic Management System (ERTMS) in the ground-transportation environment.

In the next step, requirements engineering for CKMaaS is performed based on the previous analysis. In this step, applicable security standards and regulations for the target environments are identified to define the requirements accordingly.

The next step incorporates the modification and development of logical and physical system architectures presented in static, dynamic and deployment views that meet the defined requirements. An existing virtual-image-deployed, on-premise Key Management System will serve as a baseline for creating a Key Management Service. This step will include a security analysis of the system to identify new risks, threats and corresponding

controls to mitigate or avoid the risk of a security incident.

In the last step, a Proof of Concept (PoC) of a CKMaaS is prototypically implemented according to the developed architecture.

Evaluation

The evaluation of the thesis consists of three parts:

- Comparison of as-a-service deployment scenarios and their solutions for CKMaaS.
- Comparison of the fulfillment of specific, defined requirements for CKMaaS.
- Expense estimate of procedures in Key Management use cases for different solutions (e.g. on-premise expense of key generation compared to as-a-service expense of key generation).

Result

A CKMaaS for mission-critical environments is more efficient than on-premise applications and also increases the ease of adoption for smaller organizations. As mentioned in the Design section, the thesis narrowed its scope to Key Management for the European Rail Traffic Management System (ERTMS) in for ground-transportation environment. Nevertheless, the major design choices for providing a Key Management System as a service are likely to be applicable to the other mission-critical environments as well.

Future Prospect

The as-a-service application could serve as a next step towards more cloud solutions in mission-critical environments so that mission-critical organizations benefit from the improved efficiency of their computational capabilities. Therefore, insights developed throughout this thesis may not be restricted to the domain of Key Management alone.

References and figures

- [1] Walter Fumy and Peter Landrock. Principles of Key Management. *IEEE Journal on Selected Areas in Communications*, 5:785–793, 1993.
- [2] Federal Office for Information Security Germany. Die Lage der IT-Sicherheit in Deutschland 2021. https://www.bsi.bund.de/DE/Service-Navi/Publikationen/Lagebericht/Lagebericht_node.html, 2021.
- [3] Own representation.
- [4] International Organization for Standardization. ISO 7498-2:1989 Part 2: Security Architecture. In *Information Processing Systems – Open Systems Interconnection – Basic Reference Model*. ISO, 1989.

Programmierung von Plugins zur Audiotbearbeitung und deren Einbindung in gängige DAWs

Nicolas Kahle

Reiner Marchthaler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Motivation

Für die Aufnahme und Bearbeitung von Audiodateien werden heutzutage sogenannte Digital Audio Workstations (DAWs) verwendet. Dabei handelt es sich um Programme, die üblicherweise mehrere Tonspuren enthalten können. Auf diese Tonspuren können bei Bedarf bestimmte Effekte geladen werden, die den Klang qualitativ und/oder künstlerisch aufwerten sollen. Die Effekte wiederum werden häufig als Plugin in die DAW geladen. So bleibt die grundsätzliche Funktion der DAW übersichtlich und der Nutzer muss nur so viele zusätzliche Funktionen einbinden, wie er tatsächlich braucht. So spart er sich den Speicherplatz, den unbenutzte Funktionen bräuchten.



Abb. 1: Nutzeroberfläche einer DAW (Cubase von Steinberg Media Technologies GmbH) [2]

Aufgabenstellung

Ziel dieser Arbeit ist es, mehrere Plugins mit häufig genutzten Funktionen für gängige DAWs zu programmieren. Dabei stellt sich stets die Frage: Wie können diese Funktionen besonders leistungsfähig und nutzerfreundlich implementiert werden?

Zur Programmierung wird das Framework JUCE verwendet. Damit lässt sich ein C++-Code in verschiedenen Plugin-Formaten implementieren, die mit allen gängigen DAWs kompatibel sind. JUCE bietet bereits

viele Befehle, die zur Verarbeitung von Audiodaten sehr hilfreich sind, sowie eine umfassende Bibliothek für das Erstellen der grafischen Oberfläche eines Plugins.

Dynamikplugins

Der Begriff Dynamik beschreibt in der Musik und Tontechnik den Unterschied zwischen den lautesten und leisen Stellen eines Tonsignals [4]. Plugins, die die Lautstärkeunterschiede einer Aufnahme nachträglich bearbeiten können, werden daher oft als Dynamikplugins bezeichnet. In dieser Arbeit wurden die nachfolgenden Dynamikfunktionen programmiert. Zunächst wurde ein sogenanntes Noise Gate programmiert. Noise Gates blenden leise Signalanteile einer Tonaufnahme aus. Dies dient dem Zweck, unerwünschte Hintergrundgeräusche wie beispielsweise Hall, Rauschen oder Störsignale durch elektronisches Nebensprechen zu unterdrücken. [6] [5] [3]

Als Zweites wurde ein Audiokompressor implementiert. Dieser grenzt den Dynamikbereich eines Tonsignals ein und sorgt so für ein stabileres Signal mit geringeren Lautstärkeschwankungen. [1]

Das dritte Plugin dieser Arbeit ist das Gegenstück zum Kompressor, der sogenannte Audio Expander. Anstatt die Lautstärke zu stabilisieren verstärkt er Lautstärkeunterschiede. Das kann bei musikalischen Aufnahmen zu mehr künstlerischem Ausdruck führen oder bei extremen Einstellungen auch ähnlich wie ein Noise Gate agieren und Hintergrundgeräusche reduzieren. [1]

Das vierte Dynamikplugin ist ein Limiter. Dieser begrenzt den Pegel eines Audiosignals nach oben und schützt so die Audioanlage und auch die Ohren des Zuhörers vor Überlastung. [1]

Synchronfunktion

Alle zuvor beschriebenen Funktionen haben sogenannte Attack- und Releasezeiten. Das sind die Ausregelzeiten, die das Plugin braucht, um die Lautstärke auf den

neuen Pegel zu setzen, nachdem es getriggert wurde (vgl. Abb. 2). Bei den bisher existierenden Dynamikplugins können diese Zeiten immer nur in Millisekunden eingestellt werden. In der hier vorliegenden Arbeit wurde dem Nutzer jedoch die Möglichkeit gegeben, diese Zeiten in musikalischen Zählzeiten einzugeben. So kann der Nutzer beispielsweise eingeben, dass das Plugin bei einer Pegeländerung für die Dauer einer Viertelnote regeln soll. Diese „musikalische“ Eingabemöglichkeit wird als Synchronfunktion bezeichnet. Sie wird üblicherweise in der Musikproduktion bei Plugins angewendet, die Echo- oder Halleffekte erzeugen, damit die Wiederholungen des Echos im Takt des Songs sind. Allerdings wurde bislang nie ein Dynamikplugin mit Synchronfunktion veröffentlicht.

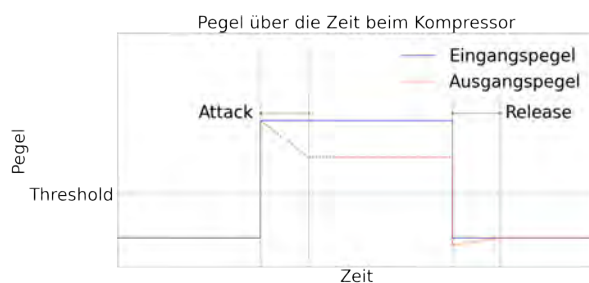


Abb. 2: Zeitlicher Verlauf von Ein- und Ausgangspegel beim Kompressor [2]

Zusammenfassung und Ausblick

Dank der Synchronfunktion ist es künftig für Musiker möglich, verschiedene Dynamikfunktionen mit einem musikalischen Ansatz einzustellen. Es ist einfacher zur Musik passende Regelzeiten zu finden, wodurch der Künstler sich mehr auf den kreativen Aspekt der Tonaufnahme konzentrieren kann, statt auf technische Details.

Dennoch bleibt ein Entwicklungspotenzial für die vorgestellten Plugins. Bei digitalen HiFi-Aufnahmen mit sehr hohen Samplerraten kommt es aufgrund des begrenzten Datendurchsatzes in der Echtzeitanwendung der Programme teilweise zu Aussetzern im Datenaustausch zwischen DAW und Plugin. Um dieses Problem zu beheben, müsste der Verarbeitungsalgorithmus optimiert werden, damit seine Zeitkomplexität reduziert wird.

Weiterhin könnte man die Features der Plugins umfangreicher gestalten. Beispielsweise bieten einige Hersteller für ihre Dynamikplugins Einstellungsmöglichkeiten für die Regelungscharakteristik an. Damit wird angegeben, in welcher Form (linear, exponentiell, ...) der Verstärkungsfaktor bei einer Änderung der Lautstärke hoch- oder heruntergeregt werden soll. In Abbildung 2 sieht man am Beispiel des Kompressors, dass in dieser Arbeit nur lineare Regelung realisiert wurde. Es gibt jedoch viele verschiedene Charakteristiken für Dynamikprozessierung, die den Klang jedes Plugins einzigartig machen, und jeder Anwender hat hier persönliche Vorlieben. Je mehr verschiedene Charakteristiken man anbieten kann, umso wahrscheinlicher ist es, dass mehr Nutzer einen Zugang zu diesem Plugin finden.

Literatur und Abbildungen

- [1] Griffin Brown. Audio Dynamics 101: Compressors, Limiters, Expanders, and Gates. <https://www.izotope.com/en/learn/audio-dynamics-101-compressors-limiters-expanders-and-gates.html>, 08 2021.
- [2] Eigene Darstellung.
- [3] Stephan Rüdiger Heydkamp. Was ist ein Noise Gate? <https://www.sabelstein.com/de/magazin/wissen/eventlexikon/n/noise-gate/>, 04 2015.
- [4] Ronald Koch. Dynamik Musik – Definition & Fachbegriffe im Überblick. <https://ronaldkah.de/dynamik-musik/>, 02 2021.
- [5] Ben Lowe. Noise Gate: What Is A Noise Gate & How Do You Use It? <https://www.musicgateway.com/blog/how-to/what-is-a-noise-gate-how-do-you-use-it>, 07 2020.
- [6] Annika Wegerle. Noise Gates: Eine Anleitung für saubere Signale. <https://blog.landr.com/de/noise-gates/>, 09 2017.

Performance und Speicheroptimierung einer bestehenden Netzwerkanalysesoftware

Kevin Kaiser

Andreas Rößler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

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

Einleitung

In der heutigen Welt nimmt Software sowohl in der Freizeit als auch in der Industrie einen immer größeren Platz ein. Umso wichtiger ist es, dass diese Software schnell und korrekt funktioniert. Da viele Projekte über einen langen Zeitraum oder unter Zeitdruck entwickelt werden, leidet häufig die Qualität des Quellcodes. Dies kann Ursache für schlechte Performance, hohen Speicherverbrauch oder Memory Leaks sein und mindert damit die Nutzerfreundlichkeit. Aus diesem Grund sind Analyse und Optimierung der Schwachpunkte wichtige Schritte zur Verbesserung der Softwarequalität, die allerdings aufgrund des potenziell hohen Zeit- und Kostenaufwandes häufig vernachlässigt werden. In dieser Arbeit wird eine industrielle Software untersucht und Schwachpunkte analysiert und optimiert.

Motivation und Ziel der Abschlussarbeit

Diese Bachelorarbeit wird in Zusammenarbeit mit der Steinbeis Embedded Systems Technologies GmbH durchgeführt. Bei der zu untersuchenden Software handelt es sich um eine Netzwerk- und Protokoll-diagnosesoftware. Diese wird von der Steinbeis-EST entwickelt und findet vorwiegend im Bereich der Industrieprotokolle ihren Einsatz.

Der IC-Monitor (kurz für Industrial Communication Monitor) wird seit vielen Jahren entwickelt und bietet einen sehr großen Funktionsumfang. Aufgrund der kontinuierlichen Entwicklung über viele Jahre hinweg sowie fortschreitender Technologien, die für die Entwicklung des IC-Monitors eingesetzt werden, besteht an einigen Stellen Optimierungspotential des Programmcodes. Ziel dieser Arbeit ist die Analyse und Optimierung der bekannten Schwachstellen und das Profiling der Software nach anderen Performanceproblemen. Dabei geht es vor allem um das Verbessern aktueller Algorithmen und um einen optimierten Speicherverbrauch. Ebenfalls wird eine Grundlage für performante Programmierung innerhalb des .NET-Umfeldes erarbeitet. Dabei werden

typische Problemstellen erläutert und grundlegendes Wissen darüber vermittelt.

Grundlagen der performanten Programmierung

Eines der wichtigsten Prinzipien in der Performance Optimierung ist das Pareto-Prinzip, das 1906 in Italien entwickelt wurde. Dieses besagt, dass circa 80% der Auswirkungen von 20% der Ursachen ausgelöst werden, es wird daher auch als 80-20-Regel bezeichnet. Übertragen auf die Softwareentwicklung bedeutet es, dass 80% der verfügbaren Ressourcen für 20% des Codes genutzt werden. Hierbei ist es wichtig zu verstehen, dass das Pareto-Prinzip nicht immer genau so zutreffen muss, sondern eher als eine allgemeine Richtlinie bezüglich der Optimierung gilt. Das Wissen darüber, dass die Performance Bottlenecks oft nur durch einen kleinen Teil des Codes ausgelöst werden, kann bei der Optimierung von Applikationen zeit- und aufwandssparend eingesetzt werden [2].

Analyse und Optimierung der Software

Für die .NET-Umgebung gibt es einige Profiling-Tools, um Programmcode zu analysieren. Im Falle dieser Arbeit wurde sich für die Profiler DotMemory und DotTrace von JetBrains entschieden. Unter anderem ist JetBrains vollkommen in Visual Studio integriert und bietet darüber hinaus noch andere nützliche Werkzeuge, die den Entwickler unterstützen.

Um Performance-kritische Stellen innerhalb der Software zu finden, wurden Szenarien erarbeitet, welche den typischen Einsatz der Software widerspiegeln. Solche Szenarien können gewisse Lasttests, der Start-up Vorgang oder das Aus- und Einlesen von Dateien sein. So wurde für einen TCP-Lasttest die Übertragung einer großen Datei innerhalb des Netzwerkes überwacht. Dabei wird jedes übertragene Paket untersucht und auf Vollständigkeit überprüft. Entscheidend hierbei ist, ob der IC-Monitor mit der Durchflussmenge und

der Überprüfung der Pakete zurechtkommt und diese zeitlich schnell genug abhandeln kann.

Der IC-Monitor wird dann in diesen Szenarien mithilfe der Profiler untersucht, wobei durch DotMemory der Speicherverbrauch über die komplette Laufzeit beobachtet und an entscheidenden Zeitpunkten ein Snapshot des Speichers aufgenommen werden kann. Ein Snapshot beinhaltet Informationen über die aktuell gehaltenen Daten, aber auch welche Datentypen am meisten Speicher benötigen (siehe Abbildung 1).

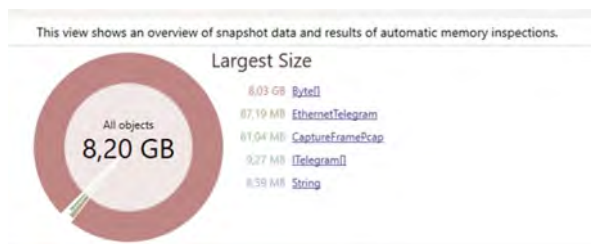


Abb. 1: Snapshot DotMemory Speicherverteilung [1]

Durch die Funktion, Snapshots miteinander zu vergleichen, lässt sich aufzeigen, welche Daten zwischen beiden Snapshots überlebt haben, bzw. welche neu hinzugekommen oder weggefallen sind. Anhand dieses Vergleiches lassen sich so z. B. Memory Leaks oder leere speicherverbrauchende Datenstrukturen finden. Mithilfe von DotTrace lassen sich die Ausführungszeiten der unterschiedlichen Methoden und Prozesse anzeigen und die zeitintensivsten Bottlenecks finden. Diese haben meistens durch schlechte Algorithmen oder falsch verwendeter Methoden einen Zeitverlust, der sich auf die Performance des Programmes auswirkt. Darin lassen sich sowohl die genauen Laufzeiten der

Methoden als auch die Anzahl der Aufrufe anzeigen, wie in Abbildung 2 zu sehen ist.

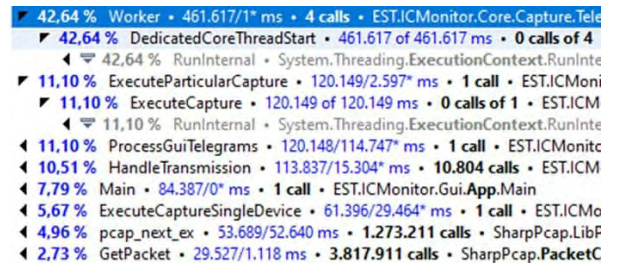


Abb. 2: Ergebnis DotTrace Analyse [1]

Nachdem Schwachstellen mithilfe der Profiler gefunden wurden, werden diese untersucht und durch Änderungen des Algorithmus oder durch Verwendung anderer Methoden optimiert. Dabei wird darauf geachtet, ob sich ein Optimierungsvorgang überhaupt lohnt, wie in Pareto beschrieben, sollte der Aufwand auch den Nutzen rechtfertigen.

Ausblick

Anhand der Analyse des IC-Monitors wird das systematische Vorgehen beim Profiling und das Einordnen der gefundenen Performanceprobleme demonstriert. Mit den Ergebnissen der Analyse werden dann Maßnahmen zur Verbesserung der Software ergriffen und hinsichtlich ihrer Auswirkung auf die Gesamtpformance untersucht. Ebenfalls ist eine Aufstellung der Schwachstellen und damit verbundene Lösungsansätze ein Ziel dieser Arbeit. Gemeinsam mit den Ergebnissen der Analyse dienen die „best practise“ Empfehlungen dieser Arbeit dazu, eine Grundlage zur kontinuierlichen Verbesserung der untersuchten Software zu schaffen.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Jim Xochellis. The impact of the Pareto principle in optimization. <https://www.codeproject.com/Articles/49023/The-impact-of-the-Pareto-principle-in-optimization>, 01 2010.

Strategie für die Testautomatisierung am Beispiel von C@P Marketing Porsche

Radika Kapour

Astrid Beck

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Porsche AG, Zuffenhausen

Einleitung

Die Bedeutung von Marketing in einem Unternehmen hat stark zugenommen, dabei spricht man von einer marktorientierten Unternehmensführung, die sich um die Bedürfnisse der Kunden kümmert. In der heutigen Zeit ist eine Software zur Unterstützung und Pflege der Kundenbeziehung ein wichtiger Bestandteil vieler Unternehmen. Aus diesem Grund informieren sich Unternehmen über die breite Auswahl an CRM-Systeme, wählen die geeigneten aus, um die Kundendaten zu erfassen und um die Beziehung stets aufrecht und gepflegt zu halten. Gleichzeitig unterstützt das CRM-System auch bei der Optimierung von Prozessen, der Steigerung der Rentabilität und auch der Automatisierung der Prozesse. Das Marketing umfasst unter anderem die Planung und Durchführung von Kampagnen, kümmert sich um die Bestandskunden und unterstützt bei der Kundenakquise. Damit die Unternehmen auch langfristig von solchen Systemen profitieren können, sollten die Prozesse der Software auf Qualität getestet werden und im besten Fall frühzeitig Fehler erkennen. Allerdings kann das manuelle Testen auf verschiedenen Ebenen sehr zeitaufwendig, kostenintensiv und fehleranfällig sein. Damit die Softwarequalität und die Testqualität gewährleistet werden können, sollte das manuelle Testen, mithilfe von geeigneten Tools, auf das automatisierte Testen umgestellt werden. Das CRM-Marketingsystem läuft über eine Datenbank, auf dem alle Kunden- und Interessentendaten gespeichert und verwaltet werden. Auf der Grundlage der Daten wird eine Segmentierung, die auf die SAP-Hybris Marketing Software beruht, durchgeführt und in Zielgruppen aufgeteilt. So lassen sich Marketing-Kampagnen planen und ausführen.

Problemstellung

Bisher gibt es für das CRM-Marketingsystem, sowohl im Front- als auch im Backend, keinerlei Tests. Es ist zu untersuchen, ob die Prozesse des CRMs getestet und automatisiert durchgeführt werden können, dabei

wird die Pyramide als Leitfaden für das Testen genutzt und untersucht, ob auf jeder Stufe der Pyramide die Testautomatisierung der Prozesse sinnvoll ist (siehe Abbildung 1). Allerdings ist darauf zu achten, ob die Tests voneinander abhängig bzw. unabhängig durchgeführt werden können und ob die Reihenfolge eine Rolle spielt.

Zielsetzung

Das Ziel dieser wissenschaftlichen Arbeit ist es, die Softwarequalität des CRM-Marketingsystems zu verbessern. In der Norm ISO/IEC 9126 wird anhand folgender Merkmale die Softwarequalität definiert: Funktionalität, Zuverlässigkeit, Effizienz, Benutzbarkeit, Übertragbarkeit und die Änderbarkeit. Um regelmäßig zu prüfen, ob all diese Merkmale, die für die Qualität ausschlaggebend sind, eingehalten wurden, spielt das Testen der Software eine zentrale Rolle. Aus Gründen wie steigende Kosten, Zeitaufwand usw. soll das manuelle Softwaretesten auf automatisiertes Testen umgestellt werden. Denn je mehr automatisiert wird, desto besser kann es, sowohl für den Entwickler als auch für das Unternehmen hinsichtlich der Produktqualität sein. [1] Schließlich soll eine Teststrategie aufgestellt werden, in der im Idealfall alle erarbeiteten Testverfahren abgedeckt und auf die Prozesse des CRM-Marketingsystems anwendbar sind. Dabei sollen die einzelnen Testverfahren untersucht und mit dem entsprechenden Tool auf die Prozesse des CRM-Systems angewandt werden. Die Testautomatisierung der Prozesse des CRM-Marketingsystems sollen unterstützen, auf Qualität zu testen, bis man zum User angelangt, denn von dort an soll durch den User Acceptance Testing die finale Entscheidung über die Qualität getroffen werden. Dieser Test ist wichtig, um sicherzustellen, dass die Software den Anforderungen der Benutzer bzw. der Kunden nachkommt und akzeptiert wird.

Vorgehensweise

Zunächst soll bestimmt werden, welche Bereiche des CRM-Marketingsystems Bedarf an Tests haben, sprich wo ist das Testen überhaupt notwendig bzw. sinnvoll? Dann soll kategorisiert werden, auf welcher Teststufe sich die einzelnen Bereiche befinden, z.B. Integrationstest. Zugleich sollen die einzelnen Testverfahren ausgearbeitet und das entsprechend geeignete Tool für das Testen ausgewählt werden. Schließlich sollen die aufgestellten Testverfahren anhand des geeigneten Tools auf das CRM-Marketingsystem prototypisch erprobt werden.

Softwaretesting

Die Softwareentwicklung ist ein mehrstufiger Prozess, in der auf jeder Entwicklungsstufe die relevanten Tests ausgewählt werden sollen. Daher ist es wichtig frühzeitig mit dem Testen anzusetzen, denn je später mit dem Testen begonnen wird, desto mehr Aufwand und Kosten können entstehen. Die Testpyramide zeigt die aufeinander folgende Teststufen, wobei die unterste Stufe die Unittest darstellen, in der überprüft wird, ob die Komponenten so funktionieren, wie diese es beabsichtigen. Darauf folgt der Integrationstest, in der das Zusammenspiel verschiedener Komponenten getestet wird. Schließlich kommt das Systemtest, die sich um die Quality Assurance kümmert, d.h. wie die verschiedenen Komponenten im vollständigen, integrierten System zusammenwirken. Die Spitze der Pyramide bilden die Abnahmetest, auch User Acceptance Test genannt, in der die Software aus Sicht des Endnutzers betrachtet wird und überprüft, ob sie wie erwünscht funktioniert. Unabhängig auf welcher Stufe man sich

befindet, können die Regressionstest durchgeführt werden, in der sich die Testfälle wiederholen, um sicherzustellen, dass die bereits getesteten Teile keine neuen Fehler verursachen. [4]



Abb. 1: Teststufen [2]

Testautomatisierung

Testautomatisierung kann Tests, die wiederholt ablaufen sollen, durch das Automatisieren vereinfachen und gleichzeitig Zeit und Kosten sparen. Manuelle Testfälle, bei denen mehrere Wochen einplanen werden müssen, können innerhalb von wenigen Stunden automatisiert durchführen werden. Erfolgreiche Automatisierung kann nur durch eine gute Planung und kompetente Umsetzung zu effizienten Tests zu führen. [3]

Literatur und Abbildungen

- [1] Stephan Augsten. Was ist Softwarequalität? <https://www.dev-insider.de/was-ist-softwarequalitaet-a-791790/>, 2021.
- [2] Eigene Darstellung.
- [3] Frank Witte. *Testmanagement und Softwaretest*. Springer Fachmedien Wiesbaden, 2019.
- [4] Frank Witte. *Strategie, Planung und Organisation von Testprozessen*. Springer Fachmedien Wiesbaden, 2020.

Provisioning von Embedded Devices in der Azure Cloud

Tunahan Kaya

Rainer Keller

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma F&S Elektronik Systeme GmbH, Stuttgart

Einleitung

Industrie 4.0 wird in der Zeit der Vernetzung immer wichtiger, da sich die Menge an Computergeräten und -modulen durchgehend vermehrt. Aufgrund dessen wird es auch umso schwieriger, für eine regelkonforme Verwaltung der Geräte zu sorgen, da die Erreichbarkeit der Geräte durch die Menge der Geräte, die im Umfeld sind, erschwert werden. Mit Einführung der Industrie 4.0, soll die industrielle Produktion mit moderner Informations- und Kommunikationstechnik versehen werden. [3] So wird dafür gesorgt, dass die Geräte schon bei der Produktion digital vernetzt werden. Das bringt den Vorteil, die Geräte in bestimmten Cloud-Computing-Plattformen zu registrieren und von dort aus zu verwalten. Die bekanntesten Unternehmen, hierfür sind Amazon, Microsoft und Google mit den Cloud-Computing-Plattform Amazon AWS IoT, Microsoft Azure IoT und Google IoT. In dieser Arbeit wird Microsoft Azure IoT verwendet, wobei hier die Geräte in sogenannten IoT-Hub's registriert werden. IoT-Hub ist der zentrale Ort, an dem die Geräte zur Überwachung und Verwaltung in Azure hinzugefügt werden können. Bei der Registrierung eines Geräts in einen IoT-Hub wird ein digitaler Zwilling für jedes einzelne Gerät erzeugt. Der digitale Zwilling ist ein digitales Abbild des Geräts, womit das Gerät seinen aktuellen Zustand (wie z.B. Temperatur, SW Stand) dem IoT-Hub mitteilen kann, indem Gerät-zu-Cloud-Telemetriedaten zur Kommunikation verwendet werden, die als JSON-Dokumente geliefert werden. So lässt sich der Gerätezustand über die Cloud in Echtzeit nachverfolgen. Der Gerätezustand wird auch mit dem Schlagwort „Predictive Maintenance“ verknüpft, wobei dieser durch Ansammlung der Gerätedaten, die Ausfallwahrscheinlichkeit des Geräts vorhersagen kann. F&S entwickelt und verkauft eingebettete Computermodule an Kunden, wobei ein digitaler Zwilling des eingebetteten Computermoduls automatisch in den gewünschten IoT-Hub angelegt werden soll, sobald sich das Board im gewünschten IoT-Hub registriert, um das Gerät anschließend überwachen und verwalten zu können. In diesem Fall sollen die Geräte in erster

Linie für Updateprozesse in IoT-Hub's registriert und bereitgestellt werden. Die Registrierung kann sowohl manuell als auch automatisiert durchgeführt werden, wobei die manuelle Registrierung bei einer großen Menge an Geräten ziemlich Zeitintensiv sein kann. Aus diesem Grund wird bei einer großen Menge von Geräten, die registriert werden sollen, eine automatisierte Registrierung empfohlen. Daher setzt auch F&S auf die automatische Geräteregistrierung. Hierfür bietet Azure eine Ressource an, dass sich Device Provisioning Service nennt. Damit sollen die Geräte automatisiert in die gewünschten IoT-Hub's zugewiesen werden.

Umsetzung

Bevor jedoch mit dem Provisioning begonnen werden kann, müssen einige Vorkehrungen getroffen werden. Dazu gehört zum einen die Geräteauthentifizierung, was bedeutet, dass die Geräte ihre Identität dem Azure Cloud nachweisen müssen. Außerdem muss auf der Cloudseite eine Registrierung für die Geräte erzeugt werden, wobei man sich vorher Gedanken machen muss, welche Art von Registrierung für den eigenen Bedarf sinnvoll ist. Sobald die Entscheidung getroffen ist, kann die Azure Ressource „Device Provisioning Service“ für die automatische Geräteregistrierung verwendet werden. Wenn die Geräte anschließend erfolgreich in die gewünschten IoT-Hub's registriert wurden, können diese zum Beispiel für Updates bereitgestellt werden. Für diesen Vorgang wird eine weitere Azure Ressource verwendet. Diese nennt sich Azure Device Update für IoT-Hub und ermöglicht Updates in IoT-Hub zu importieren und diese anschließend für Geräte bereitzustellen.

Verbindungsaufbau

Geräte können nur dann zum Device Provisioning Service eine Verbindung aufbauen, wenn diese den DPS-Endpunkt und den ID-Bereich des Device Provisioning Service wissen. Den DPS-Endpunkt verwenden Geräte, um eine Verbindung zum Device Provisioning Service aufzubauen. Der ID-Bereich ist eine Eindeutige

Identität des Dienstes, welche bei der Erstellung generiert wird und dient der Identifizierung des Bereitstellungsdiensts. Während der Entwicklung sollten diese Informationen direkt am Gerät hinterlegt werden. Die Kommunikation kann hierbei mit verschiedenen Internetkommunikationsprotokollen ermöglicht werden. Azure bietet hierfür MQTT, AMQP, HTTP, sowie auch Websockets als Kommunikationsmittel an. Bei der Registrierung mittels automatischer Bereitstellung, spricht man auch von einem Datensatz der Geräte, die registriert werden. Wird eine Verbindung vom Gerät zum Dienst aufgebaut, so wird das Gerät nach erfolgreicher Authentifizierung, in einen sogenannten Registrierungsdatensatz aufgelistet. Unter diesem Datensatz kann der verwendete Nachweismechanismus, die gewünschte Konfiguration und die Geräte-ID entnommen werden.

Geräteauthentifizierung

Die Geräteauthentifizierung kann mittels symmetrischer Schlüssel oder auch mit Zertifikatsnachweis erfolgen. Azure unterstützt hierbei ein Zertifikatsnachweis per X509-Zertifikate oder auch TPM-Chips, die sich mit sogenannten Endorsement Keys nachweisen. Da symmetrische Keys sowohl für die Verschlüsselung als auch für die Entschlüsselung ein und dieselbe Key nutzen, wird von diesem Verfahren aus Sicherheitsgründen abgeraten. Um für die F&S-Boards den bestmöglichen Schutz zu gewährleisten, kommen X509-Zertifikate als Identitätsnachweis zum Einsatz. Außerdem eignen sich X509-Zertifikate für die Bereitstellung von beliebig vielen Geräten.

Geräteregistrierungen per Device Provisioning Service

Mit der Nutzung der Ressource Device Provisioning Service gibt es zwei Möglichkeiten, Geräte in IoT-Hub's zu registrieren. Dafür muss man eine Registrierung innerhalb von Azure Device Provisioning Service erstellen. Zum einen kann man eine individuelle Registrierung erstellen, wobei diese sich dann eignet, wenn nur ein Gerät registriert werden soll. Wenn mehrere Geräte registriert werden sollen, müssten für jedes einzelne Gerät eine Registrierung auf der Cloudseite erstellt werden. Dieser Vorgang ist in der Produktionsumgebung, eher unpraktisch, da hier eine große Menge an Geräten in Einsatz kommen und die Registrierung somit viel Zeit in Anspruch nehmen würde. Auf der anderen Seite gibt es aber auch die Möglichkeit, eine Registrierungsgruppe anzulegen. Wie auch aus Abbildung 1 zu entnehmen ist, bietet die Nutzung der Registrierungsgruppe den Vorteil, dass durch diese Registrierungsgruppe beliebig viele Geräte registriert werden können. Die Geräte müssen nur die

notwendigen Zertifikate besitzen und ihre Identität dem Device Provisioning Dienst nachweisen können.

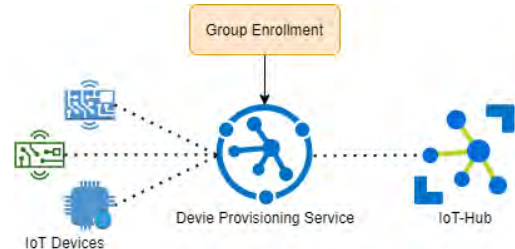


Abb. 1: Registrierungsgruppe - Device Provisioning Service [2]

Zielsetzung

Ziel dieser Arbeit ist es, Geräte reibungslos und automatisiert in den gewünschten IoT-Hub in der Azure Cloud zu registrieren und für Updates bereitzustellen. Um dies zu erreichen, werden in Azure Cloud erstmals die notwendigen Ressourcen angelegt und konfiguriert.

Vorgehensweise

An erster Stelle muss die Cloudseite eingerichtet werden, indem die jeweiligen Ressourcen erzeugt und konfiguriert werden. Die notwendigen Ressourcen sind vorerst der Azure IoT Hub und der Azure Device Provisioning Service für IoT Hub. Diese können dann anschließend miteinander verlinkt werden. Nur so kann dann der Device Provisioning Service die Geräte in die gewünschten IoT-Hub's registrieren. Außerdem müssen auch die notwendigen X509-Zertifikate ins Device Provisioning Service hinzugefügt werden und um dafür zu sorgen, dass die Geräteregistrierung von mehreren Geräten ermöglicht werden kann, muss zusätzlich eine Registrierungsgruppe erstellt werden. An nächster Stelle muss der Device Provisioning Service in den Fertigungsprozess integriert werden. So soll sichergestellt werden, dass das Board sich mit dem vorgesehenem Device Provisioning Service über den DPS-Endpunkt verbinden kann. Dabei muss das Board auch mit den X509-Zertifikaten versorgt werden. Sobald das Board den DPS-Endpunkt kontaktiert, prüft der Device Provisioning Service die Identität des Boards anhand der beigefügten Zertifikate. Kann die Identität bestätigt werden, so kommt es zu einem Verbindungsaufbau zwischen dem Gerät und dem Device Provisioning Service zustande und der Dienst kann anschließend das Gerät in den vorgesehenen IoT-Hub zuordnen. [1] In Abbildung 2 kann die Funktionsweise des Tools auch grafisch als ein Zeitdiagramm veranschaulicht werden.

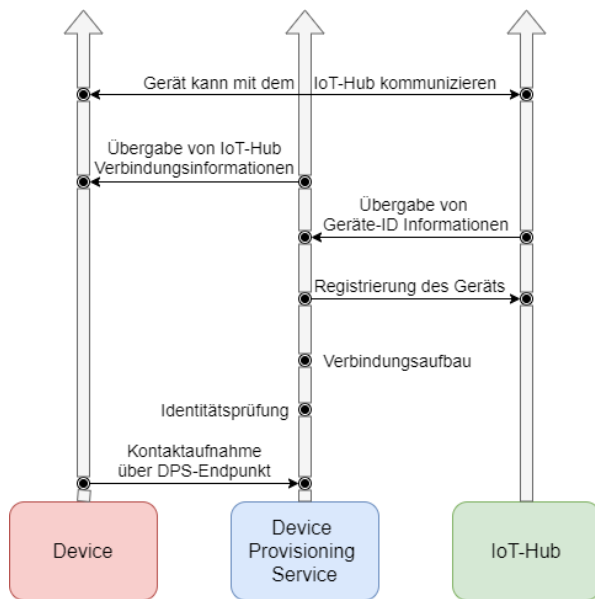


Abb. 2: Funktionsweise des Device Provisioning Service [2]

Nachdem die Geräte in die gewünschten IoT-Hub's zugeteilt wurden, können diese wie bereits erwähnt, für Updateprozesse bereitgestellt werden. Dazu bietet sich die Ressource Device Update für IoT Hub an. Dieser ist dafür da, um eine Instanz für IoT-Hub's zu erstellen. So wird dafür gesorgt, dass Updatefiles direkt in Azure hinzugefügt und für Geräte bereitgestellt werden können. Die Updates werden mithilfe des MS-Azure Clients auf den entsprechenden Boards installiert.

Ausblick

In den nächsten Schritten wird der Device Provisioning Service per Skripting automatisiert und in die F&S-Boards integriert. Dabei soll dafür gesorgt werden, dass die Geräte beim Starten automatisch die Verbindung zum Device Provisioning Service aufnehmen, so dass dieser den digitalen Zwilling des Gerätes im gewünschten IoT-Hub erstellt.

Literatur und Abbildungen

- [1] Microsoft Corporation. Was ist Azure IoT Hub Device Provisioning Service? <https://docs.microsoft.com/de-de/azure/iot-dps/about-iot-dps>, 2021.
- [2] Eigene Darstellung.
- [3] RobertRED Wikipedia. Industrie 4.0. https://de.wikipedia.org/w/index.php?title=Industrie_4.0&oldid=215965124, 2012.

Konzeption und Implementierung eines Samplingalgorithmus für den Szenariobasierten XiL-Test für Fahrerassistenzsysteme

Jannick Keller

Clemens Klöck

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Daimler Truck AG, Untertürkheim

Einleitung

Auf Europäischen Straßen passieren sehr viele Verkehrsunfälle. Deshalb hat die europäische Kommission beschlossen, bis 2050 alle Unfälle mit der Vision Zero im Straßenverkehr zu verhindern. Die Aufgabe der Automobilhersteller ist es, den Menschen bei Gefährlichen Situationen zu assistieren und mögliche Fahrfehler zu korrigieren. Um eine Unfallquote von 0% zu erreichen bedarf es einer Vollautomatisierung von Fahrzeugen. Um dies jedoch bewältigen zu können, wird diese Vollautomatisierung in einzelne, getrennt voneinander betrachtbare Systeme aufgeteilt. Diese müssen auf korrekte Funktion bei unterschiedlichen Umwelteinflüssen überprüft werden.

Motivation

Bei der Absicherung von Fahrerassistenzsystemen, ist ein erheblicher Aufwand erforderlich, da nach [2] sechs Milliarden Testkilometer benötigt werden, um eine statistische Aussage über die Funktion von Fahrerassistenzsystemen treffen zu können. Dies ist Zeitlich und Finanziell nicht wirtschaftlich. Daher wird auf eine andere Art Test ausgewichen. Es können Modelle der Umgebung und des Fahrzeugs mit allen Systemen und Sensoren erstellt werden und die Realität in einer Simulation abgebildet werden. Dies erlaubt eine effizientere Absicherung von Fahrerassistenzsystemen. Um alle möglichen Testfälle abdecken zu können werden diese Modellbasiert auf Software und Hardware in the Loop Systemen getestet und ausgewertet. Da jedoch die Simulation von tausenden verschiedenen Szenarien erhebliche Zeit in Anspruch nimmt, wird nicht jede einzelne Einflussgröße einzeln untersucht. Um jedoch herauszufinden welche Kombinationen der Einflussgrößen untersucht werden müssen, muss vor dem Test eine Teststrategie ausgearbeitet werden, um eine optimale Abdeckung zu ermöglichen. Außerdem wird eine Testautomatisierung ohne menschliche Interaktion ausgearbeitet, die es ermöglicht autark

und zuverlässig alle zuvor ausgewählten Szenarien außerhalb der Arbeitszeit simulieren zu lassen.

Ergebnisse

Die Simulationsanzahl von 19.800 Szenarien mit N-Weise Abdeckung zu simulieren ist aus Zeitgründen nicht möglich. Daher wurde die T-Weise Abdeckung, auch Pairwise Methode genannt, genutzt um die Anzahl der zu testenden Szenarien auf ein Minimum zu reduzieren. Diese Methode beschreibt den Vorgang, in einer Liste Gruppen von Parametern zu bilden, und von unten nach oben alle Szenarien zu löschen, die die selbe zuvor gebildete Gruppe besitzen.

