

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitel: 1 Entscheidungsunterstützung in Produktion und Logistik mittels Reinforcement Learning



Projektübersicht

Anzahl Studierende	1
Art (gefördertes Projekt, Projekt mit externen Partnern, Studienprojekt)	Studienprojekt
Projektverantwortung	Prof. Dr. Christian Schwede
Projektkontext	Projekt in Zusammenarbeit mit dem betreuenden Professor als Vorbereitung auf eine mögliche Konsortialforschung in dem Kontext.

Abstrakt

Zur Planung von komplexen Produktions- und Logistiksystemen werden immer häufiger Simulationsmodelle eingesetzt, um die Auswirkungen von Entscheidungen (z.B. Ressourcenpläne, Produktionsprogramme, Engpassmaßnahmen) detailliert evaluieren zu können. In dem hier beschriebenen Projekt soll mittels Reinforcement Learning die Erstellung der Entscheidungsszenarien, durch eine Künstliche Intelligenz automatisiert werden. Das Verfahren des maschinellen Lernens kann hierbei die Simulation als Trainingsumgebung verwenden.

Kurzbeschreibung

Zur Planung von komplexen Produktions- und Logistiksystemen werden immer häufiger Simulationsmodelle eingesetzt, um die Auswirkungen von Entscheidungen (z.B. Ressourcenpläne, Produktionsprogramme, Engpassmaßnahmen) detailliert evaluieren zu können. Diese Modelle werden neuerdings mit digitalen Echtzeitabbildern der Produktions- und Logistiksysteme sogenannten

„Digitalen Zwillingen“ verbunden. Dadurch können die Modelle auch in der kurzfristigen Planung und Echtzeitsteuerung eingesetzt werden. Kommt beispielsweise eine wichtige Lieferung zu spät zur Fabrik, so können die Auswirkungen mit der Simulation auf Basis der Echtzeitdaten des digitalen Zwilling ad-hoc bewertet und Gegenmaßnahmen überprüft werden.

Neben der aufwändigen Erstellung der Simulationsmodelle ist der Entscheidungsprozess bis heute im Wesentlichen manuell. Entscheidungsszenarien werden händisch im Entscheidungsunterstützungssystem erstellt, von der Simulation bewertet und dann erneut händisch angepasst.

In dem hier beschriebenen Projekt soll mittels Reinforcement Learning die Erstellung der Entscheidungsszenarien, durch eine Künstliche Intelligenz automatisiert werden. Das Verfahren des maschinellen Lernens kann hierbei die Simulation als Trainingsumgebung verwenden. Es erzeugt Eingabe und erhält durch die Simulation Feedback über deren Güte. Auf diese Weise können einerseits Entscheider durch intelligente Vorschläge unterstützt und andererseits perspektivisch weniger schwerwiegende Entscheidungen vollständig autonom getroffen werden.

Aufgabenstellung

Der/ die Studierende wird für ein vorgegebenes Produktions-/Logistiksystems ein Simulationsmodell mit der Simulationssoftware Siemens Plant Simulation erzeugen. Dieses Modell wird als Grundlage für die Entwicklung eines Verfahrens des Reinforcement Learning dienen, das Lernen soll für bestimmte Rahmenbedingungen optimale Entscheidungsszenarien zu generieren.

Bezug zum Thema Data Science

Verfahren des Reinforcement Learning sind ein Kernthema der Data Science und werden auch in den Veranstaltungen des Forschungsmasters behandelt.

Verfügbare Ressourcen

- Informationen, die zur Erstellung des Simulationsmodells benötigt werden (Systembeschreibung, Durchlaufzeiten, Kapazitäten, etc.) werden vom betreuenden Professor bereitgestellt
- Siemens Plant Simulation wird im Kontext des AI-Labs der FH Bielefeld bereitgestellt
- Hardware für das Reinforcement Learning ist über das Data Science Lab oder das CfADS der FH Bielefeld verfügbar

Projektplan

Erstes Semester: Formulierung des Forschungsexposees, Einarbeitung in Plant Simulation, Sammeln der notwendigen Information zur Erstellung des Simulationsmodells

Zweites Semester: Erstellung eines einfachen Simulationsmodells, Recherche zu relevanten Arbeiten in Themenfeld der Kopplung von Simulation und Maschinellern

Drittes Semester: Entwicklung von ersten Verfahren des Maschinellen Lernens, Entwicklung einer Schnittstelle zu Plant Simulation, Auswertung erster Ergebnisse

Viertes Semester: Erstellung eines komplexen Simulationsmodells, Implementierung und Vergleich von weiteren Verfahren, Fine-Tuning und Optimierung der Verfahren, Verbesserung der Ergebnisse, Finale Evaluierung

Eignungskriterien

Zwingend:

- Programmierkenntnisse

Optional:

- Simulationserfahrung
- Erfahrung mit Produktions- und Logistiksystemen

Erwerbbarere Kompetenzen

- Einsatz von Verfahren des Reinforcement Learnings
- Ereignisdiskrete Materialflusssimulation mit Plant Simulation
- Planung von Produktions- und Logistiksystemen

Projekttitle: 2 Der KI-Rechtsberater



Projektübersicht

Anzahl Studierende	2
Art	Studienprojekt (interdisziplinär, fachbereichsübergreifend; enge Anbindung an geförderte Projekte)
Projektverantwortung	Prof. Dr. W. Schenck, M. Eng. D. Pelkmann, M. Eng. N. Migenda, M. Sc. C. Hoppe
Projektkontext	Das Projekt findet in enger Zusammenarbeit mit dem Center for Applied Data Science (CfADS) Gütersloh und dem Fraunhofer-Institut für Software- und Systemtechnik (ISST) Dortmund statt.

Abstrakt

Rechtssysteme bilden das Fundament demokratischer Staaten. Dennoch kennt jeder das Problem, in wichtigen Lebenslagen nicht selbstständig in der Lage zu sein, aus umfangreichen Rechtsdokumenten konkrete Informationen zu entnehmen. Deshalb ist das Ziel dieses Projekts, einen Chatbot mittels maschineller Lernverfahren zu implementieren, die es auch einem Laien ermöglicht, konkrete Antworten zu komplexen juristischen Fragestellungen und die Referenz zu aktuellen, rechtskräftigen Dokumenten zu erhalten.

Kurzbeschreibung

Rechtssysteme sind eine unverzichtbare Säule im deutschen Rechtssystem. Es stellt sich die Frage, wie ein juristischer Laie Informationen aus der überwältigenden Anzahl an Gerichtsdokumenten und Gesetzestexten gewinnen soll. Um juristische Informationen effizient zu extrahieren, ist ein kollaboratives System notwendig. Traditionell bedeutete dies, einen Anwalt oder Fachspezialisten zu konsultieren,

was weder kostengünstig noch schnell ist. Mit dem Aufkommen groß angelegter Sprachmodelle (z.B. ChatGPT) und durch deren Unterstützung bei der Beantwortung von komplexen Fachfragen ergeben sich neue Möglichkeiten zur Unterstützung juristischer Laien. Mit der fortschreitenden Digitalisierung des Rechtssystems werden juristische Dokumente zunehmend digital abgespeichert. Derzeitige juristische Systeme zur Informationsbeschaffung, Datenbanken und kommerzielle Suchmaschinen speichern und verarbeiten große Mengen an juristischen Dokumenten. Leider ist es entweder sehr zeitaufwändig oder unmöglich, eine bestimmte Passage oder Antwort auf eine gestellte juristische Frage aus einem Dokumentenkörper zu extrahieren, da die bestehenden Ansätze keine Möglichkeiten für die kontextbasierte Suche in deutschen Gerichtsdokumenten bieten. Daher ist es notwendig, einen Chatbot für den deutschen Rechtsraum zu entwickeln, welcher eine präzise Antwort auf eine gestellte Frage in einem großen juristischen Korpus liefert und einen kollaborativen sowie transparenten Zugang zu juristischen Informationen ermöglicht.

Das hier beschriebene Projekt ist der direkte Nachfolger des gleichnamigen Data-Science-Forschungsmasterprojekts „Der KI-Rechtsberater“. Im vorhergegangenen Projekt konnte bereits ein funktionierender Workflow zur kontextbasierten Suche entwickelt und implementiert werden. Dies ermöglicht einerseits einen reibungslosen Projektstart und bietet andererseits eine vielversprechende Ausgangssituation zur Fertigstellung eines höchst anwendungsrelevanten Projekts. Im untenstehenden Bild ist die Funktionalität präsentiert. Der Benutzer bekommt auf eine Frage eine passgenaue Antwort mit zugehörigem Paragrafen aus dem relevanten Originaldokument.

The screenshot displays the AILA interface. On the left, there is a search input field with the question: "Wann darf die Polizei ein Fahrzeug durchsuchen?". Below the input is a "Search" button. The question is repeated in English: "Question : When may the police search a vehicle?". The answer provided is: "Answer: According to this regulation, the police may search a vehicle if there is a person in it whose identity may be established pursuant to § 26. Abs. 1, no. 4 or no. 5 PolG." On the right, under the heading "Retrieved Answers:", a score of 0.7750230729579926 is shown. Below this, a list of search results is displayed, with the top result highlighted in blue. The highlighted result includes the following details: "vg-karlsruhe-2010-05-17-9-k-151308", "Titel: vg-karlsruhe-2010-05-17-9-k-151308", "Datum: 2010-05-17", "Bez.: 9 K 1513/08", "ECLI:", "Typ.: Urteil", "Gericht: Verwaltungsgericht Karlsruhe", and "Status Gericht: None". The answer text for this result is: "findet ihre Rechtsgrundlage in § 30 Nr. 6 PolG. Nach dieser Vorschrift kann die Polizei ein Fahrzeug durchsuchen, wenn sich in ihm eine Person befindet, deren Identität nach § 26 Abs. 1 Nr. 4 oder Nr. 5 PolG festgestellt werden darf". The word "ANSWER" is visible at the end of the highlighted text.

Aufgabenstellung

Entwicklung und Validierung eines KI-gestützten Chatbots (KI-Rechtsberater) für umfassende juristische Datenbestände wie beispielsweise Urteile oder Gesetzestexte. Hierfür ist zunächst eine systematische Literaturrecherche durchzuführen und sich in die existierende Suchmaschine einzuarbeiten. Anschließend soll ein „Web Scraper“ zum automatischen Sammeln von Gerichtsdokumenten und Gesetzestexten implementiert werden. Mittels Verfahren aus dem Bereich Data Augmentation soll die Datenbasis angereichert werden. Anschließend soll mit der erweiterten Datenbasis und der existierenden Suchmaschine der Chatbot implementiert werden. In einer zunächst internen und anschließend externen Studie soll der Chatbot validiert werden.

Bezug zum Thema Data Science

Das Projekt hat in jeder Hinsicht einen sehr engen Bezug zum Thema Data Science: Von der Datensammlung über die Datenintegration bis hin zur Datenanalyse mit aktuellen Verfahren aus dem Bereich Natural Language Processing (NLP) und Large Language Models (LLMs), wie sie z.B. auch ChatGPT zugrundeliegen.

Verfügbare Ressourcen

- Der Data-Analytics-Cluster des CfADS steht über die Projektlaufzeit zur Verfügung.
- Es existieren aufbereitete und annotierte Gerichtsdatensätze.
- Auf deutsche Gerichtsurteile nachtrainierte ML-Modelle sind vorhanden.
- Aktive Unterstützung beim Erstellen von wissenschaftlichen Arbeiten, die auf (internationalen) Konferenzen präsentiert werden.
- Die Betreuung durch das Team der Ansprechpartner ist über die Projektlaufzeit gewährleistet.

Projektplan

Erstes Semester: Im ersten Semester liegt der Fokus auf der Einarbeitung in die vorhandene Infrastruktur (Data-Analytics-Cluster des CfADS) und die Schnittstelle zwischen dem Rechtswesen und der künstlichen Intelligenz. In diesem Zusammenhang ist besonders die Einarbeitung in die Themenbereiche ML, NLP und LLMs vorgesehen. Hierfür ist eine systematische Literaturrecherche durchzuführen, um den Stand der Technik aufzuzeigen. Außerdem ist es notwendig, sich mit der vorhandenen Datengrundlage vertraut zu machen. Die aktuelle Suchmaschine bezieht die Daten von einem Drittanbieter. Um unabhängig von diesem zu werden, ist außerdem ein „Web Scraper“ zu implementieren, mit dem die Daten direkt vom jeweiligen Gericht geladen werden können; mind. ein Gericht soll angebunden werden. Die Prüfungsleistung ist das Erstellen eines Forschungsexposés auf Englisch und ein dazugehöriges Kolloquium.

Zweites Semester: Das Anbinden vieler Gerichtsdatenbanken ist sehr mühselig und erfordert kontinuierliche Updates. Zudem ist ein Großteil der deutschen Gerichtsdokumente nicht digitalisiert. Um dennoch eine ausreichend große Datenbasis zu erhalten, sollen Data Augmentation-Verfahren angewendet werden. Hierbei sollten verschiedene Data Augmentation-Techniken wie Textersetzung, Paraphrasierung und Stilwechsel eingesetzt werden, um die Menge der Trainingsdaten zu erhöhen. Anschließend ist ein Benchmark durchzuführen, bei dem der originale Datensatz (1. Semester) mit dem erweiterten verglichen werden soll. Die Prüfungsleistung ist das Schreiben eines Short-Papers, welches idealerweise auf einer (internationalen) Konferenz eingereicht und präsentiert wird.

Drittes Semester: Mit der nun vorhandenen Datengrundlage und der bereits existierenden Suchmaschine ist das Ziel dieses Semesters die Erstellung eines Chatbots. Dieser soll in der Lage sein, bei einer rechtlichen Frage das dazugehörige Rechtsdokument auszugeben und den jeweiligen Paragraphen zu highlighten. Außerdem soll die Ausgabe mittels LLMs so umformuliert werden, dass es auch für einen Laien verständlich ist. Natürlich muss die inhaltliche Aussage unverändert bleiben. In einer internen Studie soll die Funktionalität validiert werden. Die Prüfungsleistung ist, aufbauend auf dem Short-Paper des zweiten Semesters, ein Full-Paper zu schreiben. Idealerweise wird das Paper auf einer Konferenz präsentiert.

Viertes Semester: In der Masterarbeit soll der Chatbot „Live gehen“. Dies bedeutet, sich aus der sicheren internen Umgebung heraus zu begeben und die Funktionalität in der realen Welt auf die Probe zu stellen. Dies birgt viele unvorhersehbare Herausforderungen, die für den Praxiseinsatz aber höchst

relevant sind. Das Thema der Masterarbeit ist grundsätzlich nur ein Vorschlag und kann nach Absprache mit den Betreuern abgewandelt werden.

Eignungskriterien

Zwingend:

- Bachelorabschluss in einer einschlägigen Fachrichtung (Informatik, Kognitionswissenschaft, Jura verbunden mit umfassenden Informatikkenntnissen, o.ä.)
- Umfassende Programmierkenntnisse in mindestens einer objektorientierten Programmiersprache
- Fließendes Englisch in Wort und Schrift

Optional:

- Programmierkenntnisse in Python
- Grundkenntnisse in NoSQL-Datenbanken
- Containerisierung

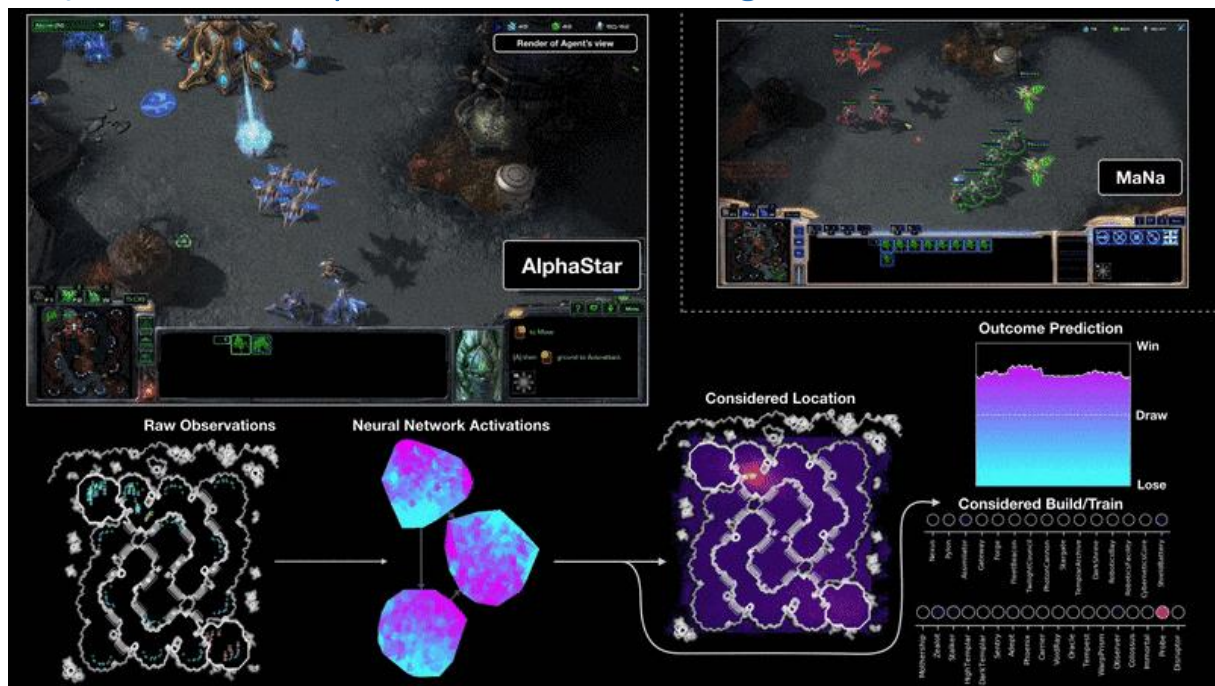
Erwerbbare Kompetenzen

Der/die Studierende ist nach Abschluss des Projekts in der Lage,

- einen Big-Data-Workflow für juristische Anwendungen aufzusetzen (Datenerfassung, -speicherung und -aufbereitung),
- die für die Verarbeitung von Texten relevanten ML- und NLP-Verfahren zielgerichtet anzuwenden und anwendungsbezogen anzupassen und ggf. weiterzuentwickeln,
- die theoretischen Anforderungen von ML-Verfahren mit den Erfordernissen der Praxis in Einklang zu bringen,
- textbasierte Dialogsysteme anzuwenden,
- die eigenen Forschungsergebnisse vor einem Fachpublikum zu präsentieren
- und wissenschaftliche Texte zu verfassen.

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: 3 Interpretierbare und übertragbare KI für StarCraft2



Projektübersicht

Startsemester (von-bis)	WiSe2019/20-SoSe2021
Anzahl Studierende	1-3
Art (gefördertes Projekt, Projekt mit externen Partnern, Studienprojekt)	Studienprojekt
Projektverantwortung	Prof. Dr. Christian Schwede
Projektkontext	Das Studienprojekt ist am Digital Labor des Campus Gütersloh angesiedelt. Das Team besteht aus einem Wissenschaftlichem Mitarbeiter und zwei weiteren Studierenden.

Abstrakt

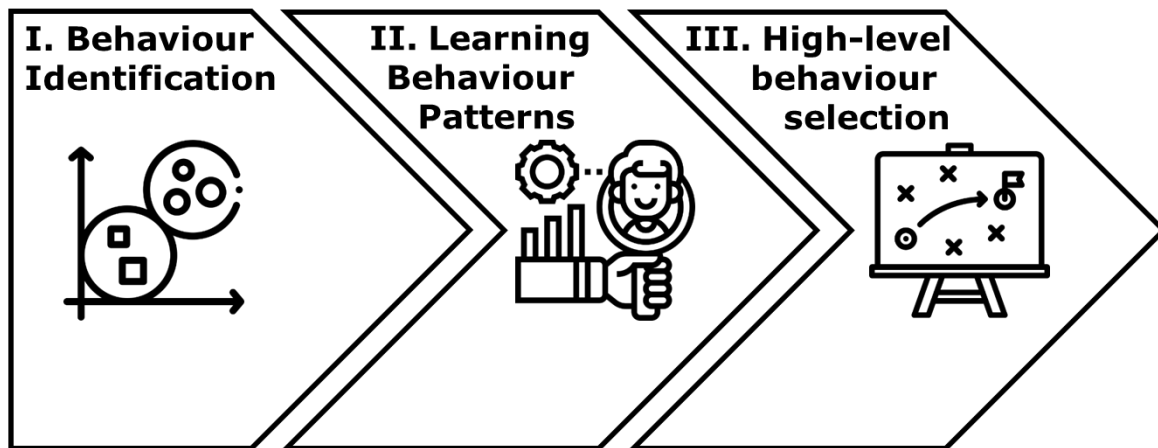
In diesem Projekt wird u.a. mittels cutting-edge Verfahren des Deep Reinforcement Learnings an einer Künstlichen Intelligenz für StarCraft2 gearbeitet die Wissen bewahrbar und übertragbar macht und zudem für Menschen nachvollziehbar ist (Explainability).

Kurzbeschreibung

Computerspiele stellen schon seit den Anfängen der Künstlichen Intelligenz eine spannende Herausforderung für die Entwicklung von Intelligenten Agenten dar. Große Durchbrüche der KI-Forschung zeigten sich häufig durch die Entwicklung von Programmen, die in der Lage waren menschliche Gegenspieler zu besiegen. Deep Blue schlug 1996 den amtierenden Weltmeister Kasparow im Schach, IBMs Watson gewinnt 2011 in „Jeopardy!“ gegen zwei gegen zwei Profis und Googles AlphaGo schlägt 2016 Lee Sedol einen der weltbesten Go-Spieler. Seit 2019 spielt Deepminds AlphaStar in den diversen Ligen von Starcraft2 und besiegt mittlerweile 99,8% der Spieler. Auch wenn diese KI hinsichtlich der Performanz unübertroffen ist, so erfüllt sie zwei andere wichtige

Anforderungen an KI kaum: Übertragbarkeit/Lerneffizienz und Erklärbarkeit. Auf dem Weg zur Starke KI ist die Lerneffizienz bzw. die Übertragbarkeit und das nicht Vergessen von Wissen entscheidend, während für uns Menschen der Einblick in das Was und Warum des Handelns von KI von entscheidender Bedeutung ist.

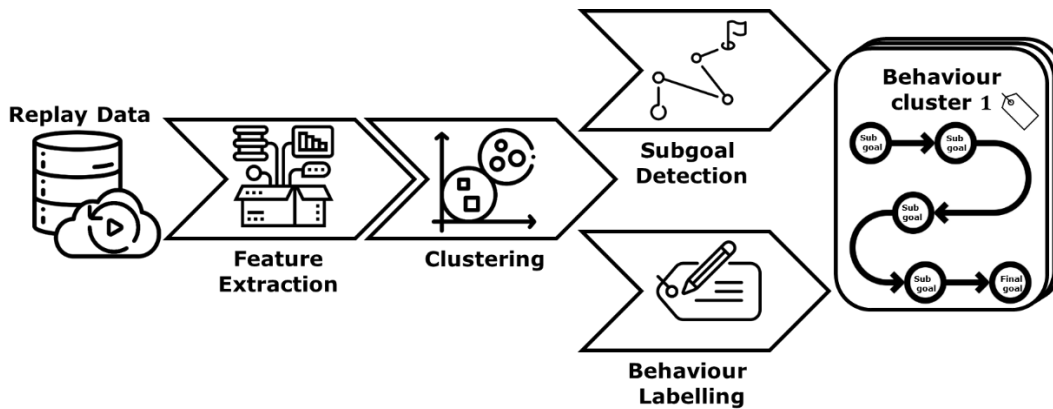
In diesem Projekt wird deshalb an einer Künstlichen Intelligenz für StarCraft2 gearbeitet die diese beiden Anforderungen erfüllt. Die Grundidee ist es eine KI zu entwickeln die lernt je nach Situation einen passenden Handlungsbaustein (Strategie) und nicht nur eine einzelne Aktion auszuwählen. So kann Wissen in den Bausteinen bewahrt werden und komplexe Handlungen durch die Kapselung in Strategien erklärbarer werden. Hierbei sollen die einzelnen Strategiebausteine auf Basis von Beispielpartien erst erkannt und dann gelernt werden. Durch die zusätzliche Konstruierung eines Subziel-Graphen, der die Zwischenziele einzelner Strategiebausteine in Reihenfolge bringt, werden diese dann für Menschen nochmal erklärbarer. Der gesamte Ansatz ist in drei Teilschritte unterteilt und in folgender Grafik dargestellt. Als erstes werden aus Beispieldaten Strategiecluster mittels Zeitreihenclusterings gelernt, dann wird auf Basis von diesen Clustern mittels Supervised oder Reinforcement Learning Strategiebausteine gelernt und schließlich findet mittels Hierarchical Reinforcement Learning die High-Level Strategie gelernt.



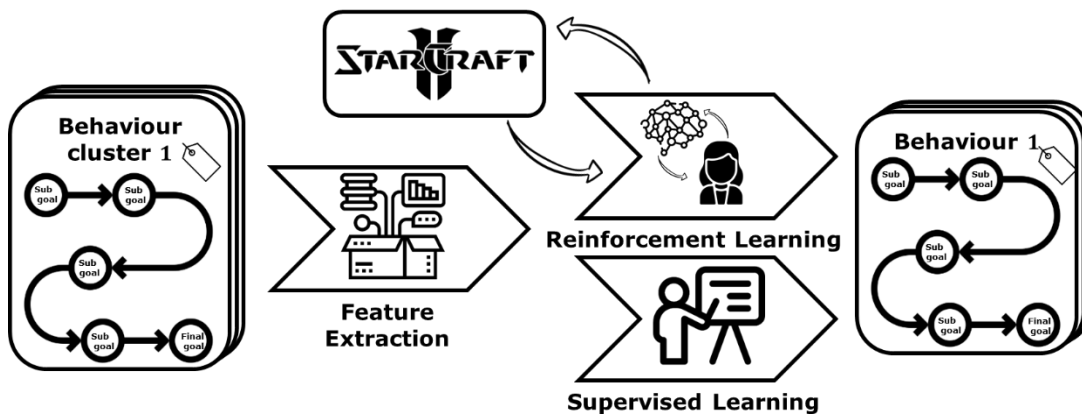
Aufgabenstellung

Die Studierenden können je nach Interesse in einem der oben beschriebenen Teilbereiche arbeiten. Der Schwerpunkt der Arbeit kann dabei einen der drei genannten Verfahrensgruppen betreffen.

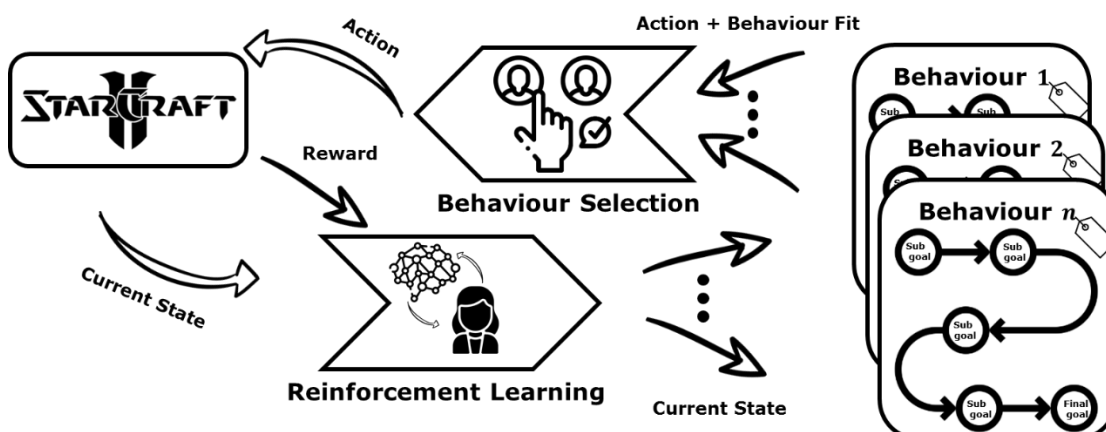
Der erste Bereich der Strategieerkennung wird hauptsächlich mit Clusterverfahren arbeiten. Aufgaben der Merkmalsextraktion im variablen und hochdimensionalen Eingaberaum von StarCraft2, der Auswahl von passenden Distanzmetriken, die Auswahl und Implementierung von Clusterverfahren für variable Zeitreihen, der Entwicklung von Verfahren für die Erkennung von Teilzielen und das automatische Labelling können bearbeitet werden.



Der zweite Bereich kann entweder Verfahren des Supervised oder des Deep Reinforcement Learnings umfassen. Überwachte Verfahren werden genutzt, um das Verhalten der geclusterten Strategien nachzuahmen, während Verstärkendes Lernen ein eigenes Verhalten anhand der Subgoals lernen soll.



Der Dritte Bereich umfasst Verfahren des Hierarchical Reinforcement Learnings. Hierbei muss erlernt werden den passenden Strategiebaustein für die jeweilige Situation auszuwählen. Außerdem muss das Auswahlverhalten derart beeinflusst werden, dass die Strategie nicht nach jeder Aktion neu gewählt wird.



Bezug zum Thema Data Science

Verfahren des Clustering, Supervised und Reinforcement Learning sind Kernthemen der Data Science und werden auch in den Veranstaltungen des Forschungsmasters behandelt.

Verfügbare Ressourcen

- Millionen von Replay-Daten und eine Python-API für StarCraft2 stehen zur Verfügung
- Hardware für das Maschine Learning ist über das Data Science Lab oder das CfADS der FH Bielefeld verfügbar

Projektplan

Erstes Semester: Formulierung des Forschungsexposees, Einarbeitung in das Spiel, den Source Code und die API.

Zweites Semester: Entwicklung eines einfachen Algorithmus in dem jeweiligen Feld für ein einfaches Spielszenario, Recherche zu relevanten Arbeiten im Themenfeld

Drittes Semester: Entwicklung von ersten komplexeren Verfahren für einfachere und komplexere Spielszenarien, Auswertung erster Ergebnisse

Viertes Semester: Erstellung eines komplexen Szenarios, Implementierung und Vergleich von weiteren Verfahren, Fine-Tuning und Optimierung der Verfahren, Verbesserung der Ergebnisse, Finale Evaluierung

Eignungskriterien

Zwingend:

- Programmierkenntnisse
- Begeisterung für Spiele/ StarCraft2 und KI

Optional:

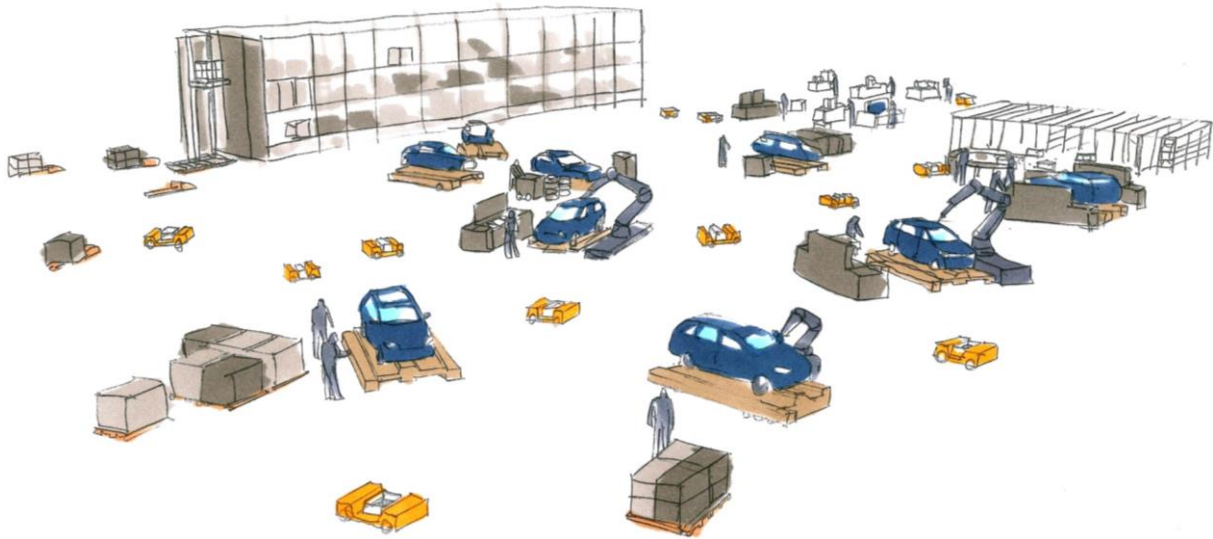
- Erfahrung in ML

Erwerbbare Kompetenzen

- Einsatz und (Weiter-)Entwicklung von State-of-the-Art Verfahren des Maschinellen Lernens in hochkomplexen Umgebungen

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: 4 KI für die hochflexible variantenreiche Massenproduktion



Projektübersicht

Anzahl Studierende	1
Art (gefördertes Projekt, Projekt mit externen Partnern, Projekt, Studienprojekt)	Projekt mit externen Partnern
Projektverantwortung	Prof. Dr. Christian Schwede
Projektkontext	Projekt in Zusammenarbeit mit dem Fraunhofer-Institut für Software- und Systemtechnik (ISST) in Dortmund und im Kontext des Center for Applied Data Science (CfADS) sowie des Institute for Data Science Solutions (IDaS) ; Anstellung als Hiwi bei Fraunhofer ist möglich;

Abstrakt

Die stetige Individualisierung der Produkte stellt nicht nur die deutsche Automobilproduktion vor Herausforderungen, die ein vollständiges Umdenken in der Produktionsorganisation erzwingen. Weg von getakteten Fließfertigung hin zur dezentralen Matrixproduktionssystemen. Schwärme von autonomen Roboter koordinieren hier die Wertschöpfungsprozesse und passen sich flexible an die aktuelle Situation an. Während die Grundidee des Konzeptes klar umrissen ist, ergeben sich im Detail noch sehr viele offene Forschungsfragen, die insbesondere das Verhalten der einzelnen Roboter betrifft. Hier setzt das Projekt an und erforscht unterschiedliche Strategien der KI in einem virtuellen Testbed.

Kurzbeschreibung

In der sogenannten Matrixproduktion wird die Produktion dezentral von Software-Agenten in einem Multiagentsystem (MAS) koordiniert. Die Ressourcen-Agenten verhandeln hierbei bspw. mit dem Auftragsagenten über die Erbringung von Wertschöpfungsprozesse, die für die Realisierung des Auftrages nötig sind. Die Freiheit der Reihenfolge der Fertigungsschritte wird hierbei nur durch

technische Eigenschaften des Produktes beschränkt. Auf diese Weise kann sich das gesamte System flexibel an nichtgeplante Vorfälle, wie bspw. einen Lieferengpass oder einen Maschinenausfall anpassen. Während dem Konzept in der Automobilindustrie zuerst mit Ablehnung und Skepsis begegnet wurde, gibt es seit ein paar Jahren intensive Bestrebungen, eine Operationalisierung zu erforschen und voranzutreiben. Grundsätzlich sind zu einem effizienten Einsatz noch zahlreiche Fragen offen die es zu erforschen gilt.

In der Forschungsgruppe der HSBI wird seit einigen Jahren in Kooperation mit dem Fraunhofer ISST in Dortmund das Softwaretool RIOTANA[®] (Realtime Internet-Of-Things ANalytics) entwickelt, das es unter anderem erlaubt mit agentenbasierter Simulation die hochflexiblen Matrixproduktionen zu modellieren und verschiedene Strategien und Designvarianten dynamisch zu bewerten. Das Tool ist außerdem als Digitaler Zwilling einsetzbar und kann über Sensordatenströme direkt mit der Produktion verbunden und zur Steuerung der operativen Prozesse genutzt werden.

Aufgabenstellung

Der/ die Studierende hat die Möglichkeit an einem der folgenden Themen zu arbeiten, um die Möglichkeiten und Limitationen der Matrixproduktion zu erforschen und neue Lösungskonzepte auf der Realisierung zum Stand der Technik beizusteuern.

Agentensteuerung für die dezentrale Produktion

Eine der entschiedensten offenen Fragen ist die der optimalen Steuerung der Agenten. Insbesondere die Synchronisation der Materialflüsse mit dem Hauptprodukt ist eine entscheidende, um Wartezeiten zu verhindern und gleichzeitig die Flexibilität zu erhalten. Die Entwicklung und Implementierung effizienter Steuerungsstrategien für das MAS ist hier die Kernaufgabe, die unter anderem die Frage beinhaltet ob und in wie weit in die Zukunft geplant und reserviert werden sollte. Methoden der Optimierung, Such und Planalgorithmen aus der KI oder Reinforcement Learning können hier eingesetzt werden.

Layoutdesign und -optimierung

Auch wenn der Begriff Matrixproduktion eine Gitternetzformige Anordnung der Produktionsstationen andeutet, so ist die Frage nach der optimalen Anordnung der verschiedenen Elemente der Matrixproduktion (Stationen, Lagerorte für Teile, Ladestationen der Transportfahrzeuge, Parkplätze zur Zwischenpufferung von Karosse usw.) noch nicht final entschieden. Optimierungsalgorithmen, die in Kombination mit der Simulation als Evaluierungsmethode eingesetzt werden müssen hier entwickelt werden, um für ein konkretes Produktionssetup das optimale Layout zu bestimmen. Auf die Frage wie stark das optimale Layout vom konkreten Auftragsmix abhängt ist relevant und könnte die Notwendigkeit einer regelmäßigen Überplanung des Layouts implizieren.

Routing und Stauanalyse

Die Entstehung von Staus vor den Produktionsstationen oder rauf den verschiedenen Wegen ist ein zentraler Aspekt der Effizienz des Gesamtsystems. Hierzu muss die Simulationsumgebung um ein Element der Routenplanung erweitert werden, das auch temporäre Routenkonflikte erkennt. KI-Suchalgorithmen aus dem Bereich der Wegfindung müssen implementiert und verschiedenen Strategien getestet werden. Schließlich müssen existierende Steuerungsstrategien und Layouts vor dem Hintergrund von Staus im System untersucht und bewertet werden.

Beschleunigung der Verhandlungen mittels Reinforcement Learning

Eine Schwachstelle von Multiagentensystemen ist das Kommunikationsaufkommen, dass durch die dezentralen Verhandlungen zwischen den Agenten entsteht. Hier gilt es Verhandlungsstrategien zu untersuchen, implementieren und zu vergleichen und insbesondere unnötige Kommunikation zu

vermeiden. Reinforcement Learning kann eingesetzt werden, um zu erkennen, wann Anfragen von Agenten nicht zielführend sind und so im Vorhinein vermieden werden können.

Agentenbasierte Wertschöpfungsnetzwerke

Die Einbindung der logistischen Prozesse aus dem Wertschöpfungsnetzwerk rund um die Produktion ist ein weiteres offenes Thema. Neben der Entwicklung geeigneter Ver- und Entsorgungsstrategien, die mit den Bedürfnissen der flexiblen Matrixproduktion übereinstimmen müssen soll die Visionäre Frage beantwortet werden, ob die dezentrale, autonome Wertschöpfung auch über die Fabrikgrenzen hinaus erweitert werden kann. Hierzu müssen die Agenten auf Basis von Geld als Zahlungsmittel Verträge mit Lieferanten und Dienstleistern verhandeln. Die entsprechenden Verhandlungsalgorithmen und Szenarien müssen umgesetzt und evaluiert werden.

Bezug zum Thema Data Science

Softwareagenten und Multiagentensysteme, Suchalgorithmen und Reinforcement Learning sind Kerngebiete der KI und werden in verschiedenen Veranstaltungen des Forschungsmasters thematisiert.

Verfügbare Ressourcen

- Das Modellierungsframework Riotana® wird von des HSBI und dem Fraunhofer ISST bereitgestellt
- Industrienähe Szenarien einer individuellen Fahrradproduktion existieren zur Validierung der Entwicklungen
- Szenarien von Industriekunden des ISST und der IoT-Factory in Gütersloh existieren ebenfalls zu Evaluierung einzelner Verfahren
- Hardware für das komplexere Maschine Learning ist über das Data Science Lab oder das CfADS der verfügbar

Projektplan

Erstes Semester: Formulierung des Forschungsexposees, Einarbeitung in die Riotana®-Umgebung, Auswahl des Themas.

Zweites Semester: Modellierung eines ersten Szenarios, Recherche zu relevanten Arbeiten in Themenfeld und Erstellung eines Reports zu Stand der Technik.

Drittes Semester: Implementierung und Evaluierung erster Verfahren, Auswertung erster Ergebnisse auf Basis des Szenarios, Verfassen eines Papers für eine (inter-)nationale Fachkonferenz.

Viertes Semester: Entwicklung weiterer Verfahren und Optimierung des Gesamtansatzes, Auswertung der Güte und Evaluierung der Ergebnisse in der Masterarbeit

Eignungskriterien

Zwingend:

- Programmierkenntnisse (optimal in Python)

Optional:

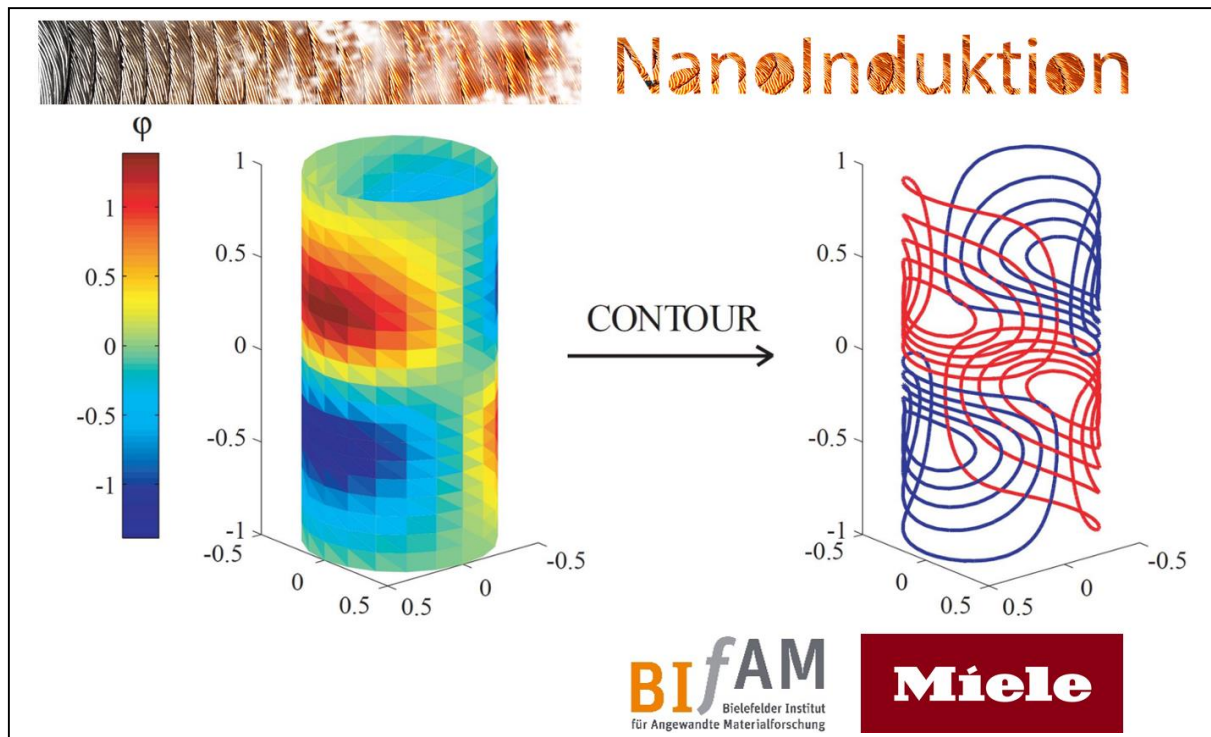
- Erfahrung mit Agentensystemen
- Erfahrung mit Simulation und Optimierung
- Hintergrundwissen zu Produktionslogistik und Produktionssystemen
- Erfahrung mit der Implementierung von Reinforcement Learning / ML-Verfahren

Erwerbbarer Kompetenzen

- Einsatz von Verfahren des Reinforcement Learnings
- Entwicklung und Bewertung von MAS
- Entwicklung von intelligenten Agenten
- Arbeiten mit Agentensimulationen
- Planung von Produktions- und Logistiksystemen insbesondere von
Matrixproduktionssystemen und Industrie 4.0

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: 5 Machine Learning-Modelle zur Lösung inverser Probleme



Projektübersicht

Anzahl Studierende	1 bis 2
Art (gefördertes Projekt, Projekt mit externen Partnern, Studienprojekt)	ehem. gefördert vom BMBF und der Fa. Miele, weiterhin in Kooperation mit der Fa. Miele
Projektverantwortung	Prof. Dr. Christian Schröder, Simon Bekemeier, M.Sc.
Projektkontext	Regelmäßiger Austausch mit Mitarbeitern der Arbeitsgruppe und anderen Studierenden des Forschungsmasters in der Arbeitsgruppe; Einbindung in laufendes Promotionsprojekt, aktuelle Forschung und Veröffentlichungen; High Performance Computing Ressourcen; AG Computational Materials Sciences and Engineering, Bielefeld Institute for Applied Material Research, mieletec, ggf. CITEC.

Abstrakt

In diesem Projekt sollen die Geometrien von Magnetspulen mittels Modellen des Maschinellen Lernens optimiert werden.

Kurzbeschreibung

Viele Ingenieurwissenschaftliche Fragestellungen sind bereits seit Jahren und Jahrzehnten gelöst, sind jedoch nach wie vor **hochaktuell** und bieten viel Raum zur Verbesserung der "herkömmlichen"

Lösungen durch **interessante und innovative**, neue Konzepte und Lösungsstrategien, die erst in den vergangenen Jahren nutzbar geworden sind!

Eine dieser spannenden Fragestellungen ist die berührungslose Übertragung von Energie mittels Induktion, die immer noch hoch aktuell ist und ein spannendes Forschungsthema darstellt. Abhängig von der zu übertragenden Leistung, findet man heutzutage eine Vielzahl von Anwendungen wie z.B. das Aufladen von Hochleistungsbatterien im Bereich der Elektromobilität, dem berührungslosen Erhitzen von Stoffen zum Schmelzen, Garen oder zur lokalen Tumorbekämpfung und der kombinierten Übertragung von Energie und Daten im Bereich RFID. Das zugrundeliegende physikalische Prinzip – die Entstehung elektrischer Felder aufgrund der zeitlichen Änderungen eines magnetischen Flusses – wurde vor fast 200 Jahren von Faraday, Henry und Ørstedt entdeckt und ist seither unter der Bezeichnung Induktionsgesetz bekannt. Die Möglichkeit mit Hilfe von Wechselfeldern auch über größere Distanzen Energie zu übertragen, ermöglicht aber nicht nur eine Vielzahl von interessanten Anwendungen, sie führt gleichzeitig zu einer Reihe fundamentaler, bis heute aktueller, Probleme in der technischen Umsetzung – und zwar unabhängig von der Art der Anwendung:

1. Die auftretenden – nicht nutzbaren – Streufelder wirken sich grundsätzlich nachteilig auf die Energieeffizienz, die Sicherheit und die elektromagnetische Verträglichkeit aus.
2. Inhomogenitäten in den erzeugten Feldern führen zu einem schlechten Wirkungsgrad bei der Energieübertragung und ungewünschten Effekten auf der Empfängerseite.
3. Aufgrund des Wechselstrombetriebs treten mechanische Kräfte auf, die sich z.B. beim Betrieb von Induktionskochfeldern als störende Geräusche bemerkbar machen oder in anderen Anwendungen sogar verschleißend und zerstörerisch wirken können.

Diesen Problemen kann mit neuen Konzepten begegnet werden, um damit die zukunftsweisende Technologie der induktiven Energieübertragung entscheidend weiterzuentwickeln. Im Projekt NanoInduktion wird dies am Beispiel einer aktuellen und weit verbreiteten Anwendung, dem Induktionskochfeld, durchgeführt.

Hierzu werden unterschiedliche Methoden zur Berechnung, Simulation und Optimierung von Magnetfeldern und den dazugehörigen Spulen, die sowohl das Sende-, als auch das Empfangssystem darstellen, untersucht. Ausgehend von einem definierten Magnetfeldverlauf auf der Empfängerseite soll mit Hilfe verschiedener Methoden diejenige Spulengeometrie ermittelt werden, die dieses Feld generiert. Analytische Verfahren zur Lösung dieses sogenannten inversen Problems sind bereits aus der Entwicklung von Magnetresonanztomografie-Scannern (MRT) bekannt, müssen jedoch für die induktive Energieübertragung im Allgemeinen oder speziell für die Entwicklung effizienter Induktionskochfelder angepasst werden. Alternativ hierzu wird die Lösung dieses inversen Problems durch Optimierungsmethoden und Machine Learning Verfahren diskutiert.

Aus dem Einsatzgebiet der Magnetresonanztomografie (MRT) sind verschiedene Verfahren bekannt, um solche inversen Probleme durch die geschickte Anwendung von Neuronalen Netzen zu lösen. Auf Grundlage dieser Ideen und Konzepte soll in diesem Forschungsprojekt ein Lösungsverfahren für die Suche nach optimalen Spulengeometrien entwickelt werden.

Aufgabenstellung

Zunächst erfordert das Forschungsprojekt eine Einarbeitung in die Grundlagen des maschinellen Lernens mit Neuronalen Netzen, auf denen die bekannten Lösungskonzepte basieren. Darauf aufbauend muss ein Überblick über die möglichen Verfahren zur Lösung inverser Probleme mittels Neuronaler Netze gewonnen werden, um anschließend einen oder ggf. mehrere geeignete Kandidaten für die weiteren Untersuchungen auswählen zu können.

Der gewählte Ansatz soll im weiteren Verlauf des Forschungsprojekts auf seine Übertragbarkeit auf den gegebenen Anwendungsfall untersucht werden. Das abschließende Ziel ist es, das Lösungsverfahren für die Anwendung zu implementieren.