	A	B	C
1	10	5	140
2	10	5	120
3	10	1	140
4	10	1	120
5	20	5	140
6	20	5	120
7	20	1	140
8	20	1	120

Abb. 1: Verfahren zur T-Weise Abdeckung in einer Liste mit Szenarioparametern [1]

In einem Fall mit drei Parametern mit je zwei Werten ergeben sich acht verschiedene Kombinationen aller. Bei einer T-Weise Abdeckung mit einer Gruppengröße von 2 Elementen lassen sich die in der Tabelle 1 in Rot und Blau markierten Kombinationen optimieren, da diese Gruppen bereits aufgetreten sind und nicht mehr getestet werden müssen. Dadurch wird die Anzahl Szenarien auf zwei reduziert.

Es werden nur noch die Kombinationen mit der Nummer sechs und acht simuliert. Eine Eigenheit dieser Methode ist die Auswahl der Kombination, die zuerst betrachtet wird. Dies ist immer das unterste und wird somit immer ausgewählt. Um das zu umgehen, wird vor der Auswahl die gesamte Liste gemischt und neu angeordnet, wodurch bei jeder Ausführung eine andere Kombination am Ende der Liste steht.

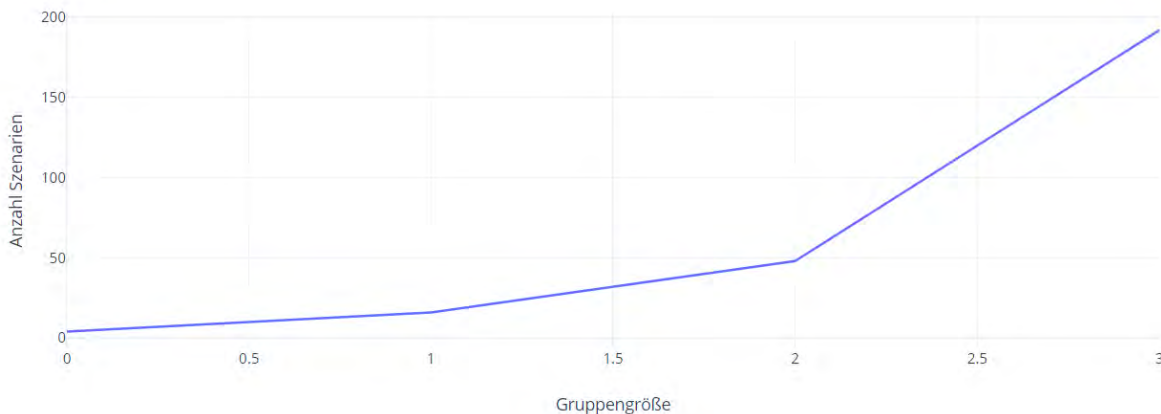


Abb. 2: Vergleich verschiedener Gruppengrößen für die T-Weise Abdeckung [1]

Eine Vorhersage der sich ergebenden Kombinationen der T-weise Abdeckung für verschiedene Gruppengrößen ist nicht möglich, da hier unterschiedliche Anzahlen von Werten pro Parameter verwendet werden. Jedoch ergibt sich bei einem Vergleich ein exponentielles Wachstum, das angenähert werden kann. Dadurch wird sichergestellt, dass alle Kombinationen, die vorkommen, einzigartig sind. Bei einer Gruppengröße von 3, ergeben sich 48 Szenarien, die simuliert werden müssen. Jedoch werden bei dieser Methode die Simulationsergebnisse nicht evaluiert, da dies ein reiner a priori Algorithmus ist.

$$T = (T_{scen} + T_{init}) * N_{scen}$$

Da die Simulationsdauer eines Szenarios stark von einzelnen Parametern wie der Geschwindigkeit des Ego Fahrzeugs abhängt, lässt sich die Gesamtdauer aller Simulationen nur schwer bestimmen. Wird jedoch von einer durchschnittlichen Simulationsdauer von 3 Minuten und einer Initialisierungszeit von 15 Sekunden ausgegangen, ergibt sich zwischen kompletter Abdeckung und T-Weise Abdeckung eine Ersparnis von 1.070 Stunden. Die Gruppengröße kann je nach Bedarf vergrößert oder verkleinert werden, um eine feinere Abstufung zu gewährleisten. Bei einer Gruppengröße, identisch zu der Anzahl der Parameter, wird sich eine vollständige Abdeckung ergeben.

Durch die obige Formel wird klar, wodurch sich die Simulationszeit verkürzen lässt. Zum einen ist die Rechenleistung des simulierenden Rechners ein großer Einflussfaktor, jedoch kann dieser nur begrenzt durch schnellere Rechner optimiert werden. Der Einflussfaktor mit dem größten Optimierungspotential ist die Anzahl an Szenarien. Dadurch verringert sich die Simulationszeit auf nur wenige Stunden, wenn davon ausgegangen wird, alle Szenarien in Echtzeit rechnen zu können. Ziel der Optimierung ist eine möglichst gleichmäßige Abdeckung aller Parameter und gleichzeitig maximal unterschiedliche Szenarien zu generieren.

Ausblick

Um die Simulationszeit maximal effizient zu nutzen, ist es unabdingbar, die Simulationsergebnisse der jeweils vorherigen Simulation zu evaluieren und entsprechende Anpassungen in den darauffolgenden Szenarien anzuwenden. Dadurch können etwaige Fehler oder Schwachstellen in den Fahrerassistenzsystemen am besten gefunden werden. Jedoch wird hier ein Maß zur Beurteilung einer Simulation benötigt, mit dem die Simulationen verglichen werden können. Mit diesem Maß kann dann während der Laufzeit von Simulationen entschieden werden, welche Parameter verändert werden müssen, um die Eigenschaften des Systems optimal abzubilden. Dies ist nicht im Umfang dieser Arbeit enthalten.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] N. Weber. Reduzierung des Parameterraums für die Freigabe von hochautomatisierten Fahrfunktionen, 2020.

Reengineering der Plugin-Verwaltung in der Desktop-Anwendung Elektra

Jan Robin Kellner

Andreas Rößler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Vector Informatik GmbH, Stuttgart

Einführung

Nach längerer Entwicklung an einer Software kann es passieren, dass durch angehende Wartung einer Software deren Komplexität steigt. Folglich werden weitere Wartungen immer aufwändiger. Ab einem gewissen Punkt muss durch Anpassung der Software die Wartbarkeit verbessert werden. Eine Möglichkeit dafür ist Reengineering.

Elektra ist eine seit 25 Jahren im Einsatz befindliche Software. Deren Entwicklung erfolgt derzeit in einem Scrum-System mit regelmäßigen Releases. Daher muss Elektra zu jedem Zeitpunkt in einem funktionierenden Zustand sein. Dies verhindert die Möglichkeit größere architektonische Veränderungen durchzuführen. Über die Jahre hat auch das Hinzufügen immer weiterer Features zu einigen technischen Schulden geführt. Eine Stelle mit Modernisierungsbedarf ist das Plugin-System, das nach dem Einloggen auf einer Server-Domain die benötigten Plugins lädt.

Zielsetzung

Das Ziel der Bachelorarbeit ist das Reengineering dieser Plugin-Verwaltung, so wie das Auswählen und Implementieren eines Frameworks, damit diese Komplexität möglichst reduziert werden kann.

Dafür muss zuerst die derzeitige Implementation analysiert werden. Aus dieser Analyse werden dann die Anforderungen an die Plugin-Verwaltung abgeleitet. Auf Basis dieser Analyse soll ein passendes Framework ausgewählt werden, das im Laufe des Reengineerings implementiert wird.

Elektra

Elektra ist ein Programm, das bei der Spezifikation des Elektrobaumes eines Fahrzeuges unterstützt. In Elektra können die Anforderung, Spezifikation, Design und Test von Software, Steuergeräte und deren Vernetzung erstellt werden.

Elektra basiert auf einer Drei-Ebenen-Architektur, siehe Abbildung 1. Das bedeutet, dass der Client zusammen mit einem Server und einer Datenbank arbeitet. Die Datenbank ist eine Oracle-Datenbank und der Server ein Tomcat-Server. Der Client selber wird nur zum Anzeigen und Bearbeiten der Objekte verwendet. Auf dem Client können eine Reihe an Jobs gestartet werden. Darunter fallen unter anderem Validatoren und Exports. Diese Jobs laufen auf dem Server, wobei nur das Ergebnis und die Statusinformationen an den Client gesendet werden.

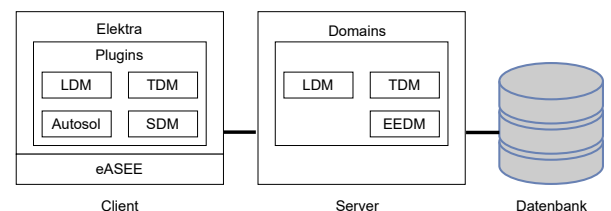


Abb. 1: Drei-Ebenen Struktur von Elektra [1]

Die Basis des Clients, *eASEE*, ist in einer Mischung aus C++ und C# geschrieben. Elektra selber besteht nur aus vier C#-Plugins, und deren Libraries, die nach Einloggen an dem Server geladen werden. Je nach Domain werden ein bis drei Plugins geladen. Die Datenstruktur besteht aus einzelnen Objekten, siehe Abbildung 2. Für jeden Objekttyp gibt es eine bestimmte Ansicht zur Konfiguration und Darstellung der Objekte. Die Ansichten der Objekte und weitere Aktionen, wie zum Beispiel Wizards, werden in dem Plugin der Domain geladen.

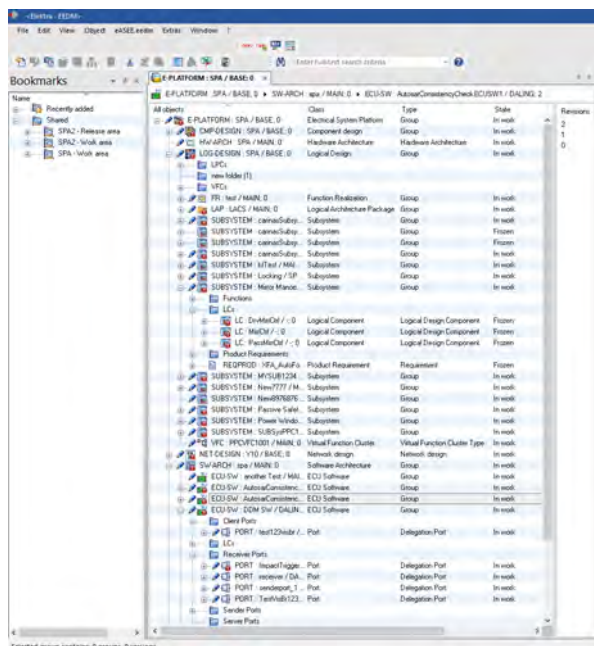


Abb. 2: Übersicht über die Oberfläche von Elektra [1]

Software Reengineering

Reengineering wird meist für den Umgang mit Legacy-Software verwendet. Hierbei handelt es sich um Software, deren Wartung zu viel Aufwand benötigt oder nicht mehr möglich ist und deren Versagen zu signifikanten Verlusten führen würde [2].

Das Ziel des Reengineerings ist die Steigerung der Qualität dieser Legacy-Software in folgenden Faktoren: Wartbarkeit, Erweiterbarkeit, Wiederverwendbarkeit und Komplexität der Software. Die Funktionalität der Software soll in diesem Prozess möglichst unverändert bleiben [2].

Reengineering beginnt mit dem Reverse-Engineering der Software. Hierbei wird durch Analyse der Software, deren Design rekonstruiert. Die Dokumentation der Software kann auch in den Prozess einfließen, oft spiegelt diese aber nicht den exakten Stand der Architektur wider [2]. Anhand dieser Analyse können Requirements und die Architektur für eine äquivalente Re-Implementierung erstellt werden.

Plugins

Mithilfe von Plugins können einzelne Teile einer Codebasis abgekoppelt werden. Dadurch wird die Implementation stark von dem Interface getrennt. Generell bieten Plugins mehrere Vorteile:

Ein Plugin kann durch ein Anderes ausgetauscht werden, das denselben Contract erfüllt. Dies hilft

besonders bei Interfaces zu externen Schnittstellen. Plugins können auch von Drittanbietern geliefert werden, wenn eine Schnittstelle existiert. Dadurch kann Funktionalität in die Software programmiert werden, die nur von einer kleinen Gruppe an Nutzern verwendet werden wird.

Auswahl eines Frameworks

Die Implementation muss in der Lage sein, die Plugins nach dem Login an dem Server zu laden. Außerdem dürfen nur die Plugins, die in einer Domain zugelassen werden, geladen werden.

Die bisherige Implementierung der Plugin-Verwaltung benutzt COM-Interfaces und soll jetzt auf C#.Net umgestellt werden.

Elektra basiert auf .Net-Framework 4. Hier gibt es zwei Möglichkeiten: Managed-Extension-Framework (MEF) [3] und die Standardimplementation in C#.

Zum Vergleich wurde mit jeder Methode ein Testprojekt angelegt. Das Testprojekt ist, wie Elektra, eine Windows-Forms-Anwendung. Das Projekt hat eine Reihe an Plugins, die jeweils ein Windows-Forms-Control liefern, das in einem Haupt-Fenster angezeigt werden kann.

Die Testprojekte zeigen, dass sich bei der geringen Anzahl von Plugins in Elektra der Programmieraufwand zwischen MEF und der C# Standardimplementation nur gering unterscheidet. Ein Vorteil von MEF ist, dass die Plugins anhand einer Annotation in der aufzurufenden Klasse ausgewählt werden können. In der Standardimplementation muss die DLL anhand ihres Dateinamens geladen werden. Praktisch ist der Aufwand dadurch in diesem Fall nicht geringer.

Wird die .Net-Version von Elektra in der Zukunft auf .Net-Core aktualisiert, dann könnte MEF nicht mehr verwendet werden.

Da die Vorteile von MEF für die kleine Menge an Plugins sehr gering sind, ist es sinnvoll die C# Standardimplementation zu verwenden.

Ausblick

Die Entscheidung über das Framework steht noch aus, die Entscheidung für die C# Standardimplementation ist aber sehr wahrscheinlich.

Im nächsten Schritt muss ein Umbauplan zum Reengineering des Plugin-Systems angefertigt werden. Im Anschluss daran wird dieser umgesetzt.

Problematisch ist hier, dass Elektra mit SVN versioniert wird. Daher kann, aufgrund von schlechtem Merging-Verhalten, der Umbau nicht in einer separaten Branch stattfinden. Der Umbau muss so geplant werden, dass Elektra zu jedem Zeitpunkt lauffähig ist.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Christian Wagner. *Model-Driven Software Migration: A Methodology: Reengineering, Recovery and Modernization of Legacy Systems*. Springer Fachmedien Wiesbaden, 2014.
- [3] Genevieve Warren, Saisang Cai, et al. Managed Extensibility Framework (MEF). <https://docs.microsoft.com/de-de/dotnet/framework/mef/>, 11 2021.

Fluchtboote mit Satellitenbildern finden

Daniel Ketterer

Markus Enzweiler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei dem AI4EO Zukunftslabor (TUM/DLR), Ottobrunn

Motivation

Immer noch ertrinken viele Migranten auf dem Weg nach Europa im Mittelmeer. Um Menschen auf der zentralen Mittelmeerroute vor dem Ertrinken zu retten muss ein etwa 175 000 km² großer Bereich regelmäßig überwacht werden. Künftig könnten hier auch Satellitenbilder eingesetzt werden. Während schon automatische Detektoren für Schiffe auf Satellitenbildern existieren, muss noch gezeigt werden, dass diese sich auch für Fluchtboote einsetzen lassen. Im Rahmen dieser Masterarbeit wird ein neuer Datensatz aus Satellitenbildern mit echten Fluchtbooten erstellt und daher eine zielgerichtete Auswertung ermöglicht. Darüberhinaus werden Referenzdetektoren auf externen Datensätzen trainiert, um zu zeigen, dass Modelle, die mit Bildern ähnlicher Boote trainiert wurden, in der Lage sind, echte Fluchtboote zuverlässig zu erkennen.

Datensatz

Ausgangspunkt dieser Arbeit ist die Sammlung von Position und Zeitpunkt von Sichtungen und Rettungsaktionen im Mittelmeer. Die teilweise öffentlichen Quellen dieser Daten sind verschiedene Organisationen, die Seenotrettung betreiben oder betrieben haben: Ärzte ohne Grenzen, Sea-Eye e.V. und Sea-Watch e.V.. Die verwendeten Satellitenbildern werden von der Firma Planet Labs bezogen. Deren PlanetScope Programm bietet einen guten Kompromiss aus kurzen Beobachtungsabständen, hoher Auflösung, Verfügbarkeit von Bildern über dem Meer und historischem Archiv. Die Kleinsatelliten befinden sich in einem sonnensynchronen Orbit und beobachten das Interessengebiet im Mittelmeer jeden Tag etwa um 10:30 vormittags. Die Satellitenbilder haben eine Auflösung von 3m pro Pixel. Da die historischen Bootsichtungen zeitlich nicht immer mit den Bildern übereinstimmen, muss mithilfe der Planet-Explorer-Vorschau jedes Boot händisch gefunden werden und das passende Satellitenbild ausgewählt werden. Je größer der zeitliche Abstand desto weiter entfernt wird auch nach passenden Booten gesucht.



Abb. 1: Rettungssituation im Januar 2018 zwischen Libyen und Malta; die Aquarius von Ärzte ohne Grenzen rettet 99 Menschen aus mehreren Kunststoffbooten. In Grün die Annotation und in Rot die Detektion mit einem Faster-R-CNN (Airbus-Datensatz). [1]

Die Satellitenbilder liegen zunächst nicht als „Bilder“ im gewöhnlichen Sinne vor, sondern müssen zuerst von Strahlungsdichte in Reflexionsgrad umgerechnet werden. Diese *Top-of-atmosphere-reflectance*-Werte werden dann mit einer statischen händisch definierten Farbkurve in natürliche Farben umgewandelt.

Annotation

Um nun einen Datensatz zu erzeugen, werden die Boote von Hand mit Polygonen und Boxen annotiert. Zu jedem Objekt werden auch noch Attribute zum Vorhandensein von Kielwasser, weißer Wellen in der Umgebung und ob das Boot als Fluchtboot eingeschätzt wird festgehalten. Da auch menschliche Annotatoren Schwierigkeiten haben, bei hohen Wellen die Boote korrekt zu identifizieren, liegt nahe dass dies einen größeren Einfluss auf die Genauigkeit eines Detektionssystems haben könnte. Aus Re-Analyse-Daten des Europäischen Zentrums für mittelfristige Wettervorhersage wird daher zu jedem Bild auch die Wellenhöhe, aus dem ERA5 Modell, hinzugefügt.

				All	Imgs w/ high waves	Objs w/o wake	Objs marked refugee
Average Precision	(AP)	@[IoU=0.20:0.40	area= all]	0.53	0.34	-	-
Average Precision	(AP)	@[IoU=0.20	area= all]	0.67	0.37	-	-
Average Precision	(AP)	@[IoU=0.30	area= all]	0.55	0.37	-	-
Average Precision	(AP)	@[IoU=0.20:0.40	area= small]	0.37	-	-	-
Average Precision	(AP)	@[IoU=0.20:0.40	area=medium]	0.66	0.34	-	-
Average Precision	(AP)	@[IoU=0.20:0.40	area= large]	1.00	-	-	-
Average Recall	(AR)	@[IoU=0.20:0.40	area= all]	0.67	0.51	0.56	0.56
Average Recall	(AR)	@[IoU=0.20:0.40	area= small]	0.37	-	0.37	0.37
Average Recall	(AR)	@[IoU=0.20:0.40	area=medium]	0.82	0.51	0.78	0.78
Average Recall	(AR)	@[IoU=0.20:0.40	area= large]	1.00	-	-	-
Average Recall	(AP)	@[IoU=0.20	area= all]	0.76	0.56	0.65	0.65
Average Recall	(AP)	@[IoU=0.30	area= all]	0.68	0.56	0.59	0.59

Abb. 2: Die vorläufigen Evaluationsergebnisse auf dem neuen Datensatz für das Faster-R-CNN, das mit den Airbusdaten trainiert wurde. Objekte kleiner 5² Pixeln gelten als small, zwischen 5² und 30² als medium und größer 30² als large. [1]

Der so zusammengestellte vorläufige Datensatz umfasst 154 Bilder mit 496 Objekten. 165 Objekte haben eine Fläche kleiner als 5² Pixel, 317 sind zwischen 5² und 30² Pixel und nur eines ist größer als 30².

Referenzdetektoren

Zwei verschiedene Deep-Learning-Netze werden als Referenzdetektoren auf jeweils zwei externen Schiffsdatensätzen trainiert. Dabei wird als einstufiges Netz ein RetinaNet und als zweistufiges Verfahren ein Faster-R-CNN verwendet. Als Trainingsdatensätze werden der Airbus-Schiffsdatensatz [2] und die Bilder mit Schiffen aus dem AI-TOD-Datensatz [3] verwendet. In allen Kombinationen werden typische Augmentierungstechniken verwendet und die Ankervorschläge mit Hilfe einer Datensatzanalyse angepasst.

Metriken

Die Evaluation der Detektoren findet auf dem neuen Datensatz mit den COCO-artigen Metriken statt. Es wird die Trefferquote (engl. recall) von Fluchtbooten und die Trefferquote unter Berücksichtigung des Kielwasserattributs analysiert. Darüber hinaus wird Genauigkeit (engl. precision) in Abhängigkeit von der Wellenhöhe untersucht, um Aussagen über die Zuverlässigkeit bei hohem Seegang treffen zu können. Letztlich sind aber auch Trefferquote und

Genauigkeit für unterschiedliche Objektgrößen von großem Interesse. Im Kontext der Fluchtbootdetektion spielt es keine große Rolle, ob eine Box genau über einem Objekt ist, sondern ob das Objekt überhaupt gefunden wird. Daher wird im Gegensatz zur COCO-Challenge nicht eine *Intersection over Union (IoU)* von 0.5 verwendet, sondern kleinere Werte zwischen 0.2 und 0.4.

Ergebnisse

Die hier präsentierten Ergebnisse müssen als vorläufig betrachtet werden. Abbildung 2 zeigt die Ergebnisse für das Faster-R-CNN das mit dem Airbus-Datensatz trainiert wurde. Wie zu erwarten steigt die Trefferquote je größer die Objekte sind. Lässt man mehr Spielraum bei der IoU steigt die Trefferquote auch. Der Einfluss der Wellenhöhe zeigt sich wie auch beim manuellen Annotieren. Auf Bildern mit einer Wellenhöhe größer 1m (N=8) sinkt die Genauigkeit deutlich und auch die Trefferquote sinkt. Die zentrale Frage, ob Fluchtboote gefunden werden können, kann bedingt positiv beantwortet werden. Mit einer Trefferquote von 0.65 werden sie bei einer geringen IoU gefunden. Mit dieser Arbeit konnte also gezeigt werden, dass Deep-Learning-Modelle, die auf externen Schiffsdatensätzen trainiert wurden, auch Fluchtboote detektieren können. Es wurden aber auch die Grenzen und Bedingungen, unter denen diese Vorhersagen verlässlich sind, aufgezeigt.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Airbus Defence-and Space. Airbus Ship Detection Challenge. <https://www.kaggle.com/c/airbus-ship-detection>, 2018.
- [3] J. Wang, W. Yang, H. Guo, R. Zhang, and G.-S. Xia. Tiny Object Detection in Aerial Images. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 3791–3798. Engineers, Institute of Electrical and Electronics, 2021.

Bewertung von IT-Unterstützungstools sowie von aktuellen Trends für Geschäftsprozesse in der Telekommunikation

Marei Kopf

Thomas Rodach

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Detecon International GmbH, Köln

Einleitung

Mit dem Ausbau der Digitalisierung in den vergangenen Jahren und auch in Zukunft steht die Telekommunikationsbranche vor der Herausforderung und der Chance sich neu zu definieren. Neben der herkömmlichen Produktpalette der Sprach- und Datenkommunikation entstehen potenzielle neue Geschäftsmodelle. Aufgrund der optimalen Ausgangsposition, einer bereits existierenden umfassenden Datensammlung, kann die Telekommunikationsbranche ihre Wettbewerbschancen ausbauen. Die Umsetzung wird zwangsläufig durch Business Process Management Tools unterstützt. Diese Bachelorarbeit zielt darauf ab, die Herausforderungen und die daraus resultierenden Trends in der Telekommunikationsbranche zu identifizieren. Anschließend werden gängige Business Process Management Tools auf Basis der identifizierten Herausforderungen mit einem Punktesystem bewertet.

Herausforderungen und die daraus resultierenden Trends in der Telekommunikationsbranche

Wie in jedem anderen Wirtschaftszweig auch muss sich ein Unternehmen der Telekommunikationsbranche von den übrigen Wettbewerbern abheben, um sich zu differenzieren. In der Telekommunikationsbranche werden jedoch meist sehr ähnliche Produkte und Dienstleistungen angeboten. Business Process Management Tools können dabei helfen, Standardprozesse reibungslos abzuwickeln und somit den Fokus auf die Kunden zu setzen. Um Kunden optimal und langfristig zu binden, erwarten diese einen guten Kundenservice.

Mit Hilfe von Business Process Management Tools fällt es den Mitarbeitern leichter, Entscheidungen im Sinne des Kunden zu treffen und so ein positives Kundenerlebnis zu schaffen. Schlechter Kundenservice führt häufig zu Kundenabwanderung. Dies kann durch die Verwendung von Customer Journey Mapping innerhalb eines Business Process Management Tools reduziert werden. Hiermit lässt sich nachvollziehen, wie der Service ankommt, an welchen Stellen es Verbesserungsbedarf gibt und wo ein Unternehmen dem Kunden mit individuellen Angeboten zuvorkommen kann. Zunehmende Technologieentwicklungen bieten neue Möglichkeiten, jedoch wächst hierdurch ebenso die Komplexität für die Telekommunikationsbranche an. Business Process Management Tools können den Umgang mit den Datenmengen erleichtern, in dem analysiert, dokumentiert und den richtigen Mitarbeitern zur Verfügung gestellt wird. In der Telekommunikationsbranche ist die Zusammenarbeit zwischen den Mitarbeitern entscheidend für den Erfolg oder Misserfolg eines Unternehmens. Umso dringlicher ist es, effizient zusammenzuarbeiten und Prozesse kontinuierlich zu optimieren, was durch Business Management Tools unterstützt werden kann. [3]

Softwareauswahl

Im Rahmen dieser Arbeit werden Business Process Management Tools vorgestellt, welche in der Telekommunikationsbranche verwendet werden. Im Folgenden werden zwei der ausgewählten Softwares kurz vorgestellt. Die unter SAP vertriebene Signavio Software bietet ein cloudbasiertes Managementsystem zum Modellieren, Optimieren und Ausführen von Prozessen. Die vier miteinander verzahnten Softwarekomponenten (siehe Abbildung 1) unterstützen hierbei vollumfänglich.



Abb. 1: Business Transformation Suite von Signavio [2]

ARIS Enterprise ist eine Prozessmanagementsoftware von der Software AG, welche dabei unterstützt KPIs untereinander auszutauschen, darzustellen und Prozesse zu optimieren. In der Abbildung 2 wird der Aufbau von ARIS dargestellt.

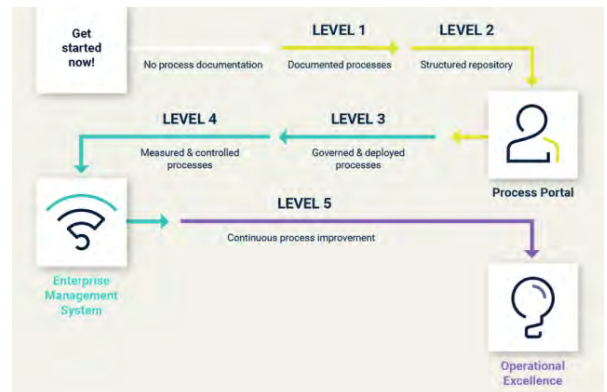


Abb. 2: Übersicht ARIS Enterprise [1]

Ausblick

Unternehmen suchen fortwährend Möglichkeiten, sich zu verbessern und zu vergleichen, zu diesem Zweck werden Business Process Management Tools langfristige Unterstützung bieten. Sie werden sich mit den Unternehmen weiterentwickeln und neue Chancen. Im weiteren Verlauf der Arbeit wird auf Basis der qualitativen Forschungsmethode und den Herausforderungen ein Bewertungsbogen für die ausgewählten Business Process Management Tools erstellt und anschließend ausgewertet.

Literatur und Abbildungen

- [1] Software AG. Mit ARIS auf ein höheres Unternehmensniveau. [https://softwareag.scene7.com/is/image/softwareagprod/ARIS-overview-diagram?\\$MediaPreset\\$](https://softwareag.scene7.com/is/image/softwareagprod/ARIS-overview-diagram?$MediaPreset$), 2021.
- [2] Wirtschaftsforum. de. Die Business Transformation Suite von Signavio: Kein anderes Unternehmen bietet die Module in dieser Kombination an. https://www.wirtschaftsforum.de/fileadmin/wirtschaftsforum/unternehmen/191512_signavio-gmbh/signavio-module.jpg, 11 2017.
- [3] Lucas de Boer. Die 5 größten Herausforderungen für Unternehmen der Telekommunikation. <https://www.signavio.com/de/post/5-groessten-herausforderungen-unternehmen-telekommunikation/>, 04 2019.

Digitalisierung im Controlling - Stand und Potentiale

Valdrin Krasniqi

Catharina Kriegbaum-Kling

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Die Digitalisierung ist ein ständig fortschreitender Prozess. Die Geschwindigkeit, mit der die Digitalisierung an Bedeutung gewinnt, ist enorm, und obwohl sich Unternehmen schon seit Jahren mit der Digitalisierung beschäftigen, stand sie noch nie sehr im Mittelpunkt wie heute. Eine Studie der Boston Consulting Group hat ergeben, dass vor allem deutsche Unternehmen zu wenig in die Digitalisierung investieren, um wettbewerbsfähig zu bleiben. [1] Erste empirische Studien zeigen, dass der Einsatz moderner Digitalisierungstechnologien einen positiven Einfluss auf den Unternehmenserfolg hat. Die Digitalisierung wirkt sich auf zahlreiche Bereiche der Wertschöpfungskette eines Unternehmens aus, von der Beschaffung über die Produktion bis zum Vertrieb. Das Controlling wird von den neuesten Technologien der Digitalisierung direkt oder indirekt betroffen und ist eine der wichtigsten Ansprechpartner für das Management eines Unternehmens. Wenn es um die Digitalisierung im Controlling geht, muss sich das Controlling selbst den neuen Herausforderungen stellen, die Chancen der Digitalisierung nutzen und durch Standardisierung und Automatisierung effizienter werden. Im Controlling können Technologien wie beispielsweise integrierte ERP-Systeme oder Big Data Analytik als Schlüsseltechnologien der digitalen Transformation identifiziert werden. Durch die Automatisierung verringern sich die Fehlerquote und steigern sowohl die Effizienz als auch die Effektivität des Controllings. Komplexe Datenanalysen, die zuvor Wochen dauerten, können jetzt in Echtzeit erstellt werden. Neben der Schnelligkeit ermöglicht der Einsatz von Big Data Analysen eine vielschichtige Analyse von Daten und die Ermittlung von Zusammenhängen, die ohne neue Technologie nicht offensichtlich wären. Mit dem Einsatz geeigneter Technologien können Unternehmen Controlling Prozesse optimieren und dadurch eine neue Auswertung erschließen. Das steigert den Nutzen des Controllings und sichert seine Existenz im Unternehmen. Die Digitalisierung im Controlling wirkt sich somit in Summe somit positiv auf den Unternehmenserfolg aus. [2]

Problemstellung

Die Industrie hat sich seit dem 19. Jahrhundert ständig fortentwickelt. So wurde der Wandel durch die Erfindung der Eisenbahn, die Nutzung der Elektrizität oder den ersten Computer geprägt. [6] Und auch heute noch befindet sich die Industrie in einem Umbruch. Begriffe wie Smart Home, Internet der Dinge oder Big Data sind inzwischen aus dem Alltag nicht mehr wegzudenken. Die durch die Digitalisierung entstehenden Überlegungen betreffen nicht nur die Entwicklung von smarten Produkten, sie betreffen auch das gesamte Unternehmen mit seinen einzelnen Abteilungen und deren Prozesse. Im Zeitalter von Industrie 4.0 beschränkt sich die Wertschöpfung des Unternehmens nicht mehr auf die Produktion, sondern wird durch die zielgerichtete Nutzung von Daten erweitert. Um den angestrebten Erfolg zu erreichen, muss sich das gesamte Unternehmen in Echtzeit mit der Auswertung von großen Datenmengen und der anschließenden Analyse beschäftigen. Eine wesentliche Rolle spielt das Controlling, das als Informationslieferant für das gesamte Unternehmen dient. Damit das Controlling mit der digitalen Entwicklung Schritt halten kann ergeben sich viele neue Möglichkeiten und Herausforderungen.

Zielsetzung

Die Relevanz des Forschungsbedarfs zum Einfluss der Digitalisierung im Controlling wird im Hinblick auf die aufgezeigte Problemstellung deutlich gemacht. Unternehmen sind bestrebt, die Gestaltung der unternehmensspezifischen Geschäftsprozesse zu optimieren und vor allem die Informationsflüsse zu verbessern, die durch die Digitalisierung entstehen können. [4] Für das unternehmensinterne Controlling ergibt sich in diesem Sinne die Anforderung, sich mit neu entstehenden Handlungsfeldern in bereits bestehenden Strukturen und Prozessen auseinanderzusetzen. Aufgabe des Controllings ist es, das Management der Digitalisierung

zu übernehmen sowie die digitale Transformation im Unternehmen zu unterstützen. Zum anderen muss das Controlling die Potenziale der Digitalisierung nutzen und sich zum digitalen Controlling entwickeln. Im Mittelpunkt dieser Arbeit steht die Analyse des Wandels des Controllings im Zeitablauf und die Rolle, die die Digitalisierung dabei spielt. Um die weitere Entwicklung des Controllings durch die Digitalisierung abbilden zu können, ist zunächst der aktuelle Stand des digitalen Entwicklungsstandes des Controllings zu erfragen, was eines der Ziele dieser Arbeit ist. Mit den daraus resultierenden Beiträgen ist es nun möglich, theoretische Implikationen für die Forschungsfrage zu formulieren.

Definition der Digitalisierung

Digitalisierung bedeutet, dass Prozesse durch den Einsatz digitaler Technologien und digitalisierter Daten ermöglicht oder verbessert werden. Die Digitalisierung erhöht die Produktivität und Effizienz und senkt gleichzeitig die Kosten. Die Digitalisierung verbessert einen oder mehrere bestehende Geschäftsprozesse, verändert oder transformiert sie aber nicht. Das heißt, sie macht einen Prozess von einem menschengesteuerten Ereignis oder einer Reihe von Ereignissen zu einem softwaregesteuerten Prozess. [5]

Big Data

Der Ursprung des Begriffs Big Data ist heute noch unklar und hat im Laufe der Geschichte verschiedene Bedeutungen angenommen. Der Begriff wurde erstmals im Jahre 1998 von John R. Mashey im Zusammenhang mit den wachsenden Datenmengen erwähnt. Schon damals erkannte er, dass das wachsende Datenvolumen nicht parallel mit der notwendigen Infrastruktur wach-

sen würde und dass Netzwerke und Speichersysteme daher in Zukunft einen Engpass darstellen könnten. Im Grunde genommen ist Big Data nichts Neues oder etwas, das erst in den letzten zwei Jahrzehnten entstanden ist. Im Laufe der Jahrhunderte haben die Menschen versucht, Datenanalyse und Analysetechniken zur Unterstützung ihrer Entscheidungsfindung einzusetzen. In den letzten zwei Jahrzehnten haben sich jedoch das Volumen und die Geschwindigkeit, mit der Daten erzeugt werden, verändert - jenseits des menschlichen Vorstellungsvermögens. Um die Entwicklung im Laufe der Zeit zu veranschaulichen, kann die Entwicklung von Big Data grob in drei Hauptphasen unterteilt werden. Jede Phase hat ihre eigenen Merkmale und Fähigkeiten. Um den Zusammenhang von Big Data zu verstehen, ist es wichtig zu wissen, wie jede Phase zur heutigen Bedeutung von Big Data beigetragen hat. [3]

Ausblick

Durch die Digitalisierung wird sich das Controlling als Unternehmensfunktion, wie wir es heute kennen, in vielerlei Hinsicht grundsätzlich verändern. Nicht nur das Controlling wird sich zu einem kaufmännischen Geschäftspartner und Wegbereiter entwickeln, der zum Beispiel in Zusammenarbeit mit Data Scientists statisch abgesicherte Prognosen zur Geschäftsentwicklung erstellt. Selbst klassische Controlling-Aufgaben werden sukzessive auf leistungsfähige, lernende IT-Systeme verlagert, die über Cloud-Lösungen auch für KMUs kostengünstiger zugänglich sind. So werden zentrale Controlling-Prozesse wie das Berichtswesen oder die Planung noch stärker als heute und in Zukunft wahrscheinlich vollständig durch IT-Lösungen unterstützt und gesteuert. Das Controlling wird die Interpretationsfähigkeit behalten, aber nur, wenn es die von den Maschinen verwendete Datenanalyse versteht.

Literatur und Abbildungen

- [1] Bcg Bcg. Ein Viertel der Unternehmen droht bei der Digitalisierung den Anschluss zu verlieren. <https://www.bcg.com/de-de/press/21april2017-beyond-the-hype>, 2017.
- [2] A Bhimani. *Accounting and Business Research*. Bhimani, A., L. Willcocks, 2014.
- [3] Bigdataframework Bigdataframework. Where does 'Big Data' come from? <https://www.bigdataframework.org/short-history-of-big-data/>, 2019.
- [4] Keimer Imke. *Die Digitalisierung der Controlling-Funktion*. Prof. Dr. Imke Keimer, Prof. Dr. Ulrich Egle, 2020.
- [5] Lubert Stefan. Was ist Digitalisierung? <https://www.bigdata-insider.de/was-ist-digitalisierung-a-626489/>, 2019.
- [6] W. Frick Thomas. Industrie 1.0 bis 4.0 – Industrie im Wandel der Zeit. <https://industrie-wegweiser.de/von-industrie-1-0-bis-4-0-industrie-im-wandel-der-zeit/>, 2017.

Untersuchung der WLAN-Konfigurationsschnittstelle eines Steuerungssystems für Brenner unter dem Aspekt der Informationssicherheit

Eva Kuffler

Reinhard Keller

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Karl Dungs GmbH & Co. KG, Urbach

Einleitung

Diese Arbeit wurde im Rahmen eines Entwicklungsprojekts für ein neues Steuerungssystem für Brenner bei der Firma Dungs erstellt. Die Firma Dungs ist ein weltweit führendes Zulieferunternehmen der Heizungs- und Prozesswärmeindustrie mit Spezialisierung auf Bauelemente und Systeme der Feuerungstechnik. Bei dem Steuerungssystem für Brenner handelt es sich um einen mikroprozessorgesteuerten Feuerungsautomaten, der umfassend die Aufgaben der Steuerung und Überwachung von Gas-Feuerungsanlagen wahrnimmt. Dies umfasst alle Prozessschritte zum Regeln, Zumessen und Modulieren von Brenngas und Verbrennungsluft, sowie die Zündung und Überwachung der Flamme. Je nach Einsatzbereich dienen diese Feuerungsanlagen zur Beheizung von großen Gebäuden, etwa Krankenhäusern oder Messehallen, oder zur Bereitstellung von Prozesswärme für gewerbliche oder industrielle Fertigungsprozesse, z.B. bei der Röstung von Kaffeebohnen oder bei der Herstellung von Glas, Papiertaschentüchern oder vielen weiteren Produkten.

Motivation

Das Steuerungssystem für Brenner verfügt standardmäßig über eine Mensch-Maschine-Schnittstelle, die allerdings nur mit einem kleinen Display und wenigen Bedienknöpfen ausgestattet ist. Um während der Inbetriebnahme, Konfiguration und Wartung einen praktischeren Zugang zum System zu ermöglichen, soll das Steuerungssystem für Brenner zusätzlich eine Konfigurationsschnittstelle bereitstellen. Mit der Verbindung durch die Schnittstelle kann über das Host-System eine Konfigurationssoftware verwendet werden, die für den Benutzer (z.B. Service-Techniker) komfortablere Eingabemöglichkeiten bietet.

Die Verbindung zwischen dem Steuerungssystem für Brenner und dem Host-System kann drahtlos hergestellt werden. Für diese Verbindung wurden unter Verwendung verschiedener drahtloser Verbindungsprotokolle mehrere Szenarien untersucht und unter dem Aspekt der Sicherheitskonzepte der IEC 62443 bewertet. Abschließend wurde eine umfassende Sammlung möglicher Sicherheitsmaßnahmen erstellt.

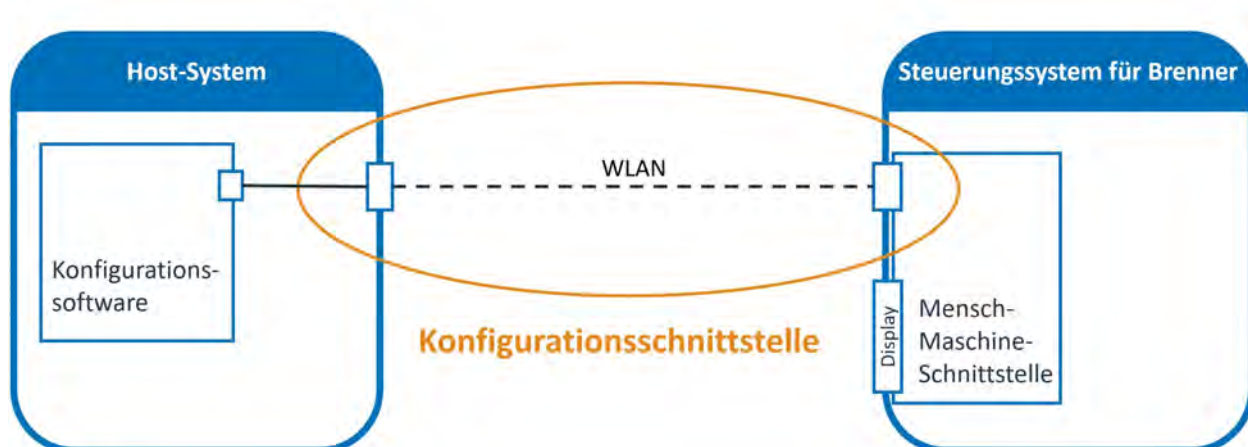


Abb. 1: WLAN-Verbindung gemäß Standard 802.11 als Konfigurationsschnittstelle [2]

Technische Gegebenheiten

Als physikalische Basis der drahtlosen Verbindung wurde eine lokale Netzwerkverbindung gemäß WLAN-Standard IEEE 802.11 untersucht. In der typischen Einsatzumgebung von Feuerungsanlagen kann normalerweise nicht auf eine bestehende WLAN-Infrastruktur zurückgegriffen werden. Allerdings wurde bei der Untersuchung vorausgesetzt, dass sowohl das Host-System als auch das Steuerungssystem für Brenner mit einer entsprechenden WLAN-Schnittstelle ausgestattet sind (vgl. Abb. 1). Die WLAN- und Sicherheitsfähigkeiten des Gesamtsystems sind stark abhängig von den Möglichkeiten der eingesetzten Mikrochips.

Industrial Security gemäß IEC 62443

Während Sicherheitsanforderungen in Bezug auf die Funktionssicherheit (Safety) in Industrieanlagen traditionell eine wichtige Rolle spielen, gewinnen neuerdings Sicherheitsanforderungen in Bezug auf die Informationssicherheit (Security) zunehmend an Bedeutung. Die IEC 62443 als junge internationale Normenreihe betrachtet umfassend die Informationssicherheit in industriellen Automatisierungs- und Steuerungssystemen. Zur Abgrenzung gegenüber der Informationssicherheit in Büro-Netzen wird hier der Begriff Industrial Security verwendet. Denn obwohl zunehmend Technologien und Protokolle aus der klassischen Informationsverarbeitung in den industriellen Bereich übernommen werden, sind doch die Voraussetzungen in vielerlei Hinsicht andere (vgl. Abb. 2).

Bewertung und Maßnahmen

Die Bewertung der untersuchten Verbindungsszenarien basierte einerseits auf den technischen Einsatzmöglichkeiten der unterschiedlichen, zur Verfügung stehenden Protokolle, andererseits auf den Requirements und Konzepten, welche die IEC 62443 für Industrial Security fordert. Die Sicherheitsziele entsprechen dabei im Wesentlichen den aus der klassischen IT-Sicherheit bekannten, nämlich Vertraulichkeit, Integrität und Verfügbarkeit. Allerdings wird im industriellen Bereich die Priorisierung dieser Ziele verändert und alles auf

eine möglichst hohe Verfügbarkeit hin ausgerichtet. Die zentrale Verteidigungsstrategie der IEC 62443 beruht auf dem Defense-in-Depth-Prinzip [1]. Es werden mehrere unterschiedliche Aufdeckungs- und Schutzmaßnahmen verwendet, um Verteidigungsringe gegen Angreifer rund um die zu schützenden Komponenten einzurichten. Dabei wird das Ziel verfolgt, das Vorankommen eines Angriffs zu verhindern bzw. zu behindern, zumindest jedoch zu entdecken. Solche Maßnahmen werden auf allen Ebenen eingesetzt und betreffen auch die untersuchte Schnittstelle. Sie reichen von physikalischen und logischen Zugangsbeschränkungen bis zur Anwendung geeigneter kryptographischer Verfahren.

Industrial Security	IT Security
Prioritäten	
Verfügbarkeit Integrität Vertraulichkeit	Vertraulichkeit Integrität Verfügbarkeit
Eigenschaften	
geforderte Verfügbarkeit	
hochverfügbar, > 99,99%	99 % ausreichend
Neustart	
schwierig	möglich
Patch Management	
große Herausforderung	automatisiert möglich
Lebenszeit der Hardware	
bis zu 30 Jahren	3 bis 5 Jahre

Abb. 2: Unterschiedliche Anforderungen an die Sicherheit im industriellen Bereich und in Büro-Netzen [2]

Literatur und Abbildungen

- [1] International Society of Automation. *Security for Industrial Automation and Control Systems. ANSI/ISA-62443-1-1-2007*. ANSI/ISA, 2007.
- [2] Eigene Darstellung.

Reduktion der Nacharbeitskosten einer Ventilinsel

Florian Lauterbach

Gabriele Gühring

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Festo SE & Co. KG, Esslingen am Neckar

Ventilinsel

Ventilinseln sind ein Zusammenschluss verschiedener Ventile, die mit einer zentralen Strom- und Luftversorgung angesteuert werden. Ventilinseln sind aus der heutigen Arbeitswelt nicht mehr wegzudenken. Durch Ventilinseln wird Platz und Komponenten eingespart, wenn man im Gegensatz die gleiche Funktion mit einzelnen Ventilen realisieren würde. Ventilinseln haben eine hohe Varianz und können für jeden Einzelfall konfiguriert werden, damit die beste Lösung für ein Problem entstehen kann. [1] Das folgende Bild zeigt eine Ventilinsel.

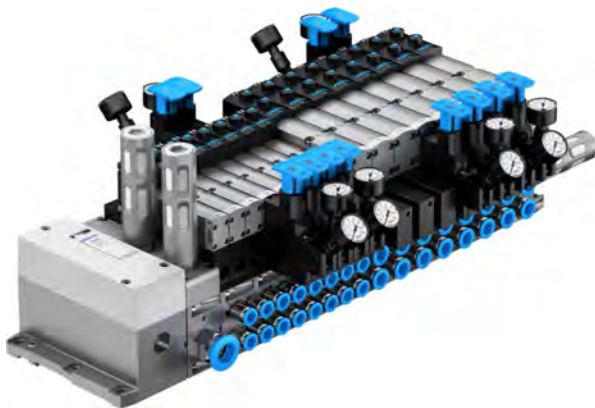


Abb. 1: Ventilinsel der Firma Festo [3]

Qualitätsmanagement

Das Qualitätsmanagement bei der Herstellung von komplexen Produkten, wie Ventilinseln, ist schwierig und aufwendig. In der Produktion müssen alle Produkte, die montiert werden, am Schluss der Montage geprüft werden. Dabei wird überprüft, ob das Produkt alle Anforderungen erfüllt, ob es zum Beispiel die gewünschte Funktion hat oder ob es im Fall der Pneumatik die vorgegebene Leckage erfüllt. Wenn das Produkt eine Anforderung nicht erfüllt, muss es nachgearbeitet werden, was bedeutet, dass die Fehlerursache bei dem Produkt gefunden werden muss. Dazu muss der Mitarbeiter oder die Mitarbeiterin die Fehlerursache erst einmal eingrenzen. Hat man komplexere Ventilinseln kann man den Fehler nicht so einfach eingrenzen, denn jeder Anschluss oder auch jedes Ventil könnte die Fehlerursache sein. Diese Fehlersuche ist zeitaufwendig und verursacht zusätzliche Kosten. Deshalb wird versucht, die Nacharbeit möglichst zu vermeiden oder wenn sie trotzdem auftritt, die Zeiten und Kosten dafür möglichst gering zu halten. Auf dem nachfolgenden Bild ist gut zu erkennen, wie die Kosten der Nacharbeit im Laufe des Herstellungsprozesses steigen.

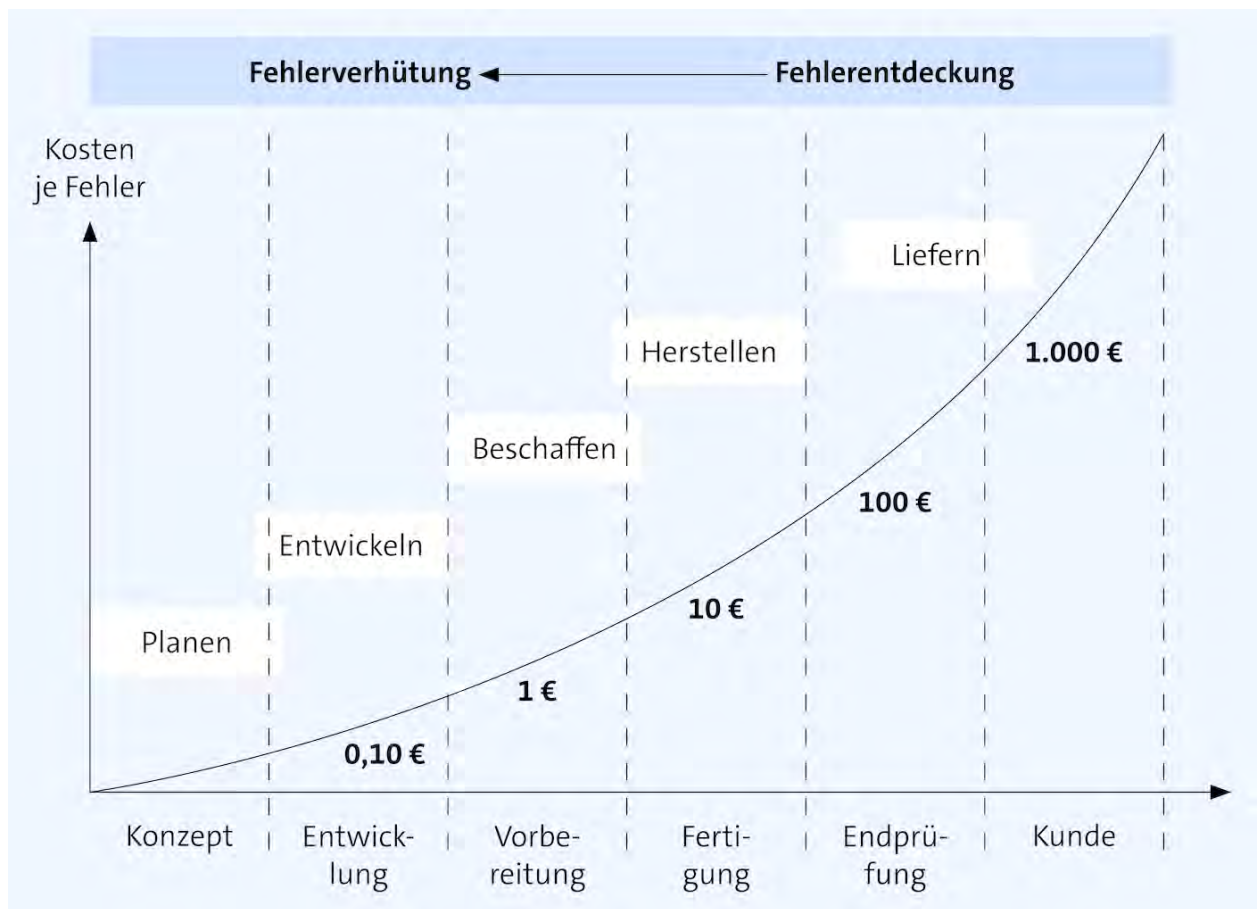


Abb. 2: Fehlerkosten [4]

Um diese Kosten so gering wie möglich zu halten, gibt es in vielen Betrieben eine Qualitätssicherung. Diese sucht nach der Fehlerursache und versucht diese so schnell wie möglich zu beheben und versucht Vorkehrungen zu treffen, damit dieser Fehler nicht noch mal vorkommen kann. Damit die Qualitätssicherung effizient arbeiten kann, benötigt sie möglichst genaue Daten zu auftretenden Fehlern. Diese Informationen kommen meistens aus zwei verschiedenen Quellen,

- Informationen, die beim Prüfen entstanden sind und

- Informationen, welche Bauteile bei der Nacharbeit getauscht wurden.

Das Sammeln und Auswerten dieser Informationen stellt eine Herausforderung dar, weil Methoden gefunden werden müssen, wie diese Daten gesammelt werden können und sie abgespeichert werden müssen, damit mit ihnen effizient gearbeitet werden kann. Das ist bei einem einfachen Produkt schon schwierig und umso komplexer das Produkt ist, umso schwieriger wird das Sammeln der Daten, weil mehrere Fehlerursachen einen Fehler verursachen können. [2]

Literatur und Abbildungen

- [1] Frank Ebel, Siegfried Idler, Georg Prede, and Dieter Scholz. *Pneumatik und Elektropneumatik*. Bildungsverlag EINS, 2017.
- [2] Tilo Pfeifer and Robert Schmitt. *Masing Handbuch Qualitätsmanagement*. Carl Hanser Verlag GmbH & Co. KG, 2021.
- [3] Festo und Co. KG SE. Universelle Ventilinseln. https://www.festo.com/de/de/c/produkte/fabrikautomation/ventile-und-ventilinseln/ventilinseln/universelle-ventilinseln-id_pim275/, 2021.
- [4] Wiwiweb. Bedeutung, Funktion und Aufgaben von Qualitätsmanagementsystemen. <https://www.wiwiweb.de/qualitaetsmanagement/einfluss-des-qualitaetsmanagements-auf-das-unternehmen-und-die-funktionsfelder/bedeutung-funktion-und-aufgaben-von-qualitaetsmanagementsystemen.html>, 2021.

Automated reverse engineering of automotive bare metal firmware

Marvin Louis

Dominik Schoop

Department of Computer Science and Engineering, Esslingen University

Work carried out at Robert Bosch GmbH, Abstatt

Motivation

Today's modern cars contain hardware components such as sensors, microcontrollers and microprocessors. They are used, for example, to control the telematics or to trigger the airbag in the event of an accident. Therefore, the firmware that configures these devices and defines their behaviour is flashed on the hardware. If a firmware contains vulnerabilities it can lead to an incident, which endangers the life of a person or costs money. For this reason, companies who create these automotive firmwares want to know if their firmware product is secure. Proving the absence of vulnerabilities is often done through code analysis, as it can reveal security flaws within the source code. Although firmware is built upon source code, it often contains more information, such as the configuration of the underlying hardware, which can also lead to vulnerabilities. As the firmware is built by combining many source code and configuration files that are scattered across different repositories and development teams, the idea of this thesis is to analyse the result of the build chain. Furthermore, the flashable firmware file is the source of truth regarding the configuration and behaviour of the microcontroller. Through automation of the reverse engineering process to deconstruct the firmware, bare metal firmware analysis becomes feasible for larger companies. This can help them, for example, to comply with parts of the ISO/SAE 21434 standard, which demands a vulnerability scan in the development life cycle. Moreover, the goal of making the firmware content accessible may generate more possibilities for bare metal firmware analysis in the future.

Firmware reverse engineering

Reverse engineering is the deconstruction of an existing product to infer the functional features and manufacturing principles. [6] The term reverse engineering is not only used for engineered physical products, but also for IT in the form of software or firmware

reverse engineering. If the definition is brought into this IT context, we can define firmware or software as a product, the whole software development process as the manufacturing principle, and the inner workings of the firmware or software as functions. Hackers and security researchers use various tools to reverse engineer a firmware with the goal of finding vulnerabilities. The most versatile tools for reverse engineering are IDA Pro [4], Ghidra [3] and BinaryNinja [5] as they are all able to disassemble the machine-executable code and provide an interactive use. These tools bring disassembly to intermediate languages, which helps to browse and annotate the firmware code in a higher C-like representation.

Automotive bare metal firmware

TCUs (Telematic Control Unit) and ECUs (Electronic Control Unit) in the automotive area use system architectures that are either microcontroller or microprocessor based. Because microcontrollers are used for small and efficient use-cases like charging devices and robots it combines memory peripherals, inputs and outputs into one chip. The small form factor leads to a constrained amount of available flash memory and computing power. Hence, firmware for microcontrollers uses the hardware interfaces directly to save memory space and to keep the overhead as small as possible. This type of firmware is called bare metal as it is tailored for a specific hardware and use-case. In contrast to that are systems based on microprocessors that usually use traditional operating systems like Linux.

Challenges in reverse engineering bare metal firmware

Reverse engineering of bare metal firmware raises different challenges compared to non bare metal firmware analysis. This is originated from the strong hardware dependency and the monolithic, non standardized software architectures which require much

more information about the device and the software architecture to understand the firmware product. For example, the absence of defined filesystems as in the unix area makes the challenge of finding firmware structures even more difficult.

Automated bare metal reverse engineering

At present there are tools that can automate the analysis and reverse engineering of firmware, but they all expect non bare metal firmware. For example, FACT [2] from Fraunhofer FKIE uses tools from the open-source community such as binwalk [8], cwe-checkers [1] and QEMU [7] to automate most of the non bare metal firmware analysis process. The tools used in FACT are specialized on non bare metal firmware and perform poorly on unspecified monolithic firmware, because the unpacking step, which is a prerequisite for the analysis with FACT, expects specified file formats. The unpacking problem arises from the variety of hardware architectures and non standardized firmware structures. Tools from the open-source community to automate the analysis are rare, because the hardware specifications in the automotive area are mostly documented in datasheets that are under a non disclosure agreement. This means that new procedures and tools need to be found or created to approach the automation of bare metal firmware analysis.

Toward automated bare metal reverse engineering

Unix based firmware defines its structure over filesystems and files, whereas bare metal firmware has custom designed memory layouts. Hence, we propose a procedure as seen in Figure 1 which enriches information pieces related to a firmware with contextual information. This can be an approach to gain firmware properties which can enable statements about the layouts. The procedure proposed in this thesis starts with the extraction of information pieces out of

information artefacts. Information artefacts thereby can be the firmware itself, datasheets, build artefacts or other external resources. In the next step, the assignment of information pieces brings them into a rough context by assigning a category and an asset. Thereby, a category is a classification within a firmware context, and an asset defines a generic definition of a firmware element containing relevant information. The last step, the composition, brings multiple context enriched information pieces together in a computer readable format, called firmware property. Performing the process described for each information artefact helps to build a collection of firmware properties that can enable future tools to automate the analysis of bare metal firmware.

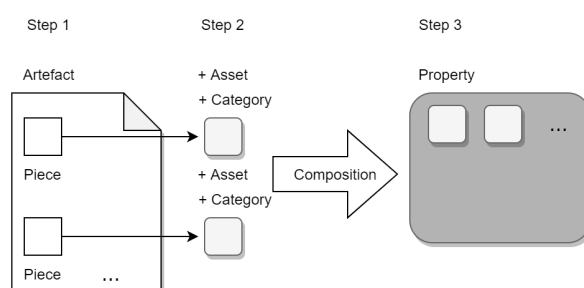


Fig. 1: From an artefact to a firmware property. [9]

Outlook: Creating the bigger picture

By connecting firmware properties we can create a bigger picture of the firmware which leads to a greater understanding of the bits and bytes. With that knowledge the configuration of the hardware or firmware structures can be inferred. This helps separating executable code and data which are found within a firmware. This separation is a prerequisite for code analysis which can partly be executed automatically by existing tools from the open-source community. By executing the code analysis and placing the results into the previously created bigger picture of the firmware, statements about the security can be made.

References and figures

- [1] Fraunhofer FKIE. cwe_checker is a suite of checks to detect common bug classes such as use of dangerous functions and simple integer overflows. https://github.com/fkie-cad/cwe_checker, 2021.
- [2] Fraunhofer FKIE. The Firmware Analysis and Comparison Tool. https://fkie-cad.github.io/FACT_core/, 2021.
- [3] Project Ghidra. A software reverse engineering (SRE) suite of tools developed by NSA's Research Directorate in support of the Cybersecurity mission. <https://ghidra-sre.org/>, 2021.
- [4] Company HexRays. A powerful disassembler and a versatile debugger. <https://hex-rays.com/ida-pro/>, 2021.
- [5] VectorThirtyFive Inc. A New Type of Reversing Platform. <https://binary.ninja/>, 2021.
- [6] Prof Dr Daniel Markgraf. Definition: Reverse Engineering. <https://wirtschaftslexikon.gabler.de/definition/reverse-engineering-45260/version-268557>, 2021.
- [7] Project qemu. A generic and open source machine emulator and virtualizer. <https://www.qemu.org/>, 2021.
- [8] Company ReFirmLabs. Binwalk is a fast, easy to use tool for analyzing, reverse engineering, and extracting firmware images. <https://github.com/ReFirmLabs/binwalk>, 2021.
- [9] Own representation.

Evaluierung und Entwicklung eines Chatbots für die Gästekommunikation im Gastgewerbe

Nico Maag

Andreas Rößler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma TourOnline AG, Wernau (Neckar)

Einleitung

Durch den Trend des Machine Learning hat auch die Bekanntheit von Chatbots einen großen Boost erlangt, da diese Themen für viele in den gleichen unbekannten Kosmos gehören. Ein Chatbot an sich ist lediglich eine Software, die es einem Benutzer ermöglicht, per Sprache oder Text mit einem Computer zu kommunizieren. Wie viel Intelligenz oder auch künstliche Intelligenz schlussendlich dahinter steckt, bleibt den Entwicklern überlassen. Dennoch finden wir sie heutzutage auf Unmengen von Webseiten vor und kommunizieren mit Chatbots auch unbewusst häufiger als wir selbst bemerken. 250 Unternehmen innerhalb der DACH Region wurden im Jahr 2020 befragt, ob sie einen Chatbot verwenden und 37% beantworteten diese Frage mit "Ja"(siehe Abbildung 1). Somit nutzen mehr als ein Drittel der befragten Unternehmen einen Chatbot.

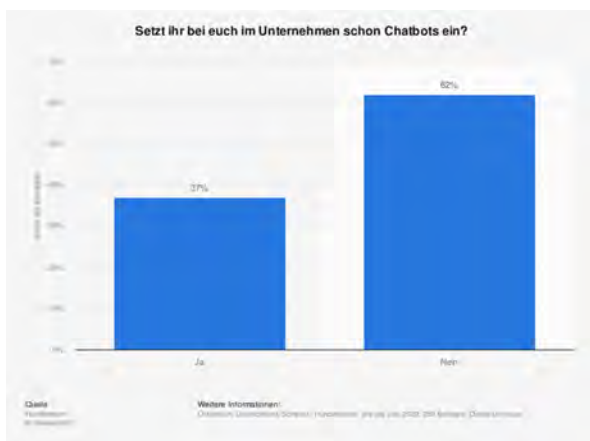


Abb. 1: Umfrage zum Einsatz von Chatbots in Unternehmen in der DACH-Region 2020 [2]

Problemstellung

In der Hotelleriebranche, aber auch im allgemeinen Umgang mit Kunden oder Gästen, gibt es viele Fragen der Kunden. Sehr viele dieser Fragen sind dieselben und wiederholen sich ständig. Kann man diese repetitiven Fragen automatisiert erkennen und beantworten oder auf Informationen hinweisen, sparen sich der Kunde und der Dienstleister viel Zeit und Nerven. Hier kommt ein Chatbot ins Spiel, der Fragen analysieren kann und die Kernaussage der Frage versteht und die entsprechende Antwort darauf liefert. Jedoch hat kaum ein Hotelier die notwendige Ausbildung, noch die technischen Möglichkeiten so einen Chatbot auf seiner Webseite oder in andere Kommunikationskanäle zu integrieren. Das Unternehmen, bei dem die Arbeit geschrieben wird, ist ein Dienstleister in der Hotelleriebranche und hat dementsprechend einige Hotels als Kunden. Da man aus wirtschaftlichen Gründen nicht für jedes einzelne Hotel den Chatbot einpflegen kann, muss man dem Hotelier eine einfache Möglichkeit bieten, ohne technisches Know-How, eine Anbindung an so einen Chatbot zu schaffen und diesen Nutzen zu können.

Zielsetzung

Die Arbeit besteht aus zwei Teilen. Einerseits die Evaluierung der aktuellen Marktsituation, um die bestmögliche Chatbot-Lösung für das Unternehmen zu finden, in Hinsicht auf den Preis, die Funktionalität, Kontingente(Limits) und weitere Parameter. Andererseits soll eine Oberfläche zur Pflege der Wissensdatenbank des Bots, die Logik zur Datenhaltung und Verwaltung im Backend, sowie die Integration in eine mobile Applikation oder in Hotelwebseiten entwickelt werden. Zusätzlich zu der Entwicklung sollen Use-Cases und Mockups erstellt werden, sodass man viele Fälle abdecken kann und die Pflege so leicht wie möglich für die Kunden hält.

Realisierung

Die Marktanalyse besteht daraus, so viele verschiedene Chatbot-Anbieter wie möglich miteinander zu vergleichen und zu testen, um sich für den zu entscheiden, welcher am besten für diesen speziellen Fall geeignet ist. Zusätzlich soll eine Umfrage an die Kunden gesandt werden, um die Interesse an so einem Chatbot zu evaluieren und die häufig gestellten Endkunden-Fragen zu kristallisieren, um Use-Cases zu definieren. Durch diese Use-Cases hat man ein besseres Verständnis für die Anwendungsfälle und kann sinnvollere Mock-Ups erstellen.

Im Vorfeld sollte man die bereits gegebene Softwarearchitektur der Produkte des Unternehmens, bei dem die Arbeit geschrieben wird - TourOnline AG (auch DIRS21) - und deren Namen verstehen bzw. gehört haben.

DIRS21 ist ein Dienstleister in der Hotelleriebranche und bietet viele verschiedene Softwareprodukte an. Für die vorliegende Arbeit sind die folgenden Produkte relevant:

- Cockpit: eine Vue.js [4] Oberfläche zur Datenpflege wie Zimmern, Raten, uvm.
- TOCC (TourOnline Content Cloud): ein [1] Backend mit Schnittstellen an verschiedene Channels wie Booking, HRS oder Expedia, in der der größte Teil der Logik verarbeitet wird
- Live App: eine App für Hotels, zur vereinfachten Kommunikation zwischen Gästen und den Hotels, aber auch dem Hotel intern

In eben genanntem Cockpit soll die Pflege der Daten integriert werden, um dem Hotelier ein einfaches Werkzeug zur Verfügung zu stellen, häufig gestellte Fragen und die entsprechenden Antworten einzupflegen. In TOCC soll die lokale Datenspeicherung, aber auch die Anbindung und Synchronisation eines, aus der Evaluierung herausgehenden, Bots entwickelt werden. In die Live App, welche noch in der Entwicklung ist, aber bereits einen Chat besitzt, soll genau dieser Bot eingebunden und die Kurznachrichten analysiert und dementsprechende Antworten an den Empfänger gesendet werden.

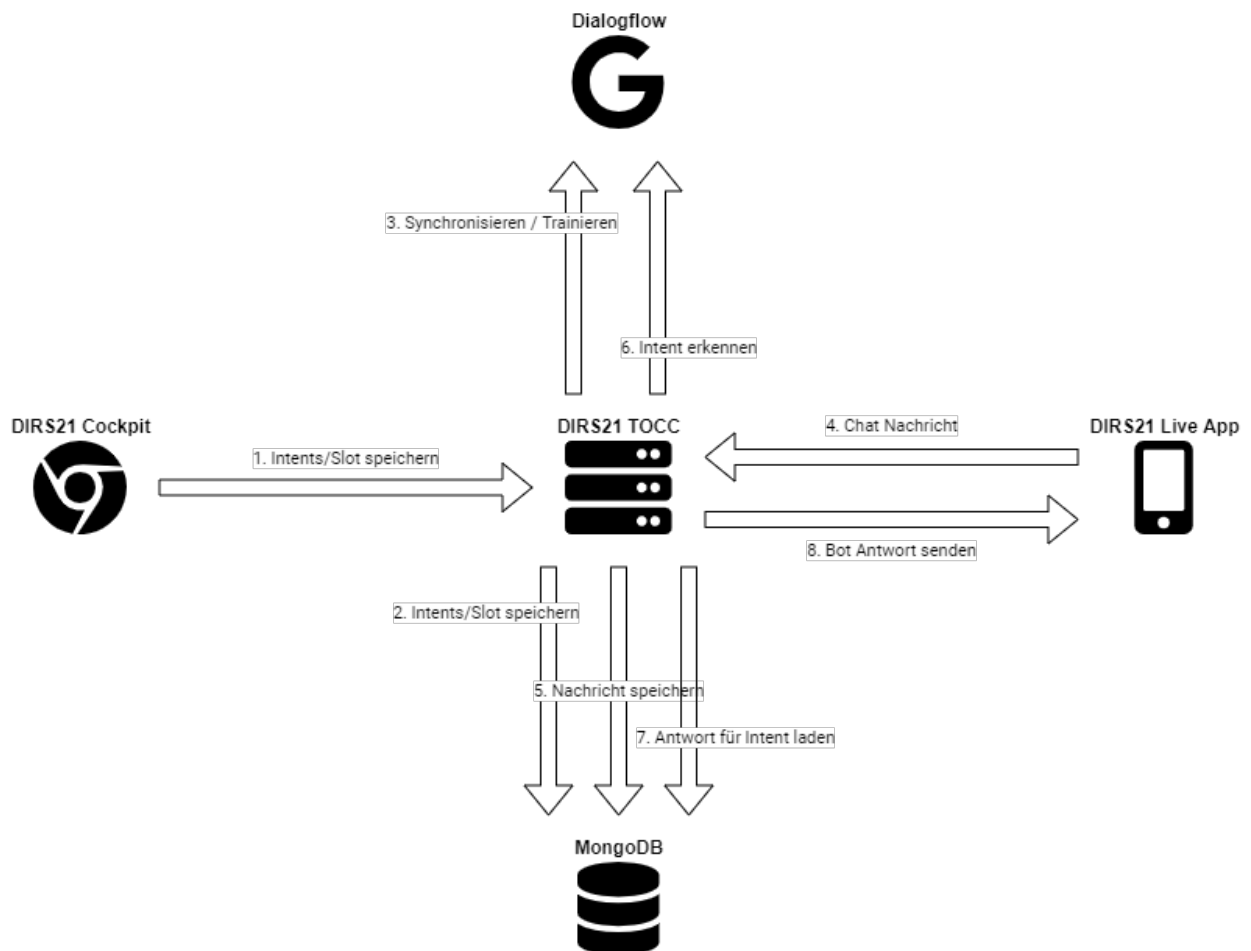


Abb. 2: Softwarekomponenten, die für den Chatbot benötigt werden [2]

Ausblick

Vorschläge

Eine sinnvolle Erweiterung wäre es, für jedes Hotel den Chat mit den Gästen zu analysieren und die meist gestellten Fragen, die nicht beantwortet werden konnten und für die noch kein Intent erstellt wurde, dem Hotelier im DIRS21 Cockpit vorzuschlagen und schon vorauszufüllen. Somit hätte der Hotelier eine Übersicht über die Fragen, die noch nicht beantwortet werden konnten und kann direkt seine Antworten auf diese Fragen hinterlegen.

Sehenswürdigkeiten

Da die Rückmeldung der Hotels häufig Fragen über die Sehenswürdigkeiten in der Nähe beinhaltete, wäre es sinnvoll eine Sehenswürdigkeiten API anzubinden und automatisch auf bestimmte Fragen mit den entsprechenden Infos über die Sehenswürdigkeiten aus der Umgebung zu antworten. Wikunia [3] wäre ein

Beispiel dafür, welches Sehenswürdigkeiten in der Nähe mit Distanz und einem Bild zurück liefert.

Bewertung

Man analysiert die Kurznachrichten in dem Chat, nachdem eine Antwort des Bots gesendet wurde und schlussfolgert, ob diese Antwort weiter geholfen hat. Aus diesen Informationen, kann man eine Bewertung der Antwort berechnen und zusätzlich zu der Analyse des Bots, seine eigenen Scores berechnen und schlussendlich die aus Erfahrungen beste Antwort senden.

Kontext

Viele der zu analysierenden Bots bieten die Möglichkeit Nutzen von Kontexten zu machen. Der Bot erkennt somit nicht nur die Intention aus dem aktuellen Satz, sondern aus dem Kontext vorhergehender Nachrichten. Das würde die Qualität des Bots deutlich verbessern, bedeutet zugleich mehr Aufwand in der Konzeptionierung der Daten, aber auch der Darstellung der Verwaltung.

Literatur und Abbildungen

[1] Stephan Augsten. .NET Core. <https://www.dev-insider.de/was-ist-net-core-a-914978/>, 11 2020.

[2] Eigene Darstellung.

[3] Sights API Wikunia. Wikunia. <http://api.wikunia.de/sights/index.php>, 11 2021.

[4] Evan You. Vue JS. <https://vuejs.org/>, 11 2016.

Modeling and Simulation of reliability and availability for autonomous driving vehicles

Moadh Mahmoud

Karin Melzer

Department of Computer Science and Engineering, Esslingen University

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

Introduction

As the world progresses toward fully autonomous driving systems (where the car is in full control), it is essential that sensors and related technologies provide reliable information and allow decisions to be made that are adapted to different conditions as well as the unpredictable behavior of other road users. In this way, it can not only provide greater access to mobility, but also help reduce the number of driving related accidents and collisions. To achieve this, the reliability of automated driving vehicles is one of the most important factors.

Objectives

The aim of this work is to increase the motivation for using reliability methods by the development of a methodical procedure. In order to meet this objective, the modeling effort should be reduced and a structured and close to reality modeling should be reached. The methodical procedure should be used as a guideline for potential users. By focusing on improving the availability and reliability of the system, the overall quality and efficiency of the process can be improved.

Reliability and availability of a technical system

Reliability is defined in the DIN 40041 standard as the following: "Condition of a unit with regard to its suitability to fulfill the reliability requirement during or after specified periods of time under specified conditions of use." [5] The reliability of a technical system depends strongly, on one side, on the conditions that affect the system from the outside and, on the other side, on the functions that the system has to perform. When both reliability (probability that the item will not fail) and maintainability (probability that the item will be successfully recovered after a failure) are considered, an additional metric is required for the probability that the component/system is

operational (i.e., has not failed or is recovered after a failure or malfunction) at a given time t . This metric is availability. Availability is a performance metric for repairable systems that takes into account both the reliability and maintainability characteristics of a component or system. It is defined as "a percentage measure of the degree to which machines and equipment are in an operational and serviceable condition when needed." [4] The use of the different domains of mechanics, electronics and software results in a much higher complexity of autonomous systems, where reliability is increasingly difficult to achieve. This work try to solve this problem by developing a methodical and consistent approach to ensure complete modeling and simulation of the reliability and availability of technical systems. The failure rate $\lambda(t)$ is one of the most important parameter for describing the failure behavior of technical systems. At a time t , it puts the number of failures or non-intact units in relation to the intact units. The density of the probability of failure $f(t)$ describes the possibility of a failure at a certain point of time. The reliability $R(t)$ quantifies the percentage of intact units at time t . The expression of the temporal course of the failure rate $\lambda(t)$ is : $\lambda(t) = f(t) / R(t)$

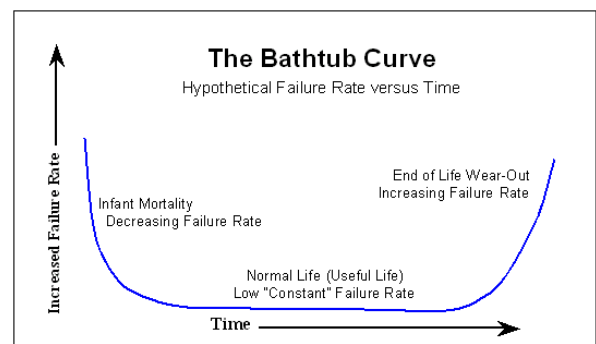


Fig. 1: The three areas of the bathtub curve [3]

The failure rate is a measure of the risk of failure

of a unit in the next considered time interval. The time-dependent curve of the failure rate can be used to describe the failure behavior of components during their entire life cycle. Figure 1 shows the bathtub curve, which describes the characteristic progression of $\lambda(t)$ over time for technical components with wear-out. Electrical components have a constant failure rate $\lambda(t)$.

Reliability Modelling methods

Reliability models or modeling methods are mathematical formulations of the relationships and links between the reliability behavior of a system and its components. Reliability analysis is used to obtain information about reliability parameters and the behavioral properties of the examined system. In this work, different modeling techniques will be discussed:

- Reliability Block Diagrams
- Fault Trees
- Bayesian Network Models
- Markov processes
- Petri nets

Redundancy

In the fields of fault tolerance and self-healing software "redundancy describes the ability to perform certain functionalities through different paths or instances. [1]

Similarly, redundancy can be described by the existence of different components that perform the same functionality. In this part of the work, a redundant braking system is introduced and analyzed as an example of a redundant system.

Monte Carlo Simulation

In reliability analysis and diagnosis of a complex system, one of the fundamental problems faced by reliability engineers is to be able to evaluate different measures such as system reliability, system hazard rate, and the probability of a component causing a system failure. Since these measures involve taking into account a large number of terms, classical calculation becomes impractical. However, for a complex system with a large number of components, a large amount of computation would become necessary to obtain all minimum path sets. The Monte Carlo method seems to be one of the most powerful alternatives to the ordinary methods. Monte Carlo simulation (MCS) is a useful technique for assessing the reliability of a system because of the modeling flexibility it offers regardless of the type and size of the problem. The method is based on repeated sampling of realizations of system configurations, which, however, rarely fail, so that a large number of realizations must be simulated in order to achieve an acceptable accuracy in the estimated failure probability, with large and expensive computation times. For this reason, techniques for efficient sampling of system failure realizations are of interest, in order to reduce the computational effort. [2]

References and figures

- [1] Antonio Carzaniga, Alessandra Gorla, and Mauro Pezzè. *Handling software faults with redundancy*. Springer, 2009.
- [2] Javier Faulin, Angel A Juan, Sebastián Martorell, and José-Emmanuel Ramírez Márquez. *Simulation Methods for Reliability and Availability of Complex Systems*. Springer, 2010.
- [3] Dennis J Wilkins. The bathtub curve and product failure behavior (part 1 of 2). <https://www.weibull.com/hotwire/issue21/hottopics21.htm>, 2021.
- [4] Vamshi K Katukoori. Standardizing availability definition. *University of New Orleans, New Orleans, La., USA*, 1995.
- [5] International Organization for Standardization. *DIN 40041:1990-12, Dependability; concepts*. Beuth-Verlag, 1990.

Single Pair Ethernet - An implementation of a 10Base-T1L physical layer for evaluation purposes

Jo Maier

Michael Scharf

Department of Computer Science and Engineering, Esslingen University

Work carried out at Steinbeis Embedded Systems Technologies GmbH, Esslingen am Neckar

Introduction

Single Pair Ethernet (SPE) describes the Ethernet based communication over a single twisted pair of conductors and is mostly defined towards usage in industrial environments. Therefore robustness of the link, especially against electromagnetic disturbance, and support for state of the art operation in hazardous environments are emphasised. The development of this new key technology in building-, process- and production automation is driven by great activity and participation within industrial stakeholder networks and in standardization committees. SPE will enable the seamless connection between the management layer of an industrial Ethernet based network, down to field level devices such as sensors or actuators via Ethernet based protocols. SPE will empower future Industrial Internet of Things and Industry 4.0 applications.