Durch die in der Literatur vorhandene Auswahl von möglichen Lösungsansätzen, ist das Forschungsprojekt für eine/n weitere/n Bewerber/in geeignet. Die grundlegende Vorbereitung zum Forschungsprojekt (Aneignung der Grundlagen Neuronaler Netze) sowie die Literaturrecherche zu möglichen Lösungsverfahren kann durch zwei Personen gemeinsam bzw. in umfangreicherer Form durchgeführt werden. Somit steht eine breitere Auswahl an Kandidaten für geeignete Verfahren für die weiteren Untersuchungen zur Verfügung. Im weiteren Verlauf des Forschungsprojekts können verschiedene Ansätze separat untersucht werden, wobei die zusätzliche Möglichkeit des Austauschs und ggf. schon frühzeitige Vergleiche der untersuchten Ansätze das Projekt sowohl für die Bewerber/innen, als auch aus wissenschaftlicher Perspektive aufwerten.

Bezug zum Thema Data Science

Zur Bearbeitung des Forschungsprojekts müssen eine vorhandene Datenbasis auf die Bedürfnisse der angewandten Methoden angepasst (Vorverarbeitung), sowie für das Ziel sinnvoll anwendbare Methoden aus dem Literatur-Korpus bekannter / veröffentlichter Verfahren ermittelt und für den Anwendungsfall angepasst werden. Diese Anforderungen gehören zu den Kernthemen der Data Science und werden in den Veranstaltungen des Forschungsmasters behandelt.

Verfügbare Ressourcen

Expertise und Informationen aus dem physikalischen Anwendungskontext des Projekts, sowie zu den behandelten Machine Learning Verfahren und der Verbindung beider Aspekte werden von den Mitarbeitern des Forschungsprojekts NanoInduktion und der Arbeitsgruppe Computational Materials Science and Engineering (AG CMSE) von Prof. Schröder zur Verfügung gestellt. Darüber hinaus besteht die Möglichkeit des Kontakts zu geeigneten Arbeitsgruppen des CITEC durch die AG CMSE.

HPC Hardware und -Systeme für komplexe Berechnungen werden durch die AG CMSE zur Verfügung gestellt. In diesem Rahmen kann überdies Unterstützung bei der Nutzung des High Performance Computing Clusters der Arbeitsgruppe gegeben werden. Weitere HPC Ressourcen stehen über das PC² der Universität Paderborn zur Verfügung. Über die Projektlaufzeit hinaus wird das Thema innerhalb des Forschungslabors mieteTec an der FH Bielefeld weitergeführt. Die Mitarbeiter des mieteTec stehen dann als Ansprechpartner zur Verfügung.

Projektplan

Erstes Semester: Einarbeitung in die notwendigen Grundlagen Neuronaler Netze, sowie das notwendige Handwerkszeug für die Arbeit mit diesen, etwa TensorFlow / Keras und Python. Erste Sichtung der Literatur zur Lösung inverser Probleme durch Neuronale Netze und Planung der weiteren Untersuchungen. Erstellung eines Forschungsexposés (ist Prüfungsleistung).

Zweites Semester: Vertiefung der Literaturrecherche, sowie Auswahl der im Folgenden zu untersuchenden Lösungsansätze. Entwicklung erster prototypischer Versuche mit den gewählten Ansätzen. Erstellung eines Papers, das einen Überblick über das jeweilige Forschungsgebiet gibt (ist Prüfungsleistung).

Drittes Semester: Verfeinerung des entwickelten Prototyps, zur Übertragung des Ansatzes auf den gegebenen Anwendungsfall. Vergleich mit Ergebnissen aus anderen Lösungskonzepten, die in der AG CMSE bereits entwickelt werden bzw. bei zwei Teilnehmer/innen auch mit den Ergebnissen der/des jeweils anderen. Erstellung eines Papers, mit ersten quantitativen Ergebnissen (ist Prüfungsleistung).

Viertes Semester: Untersuchungen zur Anwendbarkeit über den Prototypen hinaus bzw. zur Anwendbarkeit bei allgemeineren Rahmenbedingungen, bei Erweiterung des möglichen Lösungsraums (der verfügbaren Spulengeometrien, sowie komplexerer magnetischer Systeme). Dokumentation des Gesamtergebnisses unter Einbeziehung der Teilergebnisse aus den ersten drei Semestern. (Masterarbeit und Kolloquium)

Eignungskriterien

Zwingend:

- Kenntnisse der Programmierung oder die Bereitschaft und Fähigkeit sich diese eigenständig anzueignen, vorzugsweise in Python.
- Englischkenntnisse, ausreichend für die Lektüre (fast) ausschließlich englischsprachiger, wissenschaftlicher Veröffentlichungen.
- Teamfähigkeit und "Forschergeist"

Optional:

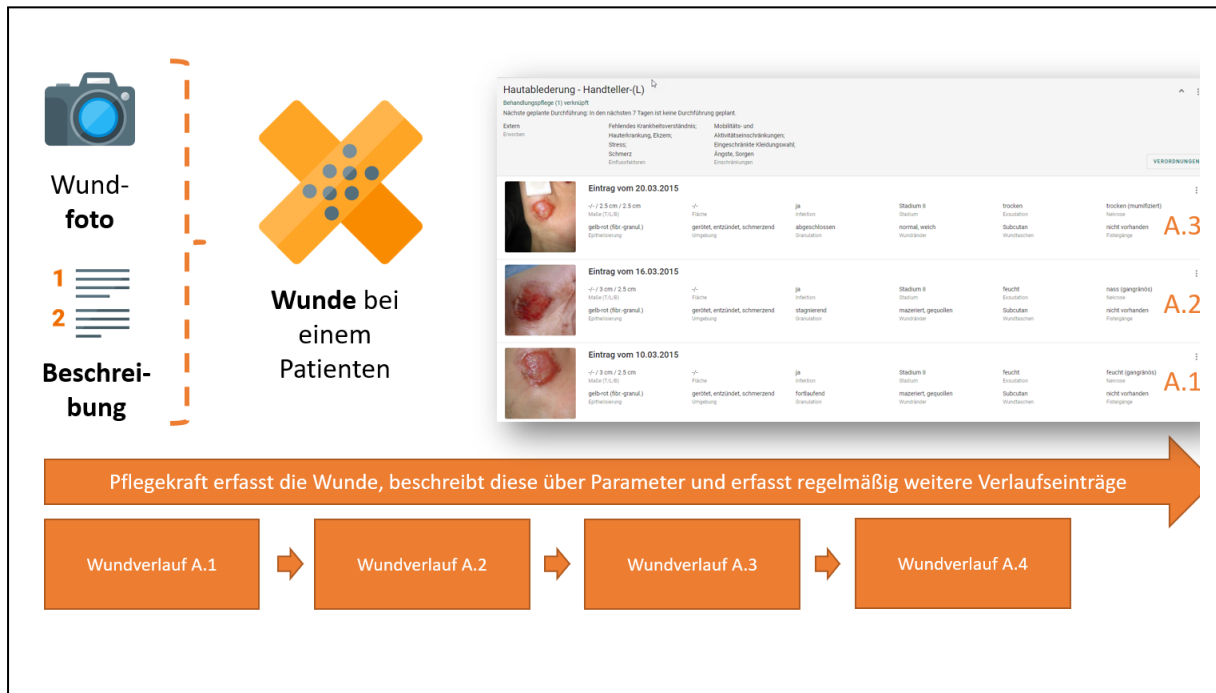
- Erfolgreich absolvierte Bachelor-Veranstaltungen oder andere Prädisposition im Kontext von Neuronalen Netzen bzw. Machine Learning.
- Kenntnisse in Python und TensorFlow / Keras oder anderen verbreiteten, nicht-proprietären Sprachen und Frameworks zur Arbeit mit Neuronalen Netzen.

Erwerbbarer Kompetenzen

- Einsatz von Verfahren des Maschinellen Lernens
- Verwendung von TensorFlow / Keras
- Methoden der Optimierung
- Verfassen wissenschaftlicher Artikel
- Präsentation wissenschaftlicher Ergebnisse
- Recherche wissenschaftlicher Veröffentlichungen
- Aufbau eigener Arbeiten auf dem bestehendem wissenschaftlichem Kenntnisstand bzw. wissenschaftlichen Veröffentlichungen

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: **6 Assistenz bei der Wundversorgung mittels Künstlicher Intelligenz zur Wundanalyse, -einschätzung und -versorgung**



Projektübersicht

Anzahl Studierende	1
Art	Projekt mit externen Partnern
Projektverantwortung	Prof. Dr.-Ing. Wolfram Schenck , Björn Gorniak (Connex Communication GmbH), medizinische Unterstützung durch Prof. Dr. med. Rena Amelung
Projektkontext	Projekt in Zusammenarbeit mit der Firma Connex Communication GmbH aus Paderborn sowie dem Center for Applied Data Science Gütersloh

Abstrakt

Das Ziel des Projekts ist es, den Heilungsverlauf von Wunden mittels optimaler Behandlung sowie fachgerechter Therapie zu verbessern, und zwar im Kontext der Wundversorgung in der Altenhilfe und -pflege. Die Aufgabe von Verfahren der künstlichen Intelligenz (KI) ist dabei, Bilder von Wunden bezüglich verschiedener Merkmale automatisch zu klassifizieren. Darauf aufbauend sollen Wundverläufe vorhergesagt und Behandlungsempfehlungen generiert werden, um Pflegekräfte in ihrer täglichen Arbeit zu unterstützen.

Kurzbeschreibung

Das Ziel des Projekts *Assistenz bei der Wundversorgung mittels Künstlicher Intelligenz zur Wundanalyse, -einschätzung und -versorgung* ist es, den Heilungsverlauf von Wunden mittels optimaler Behandlung sowie fachgerechter Therapie zu verbessern. Grundlage dafür ist die Wundversorgung in der Altenhilfe und -pflege.

In der Altenhilfe entstehen aufgrund veränderter Hautzustände und Druckbelastung sogenannte Druckgeschwüre (Dekubitus). Ein Dekubitus kann entstehen, wenn das (Haut-)Gewebe langfristig einem starken Druck und einer nicht mehr gewöhnlichen Belastung ausgesetzt ist. Durch das Zusammenpressen von Blutgefäßen werden die Zellen nicht mehr versorgt und sterben ab [<https://de.wikipedia.org/wiki/Dekubitus>]. Es entsteht eine Wunde, die im Alter in der Regel nicht mehr gut verheilt. Diese Wunden müssen nun professionell behandelt und versorgt werden.

Dabei besteht die Herausforderung, dass Wunden und deren Behandlung einen komplexen Arbeitsprozess darstellen. Besonders ausgebildete Mitarbeiter*innen übernehmen die Bewertung, Behandlungsbeschreibung und Einschätzung der Wunden in Pflegeeinrichtungen. Diese Arbeit ist vor allem durch eine Vielzahl an Erfahrungen und ‚community of practices‘ gekennzeichnet.

Oftmals stellen Abweichungen und besondere Situationen in der Wundveränderung für die behandelnden Pflegekräfte eine besondere Herausforderung dar. Ein assistierendes System soll den Mitarbeitenden in der Pflege die Unterstützung bieten, die Einschätzung des Ist-Zustandes der Wunde korrekt vorzunehmen. Anschließend kann ein intelligentes System aus dieser Einschätzung eine Behandlungstherapie ableiten.

Grundlage für dieses Vorgehen ist die Analyse des Ist-Zustands der Wunde und eine Einschätzung des Wundverlaufs. Dabei steht das Foto der Wunde im Mittelpunkt der Analyse. Es wird – in der Regel – mittels eines Smartphones erfasst und in eine standardisierte Wunddokumentation nach Expertenstandard übernommen. Diese Datenbasis soll verwendet werden, um die gewonnenen Bilddateien durch Bildanalyse einzuschätzen.

Die dabei verwendeten Techniken werden bereits in anderen Wirtschaftszweigen angewendet, z.B. um auf Satellitenbildern eine Veränderung der Infrastruktur von Landflächen [<https://computerwelt.at/news/ki-und-satellitenbilder-gegen-armut-in-afrika/>] vorzunehmen und Strukturveränderungen zu erkennen.

Aufgabenstellung

Der / Die Studierende soll im Rahmen des Projektes maschinelle Lernverfahren (ML) zur Bildanalyse auswählen, ggfs. anpassen und anwenden, die eine Wiedererkennung von Mustern und Strukturen in einem Bild einer Wunde ermöglichen. Verbunden mit den beschreibenden, strukturierten Daten von Wunden ist so ein System möglich, dass anhand eines Wundfotos eine erste Einschätzung zur Unterstützung der versorgenden Pflegekraft generiert.

Im weiteren Projektverlauf soll anhand des Clusterings von Wundverläufen eine umfassendere Bewertung von Behandlungssituationen mit ML-Techniken erreicht werden. Das System soll vor allem zur Entlastung und Unterstützung der Menschen in der täglichen Arbeit angewendet werden.

Wenn es zeitlich möglich ist, ist es wünschenswert, dass dieses Verfahren in Form einer Service-Struktur implementiert wird und als Anwendung zur Verfügung steht.

Problemstellung und Ziele der Teilaufgaben

Das Gesamtziel ist die Verbesserung der Wundheilung für Bewohner/Patienten in der Altenpflege. Da diese Branche von der Herausforderung des Fachkräftemangels besonders betroffen ist und die Versorgungssituation immer komplexer wird, soll entlastet und ein Teil der Aufgaben sinnvoll unterstützt werden.

Im Fokus steht die Betrachtung von komplexen Wunden, die durch eine Vielzahl von Ausprägungen nicht alltäglich für die einzelnen Pflegekräfte sind. Durch spezialisiertes Personal wie pflegerische Wundmanager/-experten ist hier schon eine personelle Problemlösung angegangen worden, doch ist diese durch die Vielzahl an Wundverläufen und -ausprägungen nicht zu beherrschen. Ein zentrales Element ist der Kostenfaktor, der sich aus den drei Bausteinen Hilfsmittel für die Wundtherapie, medizinische Beratung durch Fachärzte und Personalaufwendungen in der täglichen Versorgung zusammensetzt.

Ethisch zielt die Fragestellung auf den Begriff der Pflegebedürftigkeit ab, denn das Entstehen von Wunden sorgt immer für eine Verschlechterung des Allgemeinzustandes. Dieser basiert v.a. auf dem erhöhten Betreuungs- und Versorgungsbedarf einer Person und den emotionalen Implikationen einer Wunde. Zudem entstehen indirekte Kosten wie die Schmerzbehandlung und -therapie, die als Folgen nicht immer berücksichtigt werden.

Von diesem Gesamtziel lassen sich einzelne Aufgabenpakete ableiten, die als Meilensteine bei der Erreichung dienen sollten:

- Analyse der vorhandenen, strukturierten Wund-Daten
- Auswahl / Diskussion eines Analyse-Verfahrens / ML-Verfahrens zur abstrakten Wiedererkennung von Eigenschaften des Wundbildes
- Anpassung / Justierung eines solchen Verfahrens und Optimierung auf die besonderen Ansprüche des Sozial- und Gesundheitswesens
- Umsetzung und Integration in eine bestehende Micro-Service-Architektur

Um dieses zu erreichen, kann auf einen vorhandenen Datenbestand von Pflegedokumentationen zurückgegriffen werden. Diese Wunddaten wurden in den vergangenen Jahren erfasst und können historisch von Kunden der Connex Communication GmbH, die sich beteiligen, analysiert werden. Diese Daten können mittels einer API / Schnittstelle anonymisiert betrachtet werden. Darüber hinaus stehen voraussichtlich Bilddaten aus spezialisierten Behandlungseinrichtungen zur Verfügung.

Anhand dieser vorhandenen Daten sollen Mustern und Kenngrößen ermittelt werden, die wiederkehrend und gleichermaßen in der Bewertung durch die bisherigen Anwender vorgenommen wurden. Diese Kenngrößen sollen fachlich evaluiert werden (nicht durch den / die Studierende/n zu erbringen). Dafür stehen Erkenntnisse und Fachansprechpartner aus diesem Bereich zur Verfügung.

Ähnlich gelagerte Projekte und Studien legen nahe, dass es bereits existierende Verfahren und ML-Instrumente gibt, die eine Einschätzung von Bildern / eine sinnvolle Bildanalyse ermöglichen. Hier ist es wünschenswert, wenn auf bestehenden Verfahren aufgesetzt werden kann und diese etablierten ML-Methoden verwendet werden. Außerdem kann an Vorarbeiten aus einem schon abgeschlossenen Forschungsmasterprojekt zu einer ähnlichen Themenstellung angeknüpft werden.

Bezug zum Thema Data Science

Im Rahmen der Tätigkeit ist es dem Studierenden im Austausch mit den Mitarbeitern der Connex Communication GmbH möglich, die Daten systematisch auszuwerten. Die Versorgungsdaten werden dabei mittels REST-API aus den Kundensystemen bereitgestellt und durch Vorbehandlung in einheitliche Formate und anonyme Strukturen gewandelt.

Im Rahmen der Nutzung einer Cloud-Plattform können diese Daten verarbeitet und analysiert werden. Als Kooperationspartner steht dabei u.a. die Microsoft Azure Plattform zur Verfügung. Andere Plattformen, die eine sinnvollere Analyse und Bearbeitung ermöglichen, sind aber ebenfalls denkbar.

Verfügbare Ressourcen

Für die Bearbeitung des Projekts kann der Rechencluster am Center for Applied Data Science Gütersloh genutzt werden. Außerdem steht die Infrastruktur der Connex Communication GmbH zur Verfügung. Diese basiert vor allem auf den Cloud-Strukturen der Microsoft Azure Plattform, wenn es um die Analyse geht. Dabei kann dort auf verschiedene Open-Source-KI-Frameworks zurückgegriffen werden. Die Datengrundlage wird gemeinsam mit einem Modell-Anwender zu Beginn erarbeitet und gemeinsam mit der Connex Communication GmbH definiert.

Projektplan

Erstes Semester: Konkretisierung des Forschungsvorhabens und Analyse des Nutzungskontexts. Dies beinhaltet die fachliche Einarbeitung in die Thematik sowie in die technische Cloud-Infrastruktur der Connex Communication GmbH (Produktname Vivendi Assist Services), in der die Datenstrukturen und die fachlichen Prozesse dargestellt werden. Hinzu kommt eine Recherche nach vergleichbaren Ansätzen und Verfahren inklusive Modellbetrachtung anhand von Literaturquellen. Die Erarbeitung eines Forschungsexposés am Ende des Semesters ist Prüfungsleistung.

Zweites Semester: Entwicklung der Analyse und ggfs. Anpassung der benötigten Algorithmen, so dass eine erste Implementierung in Form eines Prototyps erfolgen kann. Verwendung von eingegrenzten Datenbeständen anhand einer Forschungskooperation mit einem Connex-Kunden. Die Erstellung eines Papers zum Stand der Forschung (Überblick über das jeweilige Forschungsgebiet) ist Prüfungsleistung.

Drittes Semester: Erweiterung der Analyse auf weitere Datenbestände, um eine Evaluation des Verfahrens vorzunehmen. Fachliche Bewertung anhand eines Beta-Verfahrens in der Verwendung durch den Connex-Kunden. Die Erstellung eines Papers über das Benchmarking der eingesetzten ML-Verfahren ist Prüfungsleistung.

Viertes Semester: Umsetzung des assistierenden Systems in Form als Micro-Services-Architektur zur Analyse von Wundbildern. Gemeinsame Anbindung der Connex Software an diesen Service. Ein angepasstes Einsatzszenario sowie die technische Lösung liegen vor. Die Masterarbeit inklusive Kolloquium ist Prüfungsleistung.

Eignungskriterien

Zwingend:

- Bachelorabschluss in einer einschlägigen Fachrichtung (Informatik, Elektrotechnik, angewandte Mathematik, Kognitionswissenschaft o.ä.)
- Umfassende Programmierkenntnisse in mindestens einer objektorientierten Programmiersprache
- Gute mathematische Kenntnisse
- Gute Englischkenntnisse
- Teamfähigkeit und Interesse

Optional:

- Grundlegende Kenntnisse in der Bildverarbeitung
- Grundlegende Kenntnisse im maschinellen Lernen
- Erfahrung im Erstellen, Trainieren und Evaluieren von tiefen neuronalen Netzen

Erwerbbare Kompetenzen durch das Projekt

Der / die Studierende ist nach Abschluss des Projekts in der Lage,

- einen ML- und Cloud-basierten Datenanalyse-Workflow für die Bilderkennung aufzusetzen,
- die für die Bildverarbeitung relevanten ML-Verfahren zielgerichtet anzuwenden und anwendungsbezogen anzupassen und ggf. weiterzuentwickeln,
- Deep-Learning-Architekturen anwendungsspezifisch auszuwählen und zu designen,
- die theoretischen Anforderungen von ML-Verfahren mit den Erfordernissen der Praxis in einer realen Anwendung in Einklang zu bringen,
- die eigenen Forschungsergebnisse vor einem Fachpublikum zu präsentieren
- und wissenschaftliche Texte zu verfassen.

Forschungsprojektantrag für den Forschungsmaster Data Science

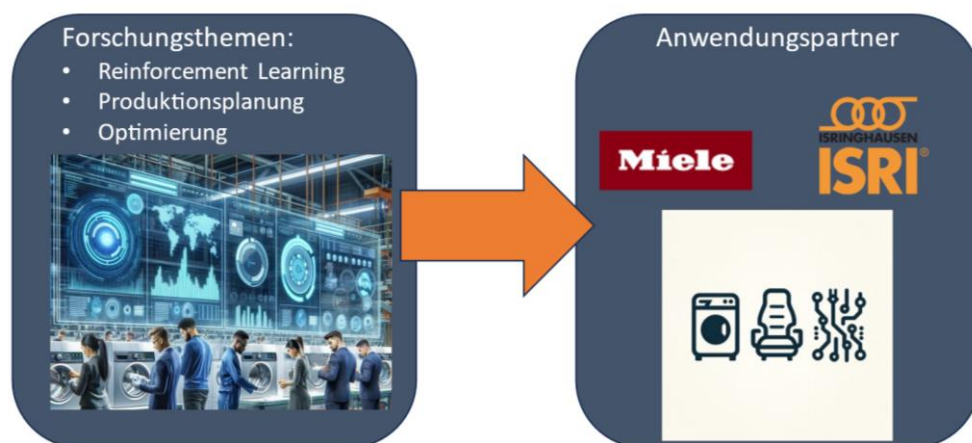
Projekttitle: 7 KI-basierte Produktionsplanung (SUPPORT: Sustainable and Human-centered Production Planning and Control Based on Reinforcement Learning Techniques)

Projektübersicht

Anzahl Studierende	1
Art	Gefördertes Projekt mit externen Partnern
Projektverantwortung	Prof. Dr. Pascal Reusch / Prof. Dr. Martin Kohlhase
Projektkontext	<ul style="list-style-type: none">• Angewandtes Forschungsprojekt in Zusammenarbeit mit Miele & Cie. KG, Istringhausen GmbH & Co. KG, Fraunhofer IOSB-INA und der FH Bielefeld (Center for Applied Data Science).• Darüber hinaus ist die Mitarbeit als wissenschaftliche Hilfskraft ggf. möglich und gewünscht.

Abstrakt

In diesem Projekt werden komplexe industrielle Produktionsprozesse mithilfe von Deep Reinforcement Learning (DRL) optimiert. Dazu wird zunächst ein datengetriebenes Verfahren entwickelt, welches Materialflusssimulationen automatisch generiert (Digitaler Zwilling der Fertigungslinie). Innerhalb dieses digitalen Zwillings lernt der zu entwickelnde DRL-Agent dann nicht nur, die Prozesseffizienz zu verbessern sondern auch nachhaltige und mitarbeiterorientierte Entscheidungen zu treffen. Mithilfe einer anwender-orientierten Analysetoolbox können menschliche Entscheider:innen mit dem DRL-Agenten interagieren und sein Entscheidungsverhalten untersuchen.



Kurzbeschreibung

Die Produktionsplanung und -steuerung (PPS) in Unternehmen beeinflusst sowohl die Wirtschaftlichkeit als auch die Materialströme, die Maschinenbelegungen sowie den konkreten Einsatz der Mitarbeiter:innen. Aktuelle Verfahren zur Planungsoptimierung stoßen bei der hohen Komplexität an ihre Grenzen und fokussieren sich zumeist nur auf die Steigerung der Produktivität eines Fertigungsprozesses. Das Ziel des Projekts, Sustainable and Human-centered Production Planning and

Control Based on Reinforcement Learning Techniques (SUPPORT) ist es, die komplexe Produktionsplanung zu vereinfachen. Dazu werden Unternehmen in die Lage versetzt, KI-Methoden in der Produktionsplanung und -steuerung einzusetzen, ohne selbst KI-Experten beschäftigen zu müssen. Konkret werden in dem Projekt die automatisierte Erstellung von Simulationsmodellen und der Einsatz von Reinforcement Learning zur Produktionsplanung entwickelt. Reinforcement Learning (bestärkendes Lernen oder verstärkendes Lernen) ist eine Form des maschinellen Lernens. Dabei werden im Unterschied zu anderen Methoden im Vorfeld keine Daten benötigt. Stattdessen beobachtet die Künstliche Intelligenz, vergleichbar wie ein Mensch, verschiedene Kausalitäten und lernt durch 'Trial-and-Error' in Zukunft unbekannte Probleme zu lösen. Der künstlichen Intelligenz wird nicht gezeigt, welche Aktion in welcher Situation die beste ist, sondern sie erhält zu einer bestimmten Zeit eine Belohnung, die auch negativ sein kann. Durch Reinforcement Learning soll es möglich sein, ohne menschliches Vorwissen komplexe Steuerungsprobleme zu lösen. Die Innovation des Projekts liegt in der nachhaltigen und zielgerichteten Produktionsplanung und -steuerung, die ökologische und ökonomische Ziele sowie die Ziele der Mitarbeiter berücksichtigt. Projektpartner sind die FH Bielefeld, das Fraunhofer IOSB-INA sowie die Unternehmen Miele und Isringhausen (<https://www.its-owl.de/news-events/news/news/news/sieben-neue-projekte/>).