10Base-T1L definition special among other modern Ethernet variants is its very long reach of up to at least 1000 m, although physical layer (PHY) chips supporting even up to 2000 m are already available. An example of such a device is the PHY chip used in this project as described in reference [1]. The link segment is a single twisted balanced pair of conductors with up to ten in line connectors distributed along the cable. This enables the deployment of Ethernet based communication in very large automation environments. A higher voltage of 2.4 Vpp may be implemented to support high cable lengths. Inside the PHY, a 4bit to 3ternary (4B/3T) encoding comes into play to keep the DC baseline in check on the cable. A 33 bit scrambler is utilized for better robustness in terms of electromagnetic compatibility. Three-level Pulse Amplitude Modulation (PAM3) symbols are used to transmit the data over the link segment at 7.5 MegaBaud (MBd). All of the above encoders and decoders are implemented in the Physical Coding Sublayer (PCS) of the PHY (see Figure 1 for reference). From the Media Independent Interface (MII) on upwards, the overall technology remains the same as it is state of the art for other Ethernet variants especially for Multi Pair Ethernet.

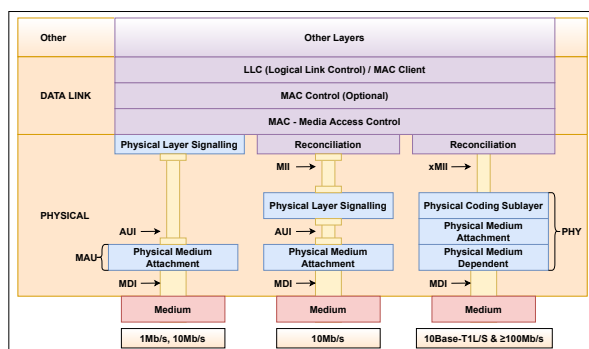


Fig. 1: Ethernet in the OSI layermodel [3]

10Base-T1L

10Base-T1L is defined by the IEEE standards committee as can be seen in the 802.3cg ammendment to the Standard for Ethernet [4]. It supports a 10 Mbps data transmission rate, full duplex operation for point to point connections, or half duplex operation for point to point, and multidrop connections. What makes the



Fig. 2: ADIN1100 shield attached to a STM32 Nucleo board [2]

Hardware

In Figure 2 the Nucleo STM32-H7 development board with a SPE shield can be seen. The development board houses a STM32-H743 ZIT6 micro controller alongside a basic set of supply circuitry and peripherals. We were able to use the breakout connector of the Nucleo board to design a piggybacking circuit board which connects the Media Access Control (MAC) module of the STM32 controller to the ADIN1100 SPE PHY on our circuit board via a MII. While the connection between a PHY chip and a controller remains the same MII/RMII as per other Ethernet technologies, the Media Dependent Interface (MDI) between the PHY chip and the medium is vastly different, since it has to support full duplex communication over a single pair of conductors. Our implementation which is suggested in reference [1] by the PHY chip manufacturer can be seen in figure 3.

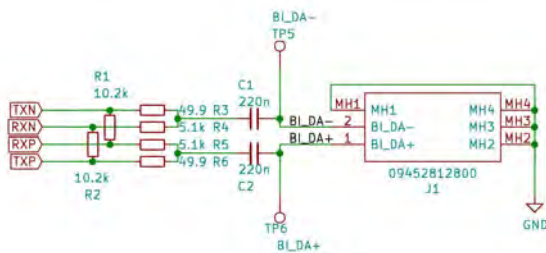


Fig. 3: Implementation of a Single Pair Ethernet Media Dependent Interface [2]

Software

The STM32 CubeIDE provides a demonstrative implementation of the iPerf application alongside the underlying lwIP stack. Fully functional code can be generated for the evaluation board and its native Multi Pair Ethernet PHY. To minimize efforts, we managed to modify this code for the ADIN1100 shield by writing a driver and changing some hardware level functions of the stack implementation. Unlike the native PHY, the ADIN1100 offers numerous configuration possibilities via extended MDIO Managable Device (MMD) registers. We made several changes to the generated code. This made it possible to use these registers as described in reference [3]. By accessing MMD registers each subsystem of the PHY, such as the

PCS or the autonegotiation module, can be configured and diagnosed separately.

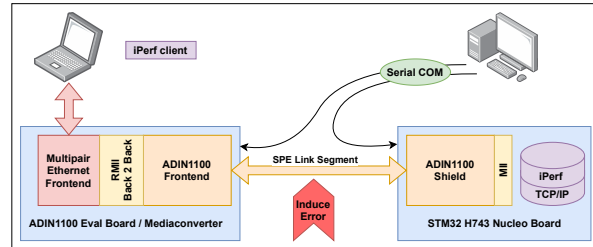


Fig. 4: Testing installation [2]

Test environment and procedure

A solid test environment is created by running an iPerf server on the STM32 controller and an iPerf client on a Linux computer. The behavior on hardware level can be observed through serial comports connections on another computer. The media converter allows insight into the SPE PHY registers via a manufacturer provided software tool. The STM32 can be monitored by using simple messages over UART, which can be displayed by a terminal application via a serial communication port. A schematic presentation of the test environment can be seen in figure 4. The reliability of the SPE connection will be tested by using different cables as link segment, and inducing various errors to it. Measurements of the performance and link quality over the course of these tests will provide the basis for a reliable evaluation.

First results and prospect

We were able to conduct first tests without error induction to get a baseline of results. Only well shielded SPE certified cablings of various lengths were used in these first experiments. By doing so we were able to verify that in unobstructed environments, the hard- and software work as expected. Possible next steps are tests with a variety of other cables such as unshielded cabling or unbalanced cabling in terms of impedance. To ensure comparability of further tests, including error induction, procedures to reproduce and quantify these errors in similar magnitudes must be set up. Possible error inductions include, but are not limited to: cable bending, cable damaging, radio frequency polluted environment generation with a software defined radio, or a combination thereof.

References and figures

- [1] Analog Devices. Analog Devices ADIN1100 - Robust, Industrial, Low Power 10BASE-T1L PHY. <https://www.analog.com/media/en/technical-documentation/data-sheets/adin1100.pdf>, 2021.
- [2] Own representation.
- [3] LAN MAN Standards Committee of the IEEE Computer Society. *IEEE 802.3 Standard for Ethernet*. IEEE, 2018.
- [4] LAN MAN Standards Committee of the IEEE Computer Society. *IEEE P802.3cg - Standard for Ethernet Amendment: Physical Layer Specifications and Management Parameters for 10 Mb/s Operation and Associated Power Delivery over a Single Balanced Pair of Conductors*. IEEE, 2019.

Herausforderungen und Potentiale der Migration auf Web-Technologien am Beispiel eines PLM-Systems

Daniel Pascal Mair

Mirko Sonntag

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Eberspächer Gruppe, Esslingen

Einleitung

Informationstechnische (IT) Systeme zeichnen sich oft durch einen langen Lebenszyklus aus - ebenso die Product-LifeCycle Management Software CIM Database PLM von Contact, welche bereits seit über 20 Jahren in verschiedenen Organisationen, darunter auch die Eberspächer Gruppe, zum Einsatz kommt. Während dieser langen Lebenszeit von Software ergeben sich für die Anwender verschiedenste Veränderungen in Form von neuen Produkten sowie Geschäftsprozessen. Dadurch entstehen neue Anforderungen an die eingesetzten IT-Systeme, welche oftmals entscheidend für den Erfolg der Organisationen sind [1]. Um in diesem, sich stetig wandelnden Umfeld, eine optimale Softwarelösung bereitzustellen, setzen Hersteller wie Betreiber auf dedizierte Softwareentwicklungsprozesse. Über eine Python-Schnittstelle lässt sich CIM Database PLM nach eigenem Bedarf anzupassen.

Problemstellung

Eine dieser Veränderungen in den letzten Jahren ist der Wandel des Client-Server Architekturmodells weg vom Rich-Client hin zum Web-Client [3]. Denn viele Softwarehersteller, darunter auch Contact, lassen den Support ihrer Rich-Client Version auslaufen. Bestandskunden wie Eberspächer werden dazu angehalten auf den neuen Web-Client zu wechseln.

Den nun anstehenden Übergang von Rich-Client zu Web-Client möchte Eberspächer für sich in Betracht ziehen. Bereits vorgenommene Softwareanpassungen lassen sich theoretisch automatisch auf den Web-Client übertragen. Allerdings divergiert die Eberspächer Instanz teilweise stark vom Standard, da es sehr viele Anpassungen gibt. Diese Adaptionen führen bereits regelmäßig zu erhöhtem Aufwand bei Versionsupdates. Deshalb liegt der Schluss nahe, dass der Web-Client-Umstieg mit den sehr vielen Anpassungen nicht automatisch funktionieren wird. Außerdem ist es nicht gewiss, ob sich jede Anpassung in den Web-Client transportieren lässt.

Zielsetzung

Das Ziel der Analyse und zugehörigen Experimente ist es ein Grundverständnis für den Umstieg auf den Web-Client bei Eberspächer zu gewinnen. Anhand dessen wird dann ein möglicher Migrationspfad hin zum Web-Client für Eberspächer diskutiert. Dieser soll dem Unternehmen eine optimale Lösung für seine zu erwartenden Probleme beim Umstieg präsentieren. Dazu wird geprüft, inwiefern sich verschiedene Kategorien von Anpassungen automatisch umwandeln lassen und wie kostenintensiv Nachbesserungen sein können. Ebenfalls wird auch die Neuentwicklung von Anpassung in Betracht gezogen, da die Anpassungs- und Erweiterungsmöglichkeiten im Web-Client aufgrund von eigenen Webapplikationen besser sind. Des Weiteren wird untersucht was sich aus Sicht der User verändert und ob sich ihre heutige Arbeitsweise im Web-Client durchführen lässt. Alles in allem ermöglicht diese Masterarbeit den Wandlungsprozess vom Rich-Client hin zum Web-Client durch eine wohl definierte Dokumentation, die auf den erprobten Grundprinzipien der Literatur sowie der gewonnen Erkenntnisse aus eigenen Experimenten basiert, zu verstehen.

Methoden und Vorgehen

Die Grundidee für das allgemeine Vorgehen stammt aus der Softwareentwicklung. Schließlich ist die aktuelle Eberspächer-Instanz sowohl Input als auch Output dieses Prozesses. Für einen besseren Überblick über die 15 Jahre Entwicklungsgeschichte wurden mit Hilfe der Eberspächer Entwickler vier Kategorien festgelegt, die typisch für die durchgeführten Anpassungen sind. Für die Kategorisierung wird die Anpassungsart als Maß eingesetzt. Durch die Durchführung eigener Implementierung anhand der Kategorien wird untersucht, welchen Aufwand die Entwickler für die Portierung haben.

Die erste Klasse umfasst kosmetische Anpassungen, die rein konfigurativer Art sind, beispielsweise Korrektur von Rechtschreibfehler. In der zweiten Klasse befinden

sich ebenfalls Anpassungen, die rein konfigurativer Natur sind, jedoch Geschäftslogik implementieren. Die dritte Klasse erweitert die Anforderungen der Zweiten um eigenen Programmcode gegen die Python-Schnittstelle. Die vierte Klasse unterscheidet sich durch die Anpassungsart, hier werden Anpassungen eingeordnet, die eigene Geschäftslogik implementieren und noch in keiner Form im Programmcode vorkommen.

Zur Veranschaulichung wird ein Use-Case aus der ersten Klasse betrachtet: Die benutzerdefinierte Operation Test Operation DM für Artikel wird häufig benötigt und soll deshalb im Kontextmenü weiter oben angesiedelt werden. Die Operation kann von jedem Benutzer im Kontextmenü eines Artikels verwendet werden. Die Nachbedingung ist, dass der Benutzer wie gewohnt weiterarbeiten kann.

Eine kosmetische Anpassung lässt sich in CIM Database PLM durch die eingebaute Konfigurationsmöglichkeiten über die Oberfläche anpassen ohne Programmieraufwand. Für die Anpassung muss man die Positionsparameter der Operation anpassen und die Anwendung neu starten. Für den Rich-Client erhält man folgendes Ergebnis:

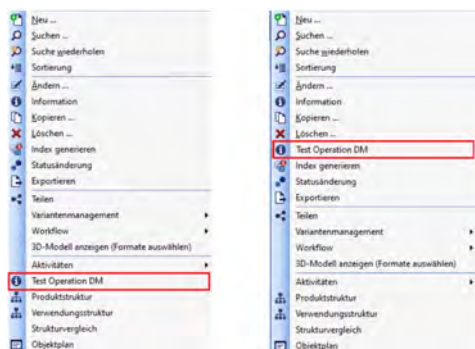


Abb. 1: Screenshot vor (links) und nach (rechts) der Anpassung im Rich-Client [2]

Der Vergleich zwischen den beiden Zuständen vor und nach der Anpassung ist in Abbildung 1 dargestellt. Wie von dem Benutzer gewünscht, befindet sich die Operation nun weiter oben und man kann sie schneller auswählen. Nun soll betrachtet werden, wie die Anpassung im Web-Client erscheint:



Abb. 2: Screenshot vor (links) und nach (rechts) der Anpassung im Web-Client [2]

Abbildung 2 zeigt einen Vergleich der beiden Zustände vor und nach der Anpassung im Web-Client. Im Zustand vor der Anpassung fällt auf, dass die zu verschiebende Operation nicht vorhanden ist. Man muss die Operation erst für den Web-Client aktivieren, bevor man sie angezeigt bekommt. Aktiviert man diese fällt auf, dass die Positionen der Operationen nicht gleich sind. Der Grund dafür ist ein Positionskonflikt zweier Operationen. Im Web-Client erhält die im Standard vorkommende Operationen den Vorrang, hingegen im Rich-Client die benutzerdefinierte Operation.

In diesem Experiment hat sich bereits gezeigt, dass selbst kleine Anpassungen manuelles Nacharbeiten benötigen, damit ein äquivalentes Arbeiten in beiden Versionen möglich ist. Hier scheint der Schluss logisch, dass in den Klassen, die Konfiguration und Implementierung benötigen, der Aufwand deutlich ansteigt.

Ausblick

Im weiteren Verlauf der Masterarbeit werden weitere Use-Cases analysiert und ausgewertet. Dadurch erhält man einen groben Einblick bezüglich des Aufwands je nach Anpassungsklasse. Ebenfalls soll untersucht werden, ob die Python-Schnittstelle sich in beiden Systemen gleich verhält und sämtliche Funktionalität vorhanden ist. Hinsichtlich dessen wird auch geprüft, ob man durch die Verwendung von selbstentwickelten Web-Apps ein besseres Ergebnis als mit der Python-Schnittstelle erzielen kann. Darüber hinaus sollen die beiden Versionen aus Sicht des Anwenders betrachtet werden und vorallem die Usability anhand von Standardszenarien untersucht werden.

Literatur und Abbildungen

- [1] Paul Alpar, Rainer Alt, Frank Bensberg, and Peter Weimann. *Anwendungsorientierte Wirtschaftsinformatik*. Springer Fachmedien Wiesbaden, 2019.
- [2] Eigene Darstellung.
- [3] Mark Richards. *Handbuch moderner Softwarearchitektur*. O'Reilly Verlag, 2020.

Markendetektion mit maschinellem Lernen

Amirhossein Marzani

Markus Enzweiler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Leuze electronic, Owen Teck

Einleitung

In den letzten Jahren ist eine neue und revolutionäre Bewegung in fast allen wichtigen Bereichen des Lebens unter anderem in der Wissenschaft, Industrie und im Alltag durch maschinelles Lernen entstanden. Das ML (Machine Learning) ist fast überall gegenwärtig, egal im eigenen Rechner, der Cloud oder in Handys ist das ML vertreten und breitet sich in weitere Themengebiete aus.

Motivation und Zielsetzung

Das Ziel dieser Arbeit ist es, die Untersuchung von ML-basierten Ansätzen zur Bearbeitung von Bildern und Informationsgewinnung aus den geschossenen Bildern. Konkret wird untersucht, ob die ML-Algorithmen in der Lage sind effizient und genau die klassischen Algorithmen in Bildverarbeitung und Mustererkennung zu ersetzen.

Maschinelles Lernen

Der Begriff maschinelles Lernen beschreibt im Allgemeinen den Versuch des Menschen, den bei sich beobachteten Prozess des Lernens mit Hilfe von verschiedenen Ansätzen und Modellen auf Maschinen zu übertragen, damit diese auch in der Lage sind sich bestimmte Eigenschaften anzueignen. Das Wort Lernen in dem oben erwähnten Begriff basiert bei uns Menschen auf Phänomenen wie Beobachten, Verstehen, Versuchen und der Erfahrungsgewinnung. Bei maschinellem Lernen basiert dieser Prozess, im Gegensatz auf Informationsgewinnung und das Bauen eines statistischen Modells durch Eingaben bzw. Informationen in der Trainingsphase auf.

Sensor und Einsatz

Die Firma Leuze electronic GmbH & Co.KG ist ein weltweit agierender Sensorhersteller mit Hauptsitz in Owen Teck. Die Firma ist spezialisiert auf optischen Sensoren und Barcodeleser. IPS 200i ist eine Baureihe der kamerabasierten Positionierungssensoren der Firma Leuze electronic. Durch den Einsatz des Sensors wird eine schnelle und einfache Positionierung von Regalbediengeräten in der Lager- und Förder-Technik ermöglicht.



Abb. 1: IPS200i [4]

Der Sensor erkennt Löcher bzw. Reflektoren in Riegeln im Regalbau und dadurch ermittelt er die Positionsabweichung des Regalbediengerätes von der Soll-Position.

Künstliche Neuronale Netze

Das menschliche Gehirn besitzt zwischen 10 bis 100 Milliarden Nervenzellen bzw. Neuronen, welche sich im Gehirn und im Körper strecken. Die hochkomplexen Abläufe und Prozesse im Alltag wie Autofahren, Laufen, Springen, Lernen und sich an Umfeld-Bedingungen anpassen, verdanken wir diesem großen komplexen Netz aus Nerven.

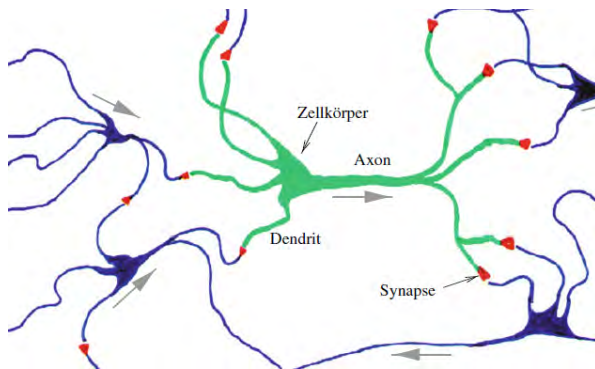


Abb. 2: Nervenzelle [2]

Eine Nervenzelle besteht aus einem Zellkörper, Dendrit, Axon und Synapsen. Der Zellkörper funktioniert wie ein Energiespeicher. Ein Dendrit ist wie eine Zweigstelle und der Empfänger der Zelle und darüber hinaus verbinden Dendriten alle Zellen mithilfe von Axonen und Synapsen. Der Zellkörper speichert die empfangenen elektrischen Impulse, die von anderen Zellen stammen und empfangen werden. Sobald die gespeicherte Menge einen bestimmten Schwellwert überschreitet, leitet die Zelle über ihre Ausgänge die gesamte Ladung an weitere Zellen. Die Synapsen sind die Verbindungsstellen zwischen den Zellen und sie sind die dynamisch veränderlichen Komponenten des Nervensystems [2].

Das mathematische Modell hinter künstlichen neuronalen Netzen

Das vorher beschriebene Modell mit Nervenzellen bzw. Neuronen wurde zum ersten Mal mathematisch in vereinfachter Form in 1943 durch McCulloch (Neurowissenschaftler) und Pitts (Logiker) in dem Artikel „logical calculus of the ideas immanent in nervous activity“ vorgestellt [5]. Dieses Modell wurde mit der Zeit unter anderem mit Addition von Aktivierungsfunktion weiterentwickelt. Das heutige Modell besteht aus Eingängen, Gewichten, Bias und Aktivierungsfunktionen.

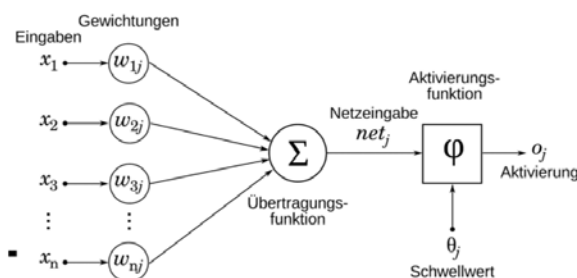


Abb. 3: Mathematisches Modell von Neuronen [7]

Die Werte $x_1 \dots x_n$ sind die Eingänge des Neuron. Durch diese Eingänge bekommt das Neuron Impulse bzw. Information, die dann jeweils mit dem zugehörigen Gewicht bzw. $w_1 \dots w_n$ multipliziert werden. Die Ergebnisse der Multiplikationen werden dann in Übertragungsfunktion zusammenaddiert und das resultierte Skalar wird mit dem Bias (Schwellwert) des Neuron akkumuliert.

$$net_j = \sum_{i=1}^n x_i * w_i + \theta_j$$

Danach wird das Ergebnis dieser Addition durch eine Aktivierungsfunktion geleitet. Die einfachste Form einer Aktivierungsfunktion kann zum Beispiel eine Funktion

$$in\ folgender\ Form\ sein: O_j = \begin{cases} 1 & \text{wenn } net_j > 0 \\ 0 & \text{wenn } net_j \leq 0 \end{cases}$$

Das Ergebnis der Additionen mit dem Bias und die Form der Aktivierungsfunktion entscheidet darüber, ob das Neuron durch seinen Eingängen aktiviert wird (*feuert*) oder nicht (*neutral*). Es gibt sehr viele Aktivierungsfunktionen, die je nach Bedarf eingesetzt werden. Die Aktivierungsfunktionen können entweder Linear oder Nicht-Linear sein. Die Nicht-Linearen Aktivierungsfunktionen werden eingesetzt, damit die nicht linearen Beziehungen in den Daten von einem neuronalen Netz erlernt werden können. Die Aktivierungsfunktionen sind mindestens in einem bestimmten Bereich differenzierbar und in diesen Bereich werden diese auch eingesetzt.

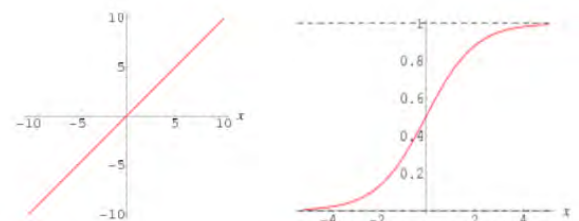


Abb. 4: Identitäts-Funktion und Sigmoid-Funktion [6]

Identitäts-Funktion $F(x) = x$, Sigmoid-Funktion $F(x) = \frac{1}{1+e^{-x}}$

Perzeptron

Das Vorherige Modell 3 stellt den Aufbau von einem künstlichem Neuron bzw. den Baustein eines Perzeptrons dar. Ein Perzeptron ist ein Netz aus einem oder mehreren Neuronen, die Informationen bearbeiten können. Das einfachste neuronale Netz besteht aus einem einzigen Perzeptron, welches nur aus einem oder mehreren Neuronen besteht. Das bedeutet, dass alle Neuronen die Eingabe als einen Vektor bekommen und es am Ende so viele

Ausgaben wie die Anzahl der Neuronen gibt. Das Zusammensetzen von mehreren Perzeptren wird verwendet, wenn die gestellte Aufgabe an Komplexität gewinnt und komplexe Zusammenhänge von einem künstlichen neuronalen Netz erlernt werden soll. Beim mehrschichtigen Perzeptron sind mehrere einschichtige Perzeptren miteinander verbunden. Das bekannteste Beispiel dafür ist das feed-forward-Netz. In feed-forward-Netzen läuft der Informationsfluss nur in einer Richtung. Dabei sind die Eingänge von jeder Schicht mit den Ausgängen von vorheriger Schicht verbunden. Im gängigen und einfachsten Fall besteht das Netz aus jeweils einer Eingabe, Ausgabe und versteckter Schicht, die jeweils unterschiedliche Anzahl von Neuronen besitzen können. Das bekannteste Beispiel dafür ist die **fully-connected-Schicht**, mit drei Schichten wie in 5 dargestellt.

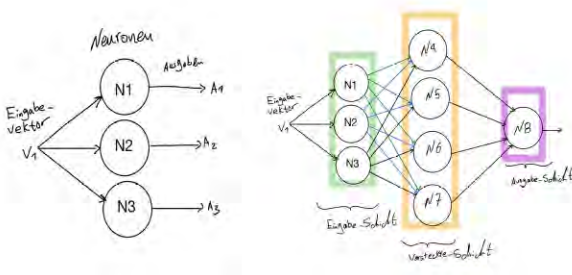


Abb. 5: ein und mehrschichtiges Perzeptron [1]

Faltungsschicht

Faltungsschichten sind in neuronalen Netzen, die ganz oder zum Teil Informationen aus Bildern gewinnen sollen, sehr beliebt. Diese Schicht besteht aus Kernen bzw. Filtern, die wesentlich kleiner als die Eingabe bzw. Bilder sind. Diese Filter werden über das Bild von der

oberen linken Seite bis zur unteren rechten Seite gelegt und der ausgewählte Bereich des Bildes wird mit dem Filter multipliziert und das Ergebnis wird zusammen addiert. Die Faltungsoperation wird beschrieben durch: $S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n)$ [3]

Das Ergebnis einer Faltungsoperation ist ein feature-map (S), welches durch verschieben der Filter (K) auf dem Bild (I) erzeugt wird. Jeder Filter erzeugt ein feature-map. Jedes feature-map beschreibt, ob und wo in der Eingabe, das gesuchte Muster vorhanden ist. Aufgrund von vorher erwähnte Faltungsdefinition wird der gesuchte Filter bzw. das gesuchte Muster in der Eingabe verkehrt-herum abgebildet.

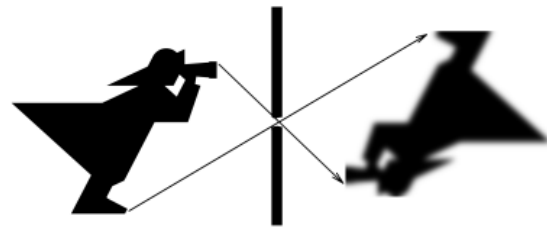


Abb. 6: umgekehrtes Abbilden des Filters in feature-maps [2]

Mithilfe der Faltungsschichten und der Trainingsprozesse sind künstliche neuronale Netze in der Lage, sowohl visuelle Muster als auch ihre räumliche Anordnung zu erkennen und zu lernen.

Fazit

Der Einsatzmöglichkeiten der Algorithmen, die auf maschinellem Lernen basieren, sind vielfältig und viel versprechend. Es gibt doch noch Einschränkungen bezüglich der Laufzeit und der Größe bei ihren Einsatzmöglichkeiten, wenn es um Embedded-Welt und Ressourcen-Einschränkungen geht.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Wolfgang Ertel. *Grundkurs Künstliche Intelligent*. Springer Fachmedien Wiesbaden, 5 edition, 2021.
- [3] Goodfellow Ian, Bengio Yoshua, and Courville Aaron. *Deep Learning : das umfassende Handbuch*. mitp : Frechen, 1 edition, 2018.
- [4] electronic leuze. IPS200i. https://www.leuze.com/de/deutschland/produkte/messende_sensoren/sensoren_zur_positionierung/kamerabasierte_positionierungssensoren/ips_2__i_11/selector.php, 2021.
- [5] Warren McCulloch and Walter Pitts. a logical calculus of the ideas immanent in nervous activity. <https://homeweb.csulb.edu/~cwallis/382/readings/482/mcculloch.logical.calculus.ideas.1943.pdf>, 1943.
- [6] Eric Weisstein. Identity Function. <https://mathworld.wolfram.com/IdentityFunction.html>, 2021.
- [7] Laurenz Wutke. Künstliche Neuronale Netzwerke. <https://datasolut.com/neuronale-netzwerke-einfuehrung>, 2020.

Requirements Engineering im Bereich Service Management zur Entwicklung einer neuen Datenquelle in einem Business Intelligence Analytics Tool

Sueleyman Omurca

Thomas Rodach

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma USU GmbH, Spitalhof, 71696 Möglingen

Einleitung

Die Bedeutung von Business Intelligence Software konnte in den letzten Jahren einen erheblichen Anstieg verzeichnen: So wuchs der Gesamtmarkt für Analyse- und Business Intelligence-Software im Jahr 2019 um 10,4% auf 24,8 Milliarden US-Dollar. [2] Insbesondere in der Wirtschaftsinformatik spielt dieser Bereich eine bedeutende Rolle, da hier die Umsetzung von IT-Lösungen im Business-Kontext gefordert ist. Als Tochtergesellschaft der USU Software AG, stellt die USU GmbH eine Analyse und Reporting Lösung für die eigene Business Service Management Software, USU Service Management (USM) her. Dabei wird USM Suite durch eine Business Intelligence Lösung ergänzt. Kunden der USU nutzen diese Business Intelligence Lösung mit dem Namen USU Analytics für Ad-hoc Auswertungen der IT-Service Management Software. Diese Bachelorarbeit befasst sich mit dem Requirements Engineering zur Erarbeitung eines Konzepts für die prototypische Entwicklung einer neuen Datenquelle für das Modul Projekt Portfolio Manager (PPM) in einer IT-Service Management Software. USU Analytics stellt für verschiedene Unternehmensprozesse (Module) nützliche Datenquellen (OLAP-Cubes) bereit, die relevanten Felder des Prozesses zusammenfassen. Auf Basis dieses Modells werden, Berichte, Analysen und Dashboards erstellt. Mit der Anforderungsanalyse und der Dokumentation der Anforderungsspezifikationen soll eine Grundlage für die Implementierung geschaffen werden.

Problemstellung

Im Digitalzeitalter generieren die Prozesse des Unternehmens kontinuierlich Daten. Durch immer umfangreicher werdende Datenbestände und immer komplexer werdenden Arbeitswelt wächst der Fokus Erkenntnisse aus diesen fortlaufend erfassten Daten zu gewinnen. Um konkurrenzfähig zu bleiben und

wirtschaftliche Vorteile bei der Datenmenge zu erzielen, die Unternehmen erzeugt haben, bedarf es professioneller Tools zur effektiven Auswertung von Informationen. Kernbestandteil bei der Erstellung eines Softwareprojekts ist die Erfassung und Spezifikation von Anforderungen. Falls im Requirements Engineering die Anforderungen nicht sauber erarbeitet oder Lücken in den Anforderungsdokumenten nicht frühzeitig erkannt werden, kann dies zu einem dauerhaften Änderungsprozess führen und den Projekterfolg gefährden. Denn jede Folgeaktivität im Entwicklungsprozess basiert auf der Anforderungsspezifikation. Demnach beeinflusst das Requirements Engineering den Erfolg des Projekts. Deutlich wird dies in der vielfach zitierten Studie der Standish Group Report, der sich seit 1994 mit dem Erfolg von IT-Projekten beschäftigt. Die Studie bestätigt, dass mehr als ein Drittel der Projekte durch falsche Vorgehensweise und Fehleinschätzungen im Requirements Engineering zurückzuführen ist. [4] Die USU Software AG tritt sowohl als Anbieter für Standardsoftware als auch Dienstleister zur Beratung und Einführung von Systemen in Kundenumgebungen auf. Daraus ergibt sich zusätzlich die Herausforderung: Die Anforderungen zu ermitteln, die für alle Stakeholder zutreffen und die Datenquelle so zu gestalten, dass eine nutzbare analytische Auswertung für den Prozess angewendet werden kann.

Zielsetzung

Ziel der Bachelorarbeit ist es, in einem geeigneten Prozess die Anforderungen an eine neue Datenquelle für das Modul Projekt Portfolio Management in einer IT-Service Management Software zu ermitteln. Zugleich beschreibt die Arbeit welche Methoden dabei zum Einsatz kommen. Um im Zuge dieser Abschlussarbeit bereits erste Ergebnisse liefern zu können, soll eine prototypische Implementierung durchgeführt werden. Dabei werden folgende Fragen definiert:

Was sind die Anforderungen an eine Datenquelle für ein Modul in einer Business Service Management Software der USU GmbH?

Wie gestaltet sich ein geeignetes Konzept für ein prototypisches Modell von Feldern, um Analysen, Berichte und Dashboards darzustellen, unter Berücksichtigung der Anforderungen?

Business Intelligence

Business Intelligence wird als ein Sammelbegriff für verschiedenste Technologien verwendet. Es gibt keine klare wissenschaftliche Abgrenzung und keine festgelegte allgemeingültige Definition. Im Folgenden wird der Begriff von Gartner wie folgt definiert:

„Analytics and Business Intelligence (ABI) is an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access and analysis of information to improve and optimize decisions and performance.“ [3]

Der BI-Prozess basiert auf der Umwandlung von Daten in Informationen, dann in Entscheidungen und schließlich in Handlungen. Es umfasst die Technologien, Anwendungen und Methoden zum Sammeln, Integrieren, Analysieren und Präsentieren von Geschäftsinformationen. Diese werden von der Industrie entwickelt und von Softwareunternehmen geprägt. Um einen interaktiven Zugriff auf Daten sowie die Manipulation von Daten zu ermöglichen und Geschäftsmanagern und Analysten die Möglichkeit zu geben, entsprechende Analysen durchzuführen. Das Hauptziel dabei ist es, wertvolle Erkenntnisse daraus zu erschließen, die es dann ermöglichen, fundierte und bessere Entscheidungen zu treffen. [8]

Requirements Engineering

Der Begriff „Requirements Engineering“ wird laut Christof Ebert wie folgt definiert:

„RE ist das disziplinierte und systematische Vorgehen (d.h. Engineering) zur Ermittlung, Dokumentation, Analyse, Prüfung, Abstimmung und Verwaltung von

Anforderungen unter kundenorientierten, technischen und wirtschaftlichen Zielvorgaben.“ [1]

Christof Ebert hebt mit dieser Definition die enthaltenen Phasen hervor. Klaus Pohl und Chris Rupp hingegen definieren das Requirements Engineering allgemeiner als einen

„[...] systematischen und disziplinierten Ansatz zur Spezifikation und zum Management von Anforderungen [...]“ [5] Und definieren die Bedeutung des Ziels im Requirements Engineering, wie folgt:

(1) *„Die relevanten Anforderungen zu kennen, Konsens unter den Stakeholdern über die Anforderungen herzustellen, die Anforderungen konform zu vergebenen Standards zu dokumentieren und die Anforderungen systematisch zu managen.“*

(2) *Die Wünsche und Bedürfnisse der Stakeholder zu verstehen, zu dokumentieren sowie die Anforderungen zu spezifizieren und zu Management, um das Risiko zu minimieren, dass das System nicht den Wünschen und Bedürfnissen der Stakeholder entspricht.“* [6]

Ausblick

Moderne Analytics- und Business-Intelligence-Plattformen ermöglichen es Kunden, komplette analytische Workflows zu implementieren. Damit kann das System nicht nur Daten extrahieren, verarbeiten und verwalten, sondern eignet sich auch zur Generierung von Erkenntnissen. Dies ermöglicht auch Nutzern, die keine große Affinität zu IT haben, Daten zu filtern oder detailliertere Ansichten zu erstellen. [7] Dabei ist die Identifizierung und Analyse der spezifischen Bedürfnisse von Kunden und Stakeholdern die Grundlage für die Entwicklung von Software. Anhand der Anforderungsspezifikationen werden von ausgewählten Anforderungen die technischen Spezifikationen definiert. Anschließend wird eine prototypische Implementierung (Proof of Concept, PoC) durchgeführt. Infolgedessen werden ausgewählten Anforderungen von der neuen Datenquelle an der Software getestet.

Literatur und Abbildungen

- [1] C. Ebert. *Systematisches Requirements Engineering. Anforderungen ermitteln, dokumentieren, analysieren und verwalten*. Ebert, C., 6 edition, 2019.
- [2] Gartner Gartner. Marktanteil: Analytics und Business Intelligence. <https://www.gartner.com/en/documents/3985421/market-share-analytics-and-business-intelligence-worldwi>, 05 2020.
- [3] Gartner Gartner. Requirements Engineering im Bereich Service Management zur Entwicklung einer neuen Datenquelle in einem Business Intelligence Analytics Tool. <https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi>, 2021.
- [4] K. Pohl and C. Rupp. *Basiswissen Requirements Engineering. Aus- und Weiterbildung zum Certified Professional for Requirements Engineering*. Foundation Level nach IREB-Standard, 4 edition, 2015.
- [5] K. Pohl and C. Rupp. *Basiswissen Requirements Engineering. Aus- und Weiterbildung zum Certified Professional for Requirements Engineering*. Foundation Level nach IREB-Standard, 4 edition, 2015.
- [6] K. Pohl and C. Rupp. *Basiswissen Requirements Engineering. Aus- und Weiterbildung zum Certified Professional for Requirements Engineering*. Foundation Level nach IREB-Standard, 4 edition, 2015.
- [7] Peter Sayer. Business-Intelligence-Software,. <https://www.computerwoche.de/a/die-besten-bi-tools,3548631>, 11 2021.
- [8] R. Sharda. *Business intelligence, analytics, and data science. A managerial perspective*, Harlow,. Sharda, R., 2018.

Lidar-based localization and mapping for autonomous vehicle

Yassine Ounajjar

Thao Dang

Department of Computer Science and Engineering, Esslingen University

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

Introduction

Autonomous vehicles are becoming a reality in the coming years. They are expected to prevent traffic deaths by eliminating human errors, reduce road congestion through better efficiency, and decrease travel time allowing greater productivity and accessibility. To enable autonomous behavior, the vehicles perceive their environment using several sensors. Then different algorithms are used to process the input and generate a driving plan. Finally, the plan have to be executed using the actuators. This process can be divided into three main categories: Perception, planning, and control. Perception refers to the ability of an autonomous system to collect information and extract relevant knowledge from the environment. Environmental perception refers to developing a contextual understanding of environment, such as where obstacles are located, detection of road signs/markings, and categorizing data by their semantic meaning [4]. Perception is the most important stage for autonomous vehicles. The more detailed the input, the more accurate the navigation.

Motivation

This thesis is conducted within project "Ameise". The project develops an autonomous bus for the city of Waiblingen. The bus will be equipped with different sensors like LIDARs, cameras, imu (inertial measurement unit), etc. A high accurate map is one way of providing the autonomous bus with crucial information on the driving environment, including the free drivable areas, surrounding obstacles locations, velocities, and even predictions of their future states [4]. In this thesis, methods to generate a high accurate map for autonomous vehicles should be evaluated, implemented, and tested. The vehicle should also be able to locate itself in the map. This computational problem is referred to as the Simultaneous Localization and Mapping (SLAM). SLAM is the process by which a mobile robot can build a map of the environment and,

at the same time, use this map to compute its location. The state-based formulation of the SLAM problem involves the estimation of a joint state composed of a robot pose and the locations of observed stationary landmarks [1]. There are mainly three types of SLAM algorithms: LIDAR-SLAM, visual SLAM, and LIDAR-vision fusion SLAM.

Approach

Significant advances have been made in visual SLAM techniques in the past several years. However, due to the fragile performance in tracking feature points in environments that lack texture, visual SLAM can hardly provide a reliable localization. Compared with visual SLAM, LIDAR-SLAM can often provide more robust localization by using 3D spatial information directly captured by LIDAR point clouds [6].

LIDAR

LIDAR refers to a light detection and ranging device, which sends millions of light pulses per second in a well-designed pattern. With its rotating axis, it is able to create a dynamic, three-dimensional map of the environment [4]. For the Ameise project, an Ouster LIDAR, shown in figure 1, will be built on top of the autonomous bus. The Ouster LIDAR will provide the SLAM algorithm with the necessary point clouds. It is also equipped with an imu to estimate the pose of the LIDAR.



Fig. 1: OS1 Ouster LIDAR [2]

SLAM

There are different LIDAR-based SLAM methods or solutions that are robust enough for this use case. A comparison of the different methods was conducted in [6]. According to this evaluation, the google cartographer SLAM algorithm scored best in comparison to other LIDAR-based SLAM methods. Cartographer is a system that provides real-time simultaneous localization and mapping in 2D and 3D across multiple platforms and sensor configurations [3].