Aufgabenstellung

Die Studierenden können je nach Interesse in einem der folgenden Teilbereiche forschen und arbeiten:

- Entwicklung des DRL-Agenten zur Steuerung der Fertigungsprozesse unter Berücksichtigung humanzentrierter Faktoren inkl. Benchmark mit herkömmlichen Methoden wie Metaheuristiken, Prioritätsregelverfahren oder ganzzahligen Optimierungsmodellen.
- Entwicklung von Verfahren zur automatischen, datengetriebenen Erzeugung von (diskreten, ereignisorientierten) Materialflusssimulationen auf Basis von Process-Mining-Techniken
- Entwicklung einer anwenderorientierten Analyse-Toolbox zur Interaktion mit dem DRL-Agenten im Kontext moderner Software- und Cloudarchitekturen, komplexer Datenmodelle und Datenpipelines, erweiterten Auswertungsmöglichkeiten (insb. im Bereich „Explainable AI“) und automatischen Hinweisen bzw. Verbesserungsvorschlägen

Bezug zum Thema Data Science

Verfahren des Reinforcement Learning, Process Mining, Metaheuristiken, Prognoseverfahren und Techniken der Explainable AI sind Kernthemen der Data Science und werden auch in den Lehrangeboten des Forschungsmasters behandelt.

Verfügbare Ressourcen

Das Projekt wird gemeinsam mit den Unternehmen Miele und Isringhausen sowie mit dem Fraunhofer Institut in realen Produktionsumgebungen durchgeführt. Über die operativen IT-Systeme der Unternehmen werden erforderliche Stamm- und Bewegungsdaten zur Umsetzung der Verfahren bereitgestellt. Die zu entwickelten Softwaremodule für den Live-Betrieb werden in einer modernen Cloudumgebung (MS Azure) implementiert. Für rechenintensive ML-Trainingsprozesse oder Optimierungsverfahren steht das Data Analytics Cluster des Center for Applied Data Science zur Verfügung.

Projektplan

Neben der Mitarbeit an der Erforschung und Entwicklung der Verfahren sind folgende wissenschaftliche Ausarbeitungen als Prüfungsleistungen vorausgesetzt:

Erstes Semester: Erstellung eines Forschungsexposés zu einem der drei Themenbereiche (siehe Aufgabenstellung).

Zweites Semester: Erstellung eines systematischen Literaturreviews über aktuell bestehende Verfahren und Konzepte auf Basis der im Forschungsexposé definierten Zielsetzung.

Drittes Semester: Erstellung eines Forschungspapers mit ersten Ergebnissen und Erkenntnissen. Dies kann z.B. die Entwicklung eines neuartigen Verfahrens oder Modells sein (konstruktionsorientierte Forschung). Alternativ können bestehende Verfahren für einen neuartigen Anwendungsfall angepasst und empirisch evaluiert werden (Benchmarking). Weiterhin möglich sind Arbeiten, wie Simulationsstudien, Surveys oder Fallsstudien.

Viertes Semester: Masterarbeit und Kolloquium. Hier kann z.B. ein bereits bearbeiteter Aspekt aus dem Projekt oder eine neuaufgetretene Forschungsfrage vertieft werden.

* Es ist ggf. möglich, darüber hinaus als wissenschaftliche Hilfskraft tätig zu werden.

Eignungskriterien

Zwingend:

- Bachelor-Abschluss mit quantitativem Schwerpunkt
- Kenntnisse in einer höheren Programmiersprache (z.B. Python, Java, C# o.ä.), die über ein Einstiegniveau hinausgehen
- Basiswissen über Statistik oder mathematische Optimierung
- Begeisterung für Data Science im Kontext von Industrie 4.0 sowie Freude am wissenschaftlichen Arbeiten im Team

Optional:

- Erfahrungen in einem oder mehreren der folgenden Bereiche sind von Vorteil:
 - Python 3.X Ökosystem inkl. einschlägiger ML-Bibliotheken
 - Machine Learning, insb. Reinforcement Learning oder Process Mining
 - Operations Research und Metaheuristiken, idealerweise im Kontext von Scheduling-Modellen
 - Materialflusssimulation bzw. diskrete ereignisorientierte Prozess-Simulation
 - Domänenwissen über Produktionsplanung/-steuerung und der systemischen Unterstützung mittels ERP, MES, APS, MRP usw.
 - Data Engineering / Cloud-Entwicklung (Datenbanken, Datenpipelines/ETL, APIs, Web-Frontend usw.)
- Interesse an einer einschlägigen Mitarbeit als wissenschaftliche Hilfskraft

Erwerbbarer Kompetenzen

- Einsatz und (Weiter-)Entwicklung von State-of-the-Art Verfahren des Maschinellen Lernens in hochkomplexen Umgebungen.
- Erweiterte Forschungskompetenzen durch die etwaige Publikation von Fachartikeln in einem hochmodernen Forschungsfeld.

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: 8 KI in der Pflegeberatung



Projektübersicht

Anzahl Studierende	1-2
Art	Projekt mit externem Partner
Projektverantwortung	Dr. Christoph Ostrau, Prof. Dr. Wolfram Schenck, Dr. Andreas Hellmann (Entyre)
Projektkontext	Das Projekt ist Teil der Arbeit von CareTech OWL und wird in enger Kooperation mit der Firma Entyre durchgeführt.

Abstrakt

Im Gesundheitswesen vorliegende Daten, wie z.B. Pflegedokumentationen, sind oft schwer systematisch auswertbar und beziehen häufig die Perspektive der Patient*innen nicht mit ein. In diesem Projekt erfolgt eine systematische Auswertung von regelmäßig bei den Patient*innen erhobenen Daten. Zur Auswertung unstrukturierter Anteile werden verschiedene Large Language Models evaluiert. Alternativ wird ein System mittels AutoML-Methoden entwickelt, das auf der Basis von Vorerkrankungen und weiteren Parametern Empfehlungen/Vorhersagen gibt.

Kurzbeschreibung

Der Übergang vom selbstständigen Leben in eine Pflegesituation, insbesondere im hohen Alter, erfolgt oft plötzlich und wird von den Betroffenen und ihren Angehörigen nur wenig vorbereitet. Gerade im Akutfall fehlt oft die Zeit und der Überblick, um alle notwendigen Hilfen und Versorgungsformen abzuwägen. Hier setzt Entyre an: Neben der Qualitätsanalyse bereits verordneter Hilfsmittel findet eine persönliche **Beratung im häuslichen Umfeld** statt, um den tatsächlichen Bedarf der Betroffenen und ihrer Angehörigen zu ermitteln. Dabei stehen die Betroffenen im Mittelpunkt: Es geht um ihre Wahrnehmung und ihre Herausforderungen. Unterstützt wird diese Erhebung im häuslichen Umfeld durch eine interaktive App, die den Pflegeberater*innen zur Verfügung steht. Erhoben werden akute Ereignisse (Sturz, Krankenhausaufenthalt etc.), Fragen zur Zufriedenheit und zum Befinden mittels „patient reported outcomes“. Viele Informationen werden in der App in strukturierter und damit maschinell gut auswertbarer Form erfasst. Einige Informationen werden in Freitextfeldern und zwischen den Konsultationsterminen ergänzt. Hier liegen die größten Herausforderungen: Wie können die strukturierten Daten vollumfänglich genutzt und ausgewertet werden? Können die Ergebnisse aktiv für die Pflegeberatung genutzt werden? Und wie können die unstrukturierten Daten so aufbereitet werden, dass diese maschinell ausgewertet werden können?

Wir wollen diesen Datenschatz heben, um einen echten **Mehrwert in der Patient*innenversorgung** zu schaffen. Nutzerorientierte Fragestellungen erheben die Sicht der Betroffenen und betrachten daher nicht nur die oft einseitigen Diagnosen von Ärzt*innen und Krankenhauspersonal. Die Ergebnisse der Datenanalyse sollen Defizite in der Hilfsmittelversorgung und andere Missstände, aber auch funktionierende Versorgungsformen aufzeigen. Akute Ereignisse und Versorgungsbedarfe sollen auf vorhergehende Erkrankungen zurückgeführt werden und Maßnahmen damit begründet werden. Darauf aufbauende Empfehlungen sollen die Fachkräfte in ihrer Arbeit unterstützen.

Ziel dieses Projektes ist es, eine Vorverarbeitung für die von der Entyre GmbH erhobenen Daten durchzuführen. Mittels einfacher statistischer Analysen soll der Datensatz charakterisiert und einfache Gesamtzusammenhänge sowie zeitliche Verläufe analysiert werden. Wahlweise kann darauf aufbauend an einem KI-Empfehlungssystem für Fachkräfte gearbeitet werden oder alternativ der unstrukturierte Teil des Datensatzes mittels großer Sprachmodelle ausgewertet werden.

Ersteres basiert auf der Analyse von Zusammenhängen zwischen durchgeführten Maßnahmen sowie äußeren Umständen und der Verbesserung bzw. Verschlechterung der Situation der Patientinnen und Patienten. Beispielsweise soll das System in der Lage sein, mögliche akute Ereignisse (schwere Stürze) auf der Grundlage von bestehenden Erkrankungen (z.B. Bluthochdruck) und der häuslichen Umgebung (hohe Türschwellen oder Teppiche) vorherzusagen.

Die Analyse unstrukturierter Daten wird zunächst mit Hilfe von vorab trainierten Modellen evaluiert. Insbesondere soll der Einsatz von Open Source Large Language Models (z.B. Llama2 und Derivate) für die Datenanalyse evaluiert werden. Quantisierte Modelle bieten hier die Möglichkeit einer effizienten und lokalen Anwendung und sind daher gerade im Bereich der Analyse sensibler medizinischer Daten von besonderem Interesse. Sollte die Qualität der Ergebnisse nicht ausreichen, wird eine Spezialisierung der Modelle auf der Basis von medizinischen Korpora oder neu zu erstellenden Datensätzen verfolgt.

Aufgabenstellung

Für beide Arbeitsbereiche werden zunächst die vorhandenen strukturierten Daten gesichtet, vorverarbeitet und bereinigt. Es wird eine Datenbank aufgebaut und einfache statistische Verfahren zur Charakterisierung der Daten angewendet.

Anschließend wird ein Empfehlungssystem auf der Basis von Interventions- und Behandlungsverläufen entwickelt. Die genaue technische Umsetzung ist Teil der Ausarbeitung. Denkbar ist hier die Anwendung von AutoML-Methoden auf einen Teil des Datensatzes und die Erweiterung der Gewinnerarchitektur.

Alternativ wird der Freitextanteil mit textverarbeitenden KIs, wie Open Source Large Language Models, ausgewertet. Eine Maßnahme zur Verbesserung der Performance soll evaluiert werden, wie z.B. Transfer Learning mit existierenden medizinischen Textdatensätzen.

Bezug zum Thema Data Science

Im vorgestellten Projekt geht es sowohl um die Datenvorverarbeitung und -aufbereitung, die Einbindung in Datenbanken und die Themen Transfer Learning, Natural Language Processing oder AutoML. Diese Bereiche sind alle direkt mit dem Thema Data Science verknüpft.

Verfügbare Ressourcen

Die benötigten Daten werden von der Entyre GmbH nach Unterzeichnung einer Vertraulichkeits- und Datenschutzerklärung in pseudonymisierter Form zur Verfügung gestellt. Als Rechenressourcen wird die TransCareTech Erweiterung des YourAI Clusters genutzt. Diese stellt 4 Nvidia A100 für das Training von KI-Methoden zur Verfügung.

Projektplan

Erstes Semester: Ausarbeitung eines Forschungsexposees, Einarbeitung in die Thematik sowie Sichtung des Datensatzes

Zweites Semester: Recherchen zum Thema KI-Empfehlungssysteme/KI-Sprachverarbeitung, Analyse des Datensatzes mittels statistischer Verfahren. Erste Veröffentlichung zum Stand der Technik.

Drittes Semester: Implementierung der KI-Verfahren und Auswertung. Zweite Veröffentlichung zur ersten Ergebnisse.

Viertes Semester: Optimierung der Verfahren, Weiterentwicklung zur Anwendung, Finale Auswertung. Masterarbeit und Kolloquium.

Eignungskriterien

Zwingend:

- Programmierkenntnisse in Python
- Begeisterung für die Motive und das Thema des Projekts

Optional:

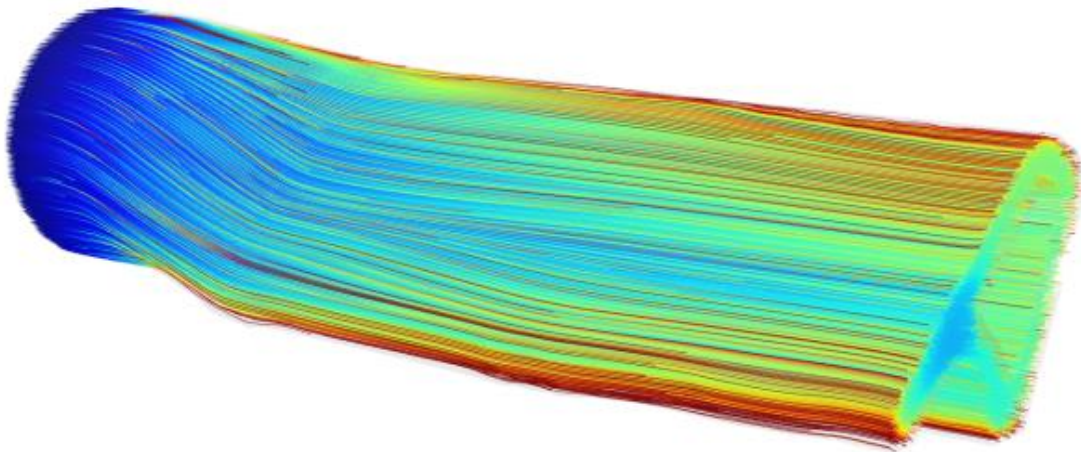
- Erfahrung mit der Anwendung von ML
- Erfahrung mit Datenbanken

Erwerbbarer Kompetenzen

- Umgang mit großen Datensätzen
- Vorverarbeitung von Daten und Feature Engineering
- Einsatz von KI-Verfahren in sozialen/gesundheitsbezogenen Berufen

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: 9 Approximation der Simulation von durchströmten Maschinen mittels tiefer Neuronaler Netze



Projektübersicht

Anzahl Studierende	1-2
Art	Projekt mit externen Partnern
Projektverantwortung	Prof. Dr. Christian Schwede
Projektkontext	Projekt mit der Firma IANUS aus Dortmund/Bielefeld. Eine Anstellung als Hiwi ist erwünscht.

Abstrakt

Mit Hilfe von Maschine Learning Verfahren soll eine Approximation einer Strömungssimulation zur Optimierung von Werkzeuggeometrien für Industrielle Anlagen berechnet werden. Die Approximierten Modelle werden dann zusammen mit Optimierungsverfahren verwendet, um die Werkzeuggeometrien über die Veränderung der Parameter ihrer Digitalen Zwillinge zu optimieren.

Kurzbeschreibung

Geometrisch und kinematisch komplexe Maschinen für die Industrie sollen in einem Rapid Prototyping Prozess optimiert werden. Dazu werden Systemlösungen eingesetzt, die den Konstruktionsprozess digitalisieren, indem sie einen Digitalen Zwilling (Digital Twin, DT) geometrisch per Parametrisierung aufbauen. Durch Variation des DT und vollautomatischer numerischer Simulation mit einem S-a-a-S Ansatz mit HPC Hardware kann man den Simulationsdurchsatz der virtuellen Maschine untersuchen und maximieren. Da die Simulationsläufe eine gewisse Zeit dauern ist eine direkte Anbindung der Simulation als Auswertung der Güte eine Geometrie mit einem Optimierungsalgorithmus nicht erfolgversprechend. Das Ziel dieses Projektes ist es deshalb die Simulation durch Verfahren des ML zu approximieren, um so eine schnelle Vorhersage der zu erwartenden Güte vieler Varianten des DT zu erlauben. Aufbauend auf diesen Bewertungen kann später eine effiziente automatische Optimierung etwa mit Evolutionären Algorithmen erfolgen.

Aufgabenstellung

In enger Kooperation mit der Firma IANUS in Dortmund und Bielefeld soll für reelles Industrieprojekt auf Basis von tiefen Neuronalen Netzen eine Approximation der bereits existierenden Strömungssimulation des Werkzeugs erlernt werden. Das StrömungsRaum System der Firma IANUS läuft auf HPC Hardware und ist als Software-as-a-Service flexible einsetzbar und von außen parametrisierbar. Die hierzu benötigten Daten können mit der Simulation erzeugt werden. Je nach Projektfortschritt kann das erlernte Modell dann in einer Metaheuristik verwendet werden um die Parameter des Werkzeugs zu optimieren. Eine Anstellung als wissenschaftliche Hilfskraft bei der Firma ist erwünscht.

Bezug zum Thema Data Science

Überwachtes lernen mit Tiefen Neuronalen Netzen ist ein Kernbereich der Data Science.

Verfügbare Ressourcen

Die Daten werden über die Simulation der Firma IANUS generiert. Die Simulation und das entsprechende Model sowie die HPC Hardware wird von der Firma bereitgestellt. Außerdem wird ein Experte aus der Firma als Ansprechpartner und Co-Betreuer des Forschungsprojektes zur Verfügung stehen.

Projektplan

Erstes Semester: Einarbeitung in die Fragestellung, den Anwendungsfall und die Simulationssoftware. Erstellung eines Forschungsexposés.

Zweites Semester: Erstellung eines Papers, das einen Überblick über das Forschungsgebiet der Simulationsapproximation durch Maschinelles Lernen gibt. Umsetzung eines ersten Modelles und Generierung von ersten Ergebnissen.

Drittes Semester: Implementierung und Vergleich von mindestens zwei Ansätzen zur Approximation der Simulation. Erstellung eines Papers, mit ersten quantitativen Ergebnissen.

Viertes Semester: Masterarbeit und Kolloquium. Verfeinerung des Modells und u.U. Einbindung in eine Metaheuristik zur Optimierung der Werkzeugparameter.

Eignungskriterien

Zwingend:

- Programmierkenntnisse in einer der großen OO Sprachen

Optional:

- Erfahrung mit Simulation (speziell Strömungssimulation)
- Programmiererfahrung in Python
- Erfahrung im Maschinellen Lernen

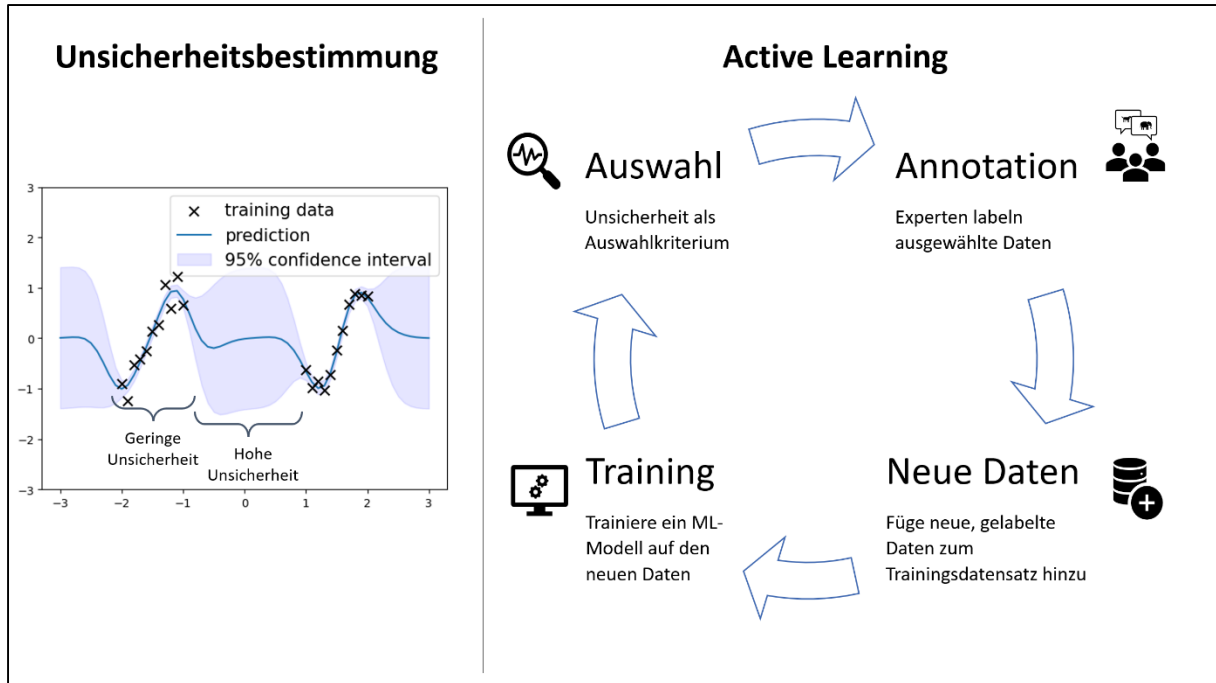
Erwerbbarer Kompetenzen

- Approximation von dynamischen Modelle mittels ML-Verfahren
- Implementierung und Anwendung Tiefer Neuronale Netze
- Implementierung und Anwendung von Metaheuristiken
- Funktionsweise und Anwendung von Strömungssimulation
- Optimierung von Werkzeuggeometrien

- Wissenschaftliches Schreiben und Veröffentlichen

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: 10 Untersuchung von Methoden zur Unsicherheitsbestimmung und ihrer Anwendung auf Active Learning



Projektübersicht

Anzahl Studierende	1
Art	Studienprojekt
Projektverantwortung	Prof. Dr.-Ing. Martin Kohlhase; M.Sc. Bjarne Jaster
Projektkontext	Das Studienprojekt ist im Projekt SAIL (Sustainable AI Lifecycle) angesiedelt. SAIL ist eine Kooperationen der Fachhochschulen und Universitäten aus OWL mit dem Ziel KI nachhaltig zu gestalten.

Abstrakt

In dem Projekt geht es darum die Unsicherheit von KI-Methoden bei Vorhersagen zu bestimmen. Dazu sollen verschiedene Methoden verglichen werden. Nachfolgend wird die Anwendung der Unsicherheitsbestimmungs-Methoden im Forschungsfeld Active Learning betrachtet.

Kurzbeschreibung

Künstliche Intelligenz (KI) wird in immer mehr Bereichen eingesetzt. Eine wichtige Grundlage für Menschen, die die KI nutzen oder mit ihr arbeiten, ist, dass die Entscheidungen/Vorhersagen der KI nachvollziehbar sind. Um das zu ermöglichen, werden Methoden entwickelt, die versuchen die Vorhersage zu begründen oder zu erklären (Interpretierbare KI). Ein weitere wichtiger und oft übersehener Faktor bei der Nutzung von KI ist die Sicherheit (Confidence) oder Unsicherheit die ein KI-Modell in seine Vorhersagen hat. Dies ermöglicht eine bessere Zusammenarbeit von Mensch und KI,

da der Mensch bei hoher Unsicherheit eingreifen kann. So kann bspw. ein Maschinist intervenieren, wenn eine KI, die eine Maschine steuert, eine hohe Unsicherheit bei einer Entscheidung aufweist. Auf diese Weise können potentielle Schäden an der Maschine verhindert werden.

Eine weitere Anwendung von Unsicherheitsbestimmungen für KI-Modelle ist Active Learning. Active Learning kommt zum Einsatz wenn für eine Problemstellung eine KI entwickelt werden soll, aber noch kein brauchbarer Datensatz vorliegt. Entweder gibt es schon Daten, die dann aber kein Label haben oder es liegen überhaupt keine Daten vor. Im ersten Fall ergibt sich die Fragestellung: Welche Datenpunkte sind besonders hilfreich, um ein KI-Modell zu trainieren? Während sich beim zweiten Fall die Frage stellt: Wo sollten Daten erhoben/gemessen werden, sodass eine möglichst leistungsfähige KI trainiert werden kann? Active Learning Methoden lösen das Problem, indem sie ein Kriterium, das den Wert eines Datenpunkts beschreibt, definieren. So lässt sich feststellen welche Datenpunkte besonders hilfreich sind (1. Fall) oder in welchen Regionen Messungen besonders nützlich sind (2. Fall).

Für das erwähnte Kriterium lassen sich auch Methoden zur Unsicherheitsbestimmung nutzen. So sind die Datenpunkte für die ein KI-Modell besonders hohe Unsicherheiten aufweist besonders hilfreich für die Verbesserung des Modells und sollten daher gelabelt werden.

Aufgabenstellung

Der / Die Studierende soll im Rahmen des Forschungsmasters Methoden zur Unsicherheitsbestimmung für verschieden Methoden künstlicher Intelligenz (Machine Learning Methoden) anwenden und vergleichen. Danach sollen diese Methoden auf Active Learning Probleme angewendet.