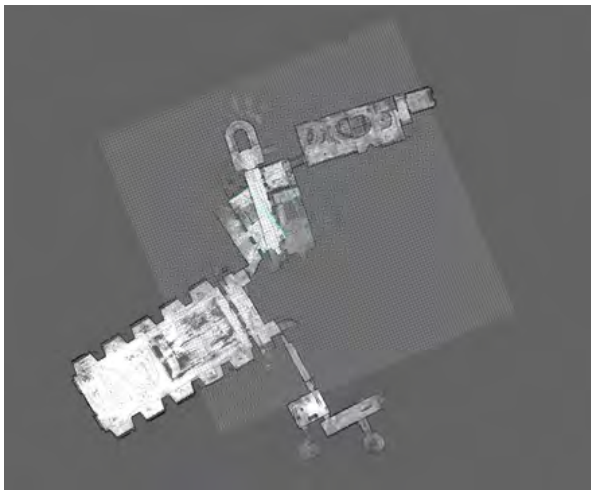


Fig. 2: A map generated with google cartographer displayed in Rviz [5]

Google cartographer can generate a map in 3D using only the LIDAR point clouds and the imu data. An example of the generated map is shown in Figure 2. Google cartographer has many advantages in comparison to other SLAM solutions, for example it can also use other sensor data like camera image, odometry or even GPS to improve the quality of the map and the localization process. In our use case, there are many requirements that must be met:

- The mapping does not have to run in real-time. The quality of the map has a higher priority.
- The localization should run in real-time.
- The algorithm should work for different sensor configurations and types.

References and figures

- [1] Tim Bailey and Hugh Durrant-Whyte. Simultaneous localization and mapping (SLAM): Part II. *IEEE robotics & automation magazine*, 13:108–117, 2006.
- [2] Ouster Company. os1 ouster LIDAR. <https://ouster.com/>, 2021.
- [3] Wolfgang Hess, Damon Kohler, Holger Rapp, and Daniel Andor. Real-Time Loop Closure in 2D LIDAR SLAM. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1271–1278. IEEE, 2016.
- [4] M. Pendleton et al. Perception, planning, control, and coordination for autonomous vehicles. *Machines*, 5:6, 2017.
- [5] Own representation.
- [6] Q. Zou et al. A Comparative Analysis of LiDAR SLAM-Based Indoor Navigation for Autonomous Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2021.

Empirische Untersuchung einer Scrum-basierten Arbeitsweise am Beispiel eines IT-Infrastrukturprojektes

Lara Rehbani

Anke Bez

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Robert Bosch GmbH, Stuttgart

Einleitung

Unternehmen sind in der heutigen schnell lebigen Zeit mit der sogenannten VUCA (Volatility, Uncertainty, Complexity, Ambiguity)-Welt konfrontiert, in dem Projekte komplexer werden, Anforderungen sich intern und extern stetig ändern und Vorhersehbarkeiten nicht mehr möglich sind. Um den neuen Rahmenbedingungen der Wirklichkeit entgegenzukommen, ist eine Neuorientierung zwingend notwendig. Klassische Vorgehensweisen gelten daher in Bereichen des Komplexen als nicht mehr zielführend.

Problemstellung

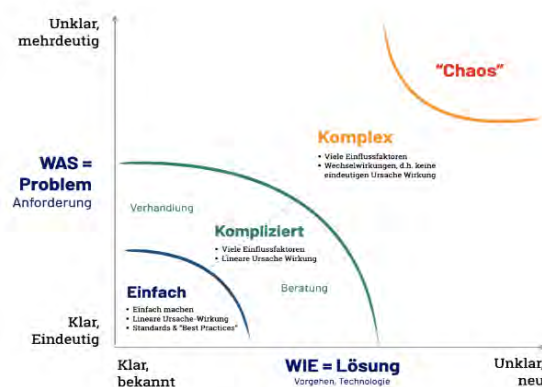


Abb. 1: Stacey-Matrix [2]

Das in der Arbeit untersuchte IT-Infrastrukturprojekt hat 2019 mit einer klassischen Vorgehensweise angefangen. Man hat gelernt, dass allzu planvolles Vorgehen und Strukturen in komplexen Bereichen (s. Abb. 1) nicht mehr funktionieren.

Daher wurde 2020 ein Umstieg auf eine agile, Scrum-basierte Arbeitsweise gewagt. Es handelt sich um eine radikale Veränderung der Neugestaltung und Neufindung der Prozesse, Strukturen und Kultur, quasi um einen grundlegenden Paradigmenwechsel.

Dabei sind neue Rollen wie Product Owner und Scrum Master, neue Wertebereiche mit kontinuierlicher Verbesserung sowie radikale Transparenz, horizontale Kommunikation, Gleichberechtigung in einer hierarchiefreien Struktur entstanden. Die Mitarbeiter des Development Team sind Netzwerker und Experten in ihren Themenbereichen. Sie haben ihr gesamtes Arbeitsleben lang klassisch gearbeitet. Gerade im Hinblick auf sie bedeutet es, die Ausrichtung ihrer bisherigen Haltung und Denkweise neu zu setzen.

Zielsetzung

Das Ziel dieser Arbeit ist es, zu überprüfen, inwieweit eine Scrum-Vorgehensweise in einem IT-Infrastrukturprojekt eingesetzt werden kann, welche Vorteile es mit sich bringt, welche Nachteile und Herausforderungen. Es soll untersucht werden, inwieweit sich das Scrum-Rahmenwerk einsetzen lässt, in welchen Bereichen Gemeinsamkeiten zur Theorie vorliegen und wo Änderungen vorgenommen werden müssen. Gerade der Umstieg von klassisch auf agil ist für das Development Team eine sehr große Herausforderung, schwierig und fordert daher eine Umstellung der Denk- und Arbeitsweise. Nun gilt es herauszufinden, wie sich das Team in der agilen Arbeitskultur zurechtfindet, was es braucht, um agiler zu werden.

Das agile Mindset

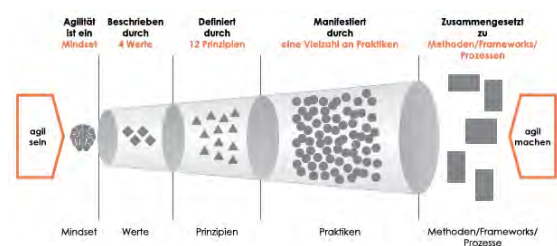


Abb. 2: Agil sein vs. Agil machen [3]

Der linke Bereich der Abbildung 2 ist der entscheidendere Teil im Kontext der Agilität. Hier ist die Entwicklung zur lernenden Einheit und das Mindset angesiedelt. Von vier Werten beschrieben, durch zwölf Prinzipien definiert und durch eine Menge an Praktiken ist das agile Machen auf der rechten Seite dargestellt. Hierunter findet Scrum seine Anwendung [3].

Es muss dem Team vor Augen geführt werden, dass es nicht ausreicht, oberflächlich nur Scrum durchzuführen. Agil machen (doing agile) beginnt in erster Linie mit dem agilen Mindset, dem agil sein (being agile).

Vorgehensweise



Abb. 3: Ablauf eines Retrospective Meeting [1]

Zunächst wird das theoretische Fundament an Wissen soweit aufgebaut, um das Projekt näher untersuchen zu können.

Um zu eruieren, wie schließlich Scrum im Projekt durchgeführt wird, werden zunächst alle Scrum-Events

als „stille Beobachter“ unter die Lupe genommen und näher studiert. Als außenstehende Person hat man einen anderen Blickwinkel auf die Geschehnisse.

Des Weiteren wird der Scrum Master und Product Owner unterstützt. Sprint Retrospectives werden vorbereitet (s. Abb.3) und moderiert, das Product Backlog bereinigt und priorisiert.

Ausblick

Ein Blick in das Innere des Projekts zeigt, wie starr einige Mitarbeiter noch an den traditionellen Arbeits- und Denkweisen hängen.

Besonders in den Retrospektiven, welches als wichtigste Zeremonie in Scrum gelten, ist der Widerstand bezüglich Teilnahme hoch. Dass es darum geht, die Iteration auf die Teamzusammenarbeit zu überprüfen und für die Zukunft anzupassen, wird nicht als solches empfunden.

Darüber hinaus gehen die Auffassungen der unterschiedlichen Rollen weit auseinander, wenn es um Transparenz, Selbstorganisation, -führung geht. Das deutet auf unzureichende Kommunikation im Team hin.

Was in der Arbeit beschrieben ist, sind Erkenntnisse, die sich über die letzten drei Monate entwickelt haben, und sich in den nächsten Monaten weiterentwickeln werden. Es handelt sich um eine Momentaufnahme. So wie das Umfeld sich stetig verändert, so entwickelt und passt sich die Agilität selbst im Projekt durchgehend an ihre Umwelt an.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Andreas Diehl. Stacey Matrix – Wann der Einsatz agiler Methoden notwendig und wirksam ist. <https://digitaleneuordnung.de/blog/stacey-matrix/>, 2021.

[3] Torsten Scheller. *Auf dem Weg zur agilen Organisation*. Vahlen, 2017.

Cloud Service Provider Evaluierung auf Basis von Infrastructure as Code Unterstützung

Julian Schallensmueller

Kai Warendorf

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Novatec Consulting GmbH, Leinfelden-Echterdingen

Einleitung und Motivation

Cloud Computing ist heute in vielen Bereichen ein fester Bestandteil des Alltags. Beim Begriff "Cloud" kommen bei vielen Menschen wohl als erstes diverse Cloud Storage Angebote in den Sinn: Dropbox, OneDrive und ähnliche Software finden sowohl privat als auch geschäftlich breite Verwendung. Neben den allgemein bekannten Cloud Storage Services sind auch die Streaming Plattform Netflix oder Microsoft Office 365 rein Cloud basierte Dienste. Alle der oben genannten Produkte gehören dem Service Modell *Software as a Service* (SaaS) an. SaaS stellt den umsatzstärksten Bereich dar, gefolgt von *Infrastructure as a Service* (IaaS) an zweiter Stelle (vgl. [1] S. 22). Durch die Auswirkungen der COVID-19-Pandemie erfährt besonders das IaaS Segment ein enormes Wachstum (vgl. [5], [1] S. 22) da Homeoffice und der diesbezügliche Ausbau von IT-Infrastruktur im letzten Jahr von vielen Firmen besonders stark vorangetrieben wurde. Dieser technologische Wandel forciert eine Anpassung der Arbeitsprozesse, Methoden und Tools. Die Umsetzung dessen stellt aktuell für viele Unternehmen eine große Herausforderung dar. Das erfolgreiche Meistern dieser Herausforderung setzt ein grundlegendes Ändern und Evaluieren der Arbeitsmethodik voraus. In der Vergangenheit bewährte Arbeitsprozesse sind im allgemeinen mit der Entwicklung Cloudbasierter Systeme nicht kompatibel, da diese eine iterative Arbeitsweise mit kontinuierlichen

Anpassungen erfordert. Insbesondere lineare Modelle wie das Wasserfallmodell stehen in starkem Kontrast zu den heutigen Anforderungen. Das volle Potential von Cloud-Technologien kann daher nur durch den Einsatz entsprechend moderner Arbeitsmethoden ausgeschöpft werden (vgl. [2] S. 3).

Zielsetzung

Die für die Industrie wichtigsten Plattformen für Cloud Computing sind Amazon AWS, Microsoft Azure und Google Cloud Platform. Viele Unternehmen besitzen bereits Infrastruktur-Systemen auf diesen Plattformen, die manuell erstellt und konfiguriert wurden. Mit wachsender Größe und Komplexität wird das Management dieser Systeme immer komplizierter und fehleranfälliger. Aus diesem Grund wird eine effektivere Arbeitsmethode durch den Einsatz von *Infrastructure as Code* Prinzipien (IaC) und den dazugehörigen Tools angestrebt. IaC umfasst die folgenden Aufgaben: Infrastruktur Templating, Bereitstellung und Management von Infrastruktur, die initiale Installation und Konfiguration von Anwendungen sowie das Anpassen der installierten Software im weiteren Betrieb. Das Tool Terraform kommt im zweiten der genannten Bereiche zum Einsatz (Abb. 1). Die Unterschiede, die beim Einsatz von Terraform mit den drei genannten Cloud Computing Plattformen festgestellt werden können, sind der zentrale Untersuchungsgegenstand dieser Arbeit.

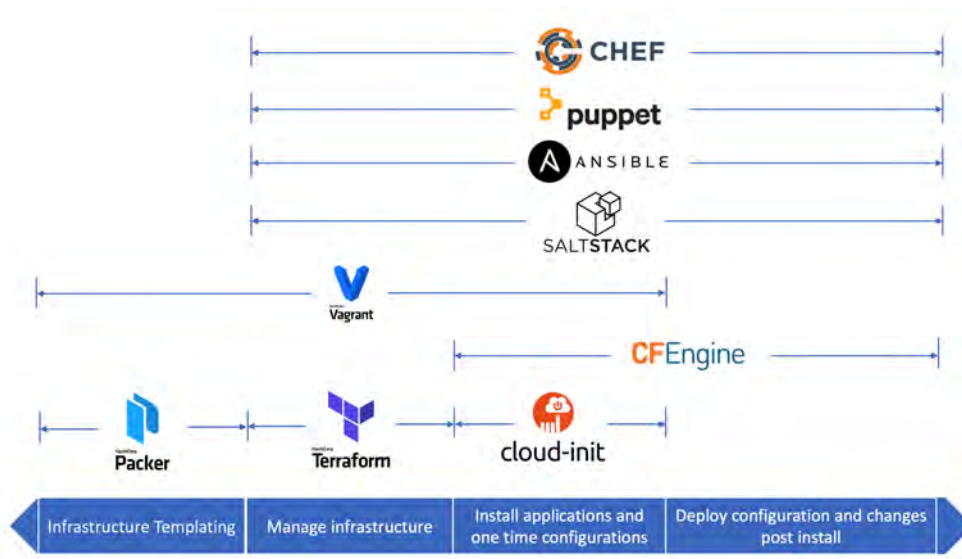


Abb. 1: Infrastructure as Code Tools Übersicht [3]

Infrastructre as Code mit Terraform

Terraform verwendet eine deklarative domänenspezifische Sprache, die sowohl Maschinenfreundlichkeit als auch Menschenlesbarkeit vereint. Der grundlegende Terraform Workflow besteht aus vier Befehlen: *init*, *plan*, *apply* und *destroy* (Abb. 2). Der *terraform init* Befehl konfiguriert die Terraform CLI und installiert die benötigten Terraform Provider. Diese Provider sind im Kern Plug-Ins welche Terraform ermöglichen mit den verschiedenen Plattformen zu interagieren, es gibt für gewöhnlich einen Provider für jede Plattform. Mit *terraform plan* wird erfasst welche Elemente des Systems bereits existieren, welche Teile neu im Code definiert wurden und ermittelt dann die Abhängigkeiten die zwischen diesen bestehen. Daraus wird ein Execution Plan erstellt, der die Schritte definiert mit welchen der gewünschte Zustand erreicht werden soll. Die Ausführung eines Execution Plans wird dann durch *terraform apply* angestoßen. Ein nicht mehr benötigtes System kann durch *terraform destroy* wieder vollständig abgebaut werden. Terraform erlaubt es große und komplexe Systeme in kleineren, voneinander unabhängigen Modulen zu beschreiben. Diese beinhalten in der Regel alle Ressourcen die für eine bestimmte Komponente benötigt werden, es gibt zum Beispiel das *eks* Modul für AWS das es vereinfacht einen Kubernetes-Cluster in AWS bereitzustellen. Solche Module werden von der Community in einer öffentlichen Public Module Registry zur freien Verwendung veröffentlicht, grob vergleichbar mit der npm Registry für Nodejs.

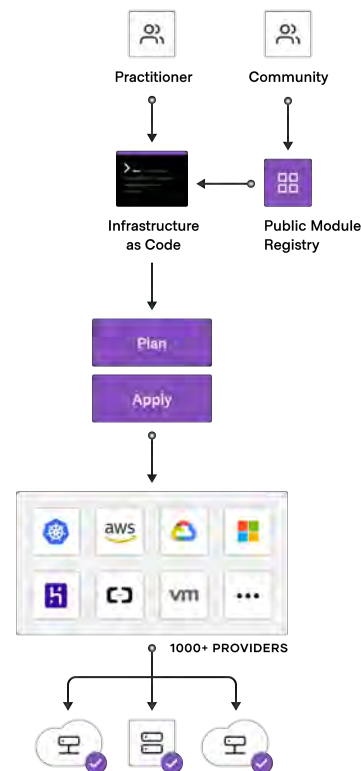


Abb. 2: Terraform Funktionsprinzip [4]

Aufbau des zu vergleichenden Systems

Der exakte Aufbau des zu untersuchenden Systems wird im Verlauf der Arbeit noch konkretisiert. Es soll in jedem Fall aus einer Reihe von typischen Computer Ressourcen zusammengesetzt sein: Netzwerk, Virtuelle Maschinen, Datenbanken, Storage und ein Kubernetes-

Cluster. Je nach Plattform werden sich die Systeme leicht unterscheiden, die High-Level Struktur, die mithilfe von Modulen definiert werden soll, wird dabei beibehalten.

Auswahl der Vorgehensweise und Vergleichskriterien

Um einen möglichst gut strukturierten und objektiven Vergleich zu gewähren, soll nach dem in der ISO/IEC 25040 beschriebenen Prozess zur Evaluierung von Software vorgegangen werden. Als Grundlage der Vergleichskriterien soll die ISO/IEC 25010 *System und Software-Engineering - Qualitätskriterien und Bewertung von System- und Softwareprodukten (SQuaRE)* - *Bewertungsprozess* dienen.

Ausblick

Im Vorfeld ist zu erwarten, dass die drei untersuchten Cloud Service Provider AWS, Azure und Google Cloud Platform sich in etwa gleichermaßen für den Einsatz in einem gewöhnlichen Szenario eignen. Es wäre daher interessant, den Vergleich in Zukunft auf weitere Anbieter zu erweitern. Alibaba, Oracle, IBM und Tencent bieten die nächst größten Plattformen für Cloud Computing an. Ein ebenfalls relevantes Thema ist der Vergleich von Terraform mit anderen IaC Tools. Das Tool Pulumi erfüllt die gleiche Aufgabe wie Terraform, ist jedoch jünger und hat daher noch nicht die selbe Verbreitung wie Terraform erreicht. Der auffälligste Unterschied besteht im Einsatz von bekannten imperativen Programmiersprachen, welche weitere Möglichkeiten eröffnen könnten.

Literatur und Abbildungen

- [1] Anders Lisdorf. *Cloud Computing Basics: A Non-Technical Introduction*. Apress Media LLC, 2021.
- [2] Kief Morris. *Infrastructure as Code*. O'Reilly Media, 2 edition, 2020.
- [3] Ramnath Nayak. When to Use Which Infrastructure-As-Code Tool. <https://dzone.com/articles/when-to-use-which-infrastructure-as-code-tool>, 05 2019.
- [4] ohne Verfasser. How Terraform Works. <https://www.terraform.io/>, 2012.
- [5] ohne Verfasser. Gartner Says Four Trends Are Shaping the Future of Public Cloud. <https://www.gartner.com/en/newsroom/press-releases/2021-08-02-gartner-says-four-trends-are-shaping-the-future-of-public-cloud>, 08 2021.

Evaluierung von Ähnlichkeitsmaßen für Zeichensequenzen im Machine Learning

Matthias Schoelpple

Gabriele Gühring

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Es stehen immer mehr Informationen zur Verfügung. Viele davon in Textform. Um diese Informationen analysieren und verwenden zu können gibt es mehrere Möglichkeiten. Einen Text von Menschen lesen zu lassen hat Vorteile. Man muss nicht viel im Voraus investieren. Sobald ein Mensch zu lesen anfängt, beginnt die Informationsübertragung. Außerdem kann die Person Zusammenhänge erkennen, die auf dem neuen Wissen beruhen. Durch die Automatisierung der Textanalyse steht das Wissen von geschriebenem Text auch zur Verfügung, wenn Mitarbeiter ein Unternehmen verlassen. Wissen, das in Datenbanken gespeichert ist, bleibt immer vollständig und vermindert sich mit der Zeit nicht. Computer können 24/7 Texte analysieren und somit kann die Flut von Informationen analysiert werden.

Aufgabenstellung

Evaluierung von Ähnlichkeitsmaßen für Zeichensequenzen im Machine Learning soll dabei helfen, Texte zu interpretieren und nützliches Wissen aus dem Geschriebenen zu generieren. Dies soll auch für nicht sprechende Texte realisiert werden. Das sind Texte, die nicht der typischen Satzstruktur und Grammatik folgen wie z. B. Produktbeschreibungen wie „apple iphone 8 plus 64gb silver“, „samsung galaxy s8 sm g950f 14.7 cm 5.8 4 gb 64 gb single sim 4g“ oder „panasonic lumix g9 mirrorless camera with lumix 12 60mm lens“. Diese nicht sprechenden Texte beinhalten auch wichtige Informationen. Wenn man auf einer Shopping-Seite im Internet Suchbegriffe eingibt, muss die Künstliche Intelligenz im Hintergrund in der Lage sein auf die gewünschten Produkte zu referieren. Woher weiß die KI, was gesucht wird? Woher weiß die KI, dass die Suchanfrage „Waschmaschine Bosch“ und „WAU28U00 Serie 6“ auf dasselbe Produkt verweisen? Damit beschäftigt sich die Arbeit zur Evaluierung von Ähnlichkeitsmaßen für Zeichensequenzen im Machine Learning.

Konzept

Prinzipiell geht es darum festzustellen in welcher Umgebung bestimmte Worte oder Zeichensequenzen auftreten. Die Worte „apple“ und „iPhone“ kommen oft in demselben Text oder Kontext vor. Daher lernt der Algorithmus, dass diese beiden Worte sich ähnlich sind. Nun ist es egal nach welchem von beiden Begriffen man sucht: Es werden beide male sehr ähnliche Produkte vorgeschlagen.

Um ein Ähnlichkeitsmaß zwischen Zeichenketten (z. B. Worte oder „64gb“), bestimmen zu können werden die Zeichenketten in Vektoren umgewandelt. Diese Vektoren besitzen bis zu 500 Dimensionen und können so jeder Zeichenkette eine einzigartige Bedeutung geben, die sie in Relation zu anderen Zeichenketten setzt. So sind sich der Vektor für „samsung“ und der Vektor von „galaxy“ sehr ähnlich. Abgesehen von den Ähnlichkeitsmaßen kann man mit dieser mathematischen Abstraktion Rechenoperationen durchführen. So findet beispielsweise folgende Rechnung statt:

„King“ – „Man“ + Woman“ = „Queen“

Dies zeigt, dass die Vektorrepräsentationen der Worte tatsächlich dem Sinn der Worte entsprechen.

Vorhandene Methoden

Es werden drei bereits existierende Methoden untersucht.

1. Word2Vec [2] von Google
2. GloVe von Stanford University
3. fastText von Facebook

Word2Vec ist die erste Methode, die zu Interpretation von geschriebenem Text entwickelt wurde (2013). Für diese Methode gibt es zwei Vorgehensweisen.

1. Continuous Bag of Words (CBOW) ist ein Vorgehen, bei dem man den Kontext benutzt, um das sich dazwischen befindliche Wort vorherzusagen.
2. Skip-gram ist ein Vorgehen, bei dem ein Wort benutzt wird, um den Kontext vorherzusagen

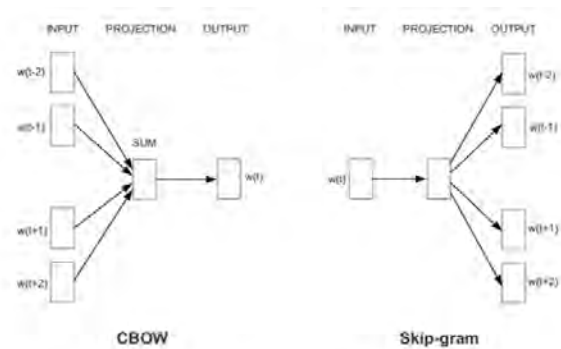


Abb. 1: Vergleich CBOW und Skip-gram [3]

GloVe und fastText bauen auf Word2Vec auf und passen das Vorgehen an die eigenen Bedürfnisse an. GloVe arbeitet mit einer globalen co-occurrence Matrix. Diese Matrix beschreibt das gemeinsame Auftreten von Worten.

	bank	career	education	employment	example	experience	Florida	home	human	insurance	job	jobs	monster.com	office	opening	opportunities	recruiters	resources	resume	resumes	retirement	search	state	unemployment	work
bank	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
career	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
education	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
employment	3	2	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
example	--	35	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
experience	42	35	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Florida	--	9	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
home	--	7	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
human	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
insurance	1	2	6	--	--	2	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
job	28	2	4	4	--	4	1	2	3	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
jobs	1	1	4	16	--	6	--	--	17	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
monster.com	--	--	--	--	--	--	--	--	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
office	--	--	--	--	--	--	--	--	3	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
opening	--	--	--	--	--	--	--	--	20	30	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
opportunities	3	20	--	--	--	1	24	15	30	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
recruiters	--	--	--	--	--	--	--	--	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
resources	1	--	--	--	--	--	--	--	26	3	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
resume	1	100	157	100	--	2	2	16	16	--	30	16	--	2	--	--	--	--	--	--	--	--	--	--	--
resumes	--	--	--	--	--	--	--	--	2	--	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--
retirement	--	--	--	--	--	--	3	1	2	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
search	1	1	7	--	--	--	--	--	30	4	--	--	--	--	--	--	--	--	1	1	--	--	--	--	--
state	2	--	15	--	--	--	3	--	6	8	--	--	--	--	--	--	--	--	--	--	1	2	--	--	--
unemployment	--	--	--	--	--	--	--	--	10	--	11	--	--	--	--	--	--	--	--	--	--	--	4	--	--
work	1	2	--	--	--	1	15	--	2	5	3	--	--	--	--	--	--	--	--	--	--	--	2	--	--

Abb. 2: GloVe Co-occurrence Matrix [1]

Mit der Kosinus-Ähnlichkeit wird die Wahrscheinlichkeit berechnet, mit der die beiden Worte gemeinsam auftreten.

FastText teilt die einzelnen Worte in n-grams auf. Das bedeutet, dass z. B. das Wort „<Elefant>“ in die Einzelteile

- <E
- ele
- lef
- efa
- fan
- ant
- nt>

aufgeteilt wird. Für die Ermittlung der Ähnlichkeit werden die Worte aus diesen n-grams wieder zusammengesetzt. Dies hat den Vorteil, dass das Wort „Elefant“ und das Wort „Elefanten“ voneinander lernen können. Dies ist bei Word2Vec nicht der Fall.

Ziel

Das Ziel dieser Arbeit ist es, eine Methode zu entwickeln, mit der man auch nicht sprechende Texte interpretieren kann. Word2Vec, GloVe und fastText sind für sprechende Texte entwickelt worden. Es wird untersucht, wie gut sich die bereits bekannten Verfahren für nicht sprechende Texte eignen und was man verändern muss, um sie anzupassen.

Literatur und Abbildungen

- [1] Gabor Melli. Co-occurrence Matrix. http://www.gabormelli.com/RKB/Word-Word_Co-Occurrence_Matrix, 2015.
- [2] Tomas Mikolov. a method that can automate the process of generating and extending dictionaries and phrase tables. <https://arxiv.org/abs/1310.4546>, 09 2013.
- [3] Tomas Mikolov. CBOW model und Skip-gram model. <https://arxiv.org/pdf/1309.4168v1.pdf>, 2013.

Batterieladesimulation zur Evaluierung von Lastverteilungsalgorithmen in einem Ladepark

Valentin Schuele

Reiner Marchthaler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma eSystems MTG GmbH, Wendlingen a. N.

Motivation

Die Verkehrs- und Energiewende ist eines der wichtigsten Themen der gegenwärtigen Zeit um dem Klimawandel entgegenzuwirken und den Menschen eine saubere Alternative zu den bisherigen Ansätzen des Verkehrs als auch der Energiegewinnung bieten zu können. Insbesondere im Hinblick auf die Elektrifizierung und dem damit verbundenen steigenden Energiebedarf muss eine Lösung, z.B. für eine dynamische Leistungsverteilung auf alle vorhandenen Lasten im Stromnetz, gefunden werden. Hierbei stellt sich die wissenschaftliche Frage, wie man das Ladeverhalten eines Batteriespeichers für E-Fahrzeuge programmatisch abbilden kann, um in einer Simulation eine effiziente Lastverteilung für mehrere Fahrzeuge zu erproben.

Aufgabenstellung

Gegenstand dieser Arbeit ist es, die Ladekurve eines Energiespeichers von Elektrofahrzeugen zu abstrahieren und programmatisch abzubilden. Hierbei soll das Laden des Energiespeichers, basierend auf den Ladekurven und dem bekannten Ladeverhalten der Fahrzeuge, simuliert werden. Basierend auf der Simulation von einem Energiespeicher soll zudem eine Simulationsumgebung für mehrere Ladesäulen und Ladeparks entworfen werden können. Diese Simulation kann im weiteren für die Erprobung von Lastverteilungsalgorithmen verwendet werden.

Grundlagen

Um das Ladeverhalten von E-Fahrzeugen verstehen zu können muss vorerst betrachtet werden, was eine Batterie ist und welche Batterietypen existieren. Batterien sind Energiespeicher, welche Energie einspeichern (laden), speichern und ausspeichern (entladen) können. Prinzipiell unterscheidet man Batterien in

Primärbatterien und Sekundärbatterien, auch Akkumulatoren genannt. Primärbatterien lassen sich, im Gegensatz zu Sekundärbatterien, nicht wiederaufladen. D.h. dass sich die ablaufende Redoxreaktion, welche eine Spannung induziert, bei Primärzellen, im Gegensatz zu Akkumulatoren, nicht wieder umkehren lässt.

Je nach eingesetzten Materialien innerhalb der Batterie eignen sich die verschiedenen Batteriearten (Li-Ion, Pb, NiCd, ...) für unterschiedliche Einsatzmöglichkeiten besser oder schlechter. Dennoch haben folgende Stressfaktoren einen erheblichen Einfluss auf das Lade- und Entladeverhalten sowie die Lebensdauer von Energiespeichern:

- Umgebungs- und Zelltemperatur
- Kalendarische Alterung
- Entladetiefe
- Lade- und Entladerate
- Lade- und Entladestromstärke

Um maximale Sicherheit und eine möglichst lange Lebensdauer einer Batterie und deren Zellen zu gewährleisten, werden in Elektrofahrzeugen aufgrund der hohen Kapazität und der hohen Leistung sogenannte Batteriemanagement-Systeme (BMS) eingesetzt. Batteriemanagement-Systeme messen und regulieren interne Abläufe, wie beispielsweise die Temperatur und das thermische Management. Bei der Erfassung der gesamten Messdaten und deren Auswertung muss das Management-System einen guten Kompromiss zwischen Effizienz, Lebensdauer und Kosten finden, was die Entwickler eines BMS vor große Herausforderungen stellt. (Vgl. [3])

Abbildung 1 zeigt eine Übersicht über verschiedene DC-Ladekurven von verschiedenen Fahrzeugen:

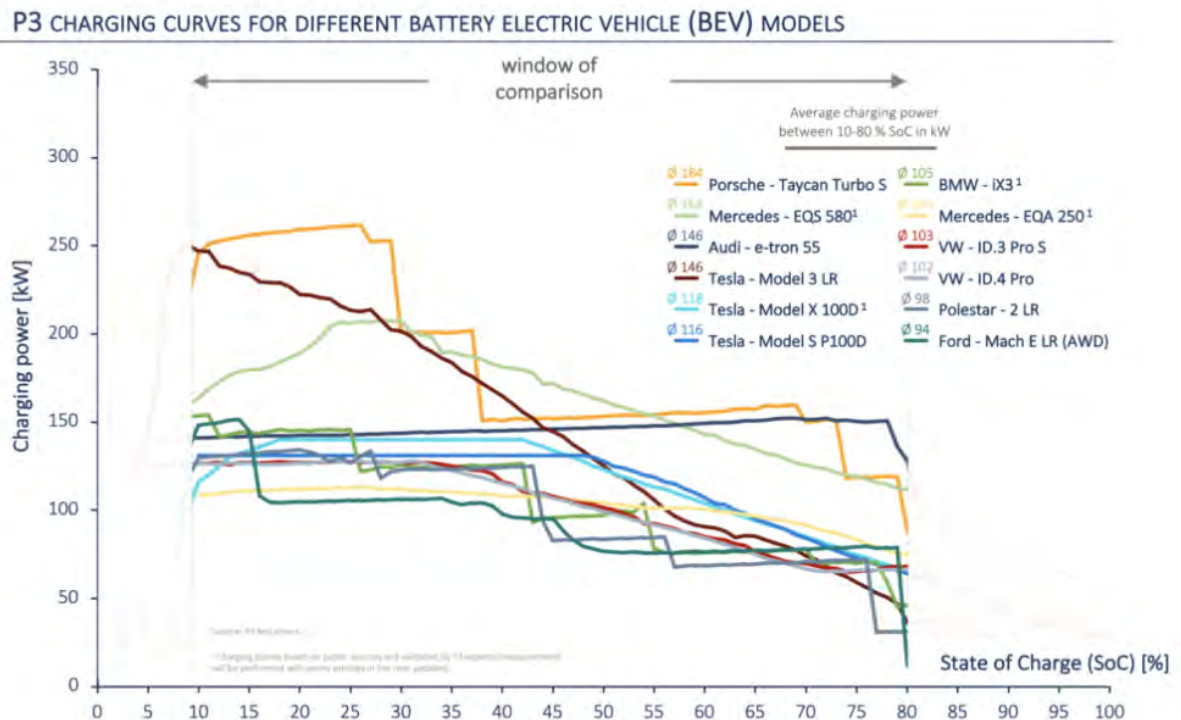


Abb. 1: Übersicht über Ladekurven verschiedener Fahrzeuge [2]

Wie in Abbildung 1 zu erkennen ist, ist das DC-Ladeverhalten der verschiedenen Fahrzeuge sehr unterschiedlich, was auf die Batteriemangement-Systeme zurückzuführen ist.

Da E-Fahrzeuge entweder über Gleichstrom (DC) oder über Wechselstrom (AC) geladen werden können und man online nur DC-Schnellladekurven findet, wird angenommen, dass die Ladekurven beim AC-Laden mit bis zu 22kW konstant und linear verläuft. Die Annahme basiert auf der Begründung der geringeren Hitzeentwicklung durch die geringere Leistungszugabe, da der Wechselstrom durch einen Gleichrichter zu Gleichstrom umgewandelt werden muss.

Anwendungsfälle

Vor der Umsetzung und Implementierung des Ladeverhaltens mussten *Use Cases* entworfen werden, in

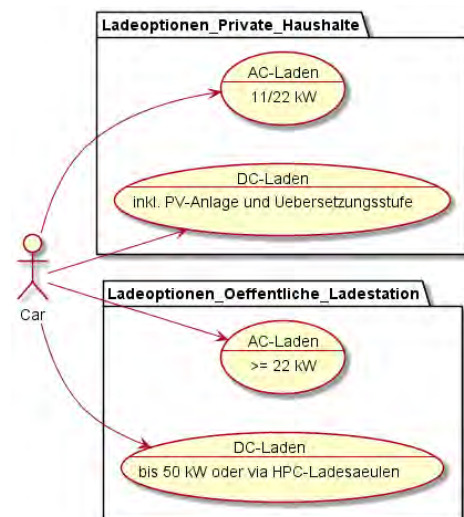


Abb. 2: Entwickelte Use Cases für Anwendung der Ladesimulation [1]

welchen die Batterieladesimulation verwendet werden können soll. Abbildung 2 zeigt die entworfenen *Use Cases*.

Vorbereitung und Umsetzung der Batterieladesimulation

Aufgrund der unterschiedlichen Steuerungen der Batteriemanagement-Systeme entschied man sich dazu, die öffentlich verfügbaren Ladekurven mithilfe eines *WebPlotDigitizers*, welcher mithilfe der Koordinaten und Skalen der x- und y-Achsen die Punkte der Kurven berechnet, zu diskretisieren.

Die diskretisierten Werte wurden anschließend gespeichert und im Laufe der Implementierung in den Programmcode, als Referenzwerte für die Simulation, eingepflegt.

Vorbereitung der Restlichen Komponenten

Da, wie in den *Use Cases* zu sehen ist, in der Simulation auch eine Photovoltaik-Komponente enthalten sein soll, um das Laden über Solarenergie miteinzuberechnen, musste vorab auch die Energieerzeugung von PV-Anlagen analysiert werden. Hierbei wurden ebenfalls Werte aus einem Bild, welches zur Verfügung stand, diskretisiert und für die spätere Implementierung normiert abgespeichert, um in der Leistungserzeugung einen beliebigen Kilowatt-Peak Wert simulieren zu können.

Zudem wurden ebenfalls die durchschnittlichen Werte des Lastgangs eines Haushalts sowie des Gesamtnetzes diskretisiert abgespeichert. Dies ist notwendig, um eine möglichst genaue Simulation realisieren zu können. Die Werte, welche vom Bundesverband der Energie- und Wasserwirtschaft (BDEW) zur Verfügung gestellt werden (siehe Standardlastprofil), mussten aufgrund der Normierung auf 1000 kWh durch den BDEW nicht selbst normiert werden.

Umsetzung der Simulation

Die Simulation sollte nach Anforderung der betreuenden Firma in der Programmiersprache GoLang geschrieben werden und den Ladevorgang sekundlich simulieren. Um die Simulation umzusetzen, musste vor Beginn der Implementierung ein zeitdiskretes Simulations-Framework gefunden werden. Die gewählte Bibliothek mit dem Namen SimGo, welche auf Github zur Verfügung steht (siehe [4]), entsprach den Anforderungen weshalb diese für die Simulation verwendet wurde.

Folgende Komponenten wurden in der Implementierung berücksichtigt:

- Auto (inkl. Fahrzeugeigenschaften wie Durchschnittsverbrauch, maximale Leistungsaufnahme, ...)

- Batterie (Fahrzeugbatterie oder Hausspeicher für PV)
- Ladepunkt/Ladesäule
- PV-Modul (mit normierten Werten)
- Lastgang (mit normierten Werten)
- Energiezusammenflussspunkt, um gesamte Energie (produziert/verbraucht) zu sammeln
- (später: Lastmanagement-Komponente)

Zu Beginn der Implementierung wurden im ersten Schritt die verschiedenen Komponenten und deren Eigenschaften erstellt sowie deren diskretisierten Werte in sog. maps, ähnlich wie ein array, als Referenzwerte abgespeichert. Ziel im ersten Schritt war es, einen kompletten Tag zu simulieren, wobei die für das Fahrzeug zur Verfügung stehende Energie von der von der PV-Anlage erzeugten Energie, der im Haushalt verbrauchten Energie sowie von dem vom Energiebetreiber vorgeschriebenen maximalen Leistungslimit abhängt. Aus diesem Grund wurde ein sogenannter *Grid Connection Point* implementiert, welcher den Tag simuliert und die gesamte Energie welche zum Zeitpunkt t erzeugt, bereitgestellt und verbraucht wird, sammelt und den verschiedenen Komponenten zur Verfügung stellt.

Ausblick

Basierend auf der Simulation eines Tages mit der zur Verfügung stehenden Energie pro Tag und Uhrzeit sollen im Laufe der Arbeit Lastverteilungs- und Lademanagement-Algorithmen getestet und evaluiert werden. Zudem sollen im Laufe der Implementierung auch Nutzer-, Nutz- und Fahrverhaltensdaten mit in die Simulation einbezogen werden, um eine möglichst genaue Näherung an das reale Verhalten eines durchschnittlichen Nutzers mit einem durchschnittlichen Verbrauch zu erhalten.

Fazit

Da der Energiebedarf stetig steigt und die Stromnetze an ihre Belastungsgrenze kommen, ist es wichtig, sich mit dem Thema Last- und Lademanagement auseinander zu setzen. Durch eine Simulation verschiedener Szenarien anhand einer dynamisch realisierten Simulation wird die Testung verschiedener Algorithmen deutlich vereinfacht, da ein gewisser Datensatz vorhanden ist, welcher auf echten Daten basiert. Die Arbeit kann somit als wichtig betrachtet werden, da sie eines der wichtigsten Themen der Gegenwart und Zukunft aufgreift und behandelt.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Markus Hackmann. P3 Charging Index - Vergleich der Schnellladefähigkeit verschiedener Elektrofahrzeuge aus Nutzerperspektive. <https://www.p3-group.com/p3-charging-index-vergleich-der-schnelllade-faehigkeit-verschiedener-elektrofahrzeuge-aus-nutzerperspektive-update-2021/>, 04 2021.
- [3] Kari-Björn Holves. Auslegungsaspekte von Batteriepacks und Batteriemanagement-Systemen, 2018.
- [4] Felix Schütz. Github - fschuetz04/simgo. <https://github.com/fschuetz04/simgo>, 2021.