Bezug zum Thema Data Science

Es werden verschiedene Methoden Künstlicher Intelligenz (Machine Learning Methoden) inklusive der Möglichkeiten zur Bestimmung von Unsicherheiten behandelt. Wie in der Kurzbeschreibung erwähnt ist das relevant für eine breite Anwendung von KI. Zusätzlich wird das Thema Active Learning behandelt was besonders für industrielle Data Science Anwendungen von Bedeutung ist.

Verfügbare Ressourcen

Für die Bearbeitung des Projekts kann die Infrastruktur des CfADS Gütersloh genutzt werden, die aus folgenden Komponenten besteht:

- Data-Analytics-Cluster: rechenstarker Computercluster auf Basis des Hadoop-Frameworks
- IoT-Factory: Modellfabrik zur Nachbildung realer Produktionsabläufe und Produktdaten
- Smart Service Lab: Labor zur Entwicklung von Smart Services und Assistenzsystemen auf Basis von Smart Devices.

Projektplan

Erstes Semester: Konkretisierung des Forschungsvorhabens. Dies beinhaltet die Einarbeitung in die CfADS-Infrastruktur, in die Methoden und Toolboxes. Ebenso wird mit einer Literaturrecherche, der Einarbeitung in die Thematik und die Formulierung von zu beantwortenden Forschungsfragen begonnen. Erstellung eines Forschungsexposés ist Prüfungsleistung.

Zweites Semester: Die im ersten Semester identifizierten Methoden zur Unsicherheitsbestimmung werden miteinander verglichen (und wenn nötig implementiert). Erstellung eines Papers, das einen Überblick über die Qualität der Methoden gibt ist Prüfungsleistung.

Drittes Semester: Die im zweiten Semester verglichenen und verwendeten Methoden werden auf Active Learning Probleme angewendet. Dabei wird besonders der Einfluss der Qualität der Methoden zur Unsicherheitsbestimmung auf die Active Learning Ergebnisse untersucht. Erstellung eines Papers, mit den Ergebnissen des Active Learnings ist Prüfungsleistung.

Viertes Semester: Im vierten Semester wird eine Methode aus den vorherigen Semestern ausgewählt und intensiv auf Stärken und Schwächen untersucht. Zusätzlich soll ein Vorschlag zur Ausbesserung der Schwächen präsentiert und untersucht werden. Die Masterarbeit inklusive Kolloquium ist Prüfungsleistung.

Eignungskriterien

Zwingend:

- Grundkenntnisse Informatik & Künstliche Intelligenz / maschinelles Lernen
- Erfahrung in Python-Programmierung
- Teamfähigkeit und Interesse an wissenschaftlicher Arbeit
- Eigenständige Arbeitsweise

Optional:

- Erfahrungen in den Bereichen der Mathematik, Statistik (Data-Science) und der Visualisierung von Daten und deren Zusammenhänge

Erwerbbarer Kompetenzen

Kompetenzen, die durch das Projekt erworben werden:

- Maschinelle Lernverfahren, speziell im Bereich der Klassifikation und Regression
- Anwendung verschiedener Active Learning Methoden
- Aufbereitung und Präsentation von Analyseergebnissen
- Wissenschaftliches Arbeiten, Schreiben und Präsentieren

Forschungsprojekttitel für den Forschungsmaster Data Science

Projekttitel: 11 Cognitive Edge Computing für KI-/ML-basierte Oberflächeninspektion



Projektübersicht

Anzahl Studierende	1
Art	Projekt mit externem Partner
Projektverantwortung	Prof. Dr.-Ing. Thorsten Jungeblut
Projektkontext	Das beteiligte Unternehmen BST Group stellt umfangreiche Testdatensätze, sowie einen prototypischen Hardwareaufbau zur Verfügung. Eine Beschäftigung als wissenschaftliche Hilfskraft ist möglich. Die BST Group ist sehr an der Gewinnung und langfristigen Beschäftigung von Nachwuchskräften interessiert.

Abstrakt

Ziel des Projektes ist die Entwurfsraumexploration von KI-/ML-Hardwarebeschleunigern in der Oberflächeninspektion. Im Zentrum steht nicht nur das Lernen der Modelle auf HPC-Systemen, sondern auch die effiziente Ausführung (Inferenz) auf eingebetteter Hardware. Ergebnis der Entwurfsraumexploration ist die Partitionierung der Anwendung, d.h. welche KI-Verfahren können auf dem Sensor ausgewertet werden, welche Verfahren können über Edge-Hardware (z.B. eingebettete GPU/FPGA) beschleunigt werden und welche erfordern leistungsfähige HPC-Hardware in der Cloud.

Kurzbeschreibung

Die Entwicklungen im Bereich der intelligenten technischen Systeme (ITS) führen derzeit zu einem Umbruch in der gesamten Wertschöpfungskette der industriellen Produktion. Die zunehmende

Leistungsfähigkeit der Informationsverarbeitung bietet für das Anwendungsgebiet der industriellen Bildverarbeitung viele neue Möglichkeiten, an denen Mensch und Maschinen bisher an ihre Grenzen stoßen. Insbesondere die Nutzung Maschinellem Lernverfahren und Methoden der künstlichen Intelligenz versprechen bisher nicht gekannte Möglichkeiten beispielsweise in der Objektklassifikation oder der visuellen Qualitätskontrolle. Etwa seit dem Jahr 2010 werden mit Deep Neuronal Networks (DNN) bzw. Convolutional Neuronal Networks (CNN) bedeutende Fortschritte erzielt.

Die Leistungsfähigkeit etablierter KI-/ML-Verfahren basiert aber bisher meist auf der Nutzung leistungsfähiger dezentraler Rechenressourcen (High-Performance Computing) in der Cloud. Nicht nur für das Lernen der Modelle, sondern auch für deren Ausführung (Inferenz) ist der Anwender auf diese leistungsfähigen Ressourcen angewiesen. Die im Bereich in der industriellen Bildverarbeitung auftretenden Anforderungen weichen jedoch aufgrund hoher Ansprüche an niedrige Latenz, Echtzeitfähigkeit oder Datenlokalität von den durch große Anbieter von KI-Know-How adressierten Fragestellungen ab. Auch Fragestellungen wie Wartbarkeit, Zertifizierbarkeit oder Privacy erschweren die Nutzung populärer Modelle wie DNNs oder CNNs in der Cloud.

Aber auch im Bereich der effizienten Ausführung von KI-/ML-Verfahren auf eingebetteten Systemen (**Cognitive Edge Computing**) wurden in der Vergangenheit große Fortschritte gemacht. Auf allen Ebenen der unterschiedlichen Verarbeitungskonzepte in der vernetzten Produktion (Edge/Fog/Cloud-Computing) finden sich geeignete Hardwarebeschleuniger, die auf einen geeigneten Kompromiss zwischen den Systemressourcen wie benötigter Leistungsfähigkeit (z.B. Klassifikationsgenauigkeit), Leistungsaufnahme/Energiebedarf oder Datendurchsatz/Latenz abzielen. Beispiele für relevante Hardwarearchitekturen sind eingebettete Mikrocontroller, Embedded GPUs, eingebettete FPGAs oder dedizierte KI-Hardwarebeschleuniger.

Die BST Group entwickelt Systeme für die Oberflächeninspektion (z.B. für die Batteriezellproduktion), die unter anderem typische Fehler Fertigungsprozess (z. B. Beschichtungsaussetzer) verlässlich und umgehend erkennen können. Durch ihren modularen Aufbau können BST Systeme an unterschiedlichste Anwendungen perfekt angepasst werden. Bilderfassung und Fehlererkennung erfolgen in Echtzeit. Die Systeme sind anwendbar auf uniforme, texturierte sowie bedruckte Oberflächen. Die umgehende und automatische Erkennung und Anzeige auch kleinster Fehler und Abweichungen ermöglichen eine schnelle und sichere Anpassung des Prozesses zur Ausschussvermeidung. Das System besteht aus mehreren optischen Sensoren, deren Sensordaten geeignet fusioniert werden. Die Anwendung maschineller Lernverfahren verspricht hier eine höhere Leistungsfähigkeit in der Fehlererkennung und Optimierung des gesamten Prozesses. Mit Bahnlaufgeschwindigkeiten von mehreren hundert Metern pro Sekunde stellen sich allerdings sehr hohe Anforderungen an die Geschwindigkeit bzw. Latenz der eingesetzten KI-/ML-Verfahren.

Aufgabenstellung

Ziel des Projektes ist die Entwurfsraumexploration von KI-/ML-Hardwarebeschleunigern für den Einsatz in der Oberflächeninspektion. Im Zentrum steht nicht nur das Lernen der Modelle auf HPC-Systemen, sondern auch die effiziente Ausführung (Inferenz). Die Betrachtung der gesamten Systemarchitektur vom intelligenten Sensor über Edge-Gateways zur lokalen Datenvorverarbeitung bis hin zur Cloud-Infrastruktur stellt hohe Anforderungen an die Abbildung der KI-Verfahren auf die Hardware. Die Wahl der Zielarchitektur hat wiederum Einfluss auf die Wahl und das Training der Modelle. Hieraus ergibt sich ein iterativer Zyklus (Model-to-Inference-to-Model), der sowohl die Auswahl geeigneter KI-Verfahren, als auch die Bestimmung der (Hyper-)Parameter des Modells beinhaltet. Diesen Ansatz einer ganzheitlichen Entwurfsraumexploration kann man auch in Anlehnung an den etablierten Begriff HW/SW-Co-Design als HW-/KI-Co-Design bezeichnen. Die Entwurfsraumexploration soll dabei unterschiedlichen Entwurfsziele, wie beispielsweise Klassifikationsgenauigkeit, Latenz oder Ressourcenbedarf der Hardware berücksichtigen. Ergebnis der

Entwurfsraumexploration ist die Partitionierung der Anwendung, d.h. welche KI-Verfahren können direkt auf dem Sensor ausgewertet werden (z.B. durch Dimensionsreduktion oder Feature-Extraktion), welche Verfahren können über Edge-Hardware (beispielsweise eingebettete GPU/FPGA oder dedizierte TPU-Beschleuniger) beschleunigt werden und welche erfordern leistungsfähige HPC-Hardware in der Cloud.

Bezug zum Thema Data Science

Die Evaluation und Anwendung von KI-/ML-Verfahren im Bereich des „maschinellen Sehens“ z.B. den Einsatz von CNNs in der Objektklassifikation sind ein Kernthema der Data Science und werden beispielsweise in den Modulen „Data Mining & Machine Learning“ sowie „Künstliche Intelligenz“ behandelt. Die KI-gestützte Bildverarbeitung stellt hohe Anforderungen an Organisation und Verarbeitung der Daten auf allen Ebenen der IoT-Verarbeitungskonzepte (Edge/Fog/Cloud). Dieses ist Kern des Moduls „Big Data Architekturen“. Die Betrachtung des gesamten Systemprozesses vom bildgebenden Sensor bis zur Cloud erfordert eine ganzheitliche Betrachtung des vollständigen Data Science Prozesses, welches im Modul „Data Science“ behandelt wird.

Verfügbare Ressourcen

- Informationen, die zur Erstellung des Szenarios benötigt werden (Systembeschreibung, Schnittstellen, Dokumentationen, relevante Kennzahlen etc.) werden von der BST Group bereitgestellt
- Die BST Group stellt umfangreiche Testdatensätze aus realen Produktionsumgebungen zur Verfügung
- Der Ansprechpartner der BST Group wird über die Projektlaufzeit zur Verfügung stehen
- Die für den prototypischen Aufbau benötigten Komponenten sowie sonstiges benötigtes Material wird von der BST Group bereitgestellt
- Hardware für das komplexere Maschine Learning ist über das Data Science, das CfADS sowiedem KI-Rechencluster yourAI der FH Bielefeld verfügbar

Projektplan

Erstes Semester: Erstellung eines Forschungsexposés als Prüfungsleistung. Einarbeitung in die das Konzept der Oberflächeninspektionssysteme von BST, die Schnittstellen der zur Verfügung gestellten intelligenten Sensoren.

Zweites Semester: Erstellung des Systemkonzepts zur Entwurfsraumexploration von Cognitive-Edge-Computing-Architekturen. Recherche zu relevanten Arbeiten im Themenfeld des Einsatzes von KI-/ML-Verfahren für die Sensordatenverarbeitung im o.g. Kontext. Erstellung eines Papers, das einen Überblick über das jeweilige Forschungsgebiet gibt, als Prüfungsleistung.

Drittes Semester: Entwicklung eines ersten Demonstrators und Proof-of-Concept zur Hardwarebeschleunigung einer Anwendung zur Oberflächeninspektion von BST. Vergleich eines KI-Verfahrens mit einer klassischen Realisierung.

Viertes Semester: Masterarbeit und Kolloquium. Implementierung und Vergleich unterschiedlicher Kombinationen aus KI-/ML-Verfahren und Hardwarebeschleunigern. Systematische Evaluation und Exploration der Effizienz der Kombinationen. Vergleich verschiedener Verarbeitungskonzepte (Embedded AI, Edge, Cloud). Finale Evaluierung durch Vergleich der implementierten Strategien. Erstellung eines Papers mit ersten quantitativen Ergebnissen als Prüfungsleistung.

Eignungskriterien

Zwingend:

- Programmierkenntnisse

Optional:

- Erfahrung mit IoT-Geräten
- Programmierung von Mikrocontrollern
- Erfahrung in der GPU-Programmierung

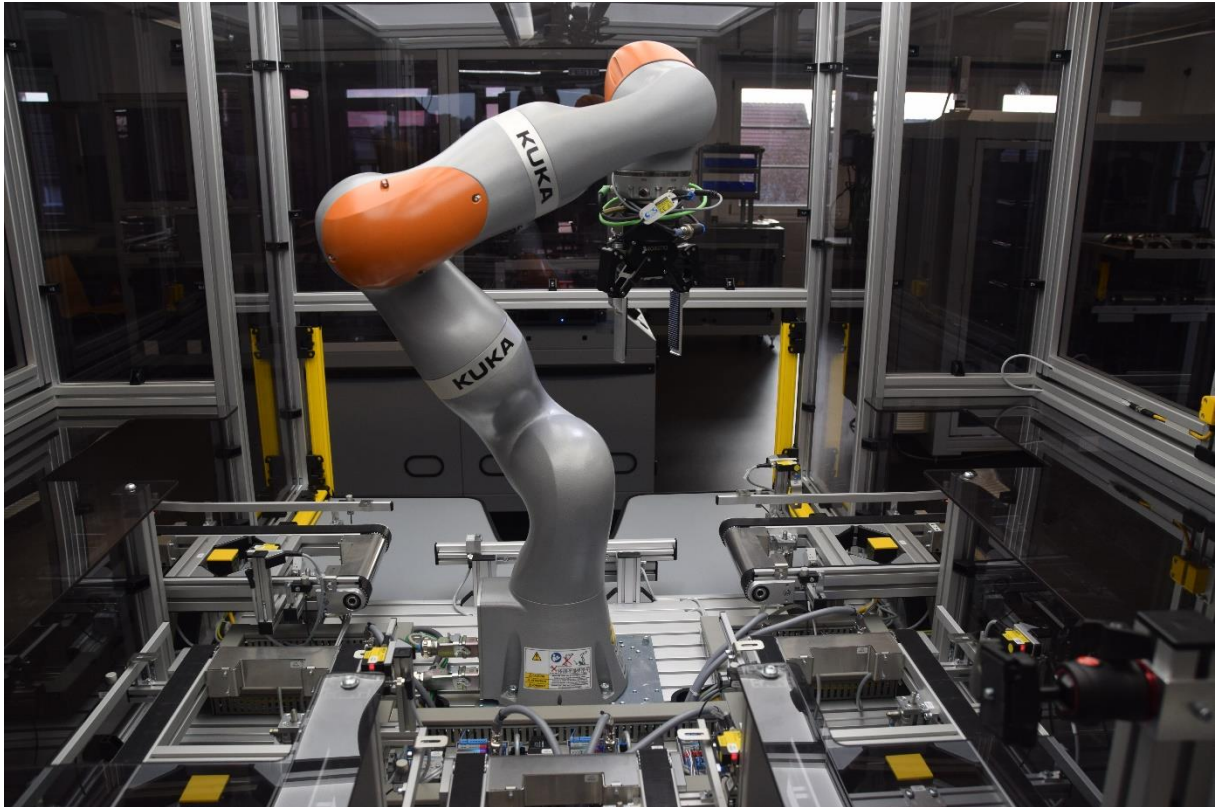
Erwerbbarere Kompetenzen

- Ressourceneffiziente Informationsverarbeitung auf den verschiedenen Ebenen (Edge, Fog, Cloud) im Sinne des IoT-Verarbeitungskonzepts
- Sensornahe Informationsverarbeitung
- KI-/ML-Verfahren
- Einsatz eingebetteter Hardware zur Beschleunigung von KI-/ML-Verfahren

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitel:

12 Entwicklung von datenbasierten Methoden zur Fehlervermeidung in Produktionsanlagen



Projektübersicht

Anzahl Studierende	1-2
Art	Gefördertes Projekt mit externem Partner
Projektverantwortung	Prof. M. Kohlhase, M.Eng. N.Migenda, M.Sc. M.Dockhorn
Projektkontext	Projektpartner: Center for Applied Data Science (Institut der HSBI) und Fraunhofer-Institut für Entwurfstechnik Mechatronik (Universität-Paderborn). Der Studierende wird Teil des Projektteams und steht in regelmäßigem Austausch mit wissenschaftlichen Mitarbeitern der HSBI und des Fraunhofer-Instituts IEM aus Paderborn. Eine Anstellung als WHK ist möglich

Abstrakt

Ziel des Projekts ist die Entwicklung und Validierung von Methoden aus dem Bereich Prescriptive-Analytics für Produktionsprozesse. Dies beinhaltet sowohl die datenbasierte Erkennung und Voraussage von Fehlern als auch die Erzeugung von Handlungsempfehlungen für den Benutzer. Für die

Implementierung und Validierung der studentischen Arbeiten steht die in Abbildung 1 dargestellte IoT-Factory zur Verfügung, welche am Campus Gütersloh der HSBI aufgebaut ist. In allen Phasen des Forschungsmasters befindet sich der Studierende im Austausch mit dem Projektteam und wird durch die betreuenden wissenschaftlichen Mitarbeiter und den Professor individuell begleitet und gefördert.

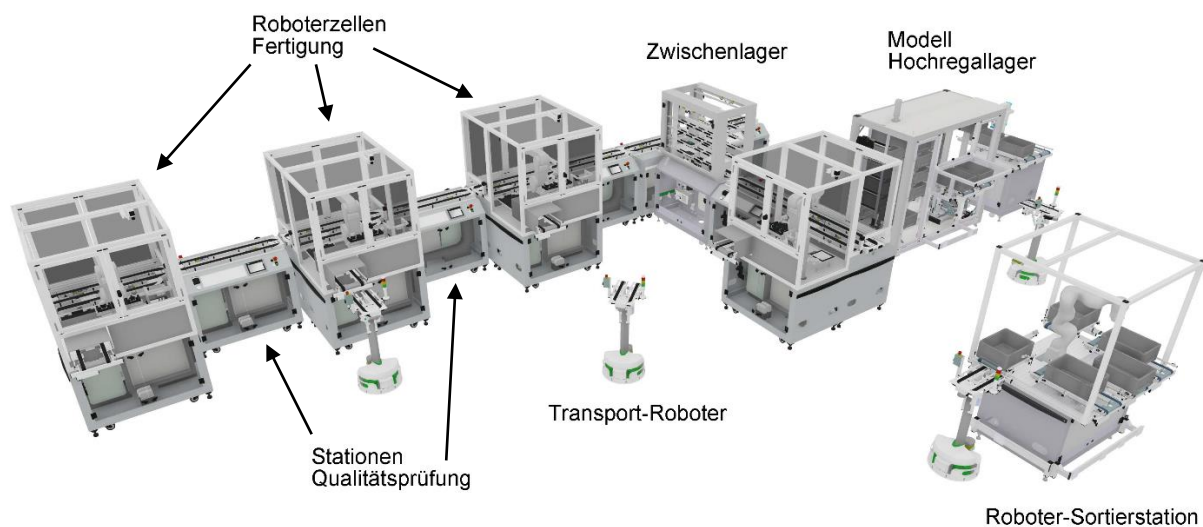


Abbildung 1: IoT-Factory der HSBI (Standort: Campus Gütersloh)

Beschreibung

Die Anforderungen an produzierende Unternehmen in Zeiten von Globalisierung und Automatisierung steigen stetig. Unternehmen jeglicher Größe müssen unter hohem Druck effizient und wirtschaftlich produzieren und schnell auf veränderte Bedarfe reagieren. Verfahren aus dem Bereich Prescriptive-Analytics und der künstlichen Intelligenz ermöglichen eine Flexibilisierung der Produktionsabläufe bspw. im Fehlerfall sowie eine Optimierung und Effizienzsteigerung der Fertigungsprozesse. Im Projekt VIP4PAPS soll eine neue, modulare und nutzerzentrierte Plattform für Anwendungen von Prescriptive-Analytics entwickelt und validiert werden, die den Entwicklungsaufwand und die Entwicklungszeit der Unternehmen senkt und durch Einsparungen von Material- und Energieressourcen zur Nachhaltigkeit in der Produktion beitragen.

Die Themenschwerpunkte und Aufgaben des Studierenden leiten sich aus den Zielen des Projekts ab. Dieses wird gemeinsam vom Center for Applied Data Science (CfaDS) und dem Fraunhofer-Institut für Entwurfstechnik Mechatronik (IEM) durchgeführt. Das langfristige Ziel des Projekts ist die Validierung einer Prescriptive-Analytics-Plattform für Produktionsanlagen. Zusammengefasst meint dies die Entwicklung eines Tools für Anlagennutzer, welches anhand von Produktionsdaten Voraussagen über zukünftig auftretende Fehler trifft und entsprechende Handlungsempfehlungen für die Vermeidung eines tatsächlichen Fehlerfalls generiert. Es wird somit Präskription angestrebt, was eine Form der Datenanalyse ist und, wie in Abbildung 2 genauer dargestellt, das Konzept von Deskription, Diagnostik und Prädiktion erweitert.

Aktuell werden im Rahmen des Projekts Methoden aus dem Bereich des maschinellen Lernens, wie z.B. Large-Language-Models und Grafdatenbanken, verwendet. Für die kommende Projektphase ist zusätzlich die Implementierung von datenbasierten Ansätzen aus dem Bereich der fehlertoleranten Systeme (Regelungstechnik) geplant. Mit diesen sollen anhand historischer Daten Bausteine zur Fehlererkennung und Erzeugung synthetischer Daten umgesetzt werden. Eine entsprechende Übersicht über das aktuelle Plattformkonzept ist in Abbildung 3 dargestellt.