Entwicklung eines Kalkulationsmodells für Softwaresysteme – exemplarische Umsetzung für Siemens anhand von intralog WMS –

Sven Schuler

Anke Bez

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Siemens AG, Stuttgart

Einleitung und Motivation

Digitale Güter sind nach Clement, Schriber et al. definiert als „immaterielle Mittel zur Bedürfnisbefriedigung, die aus Binärdaten (0, 1) bestehen und sich mit Hilfe von IKT [(Informations- und Kommunikationstechnologien) Anm. d. Verf.] entwickeln, vertreiben oder anwenden lassen.“ [1] Diese Güter unterscheiden sich im Hinblick auf die Kostenstruktur von physischen Gütern. Wie die Abbildung 1 zeigt, sind bei digitalen Gütern keine variablen Kosten vorhanden. Dies liegt daran, dass sich digitale Güter sehr einfach vervielfältigen und vertreiben lassen. Software zum Beispiel kann mit sehr wenig Aufwand per Download vertrieben werden, wodurch kaum Kosten anfallen. Dagegen sind die Forschungs- und Entwicklungskosten, die für die Erstellung des Produktes nötig sind, enorm hoch. Diese Fixkosten nennt man „First Copy Costs“. [4]

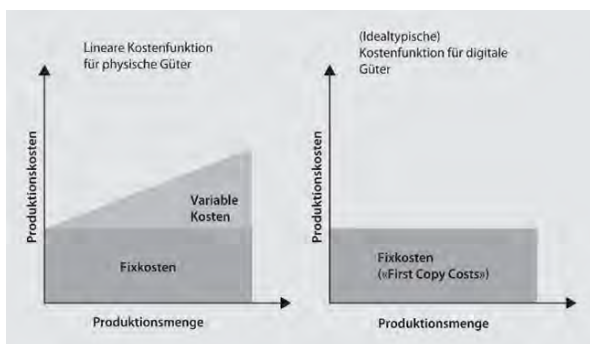


Abb. 1: Kostenfunktionen von physischen und digitalen Gütern [4]

Würde man nun den Preis auf Basis der variablen Stückkosten berechnen, erhält man ein kostenloses Angebot. [5] Somit ist für diese Art von Produkten eine andere Preiskalkulation erforderlich. Es gibt jedoch auch digitale Güter, für die das nicht gilt. Ein Beispiel hierfür ist Individualsoftware. Bei dieser fallen z.B. für die Entwicklung, Produktion oder den Absatz hohe Kosten für traditionelle Dienstleistungen an. [1] Bei Software as a Service (SaaS)-Produkten ist es durchaus sinnvoll, die Preise auf Basis der Kosten zu ermitteln. [3]

Problemstellung

Die Siemens AG hat im Bereich der Intralogistik das System ‚intralog WMS‘ entwickelt, das die Steuerung innerhalb eines Lagers übernimmt. Dieses System ist modular aufgebaut und kann somit für jeden Kunden individuell an dessen Anforderungen zusammengesetzt und angepasst werden. Für die Angebotskalkulation werden zunächst in einem Siemens-internen Gespräch die Projektdetails besprochen und dann im Gespräch mit dem Auftraggebenden abgestimmt. Diese Schritte werden so oft wiederholt, bis ein Angebot erstellt werden kann. In Abbildung 2 wird der bisherige Prozess dargestellt.

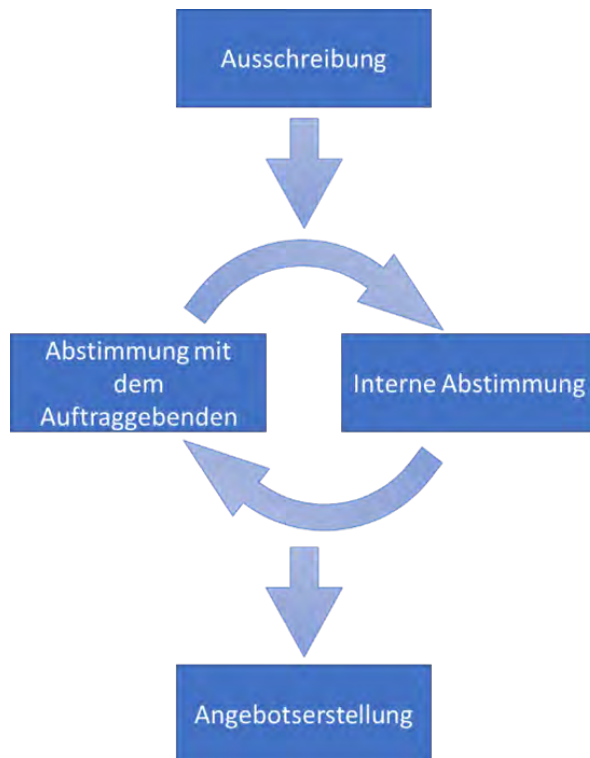


Abb. 2: Prozessablauf bisherige Kalkulationsmethode [2]

Die Schwierigkeit bei der derzeitigen Kalkulationsmethodik ist, dass die Werte auf den Erfahrungswerten und Einschätzungen einzelner Personen beruhen und somit wenig Transparenz gegenüber dem Kunden herrscht. Außerdem besteht bei den derzeitigen Mitarbeitenden ein ‚Kopfmonopol‘, da nur sie mit ihrer Erfahrung in der Lage sind, eine solche Einschätzung vorzunehmen. Aus Sicht von Siemens besteht die Gefahr, dass im Falle von Krankheit, Mitarbeiterfluktuation oder auch wenn neue Mitarbeitende hinzukommen, diese Aufgabe nicht oder zumindest nicht mit gleicher Qualität durchgeführt werden kann.

Erarbeitung eines Kalkulationsmodells

Um die beschriebene Problematik zu lösen, soll ein Kalkulationsmodell erarbeitet werden. Hierfür wurde

stellvertretend ein Kundenprojekt ausgewählt, auf Grundlage dessen das Modell entwickelt werden soll. In einem ersten Schritt wurde bereits eine Liste aller Softwaremodule mit einer funktionalen Beschreibung erstellt. Außerdem wurde eine Klassifizierung in fertige Module, Basis- und Rahmen-Module vorgenommen. Die fertigen Module können eins zu eins übernommen werden, wohingegen bei den Basis- und Rahmen-Modulen weitere Anpassungen nötig sind. Der Unterschied bei diesen Kategorien liegt im Umfang der vorhandenen Logik. Die Basis-Module liefern bereits eine gewisse Grundlogik, die bei den Rahmen-Modulen noch fehlt. Die Liste der Softwaremodule wurde in den nächsten Schritten mit Attributen angereichert, um die für die Kalkulation nötige Informationen zu liefern. Dadurch soll eine genauere Aussage getroffen werden, wie viel Aufwand für die Umsetzung der Kundenanforderungen nötig ist, um daraus die Entwicklungskosten bestimmen zu können.

Ausblick

In den nächsten Schritten sollen Fragen in das Modell mit aufgenommen werden, die für die Aufwandsschätzung relevant sind. Anschließend soll eine Kalkulation für das ausgewählte Kundenprojekt durchgeführt und bewertet werden. Das resultierende Modell soll vor allem die Vertriebsmitarbeitenden unterstützen, einen Richtpreis für das jeweilige Projekt zu kalkulieren. Darüber hinaus sollen aber auch andere Stakeholder, wie z.B. die Projektleitung oder das Management, davon profitieren. Für diese sind die Informationen, die den zeitlichen Aufwand betreffen, besonders relevant. Durch die im Modell hinterlegten Beschreibungen für die einzelnen Module ist es möglich, ohne tiefgreifende Softwarefachkenntnisse die notwendigen Module auszuwählen. In einem weiteren Schritt wäre es sogar denkbar, das Modell dem Kunden zur Verfügung zu stellen, damit dieser sich einen Überblick über den Funktionsumfang und die zu erwartenden Kosten verschaffen kann. Anhand der ausgewählten Module lässt sich dann der Preis für das System bestimmen, inklusive der zugehörigen Aufwände für die individuelle Anpassung.

Literatur und Abbildungen

- [1] Reiner Clement, Dirk Schreiber, Paul Bossauer, and Christina Pakusch. *Internet-Ökonomie*. Springer Gabler, 4 edition, 2019.
- [2] Eigene Darstellung.
- [3] Sonja Lehmann and Peter Buxmann. Preisstrategien von Softwareanbietern. *Wirtschaftsinformatik*, pages 519–529, 2009.
- [4] Jan Marco Leimeister. *Einführung in die Wirtschaftsinformatik*. Springer Gabler, 12 edition, 2015.
- [5] Ralf Peters. *Internet-Ökonomie*. Springer, 2010.

Kinematic Calibration of a 6-DOF Robotic Manipulator

Nicolas Schulz

Walter Lindermeir

Department of Computer Science and Engineering, Esslingen University

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

Introduction

The kinematic model of a robotic manipulator describes the position and orientation of the manipulator's Tool Centre Point (TCP), based on the link lengths and joint angles of the ideal robotic manipulator [6]. In order to bring the TCP to a certain pose (position and orientation), the control algorithm uses the inverse solution of the kinematics model to calculate the required angles of each joint [6]. This ideal model is often inaccurate, as geometric errors like length differences and axis misalignments are introduced during manufacturing [7]. Non-geometric errors, caused by elasticity and temperature drift, can also have an impact [7].

Poor accuracy is not only problematic when executing a delicate task, but also when programming multiple robots. The behaviour of each robotic manipulator should ideally be identical, so that two different manipulators can be used interchangeably with the same program. As inaccurate robotic manipulators have a unique behaviour, the program needs to be adjusted for each robot [7].

Kinematic calibration aims to determine the errors of the kinematic model, in order to improve the accuracy of a robotic manipulator [2]. The calibration process consists of modelling the kinematics of the robotic manipulator, measuring the positioning errors of multiple poses, identifying the error parameters and adjusting the kinematic model to compensate for these errors [7]. This process can improve the accuracy of robotic manipulators significantly [2].

Kinematic Model

A robotic manipulator can be abstracted into a kinematic chain of links and joints. The position and orientation of each joint is represented by an orthonormal reference frame. Through subsequent homogeneous transformations from one joint to the next, the relationship between the TCP pose and the global reference frame can be described:

$$T_7^0(\underline{\theta}) = T_1^0(\alpha_1) \cdot T_2^1(\alpha_2) \cdot \dots \cdot T_6^5(\alpha_6) \cdot T_7^6 \quad (1)$$

The configuration $\underline{\theta} = [\alpha_1, \alpha_2, \dots, \alpha_6]^T \in R^6$ denotes the angles of each joint, where each rotation is made around the z -axis of the respective reference frames. The last transformation, from the flange to the TCP, does not have a degree of freedom, as the end-effector is a fixed attachment [3].

Figure 1 shows the kinematic model of a 6-DOF robotic manipulator, according to the Denavit-Hartenberg convention [2].

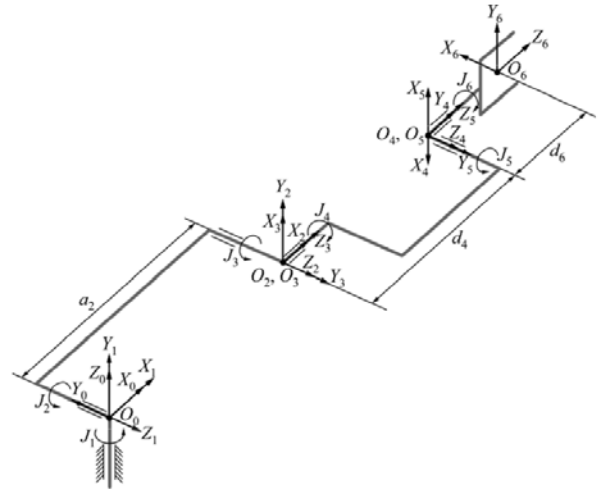


Fig. 1: Kinematic Model of a 6-DOF Robotic Manipulator [2]

Kinematic Error Model

A general model of the TCP pose is defined by

$$\underline{y} = \underline{y}(\underline{p} + \underline{\epsilon}, \underline{\theta}) \quad (2)$$

where $\underline{y} = [x, y, z, \alpha, \beta, \gamma]^T$ is the TCP pose using ZYX Euler angles for the orientation, $\underline{p} =$

$[p_1, p_2, \dots, p_n]^T$ are the n nominal kinematic parameters and $\epsilon = [\epsilon_1, \epsilon_2, \dots, \epsilon_n]^T$ are the n error parameters, which model the geometric errors of the actual robotic manipulator. [1]

Kinematic calibration aims to determine ϵ , by measuring the TCP pose \hat{y}_ν of a given measurement configuration θ_ν with a laser tracker and comparing it to the expected TCP pose y_ν . The expected TCP pose is calculated according to the nominal kinematic model, where $\epsilon = 0$. [3]

With the pose error

$$\Delta y_\nu = y_\nu - \hat{y}_\nu \quad (3)$$

and the measurement configuration θ_ν [3], the error parameters can be approximated according to the kinematic error model

$$\Delta y_\nu \approx \left. \frac{\delta y}{\delta \epsilon} \right|_{\epsilon=0, \theta_\nu} \cdot \epsilon = J_\nu \cdot \epsilon \quad (4)$$

where J_ν is the $(6 \times n)$ Jacobian matrix [1].

However, as the Jacobian of a single measurement represents an underdetermined system, multiple measurements m need to be made. The combined Jacobian J of shape $(6m \times n)$ must be at least a square matrix. Thus, the required number of measurements are calculated according to: $m \geq \frac{n}{6}$. The resulting kinematic error model is described as

$$\Delta y = J \cdot \epsilon \quad (5)$$

with $\Delta y = [\Delta y_1^T, \Delta y_2^T, \dots, \Delta y_m^T]^T$ and $J = [J_1^T, J_2^T, \dots, J_m^T]^T$ [1].

Model Reduction

The full kinematic error model has $n = 42$ different error parameters, which are not all independent, however. Therefore, a reduced kinematic error model is constructed to estimate the observable kinematic error parameters:

$$\Delta y = \hat{J} \cdot \hat{\epsilon} \quad (6)$$

with $\hat{J} = [\hat{J}_1^T, \hat{J}_2^T, \dots, \hat{J}_m^T]^T$ and $\hat{\epsilon} = [\hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_k]^T$ [1]. The number of independent parameters k is equal to the rank of J .

The independent parameters can be obtained through QR decomposition with column pivoting of J [3]:

$$J\Sigma = QR \quad (7)$$

The first $k = \text{rank}(J)$ columns of the upper triangular matrix R are independent. The independent columns of J , which relate to each error parameter, can be

ascertained through the permutation matrix Σ of the QR decomposition [3].

The QR decomposition is performed multiple times for m randomly generated measurement configurations within the robotic manipulator's workspace. The parameters that were found to be independent for most configurations are chosen to be the independent parameters of the kinematic model. The reduced model has $k = 28$ independent error parameters.

Optimal Configuration Selection

In order to reduce the impact of measurement noise and the calibration time, an optimal set of measurement configurations needs to be found [2]. The impact of measurement noise on the predicted kinematic error parameters is defined as

$$|\delta \hat{\epsilon}| \leq \frac{\sigma_1}{\sigma_k^2} \cdot |\delta \Delta y| \quad (8)$$

where σ_1 and σ_k are the largest and smallest singular values of \hat{J} , respectively [4]. The optimal set of measurement configurations can be found by maximising an observability index, specifically the *noise amplification index* $\frac{\sigma_k^2}{\sigma_1^2}$, which represents the inverse of the maximum noise amplification [4].

The laser tracker cannot measure every pose in the robotic manipulator's workspace. Therefore, the measurable workspace needs to be determined first. By discretising the workspace, a pool of feasible measurement configurations can be created.

An approximately optimal configuration set can be found with a modified DETMAX algorithm. Firstly, an initial minimum number of measurement configurations are selected randomly from the configuration pool. This initial set is then refined by the Exchange algorithm (see Figure 2). [2]

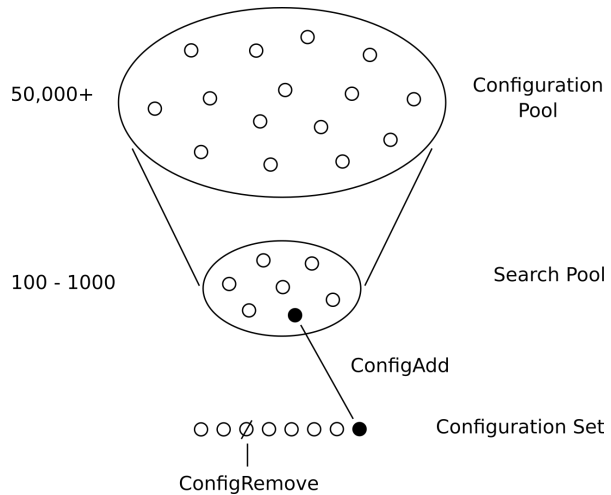


Fig. 2: Exchange Algorithm [5]

The Exchange algorithm uses two greedy algorithms, *ConfigAdd* and *ConfigRemove*. *ConfigAdd* finds the best configuration from a random subset of the configuration pool, the search pool. The size of the search pool determines the accuracy and calculation time of the algorithm. The pose, which increases the observability index of the current configuration set the most is added to the set. *ConfigRemove* removes the weakest link of the current configuration set, i.e. the configuration that decreases the observability index the least after removal. The algorithm iteratively exchanges every configuration, until the newly added configuration is found to be the weakest link. [2] Once the initial configuration set is refined, new configurations are added iteratively with *ConfigAdd*, until a given number of measurement configurations are in the set. Lastly, the configuration set is refined again with the Exchange algorithm, in order to improve the overall accuracy. [2]

Calibration

Using the optimal configuration set, the kinematic error parameters can be calculated, through linear regression using ordinary least squares

$$\hat{\underline{\epsilon}} = (\hat{\underline{J}}^T \hat{\underline{J}})^{-1} \cdot \hat{\underline{J}}^T \cdot \Delta \underline{y} \quad (9)$$

[2], or through non-linear least squares, by minimising the residual

$$\Delta \underline{y} = \underline{y} - \hat{\underline{y}} \quad (10)$$

with gradient descent [3]. The latter produces more accurate results [1], but does not guarantee finding the approximate global minimum.

Implementation

The kinematic error model was calculated symbolically, using *sympy*, for an experimental robotic manipulator. Numerical *numpy* and *scipy* functions of the TCP pose and Jacobian were then generated for the model reduction and calibration. As the measurable workspace of the experimental robotic manipulator is still being determined, valid calibration results cannot be presented at the time of writing.

Outlook

The above mentioned calibration process is expected to improve the accuracy of a robotic manipulator significantly [2]. However, not all error parameters can be compensated on the experimental robotic manipulator. The actual accuracy of the robotic manipulator is, therefore, expected to be worse than the theoretical accuracy of the 28 error parameter model. Non-geometric errors were also not regarded during calibration. Due to the small size of the experimental robotic manipulator, these errors are not expected to have a notable impact, however [3].

References and figures

- [1] Jin-Hwan Borm and Chia-Hsiang Menq. Determination of Optimal Measurement Configurations for Robot Calibration Based on Observability Measure. *The International Journal of Robotics Research*, 10:51–63, 1991.
- [2] Tian Li, Kui Sun, Zong-wu Xie, and Hong Liu. Optimal measurement configurations for kinematic calibration of six-DOF serial robot. *Journal of Central South University*, 18:618–626, 2011.
- [3] Qifeng Liu. Geometric Error Identification for 6 DOF Robotic Manipulator Calibration to Improve Absolute Positioning Accuracy, 2021.
- [4] Ali Nahvi and John Hollerbach. The noise amplification index for optimal pose selection in robot calibration. In *Proceedings of IEEE International Conference on Robotics and Automation*, volume 1, pages 647–654. IEEE, 1996.
- [5] Own representation.
- [6] Lorenzo Sciavicco and Bruno Siciliano. *Modelling and Control of Robot Manipulators*. Springer, 2 edition, 2000.
- [7] Ulrich Wiest. Kinematische Kalibrierung von Industrierobotern, 2001.

Untersuchung verschiedener Modelle des maschinellen Lernens hinsichtlich ihrer Eignung zur Unterscheidung von Anforderungstexten zu anderen Texten

Timo Schwind

Dirk Hesse

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Vector Informatik GmbH, Stuttgart

Motivation

Da die meisten Daten unstrukturiert sind, bietet die Fähigkeit, sie zu analysieren und darauf zu reagieren, eine große Chance. [4] Unstrukturiert sind Daten dann, wenn sie in einer nicht formalisierten Struktur vorliegen und somit nicht direkt in SQL-basierten Datenbanken abgelegt werden können. Natural Language Processing (NLP) Methoden helfen dabei, die Bedeutung dieser Daten zu extrahieren. Dies ist gerade in der heutigen Zeit sehr wichtig, Schätzungen zufolge macht dieser Datentyp 80–90 % des gesamten digitalen Datenuniversums aus. Zu finden sind sie in vielerlei Formen, etwa in Dokumenten, E-Mails oder auch Journalen.

Zielsetzung

Das Ziel dieser Arbeit ist es, verschiedene Texte so zu verarbeiten, dass Anforderungen in ihm erkannt und herausgefiltert werden können. Dabei sollen unterschiedliche Modelle des maschinellen Lernens verwendet und miteinander verglichen werden. Als Grundlage dafür wird eine große Textsammlung, im Linguistikumfeld als Textkorpora bezeichnet, erstellt. Diese enthält firmeninterne Anforderungen, welche schlussendlich als solche identifiziert werden sollen.

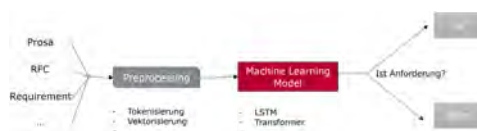


Abb. 1: Bildliche Darstellung der Aufgabe [1]

Um eine saubere Separierbarkeit und auch eine hohe Allgemeingültigkeit zu erreichen, werden den Textkorpora unterschiedliche Textformen hinzugefügt, welche nicht der Klasse der Anforderungen angehören. Beispielsweise wurden mithilfe von Web-Scrapping-Methodiken Requests for Comments (RFCs) in den

Textkorpora übernommen. Diese technischen Dokumente sind von den Anforderungstexten aufgrund ihrer semantischen Ähnlichkeit für einen Algorithmus schwer zu unterscheiden.

Die Modelle sollen hinsichtlich ihrer Funktion genau untersucht werden, sodass das Modell nicht als Blackbox betrachtet wird. Mit Methoden der erklärbaren KI (Explainable AI) im NLP-Bereich wird nachvollzogen, welche Faktoren schlussendlich zu der Klassifikation geführt haben. Ein zusätzliches Ziel ist die Weiterverarbeitung der identifizierten Anforderungen. Dabei sollen diese in unterschiedliche Klassen, basierend auf deren Inhalt, unterteilt und zugeordnet werden. Die Klassenanzahl beträgt 15 und schlüsselt die Anforderungen weiter in Unterklassen auf. Beispiele für diese sind etwa 'Test cases' oder 'Terms and definitions'.

Natural Language Processing Überblick

Das Verwenden von maschinellen Lerntechniken ist heutzutage das übliche Vorgehen bei Anwendungen im NLP-Bereich. Dabei erlernen die Systeme auf Grundlage der Trainingsdaten in Form von Textkorpora Muster. Diese Muster entsprechen im Idealfall Sprachkonzepten, die bei der Bearbeitung bestimmter Problemstellungen helfen. Ein großes Problem im NLP-Bereich ist, dass Modelle nicht direkt mit dem Text arbeiten können. Die Satzbausteine müssen zunächst einer passenden und sinnhaften Vektorrepräsentation zugeordnet werden. Nach dieser Einbettung in den Vektorraum erfolgt eine Verarbeitung mit selbstlernenden Algorithmen. Besonders gut hierbei performen sequenzbasierte neuronale Netze, die die Reihenfolge der Wörter beachten. Dafür gibt es heutzutage verschiedene Herangehensweisen. Rekurrente neuronale Netze boten dafür die Grundlagen, welche über die Jahre stetig weiterentwickelt wurden. Aktuelle State-of-the-Art Modelle wie Googles BERT (Bidirectional Encoder Representations from

Transformers) basieren auf Transformerarchitekturen, welche den Aufmerksamkeitsmechanismus nutzen. Diese sind in der Lage, Sequenzen zu verarbeiten und komplexe Abhängigkeiten zu erkennen, die dem Modell dabei helfen, anspruchsvolle Aufgaben wie Textgenerierung oder Sprachübersetzung zu meistern.

Verwendete Klassifikationsmodelle

- **Random Forest Classifier:** Klassifikationsverfahren, welches mehrere unkorrelierte Entscheidungsbäume aus verschiedenen Dateneigenschaften erstellt. Diese Bäume wachsen während des Lernprozesses. Entschieden wird durch die Mittelwertbildung der Ergebnisse der einzelnen Entscheidungsbäume.

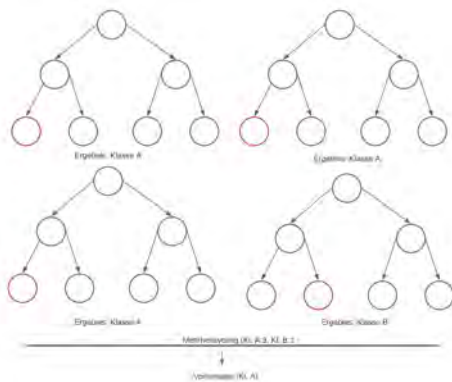


Abb. 2: Random Forest Classifier [1]

- **Naive Bayes Classifier:** Aus einer Reihe von Trainingsdaten werden Statistiken erhoben, sodass für ein neues Objekt festgestellt werden kann, wie es in diese Wahrscheinlichkeitsverteilungen hineinpasst. Der Naive Bayes Algorithmus geht davon aus, dass alle Attribute unabhängig voneinander sind.

- **Support Vector Classifier:** Das vektorbasierte Modell beschreibt (Hyper-) Ebenen, welche die gegebenen Trainingsdaten linear separieren. Ist dies in den Dimensionen der Daten nicht möglich, werden diese in einen höherdimensionalen Raum transformiert. Mithilfe einer Support-Vektor-Maschine wird ein Klassifikator gefunden, der die Datenpunkte möglichst gut voneinander trennt.

- **Feed forward neuronale Netze:** Die Topologie eines Feed Forward neuronalen Netzes lässt sich durch Knoten bzw. Recheneinheiten beschreiben, die ohne Zyklen über Schichten hinweg mit gerichteten und gewichteten Kanten verbunden sind. Jede Recheneinheit erhält und verarbeitet den gewichteten Output der Einheiten der vorherigen Schicht. Neuronale Netze sind dazu in der Lage, jede beliebige Funktion in beliebiger Genauigkeit zu approximieren. Diese Funktion kann dann zur Klassifikation verwendet werden.

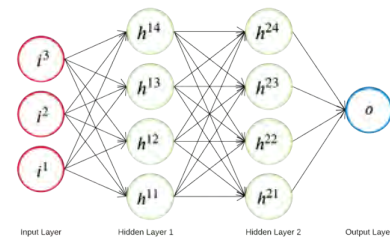


Abb. 3: Feed forward neuronales Netz [1]

- **Rekurrente Neuronale Netze und Long Short Term Memory Netzwerke (LSTMs):** Eine Erweiterung des Feed Forward Neuronalen Netzes, welche Zyklen innerhalb der Netzwerkverbindungen enthält. So können Sequenzen wie etwa Texte oder Sprache besser verarbeitet werden. Ein LSTM ist eine Form eines rekurrenten Netzes, welches Speicherzellen einführt, um Eingangssignale einzuordnen und entsprechend ihrer Bedeutung beurteilt. So können wichtige Informationen auf längere Distanz beibehalten werden.

- **Transformerarchitekturen:** Deep Learning Modelle, welche auf Zyklen verzichten und stattdessen für die Sequenzverarbeitung den sogenannten Aufmerksamkeitsmechanismus verwenden. Dieser prüft in jedem Schritt der Input Sequenz, welche anderen Teile des Satzes für den spezifischen Input von Bedeutung sind. Transformer funktionieren über einen Encoder-Decoder-Stack, wobei der Encoder die Sequenz zu einer internen Repräsentation verarbeitet. Diese wird anschließend vom Decoder benutzt, der letztendlich den Output liefert. Der von Google entwickelte und vortrainierte Encoder Bidirectional Encoder Representations from Transformers (BERT), der auch für deren Suchmaschine verwendet wird [3], hat in vielen NLP-Benchmarks neue Maßstäbe gesetzt. [2]

Zwischenstand

Ersten Ergebnissen zufolge ist die Separierbarkeit der Anforderungen mit den beschriebenen Modellen sehr gut möglich. Ein Grund hierfür ist die spezifische Sprachdomäne der firmeninternen Anforderungen, welche, gekennzeichnet durch viele charakteristische N-Grams, die Zuordnung erleichtern. N-Grams sind aufeinanderfolgende Wörter in einem Dokument, beispielsweise "Machine Learning". Dies wäre ein Bigram, da die Anzahl der aufeinanderfolgenden Wörter $N = 2$ ist. Das in der Abbildung dargestellte Streudiagramm zeigt jeweils auf den Achsen die klassenspezifisch (Anforderung/keine Anforderung) häufig auftretenden N-Grams (Unigram, Bigram, Trigram), welche durch Punkte dargestellt werden. Häufig in der jeweiligen Klasse auftretende Worte sammeln sich im "Frequent"-Bereich der dazugehörigen Achse. Besonders charakteristisch sind sie, wenn ihre Auftretensfrequenz nur bei einer Klasse hoch ist. Dies ist der Fall bei den Punkten, welche sich nah an den Achsen sammeln.

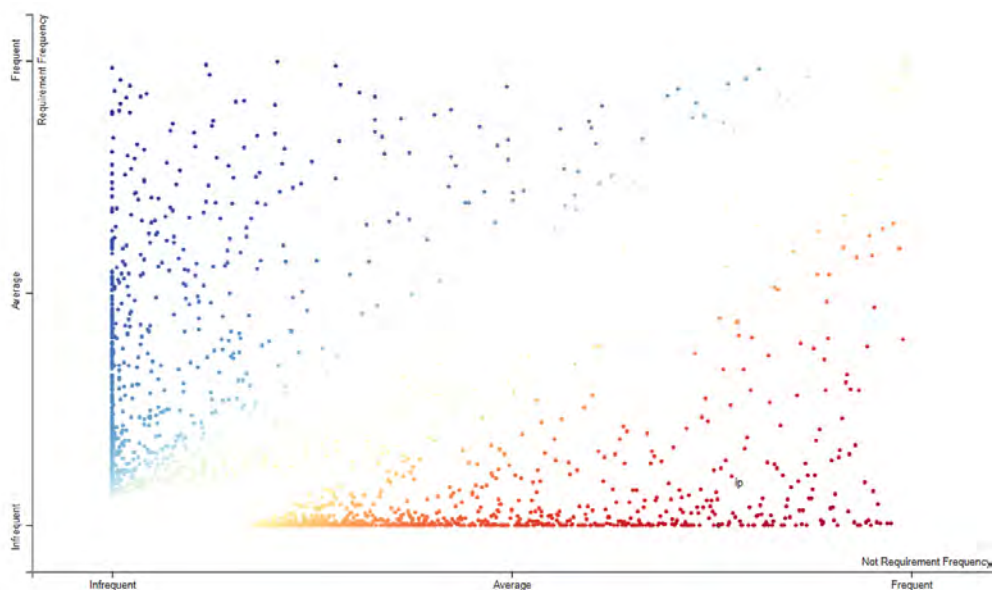


Abb. 4: Streudiagramm der Worthäufigkeiten pro Klasse [1]

Das Modell kann unter anderem durch das Erlernen eben solcher charakteristischen Stichwörter oder der Analyse semantischer Eigenschaften eine Klasseneinteilung vornehmen. Dies führt zu einer durchschnittlichen Testgenauigkeit von 99,2% auf dem Validierungsdatensatz, erreicht mit BERT als Encoder und einem kleinen neuronalen Netz, welches mit dem Encoderoutput arbeitet und auf die spezifische Klassifikationsaufgabe trainiert wurde.

Ausblick

Die Arbeit liefert die Grundlage für eine automatisierte Verarbeitung von Anforderungen. Fernziel ist es, durch künstliche Intelligenz Anforderungsdokumente zu verarbeiten. Dadurch soll eine Prüfung der Qualität realisiert werden, mit der Fehler erkannt und infolgedessen auch Verbesserungsvorschläge erstellt werden können. Durch die vorliegende Arbeit wird die grundlegende Struktur der Anforderungen erkannt und untersucht. Weitergehende Arbeiten können diese nutzen, um daraus konkrete Maßnahmen zur Verbesserung abzuleiten.

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 NAACL-HLT 2019*, pages 4171–4186. Association of Computational Linguistics, 2019.
- [3] Pandu Nayak. Understanding searches better than ever before. <https://blog.google/products/search/search-language-understanding-bert/>, 10 2019.
- [4] Mikey Shulman. Tapping the power of unstructured data. <https://mitsloan.mit.edu/ideas-made-to-matter/tapping-power-unstructured-data>, 02 2021.

Roboterassistiertes Container-Packing in der Kommissionierung

Simon Seefeldt

Steffen Schober

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Festo SE & Co. KG, Esslingen

Einleitung

In Logistikzentren ist neben dem Handhaben von einer hohen Varianz von Teilen auch das effiziente Packen von Kisten eine große Herausforderung. Umso effizienter eine Kiste gepackt werden kann, umso effizienter ist der Versand. Das hat neben den Kostenvorteilen auch einen großen Einfluss auf den CO₂-Fußabdruck eines Unternehmens. Daher wird in dieser Arbeit eine Strategie entwickelt, die es erlaubt, Kisten möglichst effizient und platzoptimal zu packen.

Problemstellung

Das zugrunde liegende Optimierungsproblem wird in der Literatur als Bin Packing, Container Packing oder Knapsackproblem bezeichnet und ist schon seit Jahren Bestand aktiver Forschung. Grund dafür ist die hohe Komplexität, die dieses Problem aufweist. Eine exakte Lösung zur Anordnung der Objekte ist in annehmbarer Zeit nicht möglich, weshalb in der Regel eine Näherungslösung mit heuristischen Verfahren errechnet wird [4]. 3D Container-Packing Probleme lassen sich in online und offline Versionen unterteilen. Bei der offline Version sind alle zu verpackenden Objekte von Anfang an bekannt und die Reihenfolge, in der die Objekte eingepackt werden, kann angepasst werden. Die Online Version zeichnet sich dadurch aus, dass Anzahl und Reihenfolge der Objekte nicht bekannt und nicht zu beeinflussen sind. Ähnlich wie beim Computerspiel 'Tetris' ist nur das Objekt bekannt, was als nächstes verladen werden soll. Im Rahmen dieser Arbeit soll ein Fall betrachtet werden, in dem die Objekte zwar von Anfang an bekannt sind, die Reihenfolge, in der die Objekte eingepackt werden, aber nicht bestimmt werden kann.

Lösungskonzept

Bei der Entwicklung einer Lösung müssen zwei Teilprobleme betrachtet werden: Die Ermittlung einer

geeigneten Ablageposition und das Aufnehmen und Ablegen von Objekten mit dem Roboter.

Ablageposition - Container Packing Algorithmus:

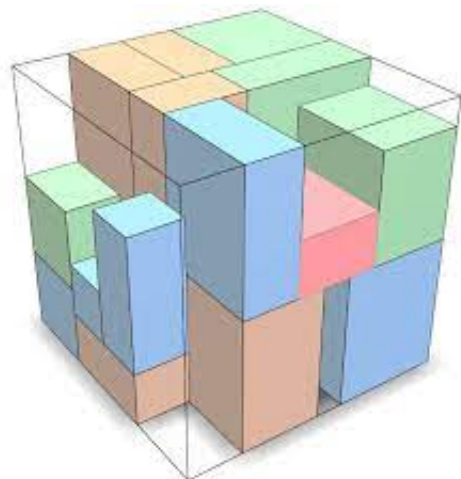


Abb. 1: Beispielhaftes Ergebnis eines Container-Packing Algorithmus [6]

Für das Problem finden sich in der Literatur zahlreiche Ansätze, die jeweils ihre eigenen Stärken und Schwächen aufzeigen. Im Großen und Ganzen lassen sich heuristische und lernbasierte Methoden unterscheiden. Heuristische Ansätze verfolgen eine vordefinierte Strategie, wie zum Beispiel der First Fit Algorithmus, der systematisch alle Plätze im Container überprüft und das Objekt an der ersten freien Stelle platziert. Es gibt auch deutlich komplexere heuristische Ansätze, deren Ergebnisse, die des FF übertreffen, jedoch ist es meistens schwer, Einschränkungen aus der echten Welt mit in den Algorithmus aufzunehmen. Lernbasierte Methoden beruhen zum größten Teil auf Reinforcement Learning Algorithmen [2] [6], die aus einer Reihe vorgegebener Aktionen lernen, die auszuwählen, die für die vorliegende Situation den höchsten Erfolg verspricht. Wie das Ergebnis aussehen kann, wird in Abbildung 1 dargestellt. Aus der Literatur geht hervor,

dass diese Methoden leicht um Einschränkungen erweitert werden können. Die meisten dieser Ansätze sind jedoch lizenziert und nur für akademische Zwecke freigegeben.

Aufnehmen und Ablegen von Objekten: In der Anlage kommen die Objekte in kleinen Kisten an und müssen in den Karton geladen werden. Um das Objekt mit dem Roboter aufnehmen zu können muss anhand eines RGBD Bilds ein Greifpunkt bestimmt werden. Nachdem der Container Packing Algorithmus den Ablageort zurückgegeben hat, fährt der Roboter an diese Position und platziert das Objekt. Besonders wichtig ist dabei, dass bekannt ist, wo genau das Objekt gegriffen wurde, um den Offset des Greifers mit dem Ablagepunkt zu verrechnen. Eine Möglichkeit dafür könnte sein, für jedes Objekt im Sortiment eine Punktwolke zu speichern, in der die Greifpunkte eingezeichnet sind. Dadurch ist gewährleistet, dass Objekte immer gleich aufgenommen werden. Durch Point Cloud Matching kann dann die Punktwolke eines Objekts, die durch die RGBD Kamera aufgenommen wurde, mit allen gespeicherten Punktwolken verglichen werden, um von dem Objekt mit der größten Ähnlichkeit Daten wie Maße, Gewicht und Greifpunkt in Erfahrung zu bringen. Alternativ dazu kann ein Neuronales Netz verwendet werden, das in der Lage ist zu lernen, die Position und Orientierung von Objekten zu erkennen und daraus einen geeigneten Greifpunkt abzuleiten. Über einen Agenten kann gelernt werden, das erkannte Objekt in einem Rahmen, gerade größer als das Objekt selbst, abzulegen. Durch Simulation können in kurzer Zeit

sehr viele dieser Pick-and-Place Versuche durchgeführt und bewertet werden. In der Praxis ist es letztlich aber nicht möglich, zu jedem Objekt den passenden Rahmen an die richtige Stelle im Container zu legen. Der Rahmen kann aber vor der Bewertung durch das Modell an die vom Packing Algorithmus bestimmte Stelle im RGBD Bild projiziert werden. Diese Methode hat zum Vorteil, dass für neue Objekte nicht erst eine Point Cloud aufgenommen und hinterlegt werden muss, setzt aber voraus, dass Maße und Gewicht anderweitig in Erfahrung gebracht werden können und dass die in der Simulation trainierten Modelle auch auf die Realität übertragbar sind.