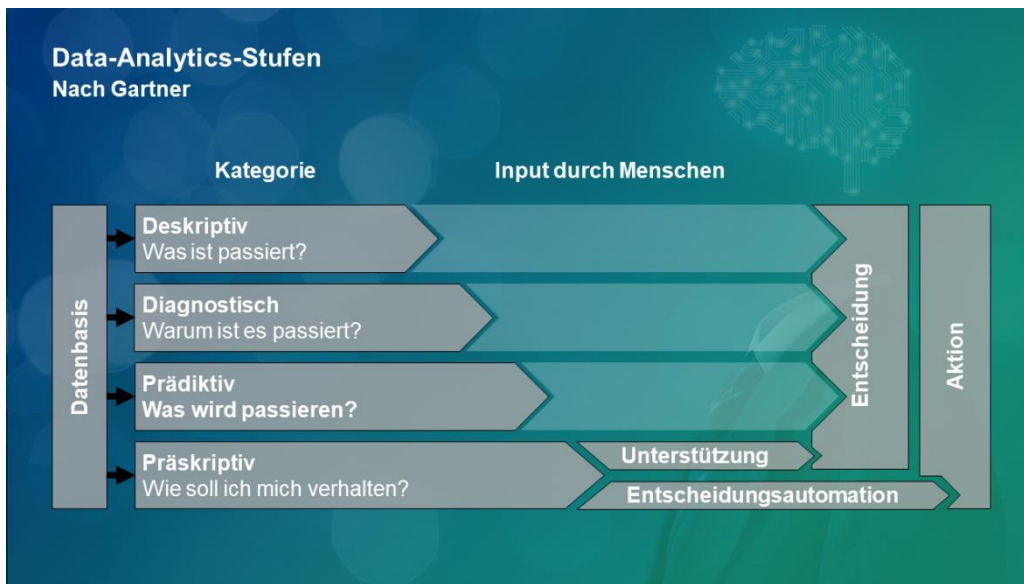


Abbildung 2: Data-Analytics-Stufen nach Gartner (Quelle: J.Weller, Fraunhofer IEM)

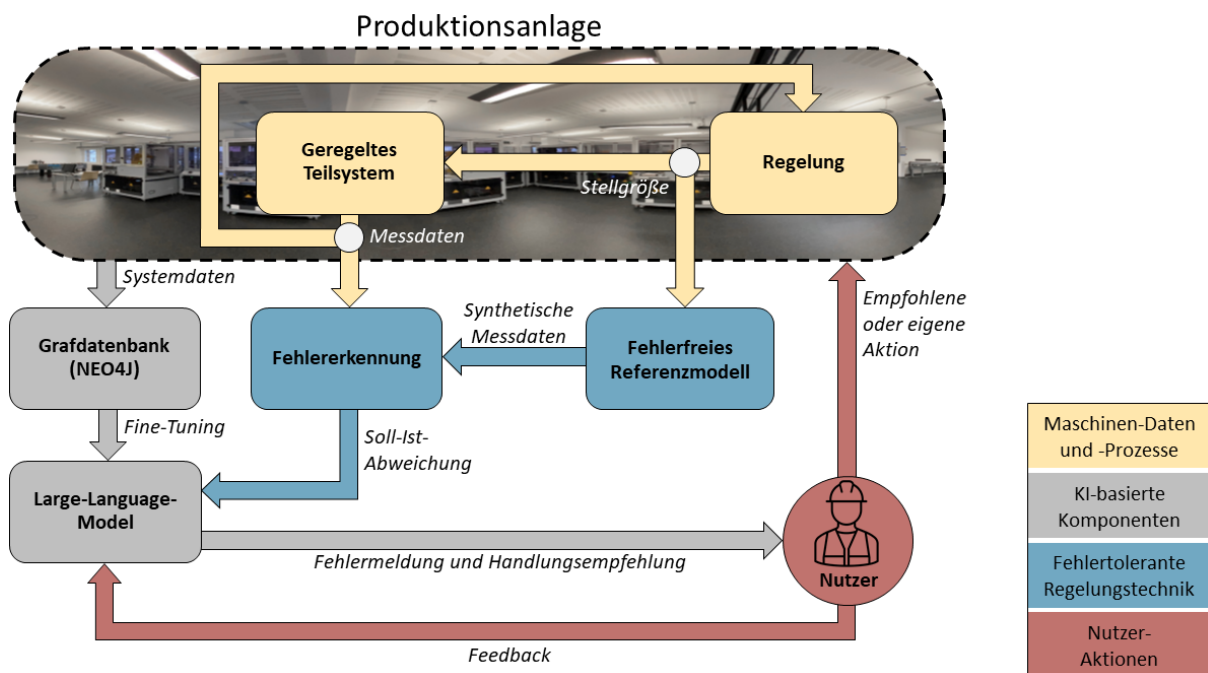


Abbildung 3: Aktuelles Konzept der Prescriptive-Analytics-Plattform

Aufgabenstellung

Vom Studierenden sollen die folgenden Leistungen erbracht werden, wobei je nach persönlichem Interesse und Entwicklung des Projektes individuelle Schwerpunkte gelegt werden können.

Bei allen Aufgaben wird der Studierende sowohl von wissenschaftlichen Mitarbeitern als auch durch den Professor begleitet und nach Bedarf gefördert.

- Einarbeitung in das Thema Prescriptive-Analytics und den aktuellen Projektstand
- Einarbeitung in datenbasierte Methoden der fehlertoleranten Systeme und des maschinellen Lernens

- Auswahl konkreter Lösungsansätze aus den oben genannten Bereichen für die Bearbeitung einer spezifischen Fragestellung im Projekt
- Implementierung und Validierung der gewählten Ansätze (zunächst simulativ, dann real an der IoT-Factory)
- Wahl eines Forschungsschwerpunktes anhand der Erfahrungen aus den zuvor aufgeführten Aufgaben
- Erarbeitung eines eigenen Forschungsbeitrages zum Themenfeld Prescriptive-Analytics
- Dokumentation und Präsentation der erzielten Ergebnisse

Bezug zum Thema Data Science

Der Schwerpunkt des Projekts liegt auf der Verknüpfung von Methoden des maschinellen Lernens und Ansätzen aus dem Bereich der fehlertoleranten Systeme (Regelungstechnik). In beiden Bereichen hat die Generierung, Auswahl, Aufbereitung und Nutzung von Produktionsdaten eine zentrale Rolle. Der Studierende bekommt somit die Möglichkeit, verschiedene Kompetenzen eines Data Scientist im realen Projektumfeld zu erlangen und anzuwenden.

Verfügbare Ressourcen

Im Rahmen des Projektes hat der Studierende die Möglichkeit

- zum regelmäßigen Austausch mit erfahrenen Mitarbeitern des CfADS und des Fraunhofer IEM
- zur Nutzung der IoT-Factory für die Datengenerierung und Validierung der eigenen Arbeit
- zur Nutzung der von der HSBI bereitgestellten Softwaretools und des Data-Analytics-Cluster

Projektplan

Erstes Semester:

Einarbeitung in das Thema Prescriptive-Analytics, den aktuellen Projektstand sowie die Grundlagen des maschinellen Lernens bzw. der fehlertoleranten Systeme. Literaturrecherche zum Stand der Forschung und Technik. Definition einer konkreten Fragestellung aus dem Projektkontext, die im zweiten Semester bearbeitet werden soll. Die Erstellung eines entsprechenden Forschungsexposés ist Prüfungsleistung.

Zweites Semester:

Literaturrecherche zu geeigneten Ansätzen für die Bearbeitung der im ersten Semester definierten Fragestellung. Auswahl konkreter Methoden auf Basis der Literaturrecherche. Beginn mit der Implementierung und Validierung einer der gewählten Methoden. Die Erstellung eines Papers, das einen Überblick über die Schwerpunkte des Semesters gibt, ist Prüfungsleistung.

Drittes Semester:

Fertigstellung und Dokumentation aller Arbeiten zu der im ersten Semester definierten Fragestellung. Auswahl eines Forschungsthemas aus dem Bereich Prescriptive-Analytics, dessen Bearbeitung einen Beitrag zu den Projektzielen leistet. Beginn der Forschungsarbeiten. Die Erstellung eines Papers, welches die im Semester erzielten Ergebnissen zusammen mit dem gewählten Forschungsthema beschreibt, ist Prüfungsleistung.

Viertes Semester:

Fortführung und Abschluss der Forschungsarbeiten sowie Anfertigung der Masterarbeit. Die Masterarbeit und das zugehörige Kolloquium sind Prüfungsleistung.

Eignungskriterien

Zwingend:

- Bachelorabschluss in ingenieurwissenschaftlicher Fachrichtung
- Programmiererfahrung in Matlab oder Python
- Freude sich tief und intensiv in eine technische Fragestellung einzuarbeiten
- Interesse Teil eines wissenschaftlichen Projektteams zu werden

Erwerbbarer Kompetenzen

Je nach gewählten Schwerpunkten hat der Studierende die Möglichkeit, einen Teil der folgenden Kompetenzen zu erwerben.

- Expertenwissen zu Prescriptive-Analytics
- Expertenwissen zu Methoden des maschinellen Lernens
- Expertenwissen zu datenbasierten Methoden fehlertoleranter Systemtechnik
- Zielgerichtete Generierung, Auswahl, Aufbereitung und Nutzung von Produktionsdaten
- Effiziente Validierung eigener Lösungsansätze
- Durchführung einer strukturierten Literaturrecherche
- Dokumentation und Präsentation der eigenen Arbeit in wissenschaftlicher Form
- Anfertigung von wissenschaftlichen Arbeiten mit Fokus auf eine potenzielle Veröffentlichung
- Softskills bezüglich der Arbeitsweise von Forschenden in Projektteams

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitel: 13 Frühzeitige Identifikation von Preisrisiken im globalen Beschaffungsmanagement



Bildquelle: VDMA (Datenquellen: DESTATIS / www.boerse.de)

Projektübersicht

Anzahl Studierende	1
Art	Projekt mit externen Partnern
Projektverantwortung	Prof. Dr. Christian Schwede
Projektkontext	Projekt in Zusammenarbeit mit ARI-Armaturen Albert Richter GmbH & Co. KG in Schloss Holte-Stukenbrock; Abteilung Einkauf / Digital Business Development; in Form einer Werkstudententätigkeit oder im vertraglichen Rahmen eines Studienprojektes der FH Bielefeld

Abstrakt

Im Rahmen dieses Projekts sollen Preisrisiken im globalen Beschaffungsmanagement frühzeitig erkannt werden. Ausgangspunkt hierfür bilden unternehmensinterne Daten, wie Rechnungen, Rahmenverträge und Preisbücher als auch externe Daten, wie bspw. die Börsenpreisentwicklung von bestimmten Rohstoffen. Mittels Methoden des maschinellen Lernens sollen basierend auf diesen Daten Preisforderungen von Lieferanten prognostiziert werden.

Kurzbeschreibung

Durch die zunehmende Volatilität des Beschaffungsmarktumfeldes wird es für Unternehmen immer wichtiger, relevante Veränderungen sowie deren Auswirkungen auf das eigene Unternehmen frühzeitig zu erkennen und entsprechende Handlungsoptionen zu entwickeln. Insbesondere im Bereich der Beschaffungsaktivitäten global-agierender Unternehmen hat sich die Relevanz durch geopolitische Ereignisse, wie die Blockade des Suez-Kanals, kriegerische Auseinandersetzungen oder den Chip-Mangel, nochmal deutlich erhöht: Wiederbeschaffungszeiten und -preise sind erheblich gestiegen.

Ziel des Projekts ist daher die frühzeitige Identifikation von Preisrisiken. Mittels KI-Technologien, wie dem maschinellen Lernen, sollen Preisrisiken, bspw. ausgelöst durch ändernde Wechselkurse und Rohstoffpreise, frühzeitig erkannt werden. Für das betrachtete Unternehmen vergrößert sich so die Handlungszeit und damit die Anzahl der Handlungsoptionen. Dies hat signifikante Auswirkungen auf die Höhe der Materialkosten und damit den Erfolg des Unternehmens.

Aufgabenstellung

Im Rahmen der vorliegenden Aufgabenstellung soll ein Verfahren zur frühzeitigen Identifikation von Preisrisiken entwickelt werden. Konkreter formuliert geht es um die Prognose von Preisforderungen von Lieferanten. Zum einen soll der Forderungszeitpunkt eines Lieferanten als auch die Forderungshöhe prognostiziert werden.

Hierzu steht eine Vielzahl interner als auch externer Daten zur Verfügung. Intern werden bspw. über einen Zeitraum von mehr als fünf Jahren Rechnungen von Lieferanten inklusive der Preisbestandteile, wie Legierungszuschläge und Arbeitspreise, journalisiert. Extern stehen bspw. Daten zu Börsenpreis-Entwicklungen von Legierungselementen und Rohstoffen oder Wechselkursentwicklungen zur Verfügung. Diese Daten können dann bspw. durch Web-Services zum Abgreifen von Informationen aus Portalen oder Schnittstellen zu Portalen in das Modell eingebunden werden.

Durch eine geschickte Kombination dieser Daten soll mittels maschinellen Lernens Preisforderungen prognostiziert werden. Durch die Nutzung der bestehenden Datensätze kann das Modell entsprechend angelernt und erprobt werden.

Bezug zum Thema Data Science

Die Aufbereitung und Analyse von Daten sowie die Entwicklung von Algorithmen zur Mustererkennung insbesondere durch Maschinelles Lernen sind Kernthemen der Data Science und werden teilweise in Veranstaltungen des Forschungsmasters behandelt.

Verfügbare Ressourcen

- Unternehmensinterne Daten, wie oben genannt
- 3 Ansprechpartner aus dem Unternehmen (1 Ansprechpartner aus dem Fachbereich, 1 Ansprechpartner aus der Unternehmens IT und 1 Ansprechpartner Digital Business Development)

Projektplan

Erstes Semester: Erstellung eines Forschungsexposés; Stand der Technik sichten; benötigte Daten ermitteln; vorhandene Daten (intern/extern) sichten; Verfügbarkeit und Eignung der Daten prüfen

Zweites Semester: Erstellung eines Papers zur Übersicht des Standes der Technik im Bereich der Preisprognose; Daten vorbereiten (Check Datenintegrität): Daten-Ausreißer und -Lücken ermitteln

Drittes Semester: Erstellung eines Papers mit ersten quantitativen Ergebnissen; Modell für einfachen Use-Case aufbauen und testen; Modell um weitere Daten(-quellen) erweitern

Viertes Semester: Masterarbeit und Kolloquium; Optimierung des Verfahrens; Auswertung der Güte und Evaluierung der Ergebnisse

Eignungskriterien

Zwingend:

- Begeisterung für das Thema Risikomanagement/global Supply-Chain-Management

Optional:

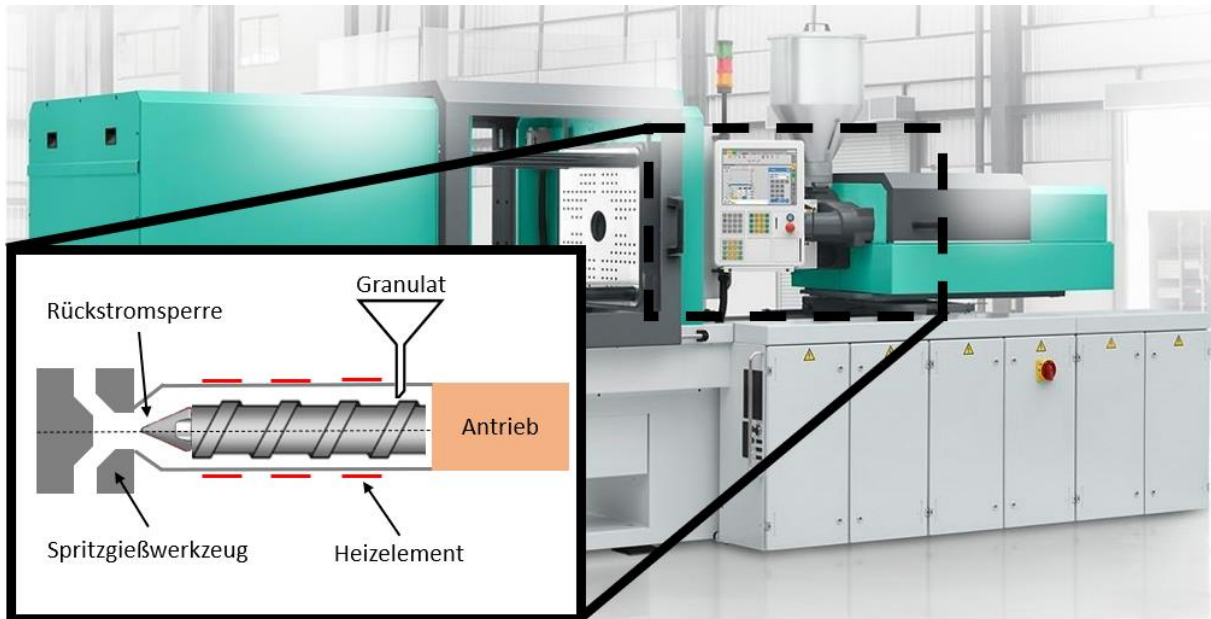
- Erste Erfahrung im Bereich des Maschinellen Lernens

Erwerbbarer Kompetenzen

- Datenvorverarbeitung von strukturierten internen und unstrukturierten öffentlichen Daten
- Einsatz und Implementierung von Verfahren des maschinellen Lernens für hochdimensionale, temporale Probleme
- Herausforderung im Beschaffungsmanagement global-agierender Unternehmen

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitel: 14 Predictive Maintenance in der industriellen Fertigung



Projektübersicht

Anzahl Studierende	1
Art	Projekt mit externem Partner
Projektverantwortung	Prof. Dr. rer. nat. Stefan Berlik, M. Eng. Nico Migenda
Projektkontext	Das Projekt findet in enger Zusammenarbeit mit der HARTING Technologiegruppe und dem Center for Applied Data Science (CfADS) Gütersloh statt. Eine Beschäftigung als wiss. Hilfskraft ist möglich.

Abstrakt

Spritzgussmaschinen werden zur Fertigung von Kunststoffbauteilen verwendet. Die Qualität der Bauteile hängt maßgeblich von der Rückstromsperre ab. Durch die große Belastung (>2500bar) während der Produktion verschleißt diese kontinuierlich, was zu erhöhten Zykluszeiten und Schwankungen in der Produktqualität führt. Mittels maschineller Lernverfahren sollen der Verschleißvorgang bei verschiedenen Produktmaterialien vorhergesagt (Predictive Maintenance) und automatisiert Instandhaltungsaufträge erzeugt werden.

Kurzbeschreibung

Das Spritzgussverfahren ist branchenübergreifend eines der wichtigsten Fertigungsverfahren zur kosteneffizienten Produktion großer Stückzahlen von Bauteilen. Der Ablauf ist schematisch in der obigen Abbildung dargestellt. Der zu spritzende Kunststoff wird in Form von Granulat über einen Trichter in eine rotierende Schnecke eingefüllt. Die Rotation der Schnecke fördert das Granulat zur Schneckenspitze. Durch Heizelemente wird das Granulat auf dem Weg aufgeschmolzen und die Schmelze sammelt sich an der Schneckenspitze, die zu diesem Zeitpunkt geschlossen ist. Hierdurch

bildet sich ein messbarer Staudruck, der in Verbindung mit der Rotation der Schnecke zur Verdichtung und Homogenisierung der Schmelze führt. Sobald sich genügend Schmelze angesammelt hat, die für das Volumen des zu fertigen Werkstücks ausreicht, wird die Rotation der Schnecke eingestellt. Anschließend wird die Schmelze hydraulisch oder elektrisch in den Hohlraum gedrückt. Damit bei diesem Vorgang kein Material zurück fließt und sich die Schmelze homogen im Hohlraum verteilt, ist eine sogenannte Rückstromsperre an der Schneckenspitze eingebaut. Von ihrer Funktion hängt die Qualität des Spritzgussteils maßgeblich ab. Ist eine Rückstromsperre beschädigt, muss der Druck nachgeregelt werden und die Zykluszeit verändert sich. Des Weiteren sind dadurch Qualitätsmängel des produzierten Produktes möglich. Während des Verschleißvorgangs schwankt der gemessene Staudruck zunehmend und ist nicht reproduzierbar. Während des gesamten Produktionsprozesses werden alle Maschinendaten (bspw. Staudruck) aufgenommen und mittels IoT-Protokolle in einer Cloud abgespeichert. Eine Datenhistorie liegt sowohl in Rohform als auch in aufbereiteter Form vor und wird den Mitarbeitern auf einem Live-Dashboard präsentiert. Dies ermöglicht es, Fehler in der laufenden Produktion zu erkennen und darauf zu reagieren (Proaktive Wartung).

Aufgabenstellung

Der / Die Studierende soll im Rahmen des Projekts unter Verwendung von maschinellem Lernen den Gesundheitszustand von Rückstromsperrern in der realen Spritzgussproduktion vorhersagen (Predictive Maintenance). Diese Vorhersagen und daraus abgeleitete Entscheidungsmöglichkeiten sind den Mitarbeitern auf intuitive Weise zu präsentieren, damit ein passgenaues Eingreifen möglich ist. Im ersten Schritt ist sich in die Funktionsweise von Spritzgussmaschinen einzuarbeiten. Anschließend ist die bereits vorhandene Datenerfassung zu studieren, damit ein grundlegendes Verständnis über das zu bearbeitende Projekt vorliegt. Auf dieser Grundlage sind potenziell für die Anwendung passende maschinelle Lernmethoden zu recherchieren. Hierbei ist zwischen Methoden der Detektion und der Vorhersage von Verschleiß/Anomalien zu unterscheiden. Basierend auf der Datengrundlage sind notwendige Datenaufbereitungsschritte durchzuführen (bspw. Merkmalsextraktion), um anschließend ein maschinelles Lernmodell zur Detektion anzutrainieren. Aufbauend sind Vorhersagemodelle zu implementieren. Hier ist zu unterscheiden zwischen verschiedenen Ansätzen (bspw. Überlebensmodell, Verschleißmodell und Ähnlichkeitsmodell). Als Minimalziel soll im Rahmen des Projekts eine Vorhersageform realisiert werden. Bei raschem Projektfortschritt bietet die Aufgabenstellung gleichzeitig eine Vielzahl an Erweiterungsmöglichkeiten. Die Ergebnisse sind abschließend so aufzubereiten, dass jeder Mitarbeiter unabhängig vom Wissensstand im Bereich Data Science in der Lage ist, die Ergebnisse nachzuvollziehen, so dass ein wichtiger Beitrag zur Entscheidungsfindung geliefert wird.

Bezug zum Thema Data Science

Moderne Spritzgussmaschinen weisen vielfältige Sensoren zur Erfassung von Betriebsdaten auf. Standardisierte Kommunikationsprotokolle (bspw. OPC UA und MQTT) erlauben die einfache Integration weiterer Sensorik zur vollumfänglichen digitalen Abbildung des Maschinenbetriebes (digitaler Zwilling). Zur Auswertung der erfassten Daten stehen Methoden des maschinellen Lernens zur Verfügung. Diese lassen automatisiert Rückschlüsse auf die Prozess- und Produktqualität zu, ohne einer aufwendigen Interaktion mit dem menschlichen Anwender. Qualitätsabweichungen können somit frühzeitig identifiziert werden. Sie erlauben des Weiteren die Erzeugung von Prognosen über die Entwicklung der Produktqualität, welche die Grundlage zur Realisierung einer bedarfsgerechten Wartung (Predictive Maintenance) darstellen.

Verfügbare Ressourcen

- Aktive Betreuung bei der Erarbeitung methodischer und praktischer Lösungen

- Es liegt bereits eine große Historie an Prozessdaten vor
- Arbeitsplatz sowohl bei HARTING als auch beim CfADS
- Ansprechpartner bei HARTING und dem CfADS sind über die gesamte Projektlaufzeit verfügbar
- Aktive Unterstützung beim Schreiben wissenschaftlicher Veröffentlichungen

Projektplan

Erstes Semester: Das Ziel des ersten Semesters ist die Einarbeitung in den Workflow von Predictive Maintenance für Spritzgußmaschinen und die dafür zu verwendenden ML-Methoden. Hierfür ist sich in die vorhandene Infrastruktur bei HARTING einzuarbeiten. Termine in der Produktion bei HARTING sind hierfür vorgesehen. Die bereits aufgenommenen Daten sind zu analysieren, damit aus den Kenntnissen von Produktion und Daten ein passgenauer Ansatz zur Anomalie Detektion und für Predictive Maintenance aufgestellt werden kann. Die Erstellung eines Forschungsexposés ist Prüfungsleistung.

Zweites Semester: Basierend auf den Ergebnissen des ersten Semesters soll in diesem Semester eine geeignete maschinelle Lernmethode zur Detektion von Verschleiß und Anomalien implementiert werden. Hierfür sind die vorhandenen Daten speziell für die ausgewählte Methode aufzubereiten. Die Methode ist anschließend bei HARTING in den vorhandenen Workflow zu integrieren und die Funktionalität zu validieren. Die Prüfungsleistung ist die Erstellung eines Short-Papers, wie Predictive Maintenance an Spritzgussanlagen durchgeführt werden kann. Idealerweise wird dieses auf einer (internationalen) Konferenz präsentiert.

Drittes Semester: Aufbauend auf den Ergebnissen der Anomalie Detektion wird in diesem Semester der Workflow weitergeführt in Richtung Predictive Maintenance. Die im ersten Semester ausgewählte Methode wird implementiert mit dem Ziel den Verschleiß und Anomalien während der Produktion vorherzusagen. Auch hier ist die Methodik in den Workflow bei HARTING zu integrieren und zu benchmarken. Das Erstellen eines Full-Papers, in dem die Ergebnisse des zweiten und dritten Semesters kombiniert werden, ist Prüfungsleistung. Idealerweise wird dieses auf einer (internationalen) Konferenz präsentiert.

Viertes Semester: Die Masterarbeit knüpft thematisch an den vorherigen drei Semestern an. Der / Die Studierende kann den Fokus je nach Projektablauf und Interesse mehr auf die reale Anwendung setzen oder mehr theoretisch vertiefen: Beim Fokus auf die industrielle Anwendung sind mögliche Themen das Übertragen der Methoden auf mehrere Spritzgußmaschinen, das Durchführen umfangreicher Benchmarks, das Durchführen von Anwenderstudien mit den Mitarbeitenden hinsichtlich Bedienbarkeit und Funktionalität. Alternativ kann der Fokus mehr auf das Wissenschaftliche gelegt werden. Hierfür ist einer der zuvor implementierten Algorithmen eigenständig weiterzuentwickeln und zu veröffentlichen. Dies könnte den ersten eigenen Schritt in Richtung Promotion darstellen. Unabhängig vom Fokus sollten beide Aspekte einen Mindestanteil von 20% haben. Die Prüfungsleistung ist die Masterarbeit.