Um das Problem dem Umfang dieser Arbeit anzupassen, werden vereinfachende Annahmen getroffen. Der Fokus liegt auf der Aufgabe, die Kartons effizient zu packen, weshalb der Aufbau auf eine statische Umgebung, bestehend aus einem zentralen Roboterarm, einer Kiste mit ankommenden Objekten links und einer Kiste, in der die Objekte einsortiert werden, rechts, reduziert wird. Nach Rücksprache mit dem Kunden und Einsicht in die Bestellstatistik, stellte sich heraus, dass ein Großteil der Artikel in Quaderförmigen Umverpackungen geliefert wird. Deshalb sollen vorerst nur Quader als Objekte zulässig sein. Ziel ist es, alle Entscheidungen und Vereinfachungen während der Entwicklung auf Basis der Kundenanforderungen begründen zu können. Abbildung 2 zeigt das Konzept mit den möglichen Optionen. Oben links das autonome Greifen auf Basis eines Machine Learning Modells [5] und unten links Point Cloud Matching [3].

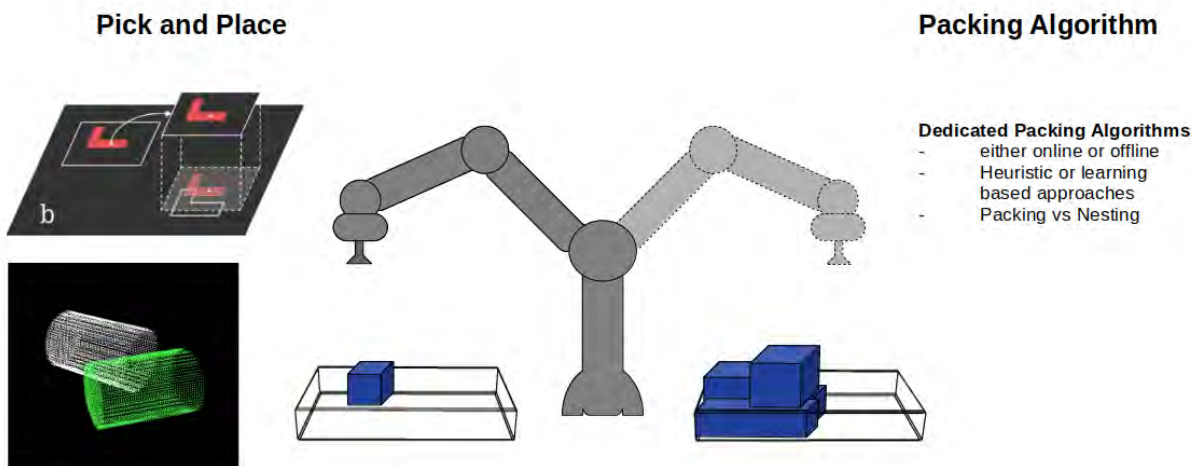


Abb. 2: Konzeptaufbau mit möglichen Optionen [1]

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Richa Verma et al. A Generalized Reinforcement Learning Algorithm for Online 3D Bin-Packing. *arXiv:2007.00463v1*, 2020.
- [3] Michael Willet. Iterative closest point. <https://github.com/michaelwillett/Iterative-Closest-Point>, 2016.
- [4] R. Yesodha and T. Amudha. A comparative study on heuristic procedures to solve bin packing problems. *DOI: 10.5121/ijfcst.2012.2603*, 2012.
- [5] Andy Zeng et al. Transporter Networks: Rearranging the visual world for robotic manipulation. *arXiv:2010.14406v2*, 2021.
- [6] Hang Zhao et al. Online 3d bin packing with constrained deep reinforcement learning. *arXiv:2006.14978v4*, 2021.

Transportverwaltung von ABAP-Coding mit GIT

Imer Spatolaj

Astrid Beck

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Quanto Solutions, Stuttgart-Vaihingen

Einleitung

Der Transport von Entwicklungsprojekten und -objekten von SAP-Systemen mit Hilfe der freien Software GIT stellt das Kernthema dieser Thesis dar. Dazu wurden mehrere Möglichkeiten in Betracht gezogen, wie man ein SAP-System mit GIT als Versionsverwaltungssystem miteinander verknüpft.

Die Produkte des Unternehmens SAP sind weltweit verbreitet. Nicht zuletzt gilt SAP als stärkstes Software-Unternehmen in Deutschland und gehört zu den Top 5 weltweit. Mit der Entwicklung auf Systemen von SAP können verschiedene Produkte entwickelt werden. Ob in den Bereichen Finanzen, Controlling, Verkauf & Vertrieb oder in den Bereichen Materialmanagement, Personalmanagement und Produktplanung.

Um Entwicklungen besser verwalten zu können, hat SAP eine Lösung gefunden, ihre Systeme mit der freien Software GIT zu kombinieren. Dadurch hat man einen großen Schritt in Richtung moderner Entwicklung gemacht, da GIT in der Programmierwelt und in der Entwicklung zunehmend an Bedeutung gewinnt.

Motivation und Ziel

Die SAP-Systeme und die Software GIT können durch zwei verschiedene Implementierungen miteinander verknüpft werden und dabei Transporte in einer SAP-Systemlandschaft durchführen. Zum einen das abapGIT, welches für die freie Implementierung verwendet wird und eine von der Community getriebene Technologie ist, und zum anderen das git-fähige Change- und Transportsystem (gCTS), welches für die native Implementierung verwendet wird und eine Technologie ist, welche von SAP entwickelt wurde.

Im Rahmen dieser Thesis, soll vor allem das gCTS implementiert werden und ggf. Begrifflichkeiten zur genaueren Erläuterung erklärt werden. Dabei soll nicht nur eine Implementierung stattfinden, sondern es soll genauer aufgeschlüsselt werden, welche Vorteile diese Implementierungen haben und was daraus gewonnen werden kann.

SAP-Systemlandschaft

Die Systemlandschaft ist eine Einheit von Systemen, welche verschiedene Zwecke erfüllen. Die häufigste verwendete Systemlandschaft ist die 3-Systemlandschaft.

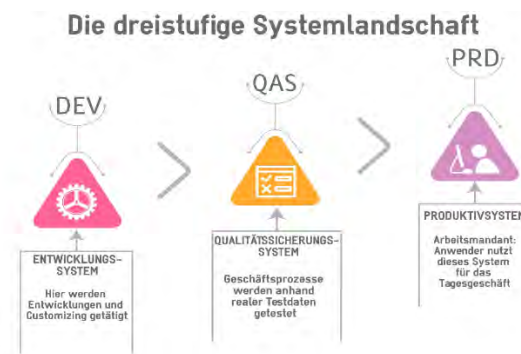


Abb. 1: 3-Systemlandschaft [3]

Die 3-Systemlandschaft besteht aus einem Entwicklungssystem (DEV), einem Qualitätssicherungssystem (QAS) und einem Produkktivsystem PRD). Im Entwicklungssystem können Entwickler ihre Einstellungen und Entwicklungen erledigen. Diese werden danach in ein Transportauftrag angehängt und in das Qualitätssicherungssystem transportiert. In dem Qualitätssicherungssystem werden die Entwicklungen getestet und nach möglichen Fehlern gesucht. Werden Fehler gefunden, so werden die Änderungen im DEV-System vorgenommen und diese werden wieder in das QAS-System transportiert. Laufen die Test fehlerfrei, so werden die Entwicklungen in das Produkktivsystem transportiert und das entwickelte Produkt wird eingesetzt.

Change- und Transportsystem (CTS)

Über Transportaufträge werden Änderungen erfasst und zwischen den Systemen verschoben.



Abb. 2: Change- und Transportsystem (CTS) [2]

Dabei wird zuerst der Transportauftrag angelegt und es wird ein leerer Transport als Container erstellt. Die getätigten Änderungen und Entwicklungen werden erfasst und in den Transportauftrag angehängt. Danach wird der Transport freigegeben und der Inhalt wird gesperrt, sodass während dem Transport keine Entwicklungen vorgenommen werden können. Dies ermöglicht Import in ein anderes System. Zum Schluss wird der Transportauftrag in das Zielsystem importiert.

Git-fähige Change- und Transportsystem

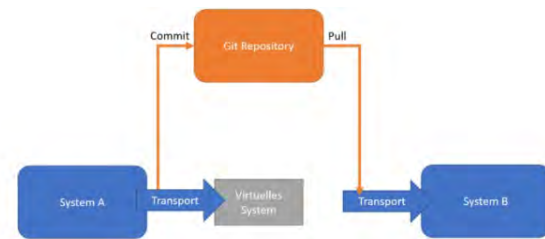


Abb. 3: Git-fähige Change- und Transportsystem [1]

Der Unterschied zwischen dem CTS und dem gCTS geschieht erst, wenn der Transportauftrag freigegeben wird. Beim CTS werden die Änderungen auf das nächste SAP-System verschoben. Beim gCTS hingegen werden die Änderungen auf eine virtuelle Transportschicht transportiert. Hier wird ein Commit auf ein externes Git-Repository erstellt. Das nächste System kann die Änderungen vom Git-Repository ziehen und importieren.

Ausblick

Die Installation und Einführung dieser Technologie bringt Vorteile mit sich mit, wie z.B. eine zusätzliche Verwaltung des Codes. Die Bachelorthesis ist eine Vorarbeit und Vorbereitung zu Einführung des gCTS im Unternehmen Quanto Solutions. Um diese Technologie einzusetzen müssen einige Konfigurationen und Installationen am SAP-Server gemacht werden, damit diese gCTS-App verfügbar ist.

Literatur und Abbildungen

- [1] L. Burkhardt. gCTS – Das Transportsystem 2.0. <https://erlebe-software.de/abap-und-co/gcts- das-transportsystem-2-0/>, 2021.
- [2] T. Harmes. SAP Transport Management. <https://rz10.de/knowhow/sap-transport/>, 2021.
- [3] D. Lees. Einführung in die Struktur von SAP-Landschaften und die damit verbundenen Herausforderungen – Teil 1. <https://www.basistechnologies.com/de/blog/einfuehrung-in-die- struktur-von-sap-landschaften-und-die-damit-verbundenen-herausforderungen-teil-1>, 2020.

Optimierung des AET-Prozesses bei Mercedes-Benz Cars, Van, AMG & G-Klasse

Steffen Spieth

Astrid Beck

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Mercedes-Benz, Sindelfingen

Einleitung

„Wer aufhört, besser zu werden, hat aufgehört, gut zu sein“ -Philip Rosenthal

Dieses Zitat von Philip Rosenthal sollte dazu inspirieren, sich ständig weiterzuentwickeln. Dieses Zitat kann man nicht nur auf die Unternehmenswelt, sondern auch auf den Alltag eines jeden Menschen beziehen. Zufriedenheit ist ein Wort, welches in erster Linie positiv erscheint. Wird das Wort Zufriedenheit jedoch im Zusammenhang mit dem Zitat von Rosenthal betrachtet, kann es kritisch hinterfragt werden. Denn wer mit sich zufrieden ist, neigt dazu, sich auf diesem Zustand auszuruhen, sich nicht weiterzuentwickeln und hört auf, das optimale aus sich herauszuholen. Dies bedeutet nicht, dass Zufriedenheit etwas Schlechtes ist. Zufriedenheit ist gut, solange sie temporär ist. Auch in der Unternehmenswelt sollte danach gestrebt werden, sich ständig weiterzuentwickeln. Es zeigt sich, Unternehmen, die sich auf ihrem Erfolg ausruhen und nicht flexibel sind sich anzupassen, geraten in Vergessenheit oder spielen nur noch eine untergeordnete Rolle in ihrem Markt. Sie können den Druck, der immer stärker werdenden Konkurrenten, nicht standhalten.

Zielsetzung

Diese Einstellung bildet die Grundlage der nachfolgenden Arbeit. Sie hat das Ziel, bestimmte Prozesse und Abläufe bei Mercedes-Benz Cars, Vans, AMG und G-Klasse anzupassen, zu optimieren und zu vereinheitlichen. Der Prozess ist ein fundamentaler Bestandteil eines Unternehmens. In jedem Geschäftsbereich, jeder Abteilung und jedem Team, werden verschiedenste Prozesse durchgeführt. In jedem der vier Bereiche befindet sich ein Prozess welche auf unterschiedlichste Art und Weise durchgeführt wird. Ziel ist es diese Prozesse zu analysieren, Schwachstellen zu identifizieren, um Handlungsempfehlungen abzuleiten und im besten Fall einen einheitlichen Prozess abzuleiten.

Umsetzung

Ein wichtiger Bestandteil des Prozessmanagements ist der Prozesslebenszyklus. Er beschreibt die verschiedenen Phasen eines Prozesses. Grob gesagt erzeugt dieser

Kreislauf einen Input an neuen Prozessen und wirft Prozesse, die überflüssig sind aus.

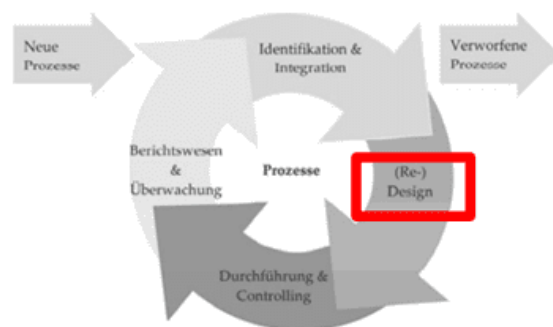


Abb. 1: Prozesslebenszyklus [2]

Die Arbeit befindet sich im Bereich des (Re-Designs) des Prozesslebenszyklus (Hier auf der Grafik rot markiert)



Abb. 2: (Re-) Design [1]

Die Grafik zeigt den Ansatz wie das (Re-Design) durchgeführt und die Prozesse angepasst werden sollen. Im ersten Schritt wird ein vorhandener Prozess modelliert, um anschließend eine Ist-Analyse durchzuführen. Die Ist-Analyse wird dabei in die Erhebung des Istzustands und die Schwachstellenanalyse unterteilt. Auf Grundlage der Ist-Analyse und der daraus resultie-

renden Verbesserungspotenziale wird ein Sollprozess generiert. Anschließend werden die Verbesserungen implementiert.

Ausblick

Betrachtet man den aktuellen Zustand und vergleicht die Prozesse miteinander ist festzustellen, dass einige Prozesse sich ähneln, viele sich aber auch voneinander unterscheiden. Abläufe, die sich in den Prozessen unterscheiden, sind vor allem für den Kunden eine Herausforderung, da sie sich ständig unterschiedlichen Gegebenheiten anpassen müssen. Die Kommunikation

zwischen den Schnittstellen läuft nicht immer optimal, sodass der Informationsfluss beeinträchtigt ist und so Zeit verloren geht. Es gibt einige Vorgänge, die ohne Software durchgeführt werden und so ein hohes Fehlerpotenzial mit sich bringen. Diese Schwachstellen sollen zukünftig verbessert werden. Software sollte dort eingesetzt werden, wo es notwendig ist. Die Kommunikation wird dort angepasst wo Informationen auf Strecke bleiben. Vereinheitlichungen werden dort eingeführt, wo sich Kunden sonst ständig anpassen müssen.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Markus Huber and Gerda Huber. *Prozess- und Projektmanagement für ITIL*. Vieweg + Teubner Verlag | Springer Fachmedien, 2011.

Passwortlose Web-Authentisierungs-Konzepte in Industrieszenarien

Patrizia Spinola

Tobias Heer

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma iss innovative software services GmbH, Stuttgart

Einleitung

Im industriellen Umfeld werden immer häufiger IoT-Geräte mit Webinterface eingesetzt. Bisher nutzen diese Webinterfaces meist eine einfache passwortbasierte Benutzerauthentisierung, da diese relativ einfach zu implementieren ist und der Anwender lediglich einen Browser benötigt, um sich einzuloggen. Diese Authentisierungsmethode hat sich aber in der Praxis als relativ unsicher und wenig nutzerfreundlich erwiesen. Benutzer wählen oftmals schwache, leicht erratbare Passwörter und verwenden das gleiche Passwort mehrfach für die verschiedenen Webanwendungen und IoT-Geräte, da sie sich nur schwer eine Vielzahl von unterschiedlichen, ausreichend langen und komplexen Passwörtern merken können [6]. Hinzu kommen weitere Gefahren wie gestohlene, öffentlich gemachte Passwort-Dateien, Phishing- oder Brute-Force-Angriffe.

Ziel der Arbeit

Ziel dieser Arbeit ist es ein besser geeignetes, sicheres und nutzerfreundliches Authentisierungskonzept für das Anmelden am Webinterface der IoT-Geräte zu entwickeln. Dabei werden verschiedene Web-Authentisierungskonzepte für Webanwendungen und IoT-Geräte mit Webinterface in einem industriellen Umfeld untersucht und verglichen. Diese werden hinsichtlich ihrer Sicherheit, Benutzerfreundlichkeit und der Eignung in Single-Sign-On-Szenarien bewertet, um herauszufinden, inwieweit sich welche Authentifizierungskonzept in den Realbetrieb integrieren lässt, damit die Kriterien einer hohen Sicherheit sowie hohen Benutzerfreundlichkeit erfüllt werden können.

Umsetzung

Untersucht wird die Authentisierung mit Passwörtern, die Zweifaktor-Authentisierung mit Passwörtern und zeitbasierten Einmal-Passwörtern (TOTPs) sowie die passwortlose Authentisierung mit FIDO2/WebAuthn

[3] [4] in Kombination mit einer PIN. Diese Authentisierungsmethoden werden sowohl in Szenarien mit dezentraler Benutzerverwaltung als auch in Single-Sign-On-Szenarien mit einem zentralen Identity-Provider und Authentifizierung über OpenID Connect [5] betrachtet. Abbildung 1 zeigt die untersuchten Szenarien.



Abb. 1: Überblick der betrachteten Szenarien [2]

Für jedes Szenario werden die genauen Abläufe aller relevanter Anwendungsfälle (engl. Use Cases) beschrieben. Hierzu gehören neben der eigentlichen Authentifizierung des Benutzers auch die Abläufe der Benutzerverwaltung und das Konfigurieren der beteiligten Geräte. Die passwortlose Authentisierung mit FIDO2/WebAuthn wird sowohl für den dezentralen Fall als auch für den zentralen Fall mit OpenID Connect prototypisch implementiert bzw. aufgebaut, um die Umsetzbarkeit in einem industriellen Szenario zu validieren und mögliche Probleme zu identifizieren. Für die Bewertung der Nutzerfreundlichkeit wird eine einfache Methode verwendet, mit welcher der Aufwand für die betrachteten Szenarien für verschiedene Fallbeispiele mit unterschiedlichen Parametern berechnet und verglichen werden kann. Bei der Sicherheitsbetrachtung werden zunächst relevante bekannte Angriffstechniken bei der Verwendung von Passwörtern aufgezeigt. Die Authentifizierungsmethoden werden dann hinsichtlich der Sicherheit bewertet, indem analysiert wird, ob und welche Authentisierungsfaktoren anfällig für diese be-

kannten Angriffstechniken sind. Hieraus wird abgeleitet, welche Angriffsvektoren unter bestimmten Voraussetzungen erfolgreich sein können. Zudem werden einige Best-Practice-Empfehlungen zum sicheren Einsatz der Authentifizierungsmethoden gegeben.

Was ist FIDO2?

FIDO2 ist die neueste Weiterentwicklung der im Jahr 2012 gegründeten Fast Identity Online (FIDO)-Allianz, die mit dem Ziel ins Leben gerufen wurde, offene und kostenlose Standards für eine sichere passwortlose Authentifizierung im Internet zu entwickeln. FIDO2 wurde gemeinsam mit dem World Wide Web Consortium (W3C) entwickelt, um die starke FIDO-Authentifizierung in allen Webbrowsern und der dazugehörigen Webplattform-Infrastruktur zu standardisieren. Der FIDO2-Standard besteht aus dem W3C-Standard Web Authentication (WebAuthn) [4] und dem Client to Authenticator Protocol (CTAP)

[1] der FIDO-Allianz, die zusammen eine passwortlose Authentifizierung zwischen externen Authenticatoren, Browsern und vertrauenswürdigen Servern (auch Relying Party genannt, kurz RP) ermöglichen. WebAuthn definiert eine Web-API, die in Webbrowsern und in Webplattform-Infrastrukturen integriert werden kann, um Methoden zur sicheren Authentifizierung bereitzustellen. Online-Dienste (z. B. Webanwendungen) können diese Methoden nutzen, um eine passwortlose Webauthentifizierung zu realisieren. CTAP ermöglicht externen Geräten wie Smartphones, Wearables, Smartcards oder USB-Sticks mit Browsern per USB, NFC oder Bluetooth zu kommunizieren und als Authenticatoren für Desktopanwendungen und Online-Dienste (z. B. Webanwendungen) zu dienen. Derzeit kann FIDO2 mit den Webbrowsern Mozilla Firefox, Google Chrome, Microsoft Edge und Apple Safari sowie auf den Plattformen Windows 10 und Android verwendet werden. Abbildung 2 zeigt eine Übersicht über die FIDO2-Architekturkomponenten.

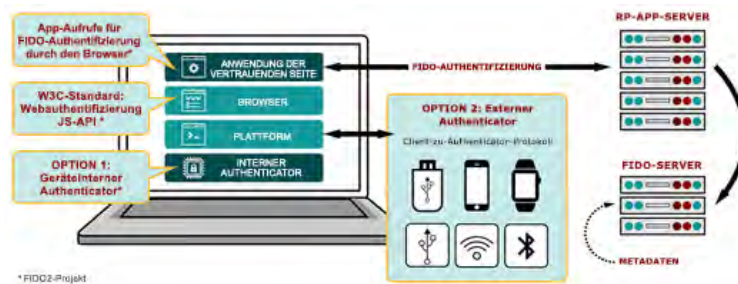


Abb. 2: Übersicht über die FIDO2-Architekturkomponenten [7]

Abschließende Betrachtung und Ausblick

Die Untersuchung der Authentifizierungsmethoden hat ergeben, dass Passwörter einige inhärente Schwächen besitzen, die aber durch den Einsatz von Passwort-Managern zum Teil abgemildert werden können. Wird die einfache Passwort-Authentisierung um einen zweiten Faktor ergänzt, kann die Sicherheit noch weiter erhöht werden. Allerdings steigt hierbei der Aufwand für den Benutzer. FIDO2 bietet im Vergleich eine deutlich höhere Sicherheit. Zudem ist der Aufwand für den Benutzer nicht signifikant höher im Vergleich zur Passwort-Authentisierung. Allerdings wird ein FIDO2-Authenticator benötigt.

Insgesamt zeigt sich, dass die passwortlose Authen-

tifizierung mit FIDO2/WebAuthn in industriellen Einsatzszenarien praktikabel einsetzbar ist, und dass mit ihr bei leicht höherem Aufwand ein signifikanter Sicherheitsgewinn im Vergleich zur passwortbasierten Authentifizierung erzielt werden kann.

FIDO2 ist eine erfolgversprechende Technologie, die Passwörter langfristig ablösen könnte. Es ist zu erwarten, dass dies zu einer deutlichen Verbesserung der Sicherheit bei der Web-Authentisierung führen wird. Dadurch können die Probleme, die schwache Passwörter verursachen, grundlegend beseitigt und nicht nur durch zusätzliche Maßnahmen wie einen Passwort-Manager oder einen zweiten Faktor abgemildert werden.

Literatur und Abbildungen

- [1] Christiaan Brand et al. Client to Authenticator Protocol (CTAP). <https://fidoalliance.org/specs/fido-v2.0-ps-20190130/fido-client-to-authenticator-protocol-v2.0-ps-20190130.html>, 01 2019.
- [2] Eigene Darstellung.
- [3] Alliance FIDO. FIDO2: WebAuthn & CTAP Moving the World Beyond Passwords. <https://fidoalliance.org/fido2/>, 2021.
- [4] Jeff Hodges et al. Web Authentication: An API for accessing Public Key Credentials Level 3. <https://w3c.github.io/webauthn/#sctn-intro>, 12 2021.
- [5] Natsuhiko Sakimura. Openid Connect Core 1.0. https://openid.net/specs/openid-connect-core-1_0.html, 11 2014.
- [6] Klaus Schmeh. *Kryptografie: Verfahren, Protokolle, Infrastrukturen*. dpunkt.verlag, 2016.
- [7] Alex Simon. Große Fortschritte auf dem Weg zur kennwortfreien Anmeldung: FIDO2/WebAuthn erreichen „Candidate Recommendation“-Status! <https://www.microsoft.com/de-de/microsoft-365/blog/2018/04/12/big-news-in-our-drive-to-eliminate-passwords-fido2-webauthn-reaches-candidate-recommendation-status/>, 04 2018.

Entwicklung eines generischen Beratungskonzeptes zur Einführung einer Predictive Analytics Lösung auf Basis der SAP Analytics Cloud

Dominik Stach

Thomas Rodach

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma bulheller consulting GmbH, Stuttgart

Einleitung

Big Data, Data Sourcing und Datenverarbeitung stehen in vielen Branchen im Fokus. Insbesondere Business Intelligence und Business Analytics sind immer prägnantere Themen, da viele Unternehmen sich nicht mehr alleine mit Excel als Analysemöglichkeit zufrieden geben. Eine Studie der Computerwoche hat herausgefunden, dass 37% der befragten Unternehmen im Analytics-Bereich Projekte umsetzen. 56% dieser Projekte können dem Teilgebiet Predictive Analytics zugeordnet werden. 94% der Unternehmen geben an, Predictive Analytics zu nutzen, um Entscheidungen fundierter treffen zu können. [2] IT-Beratungsunternehmen benötigen daher Expertise in Systemen, welche den Use-Case Predictive Analytics umsetzen. Eine solche Applikation ist die SAP Analytics Cloud. Die SAP Analytics Cloud ist hierbei eine SaaS-Lösung und bietet die Themen Business Analytics, Predictive Analytics und Business Intelligence an.

Motivation

Die bulheller consulting GmbH ist ein mittelständisches IT- und Prozess-Beratungsunternehmen mit Sitz in Stuttgart-Süd. Die Schwerpunkte der bulheller consulting sind das IT-Projekt und Servicemanagement sowie ERP-Beratung, insbesondere mit SAP-Systemen. Durch stetig steigendes Interesse der Kunden an Business Intelligence und Business Analytics Lösungen ist es für bulheller consulting interessant untersuchen zu lassen, wie die SAP Analytics Cloud aufgebaut ist und ein potentielles Beratungskonzept aussehen könnte. Damit nicht nur die Funktionalität der SAP Analytics Cloud im Fokus der Bachelorarbeit steht, soll auf Basis branchenüblicher agiler Methoden ein generisches Beratungskonzept zur Einführung und Implementierung einer solchen Lösung entwickelt werden. Ziel der Bachelorarbeit ist es, besagtes Beratungskonzept

auszuarbeiten und am Beispiel eines fiktiven Kunden mit Beispieldaten zu prüfen.

Predictive Analytics

Predictive Analytics versucht über Muster und Korrelationen Verhalten und Situationen vorherzusagen. Dies geschieht über Modelle und Algorithmen. Viele dieser Algorithmen kommen aus dem Bereich Machine Learning. Insbesondere das überwachte Lernen findet oftmals Verwendung. Die SAP Analytics Cloud bietet verstärkt Regression, Klassifikation und Zeitreihen als Predictive Analytics Lösung an. Regression ist oftmals ein gemittelter Durchschnittswert, dem weitere Einflussfaktoren hinzugefügt werden, um so viele Abweichungen wie möglich abzubilden. Mittels Regression können einfache bis komplexe Vorhersagen getroffen werden. Klassifikation hat als Ziel, neuen Daten bestimmte Eigenschaften vorherzusagen und zuzuweisen. Dies geschieht über Referenzmengen, die im selben Trefferspektrum eine hohe übereinstimmende Eigenschaftsmenge besitzen. Somit ist es wahrscheinlich, dass ein neuer Datenpunkt ebenfalls diese Eigenschaften besitzt. Unter Zeitreihenanalysen versteht man eine Beobachtung einer betriebswirtschaftlichen Größe über einen längeren Zeitraum, die dabei sich entwickelnden Werte werden als Merkmalswerte bezeichnet. Dabei versucht die Zeitreihenanalyse die Struktur und Gesetzmäßigkeit der Merkmalswerte zu erkennen und mit diesen Erkenntnissen die Entwicklung einzuschätzen. Diese Predictive Analytics Modelle werden in der Bachelorarbeit genauer erläutert. Ziel von Predictive Analytics ist es, entweder Eigenschaften von unbekannten Daten vorherzusagen, basierend auf bereits vorhandenen Daten und Ähnlichkeiten, oder Werte für die Zukunft vorherzusagen, basierend auf bereits vorhandenen Daten. [5]

SAP Analytics Cloud

Die SAP Analytics Cloud ist eine Standalone Software-as-a-Service Lösung der SAP AG, welche auf der HANA Cloud Plattform basiert. Sie kombiniert Mechanismen des Machine Learning und der Artificial Intelligence zu einer Lösung, die Predictive Analytics und damit verbunden Predictive Planning, Vorhersagen und Business Intelligence ermöglicht. Ebenfalls können die Daten visualisiert und mittels Augmented Analytics individuell von den Nutzern analysiert werden. [1]

SAP Analytics Cloud in the **Intelligent Enterprise Suite** (strategic direction)

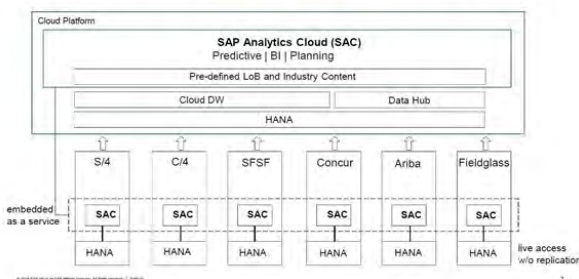


Abb. 1: SAP Analytics Cloud Architecture [3]

Dabei ist die SAP Analytics Cloud eine Lösung, die frei von grundlegenden Systemen verwendet werden kann. Sie liegt auf Systemen wie ERP-Systemen oder SAP Data Warehouse und kann sowohl Live-Analysen, als auch Offline-Analysen mit den Daten durchführen. Sichtbar ist dies in Abbildung 1.

ASAP

Accelerated SAP ist ein Implementierungsmodell der SAP AG für SAP Systeme. Es besteht aus den Phasen: [4]

- Project Preparation
- Business Blueprint
- Realization

- Final Preparation
- Go-Live and Support

Dieses Modell ist das Grundgerüst für den später entwickelten Beratungsansatz. Dies ist erkennbar in Abbildung 2. In jeder einzelnen Phase sollen die Anforderungen anhand von Beispielkundendaten durchgegangen werden, um so das generische Modell zu entwickeln. Die einzelnen Schritte werden um Elemente aus den agilen Methoden Kanban und SCRUM ergänzt, um eine Vorgehensweise zu erhalten, welche die Vorteile der einzelnen Methoden kombiniert.

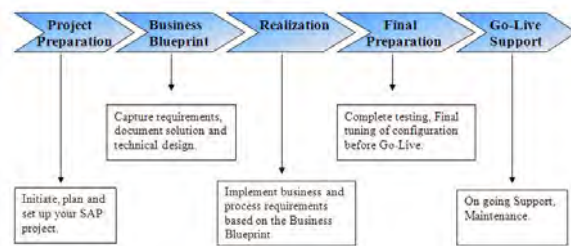


Abb. 2: ASAP Roadmap for Implementation [4]

Ausblick

In der Bachelorarbeit werden die agilen Implementierungsmethoden Kanban und SCRUM mit ASAP verglichen, um ein generisches Beratungskonzept zu erarbeiten, welches die Vorteile aller Methoden kombiniert. Mit diesem Konzept soll anschließend beispielhaft an einem möglichen Kundenszenario die SAP Analytics Cloud implementiert werden. Dabei soll die SAP Analytics Cloud Chancen aufzeigen, die einem möglichen Kunden dargestellt werden können, um ihm ein Predictive Analytics und Business Intelligence Tool zur Verfügung zu stellen. Zu erwarten ist, dass im Prozess der Implementation Kanban als Unterstützung hinzugezogen werden kann und nach der Implementierung, im Bereich der kontinuierlichen Verbesserung, das Scrum-Modell seine Stärken einbringen kann.

Literatur und Abbildungen

- [1] Sabil Nararya et al. Automation in Financial Reporting by using Predictive Analytics in SAP Analyt-ics Cloud for Gold Mining Industry: a Case Study. In *Automation in Financial Reporting by using Predictive Analytics in SAP Analyt-ics Cloud for Gold Mining Industry: a Case Study*. ICISS, 2021.
- [2] Christiane Pütter. Key-Findings in der Predictive Analytics Studie. *Computerwoche*, 2018.
- [3] SAP SE. SAP Analytics Cloud Architecture. <https://infocient.de/wp-content/uploads/2019/05/SAC-Architecture.jpg>, 2018.
- [4] Manish Shankar. <https://blogs.sap.com/2012/08/20/sdlc-vs-asap-methodology/>. <https://blogs.sap.com/2012/08/20/sdlc-vs-asap-methodology/>, 2012.
- [5] Felix Weber. *Künstliche Intelligenz für Business Analytics*. Springer Vieweg, 2020.

Möglichkeiten und digitale Umsetzung eines Handels mit Emissionszertifikaten für ein Tiny House

Jan Trillinger

Anke Bez

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Problemstellung

Klimawandel und Wohnraumproblem sind zwei Begriffe, die heutzutage jeder kennt. Unternehmen stehen jeden Tag vor der Herausforderung, ihre Emissionen zu kompensieren. Deutsche Unternehmen bekommen die meisten Vorgaben zur Einhaltung von Klimazielen von europäischen Gesetzen. Hierfür wurden die Grundlagen der Umwelt- und Klimapolitik im sogenannten 2020-Rahmen festgelegt. In diesen Richtlinien ist unter anderem der Ansatz für ein Emissionshandel sowie sämtliche Klimaziele der EU niedergeschrieben. Der Rahmen der Europäischen Union wurde 2020 überprüft und neue Ziele bis 2030 festgelegt. Da es sich hierbei um ein globales Thema handelt trat am 16. Februar 2005 das Kyoto-Protokoll in Kraft. 191 Staaten haben das Abkommen unterschrieben und versuchen die Ziele wie die Reduzierung der Treibhausgase einzuhalten. Anhand der Tabelle 1 lässt sich die Auswirkung von Kohlendioxid (CO₂) auf die Erderwärmung beobachten. Das GWP (Global Warming Potential) ist ein Vergleichswert mit einem anderen Gas, das ausgestoßen werden müsste, um den gleichen Einfluss auf die Erderwärmung zu erreichen. Diese Tabelle wird für die Berechnung der Emissionszertifikate verwendet und zeigt, wie hoch der Einfluss von Kohlendioxid ist. [4]

Tabelle 1. Erderwärmungspotenziale nach UNFCCC, Climate Change 1995

Treibhausgas	GWP ₁₀₀ [tCO ₂ e/t]
Kohlendioxid (CO ₂)	1
Methan (CH ₄)	21
Distickstoffoxid/Lachgas (N ₂ O)	310
Perfluorierte Kohlenwasserstoffe (FKW)	Bis zu 9200
Teilfluorierte Kohlenwasserstoffe (H-FKW)	11700
Schwefelhexafluorid (SF ₆)	23900

Abb. 1: Tabelle Treibhausgas [4]

Das Kyoto-Protokoll ist 2020 ausgelaufen und wurde bei neuen Verhandlungen durch das Pariser Überein-

kommen abgelöst. Das Ziel ist klar „die Erderwärmung auf deutlich unter 2 Grad Celsius - möglichst unter 1,5 Grad Celsius - zu begrenzen.“ [1] Nicht nur für Unternehmen, sondern auch für Privatpersonen ist der Umweltschutz ein wichtiges Thema. „So spielt für 91% der Deutschen Umweltschutz eine wichtige Rolle“. [4] Mitarbeiter können sich besser mit Unternehmen identifizieren, wenn diese Maßnahmen für den Umweltschutz ergreifen.

Der Emissionszertifikatshandel

Der Emissionszertifikatshandel wurde von der Europäischen Union beschlossen und trat am 1. Januar 2005 in Kraft. Hierbei wird jedes Jahr eine Obergrenze für die Menge an emittiertem CO₂ festgelegt und in Form von Zertifikaten erstellt. Unternehmen, die eine große Menge an Treibhausgasen emittieren, bekommen eine definierte Anzahl an Zertifikaten zugeteilt, dabei steht ein Zertifikat für eine Tonne emittiertes CO₂. Benötigen Unternehmen mehr Zertifikate, können diese in einer Auktion ersteigert oder mit anderen Unternehmen gehandelt werden. Dieses Prinzip wird "cap-and-trade" genannt. Hierbei werden die Zertifikate erst auf die Länder aufgeteilt und die Zertifikate, die nicht benötigt werden, können gehandelt werden. [4] Die Obergrenze für die Zertifikate wird jedes Jahr gesenkt, dadurch soll sichergestellt werden, dass die Minderung gemäß dem Kyoto-Protokoll bzw. dem Pariser Übereinkommen erreicht wird.

Der Emissionszertifikatshandel ist umstritten, da die Festlegung der Obergrenze schwierig ist, wodurch die Preise für die Zertifikate stark variieren und teilweise weit sinken. Zudem ist noch nicht klar, ob die generierte Menge Geld für die Investition in klimafreundliche Alternativen ausreichend ist. Die Auswertungen des Bundesministeriums für Wirtschaft und Energie 2 zeigen aber eindeutig, dass die CO₂-Emissionen in den letzten Jahren deutlich zurück gegangen sind. Insbesondere die Belastung durch Unternehmen ist unter anderem durch die Einführung des Emissionszer-

tifikatshandel im Jahr 2005 bis 2018 um 29% deutlich zurückgegangen.

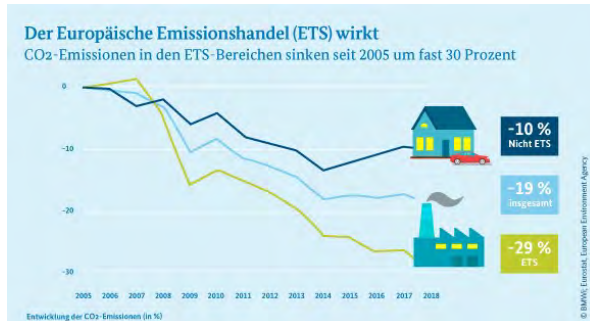


Abb. 2: Emissionshandel mit Erfolgsbilanz [2]

Tiny House

Durch den Handel von Emissionszertifikaten besteht zu einem die Möglichkeit, innerhalb von Unternehmen die Zertifikate zu handeln und zum anderen die Erstellung von Teilnehmerzertifikate, die in dieser Arbeit untersucht werden. Teilnehmerzertifikate sollen für die Nutzung eines Tiny House als Co-Working Space, das einen positiven Einfluss auf die Klimaziele hat, ausgestellt werden. Dieser Gedanke bringt viele Vorteile mit sich, nicht nur für die Umwelt, sondern auch für Unternehmen. Ein Tiny House ist ein kleines, energieautarkes Haus, meist in der Größe eines Schiffcontainers, welches auf einem Anhänger montiert ist. Es ist standortunabhängig, kann problemlos im Grünen platziert und Mitarbeitern als Alternative zum klassischen Büro mit vielen Menschen geboten werden. Die Stromversorgung erfolgt meist durch eine eigene Photovoltaikanlage und ein Tiny House ist speziell gedämmt, sodass möglichst energieeffizient geheizt werden kann.

Viele Unternehmen bringen immer mehr Wohlfühlatmosfera zu den Mitarbeitern, von Grünanlagen bis hin zu ganzen Wasserfällen wie im Google Firmensitz. Ein Architekturbüro aus Belgien baut aus alten Schiffcontainern ein Tiny Office, mit Arbeitsplätzen und Besprechungsraum, aber das alles im Grünen wie in 3 zu sehen ist. [3]



Abb. 3: Tiny Office [3]

Dadurch kommen Mitarbeiter aus großen Städten raus in die Natur und können ihre Kreativität ausleben und mit neu gewonnener Produktivität arbeiten. Ein Tiny House bietet zudem einen guten Ausgleich zwischen Home Office und Großraumbüro, Mitarbeiter haben kürzere Wege und müssen nicht mit überfüllten öffentlichen Verkehrsmitteln oder gar mit Stau in die Stadt fahren und haben trotzdem einen gut eingerichteten Arbeitsplatz ohne Ablenkung von daheim. Die Steigerung der Produktivität hat somit einen Mehrwert für das Unternehmen, zusätzlich zum Ausgleich der Teilnehmerzertifikate.