Eignungskriterien

Zwingend:

- Fließendes Englisch in Wort und Schrift
- Programmierkenntnisse

Optional:

- Kenntnisse in der Programmiersprache Python
- Erfahrungen im Umgang im Bereich der Datenerfassung/Übertragung

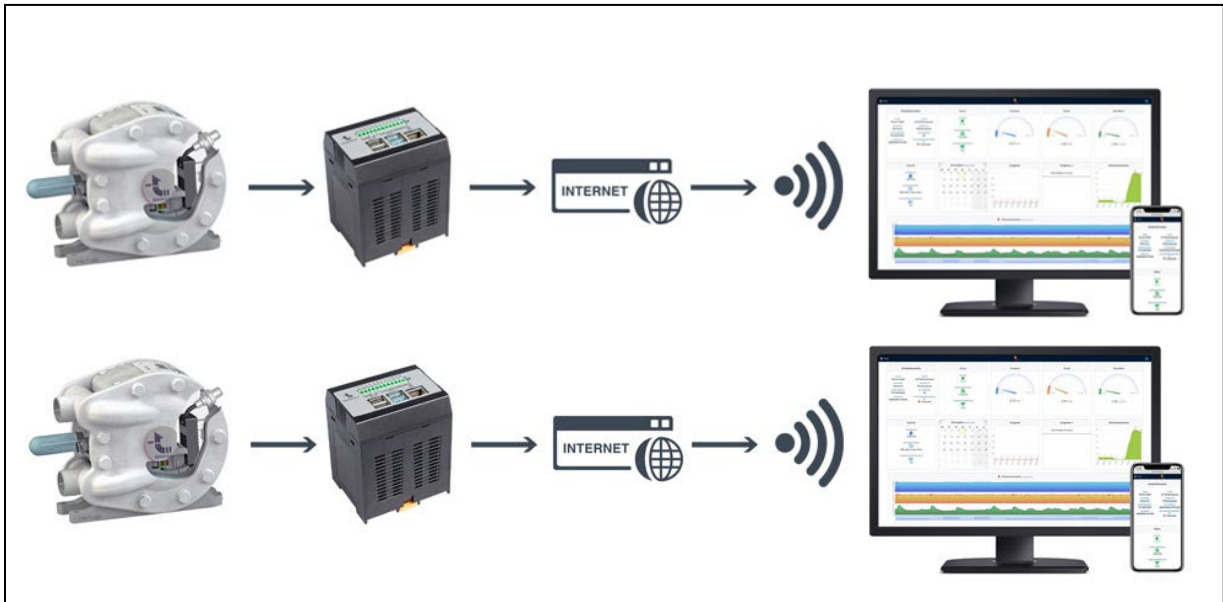
- Praktische Erfahrung mit mechanischen Aufbauten

Erwerbbarere Kompetenzen

- Verknüpfen von Theorie und Praxis: anwenden theoretischer Grundlagen an realen Spritzgussmaschinen
- Durchführung eines eigenen und umfangreichen Projektes
- Wissenschaftliches Arbeiten in einem Forschungsteam
- Teilnahme an Konferenzen mit Präsentation der eigenen Forschungsergebnisse
- Schreiben und Entwickeln von Algorithmen für künstliche Intelligenz
- Testen und Trainieren von künstlichen bzw. neuronalen Modellen
- Auswertung und systematische Analyse von KI-generierten Daten

Forschungsprojekttitel für den Forschungsmaster Data Science

Projekttitel: 15 Verteilte KI-basierte IoT- Zustandsüberwachung zum Störungsmanagement von intelligenten Pumpen Lösungen



Projektübersicht

Anzahl Studierende	1
Art	Projekt mit externem Partner
Projektverantwortung	Prof. Dr.-Ing. Thorsten Jungeblut
Projektkontext	Das beteiligte Unternehmen Timmer GmbH stellt umfangreiche Testdatensätze, sowie einen prototypischen Hardwareaufbau zur Verfügung. Eine Beschäftigung als wissenschaftliche Hilfskraft ist möglich. Die Timmer GmbH ist sehr an der Gewinnung und langfristigen Beschäftigung von Nachwuchskräften interessiert.

Abstrakt

Aufgabe im Projekt ist es, einen prototypischen Aufbau eines verteilten Sensorsystems basierend auf dem tim[®]IOT der Timmer GmbH aufzubauen, und geeignete KI-/ML-Verfahren für die Zustandsüberwachung und die vorausschauende Wartung zu evaluieren. Beispielsweise sollen mehrere mechatronische Membranpumpen über die bei Timmer entwickelten iHZ-Sensoren ausgestattet werden. Die iHZ-Sensoren verfügen über leistungsfähige Mikroprozessoren, die die erfassten Pumpenhübe einer hochgenauen Zeitbasis zuordnen. Zusätzlich können weitere Messgrößen (z.B. Strom/Spannung) aufgenommen werden.

Kurzbeschreibung

Die Entwicklungen im Bereich der intelligenten technischen Systeme (ITS) führen derzeit zu einem Umbruch in der gesamten Wertschöpfungskette der industriellen Produktion. Die zunehmende Leistungsfähigkeit der Informationsverarbeitung bietet für das Anwendungsgebiet der industriellen Bildverarbeitung viele neue Möglichkeiten, an denen Mensch und Maschinen bisher an ihre Grenzen stoßen. Insbesondere die Nutzung Maschinellem Lernverfahren und Methoden der künstlichen Intelligenz versprechen bisher nicht gekannte Möglichkeiten beispielsweise in der Objektklassifikation oder der visuellen Qualitätskontrolle. Etwa seit dem Jahr 2010 werden mit Deep Neuronal Networks (DNN) bzw. Convolutional Neuronal Networks (CNN) bedeutende Fortschritte erzielt.

KI-gestützte **Sensordatenfusion** fasst die Daten einer Vielzahl von Sensoren zusammen, lernt statische gegenseitige Abhängigkeiten der Datenquellen und extrahiert die für den Prozess relevanten Informationen. Die Kombination unterschiedlicher Sensortypen kann dazu beitragen die Genauigkeit in der Auswertung, beispielsweise in der Zustandsüberwachung oder vorausschauenden Wartung, zu erhöhen, oder Mehrdeutigkeiten zu verringern. Sensorfusion kann dazu beitragen, die enormen Datenmengen, beispielsweise in komplexen Produktionsprozessen auf die wesentlichen Informationen zu reduzieren und die darauf aufbauenden Perzeptionsverfahren handhabbarer und effizienter zu gestalten.

Aufgabenstellung

Ziel des Projektes ist die Evaluation von Verfahren künstlichen Intelligenz bzw. maschinelle Lernverfahren für den Einsatz in der Zustandsüberwachung und vorausschauenden Wartung. Anhand der Auswertung hochauflöser und kontinuierlich anfallender Messdaten wie beispielsweise Spannung, Strom oder Ist-Position einer Doppelmembranpumpe soll auf den aktuellen Betriebszustand oder die momentane Leistung zurückgeschlossen werden. Über die zeitliche Analyse der Messdaten lassen sich „normale“ Betriebszustände erkennen und von Anomalien oder schleichenden Änderungen unterscheiden.

Solche Änderungen können auf in naher Zukunft zu erwartenden Schäden hinweisen. Die Auswertung der Messdaten soll sich aber nicht nur auf einzelne Sensoren beschränken, sondern es sollen auch die Daten einer Vielzahl von Sensoren fusionieren werden, um so weitaus genauere Informationen zum aktuellen Systemzustand des gesamten Prozesses zu erhalten. KI-/ML-Verfahren bieten hier neue Möglichkeiten, ggf. auch ohne Annotation durch menschliche Nutzer, Korrelationen zwischen verschiedenen Messgrößen zu erkennen, Anomalien bzw. Fehlerzustände vorherzusagen oder Prozesse effizienter zu organisieren.

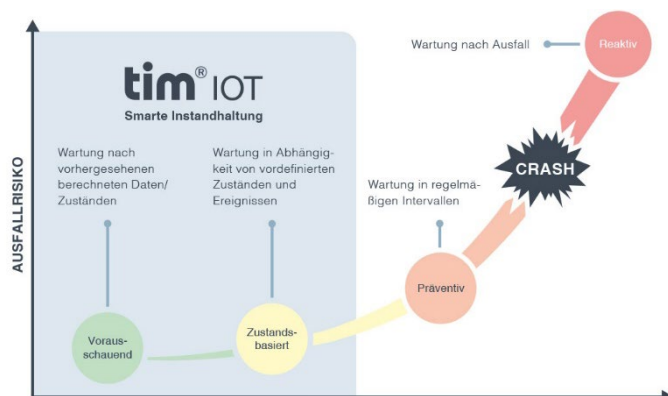


Abbildung 1 - Die rechtzeitige Erkennung von drohenden Ausfällen kritischer Systemkomponenten ermöglicht die den frühzeitigen Austausch und verhindert so kostspielige Ausfallzeiten

Aufgabe im Projekt ist es, einen prototypischen Aufbau eines verteilten Sensorsystems basierend auf dem tim[®]IOT der Timmer GmbH aufzubauen, und geeignete KI-/ML-Verfahren für die Zustandsüberwachung und die vorausschauende Wartung zu evaluieren. Beispielsweise sollen mehrere mechatronische Membranpumpen über die im Hause Timmer entwickelten iHZ-Sensoren ausgestattet werden. Die iHZ-Sensoren verfügen über leistungsfähige Mikroprozessoren, die die erfassten Pumpenhübe einer hochgenauen Zeitbasis zuordnen. Zusätzlich können weitere Messgrößen, wie z.B. Strom/Spannung aufgenommen werden.

Bezug zum Thema Data Science

Die Evaluation und Anwendung von KI-/ML-Verfahren zur Zustandsüberwachung und Vorausschauenden Wartung sind ein Kernthema der Data Science und werden beispielsweise in den Modulen „Data Mining & Machine Learning“ sowie „Künstliche Intelligenz“ behandelt. Die Aufnahme hochauflöster Messdaten von einer Vielzahl an Sensoren in komplexen Prozessen stellt hohe Anforderungen an Organisation und Verarbeitung der Daten. Dieses ist Kern des Moduls „Big Data Architekturen“. Die Betrachtung des gesamten Systemprozesses vom Sensor bis zur Cloud erfordert eine ganzheitliche Betrachtung des vollständigen Data Science Prozesses, welcher im Modul „Data Science“ behandelt wird.

Verfügbare Ressourcen

- Informationen, die zur Erstellung des Szenarios benötigt werden (Systembeschreibung, Schnittstellen, Dokumentationen, relevante Kennzahlen etc.) werden von der Timmer GmbH bereitgestellt
- Die Timmer GmbH stellt umfangreiche Testdatensätze zum Training und zur Evaluation zur Verfügung
- Der Ansprechpartner der Timmer GmbH wird über die Projektlaufzeit zur Verfügung stehen
- Die für den prototypischen Aufbau benötigten Komponenten (tim[®]IOT) sowie sonstiges benötigtes Material wird von der Timmer GmbH bereitgestellt
- Hardware für das komplexere Maschine Learning ist über das Data Science Lab, das CfADS sowie dem KI-Rechencluster yourAI der FH Bielefeld verfügbar

Projektplan

Erstes Semester: Erstellung eines Forschungsexposés als Prüfungsleistung. Einarbeitung in die das Konzept von tim[®]IOT, die Schnittstellen der zur Verfügung gestellten intelligenten Sensoren, der tim[®]IOT-Smartbox sowie die tim[®]IOT-Cloud-Plattform.

Zweites Semester: Erstellung des Systemkonzepts zur Zustandsüberwachung und prädiktiven Wartung von intelligenten IoT-Geräten. Recherche zu relevanten Arbeiten im Themenfeld des Einsatzes von KI-/ML-Verfahren für die Sensordatenverarbeitung im o.g. Kontext. Erstellung eines Papers, das einen Überblick über das jeweilige Forschungsgebiet gibt, als Prüfungsleistung.

Drittes Semester: Entwicklung eines ersten Demonstrators basierend auf dem tim[®]IOT-System und einem einzelnen intelligenten Sensor. Implementierung und Vergleich unterschiedlicher KI-/ML-Ansätze zur Zustandsüberwachung. Visualisierung des Systemzustandes.

Viertes Semester: Masterarbeit und Kolloquium. Erweiterung des Systems auf mehrere intelligente Sensoren und Fusion der Sensordaten. Evaluierung KI-/ML-basierter Verfahren für die Zustandsüberwachung und vorausschauende Wartung. Finale Evaluierung durch Vergleich der implementierten Strategien. Erstellung eines Papers mit ersten quantitativen Ergebnissen als Prüfungsleistung.

Eignungskriterien

Zwingend:

- Programmierkenntnisse

Optional:

- Erfahrung in der Elektronikentwicklung

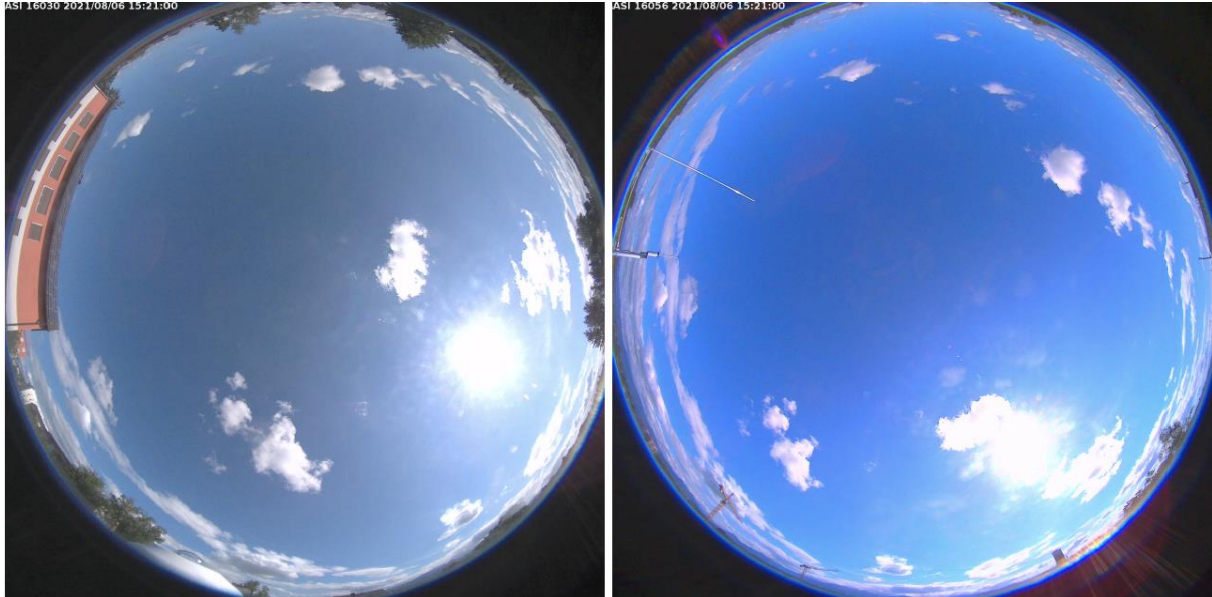
- Erfahrung mit IoT-Geräten
- Programmierung von Mikrocontrollern

Erwerbbarere Kompetenzen

- Ressourceneffiziente Informationsverarbeitung auf den verschiedenen Ebenen (Edge, Fog, Cloud) im Sinne des IoT-Verarbeitungskonzepts
- Sensornahere Informationsverarbeitung
- KI-/ML-Verfahren
- Predictive Maintenance

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: 16 Vorhersage der Sonneneinstrahlung aus Wolkenbildern zur verbesserten Integration von Solarenergie im Stromnetz



Projektübersicht

Anzahl Studierende	1
Art	DBU - gefördertes Projekt Helios mit externen Partnern (Laufzeit vorr. 01.2024-12.2026)
Projektverantwortung	Prof. Dr.-Ing. Grit Behrens
Projektkontext	Das Teilprojekt Data-Mining für Wolkenerkennung- und -klassifikation für Energieertragsprognosen ist im SolarComputingLab am Campus Minden angesiedelt und wird von Prof. Behrens geleitet. Das Team besteht weiterhin aus einem wissenschaftlichen Mitarbeiter beim Projektpartner im Labor für Solare Energiesysteme an der TH Rosenheim unter Leitung von Prof. Mike Zehner. Es ist eine Anstellung als WHK mit bis zu 17Std./Woche möglich (unter Vorbehalt der endgültigen Mittelfreigabe)

Abstrakt

In diesem Projekt werden All-Sky-Images unter Anwendung u.a. von Convolutional Neural Network und Deep-Learning Verfahren analysiert, um die solare Einstrahlung auf Photovoltaikfelder und die zu erwartenden Erträge prognostizieren zu können.

Kurzbeschreibung

Die Schwankungen und Überhöhungen der Solarstrahlung werden maßgeblich durch Wolkenzüge beeinflusst. Diese Volatilität hat einen erheblichen Einfluss auf die Leistung von Photovoltaik (PV)-Anlagen. Um den Eintrag von PV-Anlagen ins Stromnetz möglichst präzise vorhersagen zu können, ist eine zeitlich und räumlich hochauflösende Vorhersage der Solarstrahlung erforderlich. Hierbei spielen All Sky Imager (ASI) eine wichtige Rolle. Dabei ist es insbesondere von Bedeutung, Wolkenformationen in ASI-Images mehrerer in großen PV-Feldern verteilter Kameras zu erkennen und verfolgen zu können, und sie nach verschiedenen Kriterien zu bewerten und zu klassifizieren.

Aufgabenstellung

Der Studierende soll in ASI-Images mit Deep-Learning Verfahren (u.a. CNN) Wolken segmentieren, verfolgen, unter Anwendung von Clustering-Verfahren (z.B. kmeans) klassifizieren und auf der Grundlage einer zeitlich und räumlich hochaufgelösten Vorhersage der solaren Einstrahlung und eines Digital Twins des PV-Feldes unter Anwendung von RNN (ggf. LSTM, BiLSTM) die Energieerträge für die nächsten 5-15 Minuten möglichst präzise prognostizieren.

Bezug zum Thema Data Science

Verfahren des Deep-Learning, des Clustering sowie der zeitreihenbasierten Prognose und Simulation mit Digital Twins sind Kernthemen der Data Science und werden in den Veranstaltungen des Forschungsmasters behandelt.

Verfügbare Ressourcen

Im Rosenheimer Technologiezentrums Energie und Gebäude sind zwei Wolkenkameras (ASI-16/51) und meteorologische Sensoren installiert, die seit August 2021 Bilder und Daten aufzeichnen. Die verfügbare Datengrundlage wird auch weiterhin kontinuierlich alle 20 Sekunden mit parallel jeweils einem pro Bild pro Kamera erweitert. Rechnerressourcen stehen im Solar Computing Lab in Form eines Machine Learning Servers zur Verfügung.

Projektplan

Erstes Semester: Sammeln aller Datenquellen aus Sensoren, verfügbaren API's (z.B. open meteo) und der ASI-Images. Konsolidierung des Datensatzes, Datenanalyse. Detektion von Wolken und Verfolgung von Wolkenobjekten in mehreren ASI-Kameras. Erstellung des Forschungsexposés. Optional Erstellung und Einreichung eines Papers für eine „Work-in progress abgabe“ auf einem nationalen Workshop (z.B. im AK de GI AI & Sustainability).

Zweites Semester: Optimierung der Erkennungsleistung zur Wolkendetektion. Erstellen von Prototypen für Wolkenklassen. Erstellen eines ersten gelabten Lerndatensatzes für Wolkentypen, erste Klassifikationsergebnisse. Erstellung und Einreichung eines Papers auf einer Nationalen Konferenz (z.B. Photovoltaisches Symposium). das einen Überblick über das jeweilige Forschungsgebiet gibt ist Prüfungsleistung.

Drittes Semester: Konsolidierung des Datensatzes mit gemessenen Ertragsdaten und simulierten Ertragsdaten aus dem Digital Twin der PV-Anlage (Nutzen von Daten der Projektpartner. Erstellen von Prognose der Solaren Einstrahlung und von Prognosen der PV-Erträgen. Erstellen und Einreichen eines Papers mit Vortrag auf einer internationalen Konferenz (z.B. EUPVSEC oder ENVIROINFO).

Viertes Semester: Masterarbeit und Kolloquium mit an den im Verbundprojekt erreichten Forschungsstand Forschungsfragen.

Eignungskriterien

Zwingend: Programmierkenntnisse

Optional: Kenntnisse in Photovoltaik und Energiemeteorologie.

Erwerbbare Kompetenzen

Einsatz von State-of the Art Technologien des maschinellen Lernens auf Bild- und Sensordaten im technischen Anwendungsfeld. Spezialkenntnisse im Bereich der Energiemeteorologie, der Photovoltaik und des SolarComputing. Selbständiges wissenschaftliches Arbeiten im Verbundprojekt mit Firmen und Forschungslaboren.

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: 17 Predictive Health



Projektübersicht

Startsemester (von-bis)	WiSe 22/23 – SoSe 24
Anzahl Studierende	1
Art	Studienprojekt (interdisziplinär, fachbereichsübergreifend, praxisnahe Anbindung an ein Reallabor)
Projektverantwortung	Prof. Dr. W. Schenck, Dr. Christoph Ostrau Beraterin: Prof. Dr. Annette Nauerth
Projektkontext	Das Projekt findet im Rahmen des TransCareTech und dessen Reallabor Geriatrie statt. Außerdem ist das Vorhaben an das Center for Applied Data Science (CfADS) Gütersloh und den Fachbereich Gesundheit angebunden.

Abstrakt

In der ambulanten Pflege älterer Patient*innen herrscht akuter Personalmangel mit entsprechender Auswirkung auf die Qualität der Versorgung. Der effiziente Einsatz von Pflegekräften könnte diesen Fachkräftemangel etwas abmildern. Das Reallabor Geriatrie im TransCareTech bietet die Möglichkeit, echte Daten von Patient*innen zu erheben. In diesem Projekt werden Methoden der Data

Augmentation und der Predictive Maintenance genutzt, um Pflegebedarfe vorherzusagen und den Gesundheitszustand unter Positive Health Aspekten zu bewerten.

Kurzbeschreibung

Die Pflegesituation von **geriatrischen Patient*innen** als auch pflegebedürftigen älteren Menschen hat sich in den letzten Jahren verschlechtert. Dabei spielen akuter Fachkräftemangel und finanzieller Druck die Hauptrollen. Aufgrund von fehlender technischer Ausstattung und Standardisierung bleiben Chancen und Möglichkeiten der Digitalisierung oftmals ungenutzt. Im **TransCareTech** wollen wir uns genau diesen Themen widmen, um so langfristig die Situation pflegebedürftiger Mitmenschen zu verbessern.

Das vorgeschlagene Projekt ist eng verzahnt mit dem **Reallabor Geriatrie**, das sich gegenwärtig noch im Aufbau befindet. Dort werden Patient*innen in einer Beispielwohnung nach einer stationären Behandlung geriatrisch rehabilitiert. Neben der technischen Ausstattung der Wohn- und Behandlungsräume, die die Aufnahme von Vital- und Bewegungsdaten ermöglicht, werden die Probanden regelmäßig zu diversen Aspekten befragt. Ein Ziel dieser Arbeit ist die Berechnung eines **Gesundheits-Scores**, der sich nicht nur auf die körperliche Gesundheit bezieht, sondern auch Kategorien wie die mentale Gesundheit, die Lebensqualität und die soziale Einbindung (Stichwort **Positive Health**) mit in den Blick nimmt. Auf Basis der vorliegenden Datenquellen soll bewertet werden, inwieweit ein solcher Score sinnvoll gebildet werden kann. Eine weitere Aufgabe ist es, den sinnvollen Einsatz von Pflegepersonal durch die **Vorhersage von Pflegebedarfen und Pflegenotfällen** zu evaluieren. Hierbei werden KI-Methoden der „**Predictive Maintenance**“ auf vorliegende Fälle übertragen. Da ein solches Vorhaben nur unter der Nutzung großer Datensätze möglich wird, muss dafür zunächst ein **Data Augmentation-Ansatz** verfolgt werden, um den zur Projektzeit vorliegenden Datensatz entsprechend zu vergrößern. Diese Arbeit ist damit eng an die noch aufzubauende **Care-Dateninfrastruktur** angebunden und bildet einen essenziellen Baustein ebendieser. Eine Rückkommunikation der Scores und deren zeitlichen Verlauf an die Probanden und die behandelnden Fachkräfte verbessert das Selbstmanagement der Patient*innen. Die Einbeziehung in die Aufbauarbeiten ermöglicht eine Teilhabe der Patient*innen an dem Forschungsprojekt.

Aufgabenstellung

Das Projekt besteht aus zwei größeren Zielen: Zum einen dient ein zusammenfassender Score des allgemeinen Gesundheitszustands als Indikator für den Bedarf und Optimierungserfolg ambulanter Pflege. Zum anderen ermöglicht die Datenerweiterung in Bezug auf das Reallabor die anschließende Anwendung maschineller Lernverfahren zur Vorhersage von Pflegebedarfen/-notfällen. Außerdem wird zum Aufbau der Care-Dateninfrastruktur beigetragen. Daher umfasst das Aufgabenspektrum die folgenden Bereiche:

- Datenvorverarbeitung und Katalogisierung der Daten.
- Evaluation der Datenquellen als Grundlage für die Bildung eines Gesundheits-Scores zur Bewertung der Gesamtsituation einer Proband*in unter Positive Health Aspekten.
- Erweiterung des vorhandenen Datensatzes zur Vorhersage von Pflegebedarfen und -notfällen.
- Umsetzung einer KI zur Vorhersage von Pflegebedarfen

Bezug zum Thema Data Science

Die Arbeit ist eng an den Aufbau einer Dateninfrastruktur angebunden. Daher werden die Teilbereiche der Vorverarbeitung und der Archivierung/Strukturierung von Datensätzen behandelt. Bei der Datenverarbeitung spielt die Bewertung der Datenquellen zur Bestimmung der Gesundheit eine große Rolle. Um maschinelle Lernverfahren, die auf Big Data angewiesen sind, im Reallabor anwenden zu

können, steht Data Augmentation im Mittelpunkt. Zuletzt werden Lernverfahren aus der Predictive Maintenance auf die erhaltenen Datensätze angewandt.