Thema dieser Arbeit ist die Ausstellung und Übertragung von Teilnehmerzertifikaten bei der Nutzung eines Tiny House und stellt eine Pilotstudie dar. Unternehmen sollen ein Tiny House anmieten können und die dadurch eingesparten Emissionen als Zertifikat ausgestellt bekommen. Hierbei wird die Übertragung von Daten für die Erstellung der Zertifikate für die Einsparung der Unternehmen betrachtet.

Ausblick und Zusammenfassung

Klimawandel ist ein bedeutendes Thema, welches uns noch vor weitere Herausforderungen stellen wird. Der Gedanke, ein Tiny House als Co-Working Space zu nutzen, bringt viele Vorteile mit sich, aber natürlich auch Kosten für Unternehmen. Es steht außer Frage, dass Unternehmen Maßnahmen in Richtung Klimaziele durchführen müssen, aber die Lösung muss auch praktikabel und wirtschaftlich sein. Die weitere Datenerfassung und Erstellung von Teilnehmerzertifikaten wird die Nutzbarkeit für die Praxis aufzeigen.

Literatur und Abbildungen

- [1] Internetredaktion der LpB BW. Kyoto-protokoll. <https://www.lpb-bw.de/kyoto-protokoll#c19513>, 04 2021.
- [2] European Enviroment Agency Eurostat. Emissionshandel mit Erfolgsbilanz. https://www.bmwi-energiewende.de/EWD/Redaktion/Newsletter/2020/08/Meldung/direkt-erfasst_infografik.html, 2021.
- [3] Gertraud Gerst. Ein Büro im Kornfeld. <https://www.ubm-development.com/magazin/neuer-trend-tiny-office/>, 2021.
- [4] Bastian Mutschler. *CO2-Kompensation als Argument im Marketing*. Diplomica Verlag GmbH, 2018.

Design and implementation of a Data Quality Management Framework

Bahar Uzun

Reinhard Schmidt

Department of Computer Science and Engineering, Esslingen University

Work carried out at Bank Vontobel AG, Zürich, Switzerland

Introduction

The digital transformation results in ever more significant technological advances that are unlocking new opportunities and fueling the economy. However, the full potential can only be realized if we harness the potential of the data generated in the course of digitalization. Due to this quantitative alteration, quality is becoming more and more important. [2] In recent years, data-driven business with its respective decision-making resulted in the need for good data quality. Vontobel's Lighthouse 2020 strategy includes the mission to seamlessly interweave client and investment processes with digital data and analytics and to make them the core of the business model. [1] To enable the client units Wealth Management, Asset Management, as well as Digital Investing the best possible decision-making, knowledge must be gained from the available data. Therefore, holistic

data management is introduced, whereby the focus of this work is on data quality.

Approach

On an operational level, various roles are identified based on data governance, master metadata management is operated and processes are defined.

Data quality is often understood as a multidimensional concept, that is illustrated in 1 and comprises the dimensions Completeness, Consistency, Uniqueness, Validity, Accuracy, and Timeliness. These dimensions reflect a measure for the suitability of data for specific requirements in business processes. [3] First, the status quo regarding data quality must be determined. For this purpose, interviews are conducted with various experts, use cases are identified, and data quality metrics that are already in use are identified.

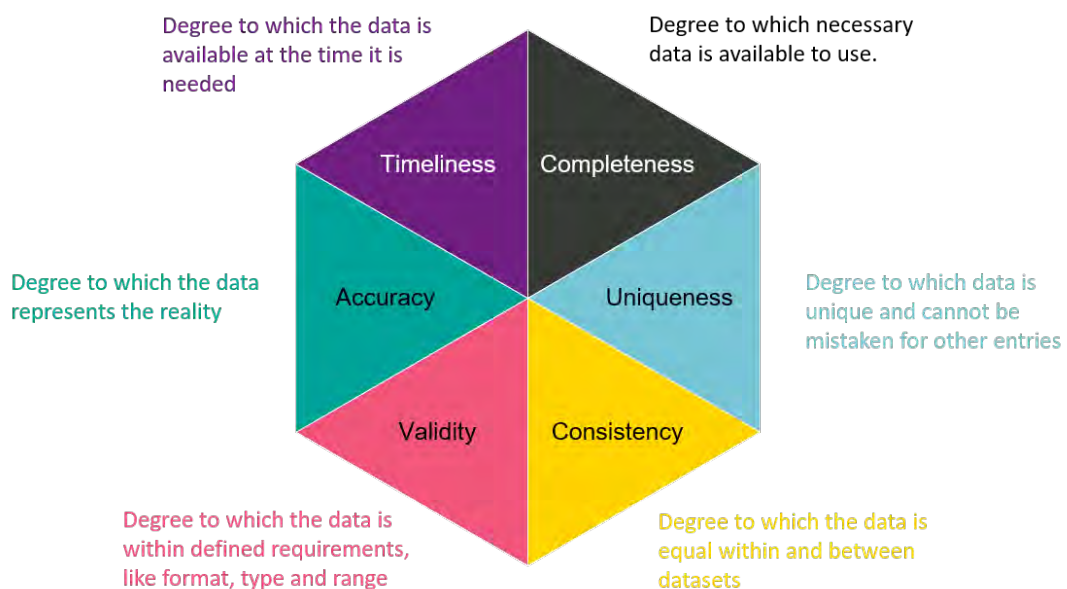


Fig. 1: The six dimensions of Data Quality [4]

This is followed by the design of the framework. It is important to note that preventive measures are used. Data quality checks are to be automated and made measurable to ensure the usability of the data. On a technical level, data quality processes are implemented in the Data Quality Tool. First, the possibilities and limitations of this tool are explored. Essentially, it is possible to connect various databases, use quality metrics, store results, ensure automation, and create various data views and dashboards. One measure for monitoring data quality is a data quality score. This

is to be integrated into the data governance tool via an interface and be visible to the data users. At the same time, case management is to be introduced in the data quality management framework. Thus, one requirement is the central possibility for data users to request quality metrics that are meaningful for their handling of data. Another requirement for the case management is an automated process start to resolve data quality issues. A simplified overview of the interaction between the two tools is shown in 2.

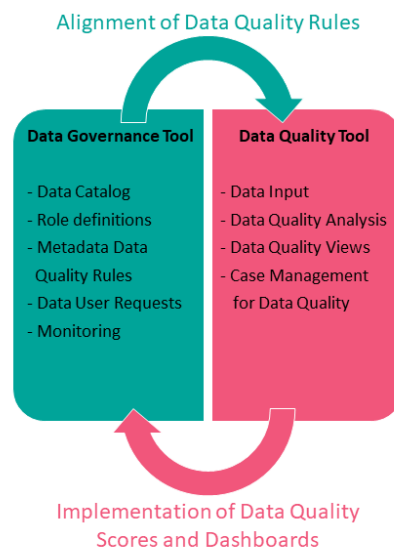


Fig. 2: Interaction of the Data Governance tool and Data Quality tool [4]

Outlook

Based on the results, further data quality checks can be introduced centrally for future use cases. Furthermore,

dashboards for different stakeholders can be built and shared with different granularity. Data quality can thus be monitored and improved in the long term.

References and figures

- [1] Vontobel Holding AG. Geschäftsbericht 2020. https://www.vontobel.com/siteassets/about-vontobel/downloads/2021-02-11_fr_geschaeftsbericht_de.pdf, 02 2021.
- [2] Viktor Mayer-Schönberger, Kenneth Cukier, et al. *Big Data: Die Revolution, die unser Leben verändern wird*. Redline Verlag, 2013.
- [3] Boris Otto and Hubert Österle. *Corporate Data Quality: Voraussetzung erfolgreicher Geschäftsmodelle*. Springer Berlin Heidelberg, 2015.
- [4] Own representation.

Software-Geschäftsmodellentwicklung im Kontext der Bauteileoptimierung durch künstliche Intelligenz

Lora Venkova

Astrid Beck

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma evia innovation GmbH, Stuttgart

Einleitung

Künstliche Intelligenz (KI) wird immer mehr ein Teil von unserem Leben und vor allem bei der Entwicklung neuer technologischer Innovationen. Zu den großen Teilgebieten der KI zählen maschinelles Lernen, tiefes Lernen und neuronale Netze (siehe Abb. 1), mit denen Algorithmen und Techniken viele reale Probleme gelöst werden können.

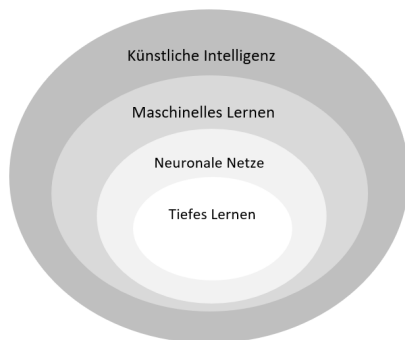


Abb. 1: Abstufung von KI (in Anlehnung an Gavrilova, 2021) [3]

Dadurch eröffnen sich enorme Möglichkeiten, Aufgaben, die früher menschliche Intelligenz erforderten, besser, schneller und effizienter zu erledigen. Mit Hilfe von KI ist es möglich, Situationen, Daten und Informationen zu analysieren und auch vorherzusagen. Auf diese Weise werden Erkenntnisse bereitgestellt, die dazu beitragen, den Betrieb und seine Leistung zu verbessern. Mit Hilfe dieser Vorhersagen können immer mehr Entscheidungsfindungen berücksichtigt werden, was zur Festlegung realistischer Geschäftsziele für die Einführung von KI führt. Dadurch können beispielsweise Optimierungen in verschiedenen Bereichen angeboten werden, eine davon ist die montageorientierte Produktion. Die Hauptziele von der Prozessoptimierung mithilfe KI sind Kostenersparnis, Zeiteinsparung, Minimierung von Produktionsausfällen, erhöhte Effizienz

und Sicherheit in der erzielten Ergebnissen.

Zielsetzung

Die nachfolgende Arbeit wird in Zusammenarbeit mit der evia innovation GmbH und dem Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA erstellt. Das Ziel ist die Entwicklung von einem Software-Geschäftsmodell in Betrachtung des KI-Mont Projekts. Dabei werden die Hauptmethoden- und Komponenten der KI und der intelligenten Automatisierung berücksichtigt, und was hinter der Technologie von KI-Mont steckt. Es wird geforscht, welche Zielgruppen/Branchen und Märkte am besten für das Projekt geeignet sind. Es wird eine Analyse der unterschiedlichen Einflussfaktoren darunter die Bedürfnisse, das Kundensegment, das Produkt/Angebot, mögliche Implikationen und weitere Aspekte erstellt.

KI-Toolchain bzw. KI-Mont

KI-Mont ist ein Online-Tool, das effiziente Bauteiloptimierung für unterschiedliche Branchen in der Industrie bietet. Die Methoden, die dahinterstecken, ermöglichen CAD-Bauteile anhand ihrer Eigenschaften zu visualisieren und an Basis von denen eine Verbesserung bzw. Optimierung anzubieten. Mit diesem Tool existiert die Möglichkeit, ohne die Hilfe von Spezialisten die Erkennbarkeit von Bauteilen und danach die Vereinzelbarkeit zu analysieren und zu bewerten. Für das Ziel werden STEP-Files genutzt, die hochgeladen werden können (siehe Abb. 2). Mit deren Hilfe wird ein Diagramm ausgestellt, wo die notwendige Informationen enthalten sind. Die CAD-Dateien werden nach Orientierungsfähigkeit, Trennbarkeit und Greifbarkeit geforscht. Dazu wird eine Bewertung nach einer Skala von eins bis zehn durchgeführt. Dabei wird auch die Wahrscheinlichkeit angegeben, wie korrekt die Bewertung durchgeführt ist. Um die Funktionsweise von KI-Mont werden verschiedene Algorithmen und xAI Methoden verwendet. Eine davon ist die Layer-

wise Relevance Propagation (LRP) Teil von dem Salient Relevance (SR) map, die darauf abzielt, wie tief CNNs Bilder erkannt werden und Merkmale aus bestimmten Bereichen verstanden werden. Und DeepDream Algorithmen/Sensitivitätsanalyse, die nach dem Prinzip arbeitet, ein Bild überzuarbeiten und seine Eigenschaften zu verbessern.

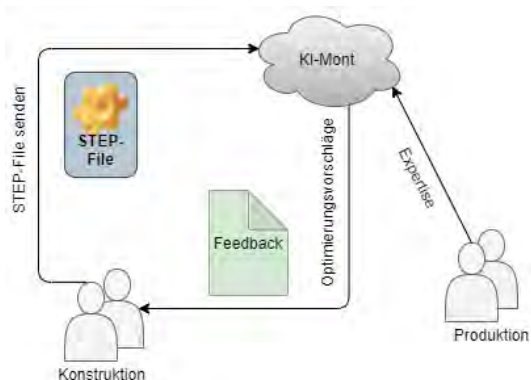


Abb. 2: KI-Mont Tool [1]

Künstliche Neuronale Netze

KI-Mont Struktur und Arbeitsweise wird auf Basis künstlicher neuronaler Netze (KNN) erstellt (siehe Abb. 3).

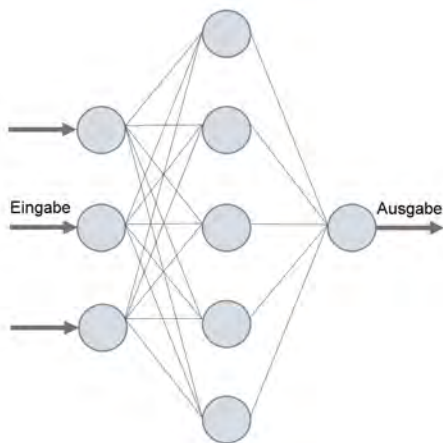


Abb. 3: Schematischer Aufbau eines einfachen neuronalen Netzes (in Anlehnung an Konertz und Schönhof, 2020, S. 47) [4]

Die neuronale Netze bestehen im Wesentlichen aus drei Schichten: Eingabeschicht, Zwischenschicht (versteckte Schicht) und Ausgabeschicht. Dabei hat jede Verbindung ein eigenes Gewicht. Die erste Schicht erhält die Daten als Eingabe. Wenn dieser Eingabewert ankommt, wird er mit einem Gewichtswert multipliziert. Jede nachfolgende Schicht erhält die Eingabedaten als Ausgabe von der vorherigen Schicht. Die versteckte Schicht liegt zwischen der Eingabe- und der Ausgabeschicht, und hat die Aufgabe alle notwendigen Berechnungen durchzuführen. Die Ausgabeschicht erzeugt dann das Ergebnis für die gegebene Eingaben. Das passiert, wenn die Eingaben mit den entsprechenden Gewichtungen multipliziert werden. In KI-Mont wird das mehrlagiges Perzeptron (MLP) mehrschichtiges neuronales Netzwerk (tiefes neuronales Netz) [5] verwendet.

Software-Geschäftsmodellentwicklung

Das Geschäftsmodell wird nach dem Prinzip des St. Galler Business Modell Navigator entwickelt. Dabei wird auf die 4 Punkten: Initiierung (Umfeldanalyse), Ideenfindung (Wahl zwischen Ähnlichkeitsprinzip oder Konfrontationsprinzip), Integration (interne und externe Komponenten), Realisierung (Prototyperstellung und Optimierungen) [2] eingegangen. Mit Hilfe des Modells wird die Marktleistung verfolgt und analysiert. Durch die unterschiedlichen Phasen werden alle Aktivitäten im Zuge des Produktkonzepts untersucht, was zu der allgemeinen Situationsbewertung hilft.

Ausblick

Im weiteren Verlauf der Abschlussarbeit wird auf das technische Konzept in Verbindung KI, das Produktkonzept und der Entwicklung von dem Software-Geschäftsmodell näher eingegangen. Dazu wird die Durchführung von der Fallstudie und Marktanalyse in Betrachtung genommen. Am Ende werden die Ergebnisse, die Vorausschau über zukünftige Chancen für das etablierte Geschäft zu erkennen und zu berücksichtigen.

Literatur und Abbildungen

- [1] Eigene Darstellung.
- [2] Oliver Gassmann, Karolin Frankenberger, and Michaela Csik. *Geschäftsmodelle entwickeln. 55 innovative Konzepte mit dem St. Galler Business Model Navigator*. Hanser Verlag GmbH & Co. KG, München, 2013.
- [3] Yulia Gavrilova. Artificial Intelligence vs. Machine Learning vs. Deep Learning: Essentials. <https://serokell.io/blog/ai-ml-dl-difference>, 2020.
- [4] Roman Konertz and Raoul Schönhof. *Das technische Phänomen „Künstliche Intelligenz“ im allgemeinen Zivilrecht*. Nomos Verlagsgesellschaft mbH & Co. KG, 2020.
- [5] Raoul Schönhof and Manuel Fechter. Towards automated Capability Assessment leveraging Deep Learning. *Procedia CIRP* 91(2), pages 433–438, 2020.

Konzeption und Implementierung einer Website zur Evaluation von 3D-Interaktionstechniken

Simon Weber

Andreas Rößler

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einführung

In den letzten Jahren ist die Bedeutung von 3D-Benutzeroberflächen und -Interaktionen immer wichtiger geworden. Sei es durch populäre Medien, wie "Ready Player One", die Nutzung in der Industrie zum Beispiel für die Darstellung von komplexen 3D-Modellen, oder in jüngster Vergangenheit durch die Umbenennung von Facebook in Meta und die anschließende Enthüllung des Metaverse. Dabei handelt es sich um das kommende Virtual-Reality-Konzept von Meta. Auch ist die zugrunde liegende Technologie nach Jahren der Forschung so ausgereift, dass sie als Unterhaltungselektronik verkauft werden kann.

Motivation

Bei der Entwicklung eines 3D User Interface (UI) gibt es verschiedenste Probleme. Zum einen ist die 3D-Interaktion schwierig [3]: Der Mensch hat Probleme sich im dreidimensionalen Raum zurecht zu finden, da die physische Welt weit mehr Hinweise für die Interaktionen und Beschränkungen bietet. Auch muss bei der Konzeption dieser UIs auf die Usability geachtet werden, da traditionelle 2D-Interfaces (WIMP - "Windows, Icons, Menus, Pointer") nicht eins zu eins übernommen werden können. An diesem Aspekt setzt diese Arbeit an: Die Bereitstellung einer im Internet besuchbaren Website, die verschiedene 3D-Interaktionstechniken

dem User bereitstellt und dessen Interaktionen auswertet. Somit können einfache, kostengünstige und einer breiten Masse zugängliche Usability Tests für 3D-Interaktionstechniken bereitgestellt werden.

Konzeption und Implementierung

Zur Umsetzung der benötigten Funktionalitäten werden beispielhaft folgende Technologien verwendet:

- Oculus Quest VR-Brille, Entwicklung und Testing
- three.js für die Cross-Browser 3D-Darstellung und VR-Implementierung
- Node.js mit express und socket.io zur Übertragung der erhobenen Interaktionsdaten an das Backend

Die three.js-Bibliothek bietet verschiedenste Methoden um 3D-Modelle im Browser anzuzeigen, aber auch eine Einbindung der WebXR-API. Diese ermöglicht die einfache Erstellung und Darstellung von immersiven VR-Inhalten. Auf Grund der minimalen Anforderungen an das Backend und die wird Node.js mit express verwendet, um die Website bereitzustellen. Für die Kommunikation zwischen Client und Server wird Socket.IO verwendet. Alle verwendeten Technologien stehen als Open-Source-Projekte frei zur Verfügung.

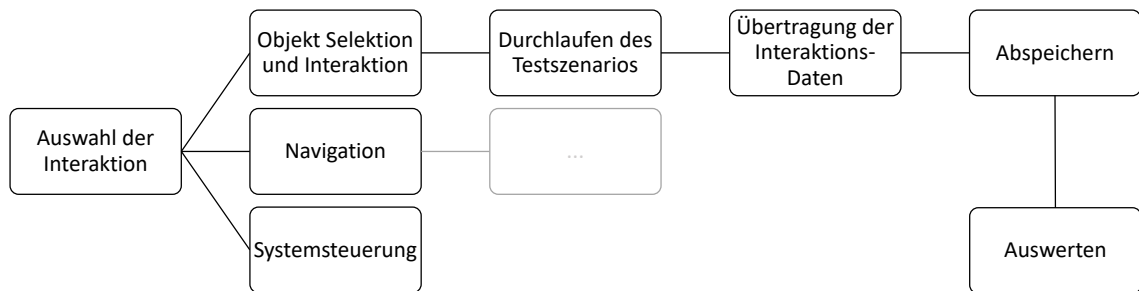


Abb. 1: Beispielhafter Ablauf [2]

Abbildung 1 zeigt einen beispielhaften Ablauf der Nutzerinteraktion mit der Website. Die grundlegenden Interaktionsarten, wie z.B. Objekt-Manipulation oder die Navigation im Raum werden in einem Spiel-ähnlichem Szenario vom Nutzer durchlaufen (kognitiver Walkthrough [1]), während relevante Usability-Daten an das Server-Backend gesendet werden. Diese Daten beinhalten neben verstrichener Zeit, auch Position der Pointer bzw. Hände der Nutzer. Diese werden dort abgespeichert und können am Ende der Evaluation ausgewertet werden.

Ausblick

In vielen Bereichen des Lebens hat die virtuelle oder augmentierte Realität Einzug gehalten und wird, auch durch die Promotion von Branchenriesen wie Meta/Facebook, weiter wachsen. Deshalb wird es für Forschung und Entwicklung für das immersive Web, aber auch generische VR-Anwendungen, hilfreich sein, mit Hilfe von einfachen Websites, die Interaktion der Nutzer evaluieren zu können. Das entwickelte System ist als eine Art Prototyp oder Proof-Of-Concept zu sehen, das nicht in der Produktion benutzt werden sollte. Mit ein paar kleinen Anpassungen wäre es jedoch fit dafür. Auch könnte eine automatisierte Testauswertung implementiert werden, um großflächige Nutzer-Tests im Internet, noch einfacher zu gestalten.

Literatur und Abbildungen

- [1] Doug A Bowman, Joseph L Gabbard, and Deborah Hix. A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence*, 11, 2002.
- [2] Eigene Darstellung.
- [3] Joseph J LaViola Jr, Ernst Kruijff, Ryan P MacMahan, Doug A Bowman, and Ivan Poupyrev. *3D User Interfaces : Theory and Practice*. Addison-Wesley, 2 edition, 2017.

Analyse für den Einsatz von Machine Learning für die Security Log Korrelation und Alarmierung

Tobias Winkler

Steffen Schober

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Thales Deutschland GmbH, Ditzingen

Einleitung & Motivation

Integrität und die Verfügbarkeit sind in sicherheitsrelevanten Systemen von hoher Bedeutung. Die Grundlage für die Beobachtung der Sicherheit, sowie der Verfügbarkeit, erfolgt in der Regel über die Auswertung von Logdaten. Logdaten eines System können durch entsprechende Auswertung wichtige Informationen über den System- und Gerätestatus bereitstellen. Durch die passende Auswertung können Anomalien bzw. Probleme erkannt und gemeldet werden. Für eine automatische Erkennung gibt es verschiedenste Gründe. Ein wesentlicher Punkt ist der Zeitfaktor. Sowohl die Erkennung, als auch die Auswertung muss in sicherheitsrelevanten Systemen schnell erfolgen, um auf Anomalien entsprechend reagieren zu können. Eine manuelle Analyse von Logdaten wäre mit hohem Zeitaufwand und hoher Komplexität verknüpft. Daher benötigen entsprechende Systeme ein Monitoringsystem für Logdaten.

Thales Deutschland entwickelt im Bahnbereich verschiedenste Systeme und Geräte für die Steuerung und Sicherung des Bahnbetriebs. Diese fallen in einen sicherheitsrelevanten Bereich, sodass Sicherheit und Verfügbarkeit eine wichtige Rolle spielen. Um sämtliche sicherheitsrelevante Logdaten zu sammeln und auszuwerten, entwickelt Thales eine Central Logging Application, welche als Observer bezeichnet wird. Der Observer basiert auf dem ELK-Stack und einer Alerting Engine. ELK steht für die drei Komponenten Elasticsearch, Logstash und Kibana. Logstash ist für die Erfassung und Normalisierung, Elasticsearch für die Speicherung und Kibana für die Visualisierung, der eintreffenden Logs, verantwortlich. [4] Die Alerting Engine ist für das Erkennen von Anomalien zuständig. Die Alerting Engine benötigt Regeln, um Anomalien zu erkennen. Diese Regeln müssen aktuell statisch durch eine Person definiert werden, sodass die Alerting Engine eigenständig Anomalien erkennt und als Alarme meldet.

Problemstellung & Zielsetzung

Aufgrund der hohen Datenmenge und einer lediglich statisch definierte Alerting Engine ergibt sich die Frage, inwiefern Machine Learning, das beschriebene System verbessern bzw. erweitern kann. Hierfür muss zunächst das bestehende System analysiert werden und ein passender Ansatz für Machine Learning Modelle gefunden werden. Die Kernfrage der Arbeit bezieht sich auf die Analyse der Erweiterung des Observers bzw. der Alerting Engine um einen Machine Learning Ansatz. Es soll mithilfe eines passenden Machine Learning Models die Kategorisierung und Erkennung von Anomalien verbessert werden. Hieraus ergeben sich folgende Forschungsfragen, mit welchen sich die Abschlussarbeit befasst.

1. Wie kann Machine Learning die Arbeit der Alerting Engine erweitern, verbessern oder ersetzen?
2. Funktioniert eine KI-gestützte Erkennung besser als die bisherige Alerting Engine?
3. Wie kann eine KI-gestützte Erkennung in das bestehende System integriert werden?
4. Ist es für Thales sinnvoll einen Machine Learning Ansatz weiterzuentwickeln?

Ansatz & Grundlagen

Für die Umsetzung müssen zunächst verschiedene Ansätze herausgearbeitet werden, um Logdaten mithilfe von Machine Learning Modellen zu analysieren. Abbildung 1 beschreibt in fünf Schritten wie eine Log-Analyse abläuft. Die ersten drei Schritte beziehen sich auf die Datenanpassung bzw. Datenvorbereitung, um sie für Machine Learning Ansätze nutzen zu können. Darauf folgt die eigentliche Log-Analyse und die daraus resultierende Ergebnisse. [1]

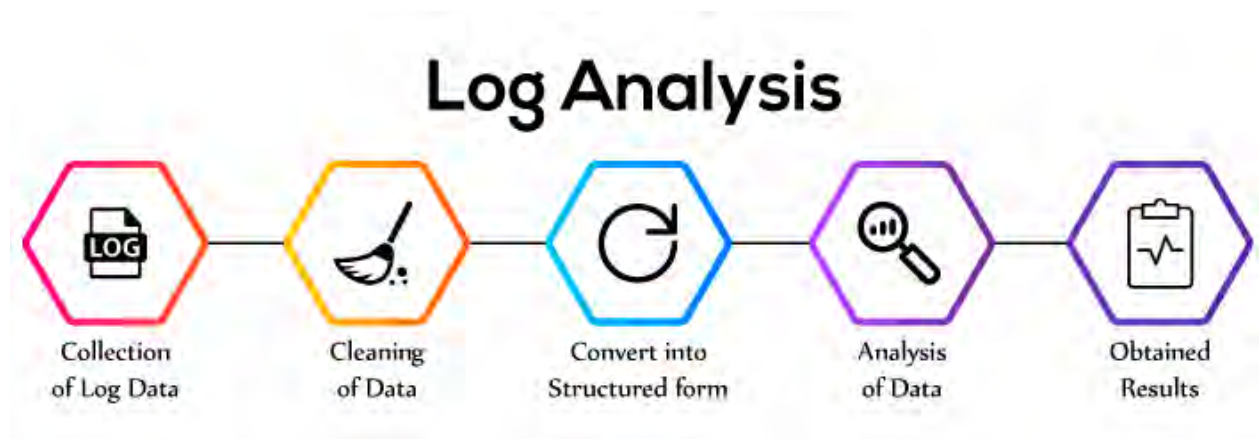


Abb. 1: Abbildung 1: Log Analysis [2]

Da Logdaten größtenteils textuelle Informationen sind, die in unterschiedlichen Formaten und Inhalten vorkommen, ist die Vorbereitung der Daten eine wichtige Grundlage. Hierfür gibt es im Bereich des Text Mining verschiedenste Ansätze. Ein Teilgebiet stellt Natural Language Processing (NLP) dar. Hier wird versucht Text durch numerische Werten darzustellen und diesen eine Gewichtung mitzugeben, sodass auch die Bedeutung bzw. der Kontext, des Textes, miteinbezogen wird. Dabei gibt es verschiedenste Ansätze, die je nach gewähltem Machine Learning Modell genutzt werden können. [1] Beispiele wären hierfür Bag of Words (BAG), Term Frequency - Inverse Document Frequency (TF-IDF) oder Word2Vec. [5]

Für die Machine Learning Modelle eignen sich aufseiten des Supervised Learning verschiedenste Klassifizierungsansätze wie Super Vector Machine (SVM) oder Decision Trees. [6] Bei einem solchen Ansatz wird mithilfe durch den Observer kategorisierter Logdaten ein Modell antrainiert, um so unbekannte Logdaten mithilfe des Modells zu kategorisieren und Anomalien zu erkennen. Da die Trainingsdaten auf Grundlage der Alerting Engine gelabelt sind, wird mithilfe dieses

Ansatzes vermutlich lediglich eine generische Alerting Engine nachgebildet. Ein weiterer Ansatz ist ein Deep Learning Ansatz, wie beispielsweise DeepLog. Hier wird ein Modell mithilfe von Trainingsdaten erzeugt. Die Daten repräsentieren Logdaten unter normaler Ausführung des Systems. Das Modell erlernt Log Patterns ohne Anomalien und kann so Logs welche nicht in die erlernten Patterns passen entdecken. [3]

Ausblick

Durch die Erkenntnisse dieser Abschlussarbeit soll analysiert werden, ob ein Machine Learning Ansatz, für den beschriebenen Observer, eine Verbesserung in der Erkennung von Anomalien bewirkt. Hierfür werden die gewählten Modelle und deren Ergebnisse, mit den aktuell erkannten Anomalien, des Bestandssystems, verglichen. Hierdurch wird deutlich, inwieweit eine Verbesserung des Bestandssystems erzielt werden kann. Für diesen Vergleich müssen als nächste Schritte geeignete Modelle ausgewählt werden und mithilfe von verschiedensten Logdaten, Ergebnisse gesammelt werden.

Literatur und Abbildungen

- [1] Menzli Amal. A Machine Learning Approach to Log Analytics – How to Analyze Logs? <https://neptune.ai/blog/machine-learning-approach-to-log-analytics>, 2021.
- [2] A. Arnaud. XenMobile – Troubleshooting Log Collection Guide. <http://arnaudpain.com/2017/12/14/xenmobile-troubleshooting-log-collection-guide/#sthash.xB4MFuuK.dpbs>, 2017.
- [3] Min Du, Feifei Li, Guineng Zheng, and Vivek Srikumar. DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning. <https://www.cs.utah.edu/~lifeifei/papers/deeplog.pdf>, 2017.
- [4] B.V. Elasticsearch. Was ist der ELK-Stack? <https://www.elastic.co/de/what-is/elk-stack>, 2021.
- [5] Purva Huilgol. Quick Introduction to Bag-of-Words (BoW) and TF-IDF for Creating Features from Text. <https://www.analyticsvidhya.com/blog/2020/02/quick-introduction-bag-of-words-bow-tf-idf/>, 2020.
- [6] M. IKONOMAKIS, S. KOTSIANTIS, and V. TAMPAKAS. Text Classification Using Machine Learning Techniques. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.9153&rep=rep1&type=pdf>, 2005.

Entwicklung eines Vision-Objekterkennungs-Stacks für Autonomes Fahren

Muhammed Yildirim

Clemens Klöck

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Ziel der Arbeit

Das Ziel der vorliegenden Arbeit ist es, eine Software zu entwickeln, die mithilfe von Daten aus einer Monokamera, eine Objekterkennung durchführen kann. Dies wird durch die Anwendung von Machine Learning Algorithmen ermöglicht. Ein Objektdetektor soll mithilfe der Objekterkennungs-API von Tensorflow erstellt und mit ausgewählten Datensätzen trainiert werden. Im Anschluss wird das trainierte Netz ausgewertet und verbessert. Sind die Ergebnisse positiv, wird die Software in einem Forschungsfahrzeug getestet.

Neuronale Netze

Bei einem Neuronalen Netz (NN) wird die Funktionsweise des menschlichen Gehirns nachempfunden. Es ist ein wichtiger Entwicklungszweig im Bereich der Künstlichen Intelligenz. Mithilfe von Deep Learning Algorithmen können NN Muster in großen Datenmengen hocheffizient erkennen. NN findet in wichtigen Forschungsbereichen wie Gesichts-, Objekt- und Spracherkennung Anwendung, da hier Muster in großen Datenmengen analysiert werden müssen. [5]

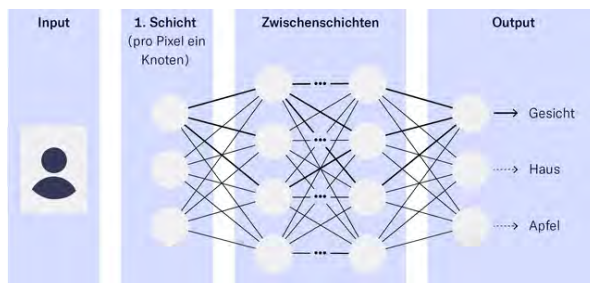


Abb. 1: Aufbau eines neuronalen Netzes [4]

NN haben folgende Struktur: Einen Input, Hidden und Output Layer. In jedem Layer befinden sich eine bestimmte Anzahl von Neuronen, die mit den Neuronen in der vorherigen und folgenden Layer mit Gewichtungen vollständig verbunden sind, außer im Input und Output

Layer. Die Gewichtungen werden während des Trainings so angepasst, dass der erwartete Output ausgegeben wird. Im Input Layer ist beispielsweise jedes Neuron ein Pixel eines Bildes. Was im Hidden Layer passiert sieht man nicht, da es abstrakt gehalten wird. Im Output Layer geben die Neuronen die Werte aus. Sind alle Neuronen miteinander verbunden, spricht man von einem Fully-connected Layer. Hat man sehr viele Hidden Layer, handelt es sich um Deep Learning.

CNN

Die Architektur eines CNN unterscheidet sich von den klassischen NN. CNN sind aus verschiedenen Schichten zusammengesetzt. Die Convolutional-Schicht erkennt und extrahiert einzelne Merkmale eines Bildes. Die Pooling-Schicht reduziert die Auflösung eines Bildes, indem sie überflüssige Informationen entfernt und dabei die Berechnungsgeschwindigkeit erhöht. [7]

Objektdetektion

In der Objektdetektion geht es darum, Objekte in einem Bild oder Video zu identifizieren und lokalisieren. Sogenannte bounding boxes markieren dabei die detektierten Objekte und kennzeichnen sie mit deren Klasse und Position im Bild. Der Unterschied zwischen der Objektdetektion und Bilderkennung ist, dass bei der Bilderkennung nur ausgegeben wird, um was für eine Klasse, z.B. "Dog" es sich handelt. Bei der Objektdetektion werden in einem Bild oder Video Objekte zusätzlich auch lokalisiert. Damit ist es möglich, Objekte zu zählen und diese zu verfolgen. Die Anwendungsgebiete sind vielfältig, beispielsweise im Bereich Videoüberwachung, Autonomous Driving und Gesichtserkennung. [1]

TensorFlow Object Detection API

TensorFlow Object Detection API ist ein Open-Source Framework, um leichter Objektdetektion Modelle zu

konstruieren, trainieren und zur Verfügung zu stellen. Es gibt eine Sammlung von vortrainierten Modellen, die auf dem COCO-Datensatz, KITTI-Datensatz und Open Images Dataset trainiert wurden. Beim Aufbau der API mussten Bilder aufgenommen und mit dem Tool Labellmg gelabelt und abgespeichert werden. Diese werden als XML-Files abgespeichert. Daraufhin werden die XML-Files zu TFRecord Files konvertiert. TensorFlow bietet durch TensorFlow Zoo Garden eine

Auswahl an vortrainierten Modellen. Nachdem diese heruntergeladen wurden, können in einem config file Parameter verändert werden wie die batch size, etc. und das Modell kann trainiert werden. Im Anschluss ist es möglich, über das TensorBoard nachschauen, wie die Loss Function war. Zuletzt wird mit Testbildern überprüft, wie gut das trainierte Modell Objekte detektiert. [8]

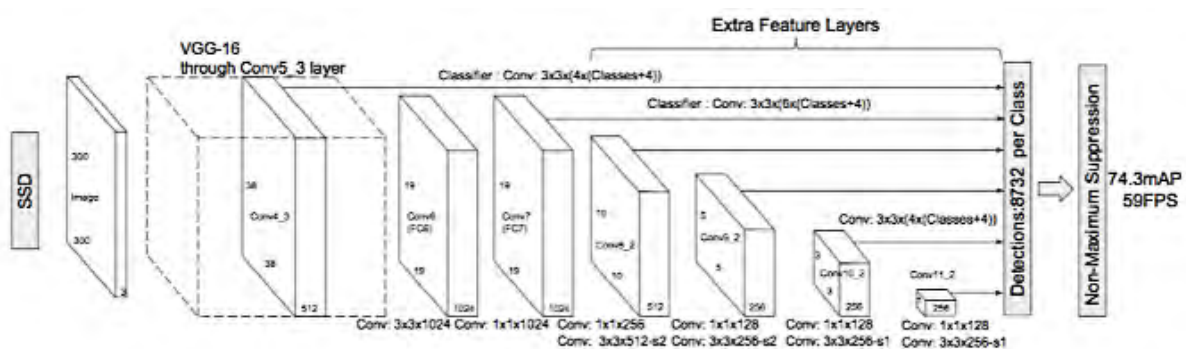


Abb. 2: Architektur eines single-shot detectors mit Backbone (VGG-16) [3]

SSD

SSD steht für "single-shot detector" und ist ein populäres Objekterkennungsmodell. Es hat zwei Komponenten, ein Backbone-Modell (z.B. VGG-16), das ein vortrainiertes Bildklassifizierungsnetz als Merkmalsextraktor hat und ein SSD-Head, der ein oder mehrere Convolutional Layers hat, die zum Backbone hinzugefügt werden. Die SSD benötigt für das Training nur ein Bild und Ground-Truth-Boxen für jedes Objekt während des Trainings. Somit ist es in der Lage mit nur einer Aufnahme ein Objekt in einem Bild zu erkennen. [2]

Datensatz

Um ein Modell gut zu trainieren, benötigt es vorgelabelte Bilder. Je besser die Datensätze sind, desto besser kann das Modell die Objekte detektieren. In meinem Fall habe ich mich für den Waymo Datensatz entschieden. Dieser Datensatz besteht aus 11.8 Millionen 2D bounding box labels mit IDs von Kameradaten. Die Objektklassen sind Fahrzeuge, Fußgänger, Fahrräder und Schilder. Es hat Bilder aus fünf Perspektiven (links, vorne-links, vorne, vorne-rechts und rechts). In meinem Fall sind nur die Bilder von vorne relevant. [6]

Literatur und Abbildungen

- [1] Fritz Ai. Object Detection Guide. <https://www.fritz.ai/object-detection/>, 2021.
- [2] ArcGIS Developers. How single-shot detector (SSD) works? <https://developers.arcgis.com/python/guide/how-ssd-works/>, 2021.
- [3] Eddie Forson. Understanding SSD MultiBox — Real-Time Object Detection In Deep Learning. <https://towards-datascience.com/understanding-ssd-multibox-real-time-object-detection-in-deep-learning-495ef744fab>, 2017.
- [4] Ruth Fulterer and Eveline Geiser. Was ist künstliche Intelligenz? Wie funktioniert Deep Learning? Ein Überblick. <https://www.nzz.ch/technologie/kuenstliche-intelligenz-ki-deep-learning-einfach-erklaert-ld.1632034#subtitle-was-sind-k-nstliche-neuronale-netze-second>, 2021.
- [5] retresco GmbH. Neuronale Netze. <https://www.retresco.de/lexikon/neuronale-netze/>, 2021.
- [6] Waymo LLC. About Waymo Open Dataset. <https://waymo.com/open/about/>, 2021.
- [7] Stefan Luber and Nico Litzel. Was ist ein Convolutional Neural Network? <https://www.bigdata-insider.de/was-ist-ein-convolutional-neural-network-a-801246/>, 2019.
- [8] Hussain Mujtaba. Real-Time Object Detection Using TensorFlow. <https://www.mygreatlearning.com/blog/object-detection-using-tensorflow/>, 2020.