Verfügbare Ressourcen

Für die rechenintensiven Anteile des Projekts stehen die Infrastruktur des CfADS, insbesondere des Data-Analytics-Cluster, zur Verfügung. Die Datengrundlagen bilden Erhebungen im Reallabor Geriatrie des TransCareTech Forschungsprojektes.

Projektplan

Erstes Semester:

Einarbeitung in die Infrastruktur des CfADS, Analyse der im Zusammenhang mit dem Reallabor Geriatrie anfallenden Daten, Literaturrecherche zur Bewertung von ganzheitlichen Gesundheitsaspekten, insbesondere des Positive Health Ansatzes, und zum Thema „Predictive Maintenance“, Überblick über domänenspezifische Datensätze.

Zweites Semester:

Vorverarbeitung der Daten des Reallabors und Einbindung in die Care-Dateninfrastruktur, Einarbeitung in das Thema Data Augmentation.

Drittes Semester:

Augmentieren der Daten des Reallabors, Herausstellung entscheidender Charakteristika von Pflegebedarfen und Generierung synthetischer Daten, Berechnung eines wochengenauen Gesundheits-Scores.

Viertes Semester:

Umsetzung der Vorhersagen (hinsichtlich Versorgungsbedarfe/akuter -notfälle) und Evaluation in dem Reallabor Geriatrie, Vergleich mit dem State-Of-The-Art und Potentialanalyse.

Eignungskriterien

Zwingend:

- Programmierkenntnisse in mindestens einer objektorientierten Sprache
- Englischkenntnisse zur Lektüre englischsprachiger Veröffentlichungen
- Teamfähigkeit

Optional:

- Programmierkenntnisse in Python
- Grundkenntnisse in Hadoop
- Grundkenntnisse in HBase

Erwerbbarer Kompetenzen

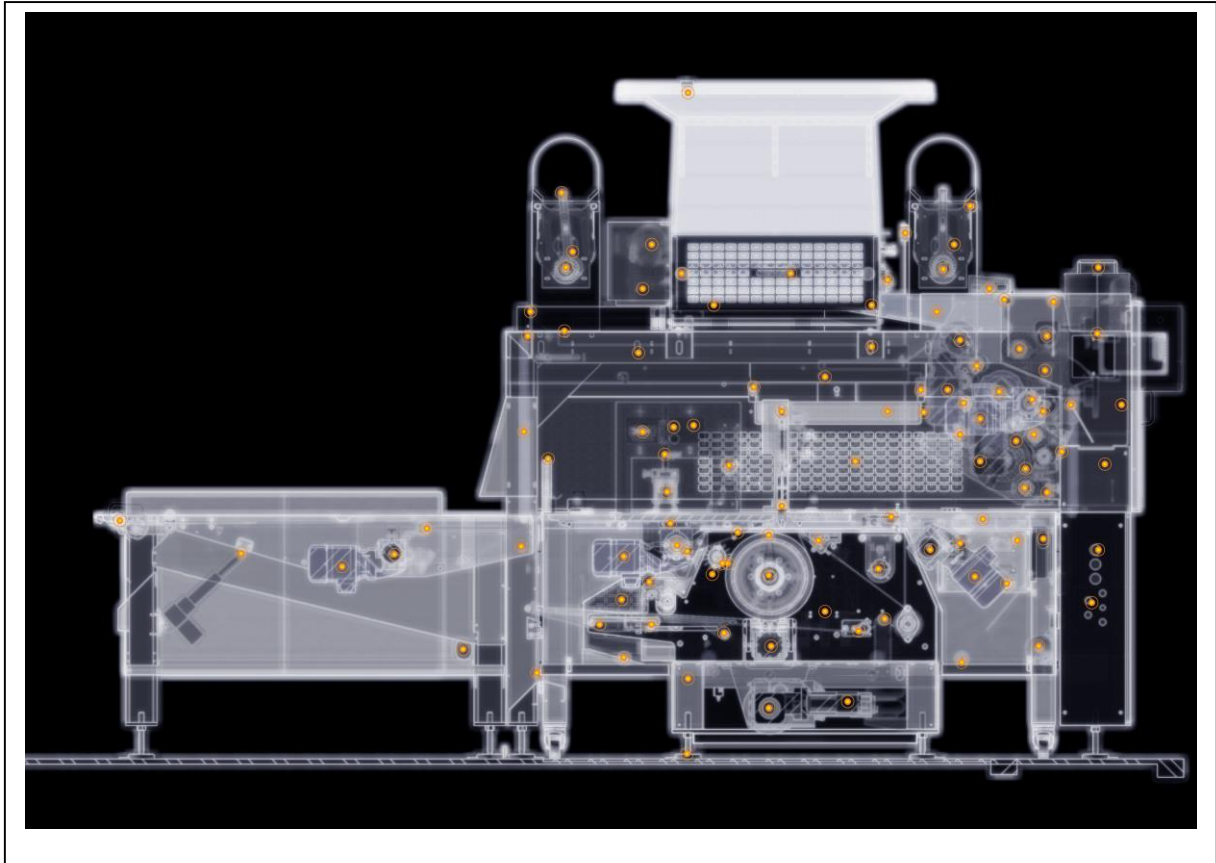
Der/die Studierende ist nach Abschluss des Projekts in der Lage,

- ML-basierte Verfahren auf großen Datensätzen anzuwenden.
- Datensätze künstlich zu erweitern, um Big Data Methoden in Bereichen mit Sparse Data zu benutzen.
- ML-Verfahren auf sequenziellen Datensätzen anzuwenden.
- wissenschaftliche Publikationen zu verfassen.

- die eigenen Ergebnisse vor einem Fachpublikum zu präsentieren.
- im Team zusammenzuarbeiten und fachbereichsübergreifend zu kommunizieren.

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: 18 KI-gestützte Qualitätsoptimierung in der Kleingebäcksproduktion



Projektübersicht

Anzahl Studierende	1
Art	Projekt mit externen Partnern
Projektverantwortung	Prof. Dr. Christian Schwede Dr.-Ing. Viktor Just (WP Digital) Dr.-Ing. Felix Oestersötebier (WP Kemper)
Projektkontext	Projekt in Zusammenarbeit mit WP Digital GmbH und WP Kemper GmbH in Rietberg; Eine Anstellung in Form einer Werkstudententätigkeit ist möglich.

Abstrakt

Das Thema „Erfassung, Dokumentation und Optimierung der Produktqualität“ erfährt in der Bäckereibranche zunehmende Bedeutung. Ziel ist es, die Qualität der Ausbringung zu erhöhen und den Ausschuss zu minimieren. In diesem Projekt soll eine Qualitätsbeurteilung der Teiglinge mittels KI-Objekterkennung installiert werden. Dann sollen alle Informationen aus Sensorik und Prozessparametern zu virtuellen Teigling-Objekten aggregiert werden. Diese Datenobjekte werden anschließend genutzt, um eine übergeordnete KI zur Erkennung von Ursache-Wirkungs-Zusammenhängen zu trainieren, die zur Optimierung der Produktions-Qualität dient.

Kurzbeschreibung

Im Zuge der zunehmenden Digitalisierung in der Bäckereibranche gewinnt das Thema automatische Erfassung, Dokumentation und Optimierung der produzierten Produktqualität – gerade in großhandwerklichen und industriellen Betrieben – mehr und mehr an Bedeutung. Übergeordnetes Ziel ist es dabei u.a., die Qualität der Ausbringung zu erhöhen und den Ausschuss zu minimieren. Als eigenständige Gruppe bündelt WP Digital die Aufgaben rund um die Digitalisierung innerhalb der WP Bakery Group und treibt diesen Bereich u.a. bereits durch die Produkte WP DIGITAL PORTAL, WP CONNECT und WP BAKERYCONTROL erfolgreich voran.

In diesem Projekt sollen in vier wichtigen Schritten folgende Aufgaben in Angriff genommen werden:

1. Automatische Erkennung der **Produktqualität**
2. Punktgenaue **Daten-Aggregation** aller Informationen aus Sensoren, Prozesseinstellungen, Qualitätsbewertung und Teig-Eigenschaften auf die einzelnen Teiglinge
3. Optional: Auffinden der **Ursache-Wirkungszusammenhänge** bei Qualitätsabweichungen.
4. **Optional: Optimierung der Produktqualität** durch Hinweise an den Maschinenbediener

Auf Basis dieser gefundenen Ursache-Wirkungszusammenhänge soll der Maschinenbediener in Zukunft bei auftretender Qualitätsabweichung Hinweise zur Anpassung der Prozess-Parameter der Maschine oder Hinweise zur Verbesserung der Teigqualität bekommen. Sobald diese Hinweise sich in der Praxis als verlässlich und genau genug erwiesen haben, soll die Anlage automatisiert die Prozessparameter dahingehend optimieren, dass kontinuierlich eine gleichbleibend hohe Produktqualität erreicht wird.

Aufgabenstellung

Die Erfassung von aktuellen Prozess- und Qualitätsdaten wird heute auch in der Bäckerei immer mehr zum Standard. Diese Daten sollen in diesem Projekt zur Identifikation und Quantifizierung von bisher nicht bekannten bzw. wenig untersuchten Ursache-Wirkung-Zusammenhängen innerhalb des Herstellungsprozesses von Kleingebäck genutzt werden. Hierzu sollen geeignete Methoden der künstlichen Intelligenz identifiziert und genutzt werden.

In diesem Projekt sollen in vier wichtigen Schritten folgende Aufgaben in Angriff genommen werden:

1. Automatische Erkennung der Produktqualität

Zusätzlich zu den bereits zur Verfügung stehenden Prozessdaten soll ein optisches System zur Erkennung von Teiglingen und deren Eigenschaften entworfen werden. Dieses soll die Lage der Teiglinge sowie Form und Größe optisch analysieren können und spezifische künstliche neuronale Netze (CNNs) nutzen. Dabei erscheint z.B. die Nutzung von vorhandener Open Source Software in Python (Tensorflow, Roboflow, etc.) und deren Modelle vielversprechend. Die Intention ist es, genauere Aussagen über die Produktqualität treffen zu können. Dazu sollen Informationen über den Status der Teiglinge anhand von zu definierenden Kennzahlen erhoben werden können. Die Produkterfassung und Qualitätsbeurteilung sollen in Form eines Moduls erfolgen, welches nach Bedarf einer Anlage hinzugefügt werden kann. Durch den Aufbau bzw. Training eines Modells zur Erkennung der Teiglinge sowie durch entsprechende, zu definierende Tests wird die prinzipielle Funktionsfähigkeit des konzipierten Moduls nachgewiesen. Weiterhin sollen Kennziffern zur Beurteilung der Teiglinge anhand der optischen Auswertung erarbeitet werden.

2. Punktgenaue Daten-Aggregation aller Informationen auf die einzelnen Teiglinge

Die Sensordaten, Prozessparameter und Stellgrößen, die während der Teigling-Herstellung angefangen von der Knetung bis hin Schneiden und Rundwirken entstehen, werden durch WP

CONNECT in einer Zeitreihen-Datenbank (InfluxDB) im Sekundentakt abgelegt. Auch nach jeder Knetung wertet das Tool MIXCONTROL automatisiert den Knetvorgang aus und erkennt, ob ein Teig zu Ende geknetet, überknetet oder überhitzt wurde. Die Ergebnisse der Knetung werden mit Zeitstempel am Kneiter abgespeichert.

Ziel ist es nun, dass alle diese Daten zusammen mit den neu hinzugekommenen Qualitätsmerkmalen der Teiglinge (s. Aufgabe 1) punktgenau auf die jeweiligen Teiglinge aggregiert werden und zu jedem Teigling ein virtuelles hierarchisches Daten-Objekt erzeugt wird. Das virtuelle Teigling-Objekt enthält dabei hierarchisch gegliedert nach den durchlaufenen Einflussbereichen nur die Daten, die auf genau diesen Teigling eingewirkt haben.

Um das zu erreichen, muss zunächst jeder Teigling durch seinen Produktionsprozess verfolgt werden. Dazu soll ein modellbasierter Teigfluss-Beobachter aufgebaut werden, der den Teigtransport auf Basis der vorhandenen Stell- und Messsignale simuliert. Das Beobachtermodell soll dann anhand vorhandener Lichtschranken fortlaufend eine Korrektur-Rückführung erhalten.

3. Optional: Auffinden der Ursache-Wirkungszusammenhänge bei Qualitätsabweichungen.

Die virtuellen Teigling-Objekte (s. Aufgabe 2) enthalten nun alle relevanten Informationen, die diesen Teigling betreffen und sollen als Datenbasis zum Training der Ursachen-Wirkungszusammenhänge dienen. Dazu soll eine KI aufgebaut werden, die auf Merkmale von schlechter Produktqualität (s. Aufgabe 1) reagiert und das Teigling-Objekt nach Mustern durchsucht, die mit diesem Qualitätsmerkmal in Zusammenhang stehen.

Zum Training und zur Absicherung der Ursachen-Wirkungszusammenhänge ist es wichtig, eine breite Datenbasis aus einer realen Produktion zu erhalten. Dazu soll die Qualitätserkennung aus Aufgabe 1 und die automatische Datenaggregation aus Aufgabe 2 bei einem Test-Kunden installiert werden, um laufend frische Trainingsdaten zu erhalten. Die Daten mit spezifischen Qualitätsmerkmalen sollen automatisch abgespeichert und der KI zum Training zugeführt werden.

4. Optional: Optimierung der Produktqualität

Auf Basis dieser gefundenen Ursache-Wirkungszusammenhänge soll der Maschinenbediener in Zukunft bei auftretender Qualitätsabweichung Hinweise zur Anpassung der Prozess-Parameter der Maschine oder Hinweise zur Verbesserung der Teigqualität bekommen. Dazu in dem vorhandenen Tool WP CONNECT eine Hinweis-Liste implementiert werden, die auftretende Qualitäts-Abweichungen dokumentiert und mit konkreten Verbesserungs-Hinweisen anreichert.

Sobald diese Hinweise sich in der Praxis als verlässlich und genau genug erwiesen haben, soll die Anlage automatisiert die Prozessparameter dahingehend optimieren, dass kontinuierlich eine gleichbleibend hohe Produktqualität erreicht wird. Dazu soll ein Konzept erstellt werden, dass die automatisierten Parameteroptimierung zur Sicherstellung einer gleichbleibenden Produktqualität vorbereitet.

Bezug zum Thema Data Science

Bilderkennung und-verarbeitung mit CNNs sowie die Erstellung von Digitalen Abbildern von Produkten und datenbasiertes Aufdecken von Ursache-Wirkungszusammenhängen sind Kernthemen der Data Science.

Verfügbare Ressourcen

- Teigband-Anlage „PANE“ im Backzentrum bei WP Kemper in Rietberg zur Aufnahme von Messdaten

- Anschaffung einer Industriekamera und benötigter Hardware (Jetson Xavier von NVIDIA) Objekterkennung der Teigling-Qualität
- WP CONNECT-PC mit integrierter Datenbank zur Aufnahme und Verarbeitung von Messdaten
- Unterstützung mittels benötigter Informationen sowie durch Maschinen-Entwickler und Backmeister

Sicherstellung der Verfügbarkeit von Daten, Rechenressourcen, Hardware, Anwendungsexperten

Projektplan

Neben der Mitarbeit an der Erforschung und Entwicklung der Verfahren sind folgende wissenschaftliche Ausarbeitungen als Prüfungsleistungen vorausgesetzt:

Erstes Semester:

- Erstellung eines Forschungsexposés zur Detailplanung der Arbeiten.
- Auswahl und Einrichtung Training einer Industriekamera Erzeugung von Datensätzen.

Zweites Semester:

- Ermittlung des Standes der Technik auf dem Gebiet der Qualitätsmessung durch Bildanalyse
- Auswahl und Implementierung von Verfahren. Gegenüberstellung anhand des Datensatzes
- Erstellung eines Technical Reports zum Stand der Technik

Drittes Semester:

- Verfeinerung der Analyse und Adaption der Methoden, um die Ergebnisqualität zu erhöhen
- Erstellung eines Papers zur KI-basierten Qualitätsmessung von Teiglingen.

Viertes Semester:

- Aufbau eines modellbasierten Teigfluss-Beobachters für das Tracking und Tracing der Teiglinge
- Datenselektion und Aggregation aller Informationen aus Sensoren, Prozesseinstellungen, Qualitätsbewertung, und Teigeigenschaften zu einem virtuellen Teigling-Objekt.
- Masterarbeit und Kolloquium

Eignungskriterien

Zwingend:

- Programmierkenntnisse
- Erfahrung mit SPS-Steuerungen
- Teamfähigkeit und Interesse an wissenschaftlicher Arbeit
- Selbstständige Arbeitsweise

Optional:

- Simulationserfahrung
- Prozess-Kenntnisse der Kleingebäcksproduktion
- Grundlagenkenntnisse in Objekterkennung und KI-Einsatz

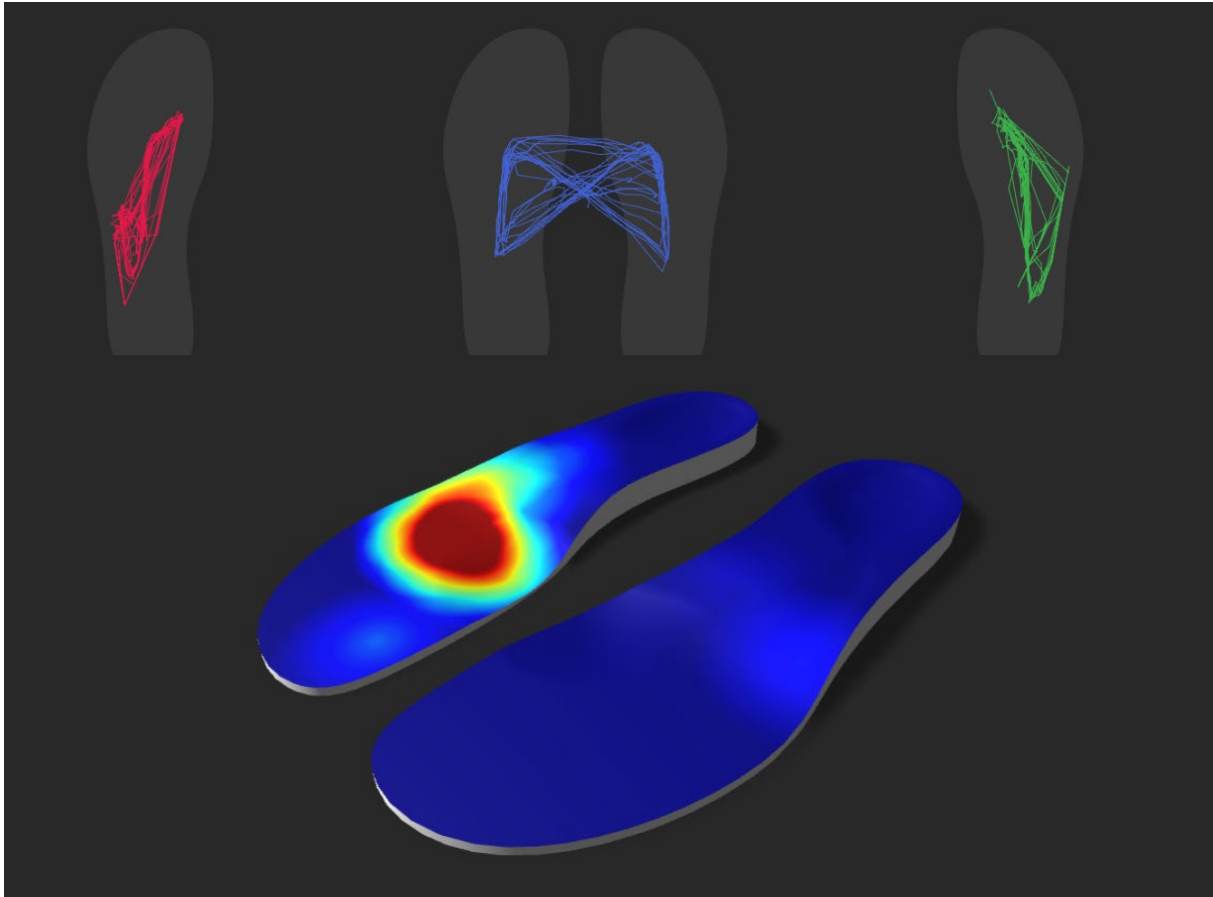
Erwerbbarer Kompetenzen

- Einsatz von visueller, KI-basierter Objekterkennung
- Einsatz eines modellbasierten Teigfluss-Beobachters
- Umgang mit Big-Data: Selektion und Aggregation

- Einsatz neuronaler Netze zum Training von Ursache-Wirkungs-Zusammenhängen
- Anwendung von Ursache-Wirkungs-Zusammenhängen zur Prozessoptimierung

Forschungsprojekttantrag für den Forschungsmaster Data Science

Projekttitel: 19 Entwicklung einer KI zur Fußdruck- und Ganganalyse für Patient*innen mit diabetischem Fußsyndrom.



Projektübersicht

Anzahl Studierende	1-2
Art	Gefördertes Projekt
Projektverantwortung	Prof. Dr.-Ing. Martin Kohlhase, M. Eng. Jan Finke
Projektkontext	Projekt in Zusammenarbeit mit dem Center for Applied Data Science Gütersloh (CfADS) sowie externen Partnern aus einem geförderten Forschungsprojekt.

Abstrakt

Diabetes ist eine weit verbreitete Zivilisationskrankheit. Eine Folgeerkrankung dabei ist das diabetische Fußsyndrom (DFS), welches Einschränkungen verursacht und einen enormen Verlust an Lebensqualität für Patientinnen und Patienten bedeuten kann. In dem hier beschriebenen Projekt soll mittels maschinellen Lernens (ML) eine Fußdruck- und Ganganalyse auf Basis einer Fuß-Sensorik entwickelt werden.

Kurzbeschreibung

Diabetes ist eine weit verbreitete Zivilisationskrankheit, die bei einer unzureichenden Überwachung und Behandlung dramatische Folgeerkrankungen hervorrufen kann. Eine dieser möglichen Folgeerkrankungen ist das Diabetischer-Fuß-Syndrom (DFS). Das DFS ist ein schleichender Prozess, der schwerwiegende Folgen für die betroffenen Personen hat. Von anfänglichen Empfindungsstörungen und kleineren Verletzungen des Fußes führt dieses Krankheitsbild im schlimmsten Fall zu teilweisen oder vollständigen Amputationen des betroffenen Fußes. Die dadurch verursachten Einschränkungen sind ein enormer Verlust an Lebensqualität für Patientinnen und Patienten. Durch den Einsatz einer Sensorik soll der Fuß überwacht werden, um eventuelle Überlastungen zu erkennen. Als Richtwert gelten die international anerkannten Vorgaben der *International Working Group in the Diabetic Foot* (IWGDF).

Betroffene Personen werden in der Regel mit orthopädischen Schuhen und in manchen Situationen mit Orthesen ausgestattet. Diese sollen die Füße optimal entlasten und für eine ausreichende Druckverteilung sorgen. Mit aktuellen Messverfahren kann lediglich gesagt werden, ob eine punktuelle Überbelastung stattgefunden hat. Über die durchgeführte Bewegung – stehen, gehen, Treppen steigen etc. – kann derzeit keine Aussage getroffen werden. Somit ist es nur in entsprechenden Laboren möglich, das Schuhwerk hinsichtlich seiner Funktion und Druckentlastung zu untersuchen. Der Alltag der Personen bleibt derzeit unüberwacht. Dies soll sich durch den Einsatz einer Fuß-Sensorik ändern, welche im Alltag in das Schuhwerk oder in eine Orthese eingebaut wird. Anhand der gewonnenen Messdaten sollen Algorithmen erstellt werden, die in der Lage sind, die unterschiedlichen Druckbelastungen und Bewegungsmuster zu erkennen.

Das hier beschriebene Projekt umfasst die Anwendung von ML-Verfahren in einem realistischen Umfeld. Die Planung von Versuchen zur Generierung passender Messdaten und deren Analyse zur Erkennung der zuvor durchgeführten Bewegungen ist zur Optimierung der Methoden des maschinellen Lernens ebenfalls notwendig.

Aufgabenstellung

Der Studierende / die Studierende entwickeln Lösungen, welche mit Hilfe von künstlicher Intelligenz in der Lage sein soll, Messreihen auszuwerten sowie Druckbelastungen bzw. Bewegungsmuster zu erkennen.

Bezug zum Thema Data Science

Die Analyse von Messsequenzen unter Anwendung von Verfahren des maschinellen Lernens mit vorangehender Generierung geeigneter Trainingsdaten sind wichtige Bausteine der Data Science.

Verfügbare Ressourcen

- Die Fuß-Sensorik in passender Schuhgröße kann beschafft werden. Größe 43 ist bereits vorhanden.
- Die / der Studierende erhält Zugriff auf die Infrastruktur des Center for Applied Data Science (CfADS), wie z.B. einen leistungsstarken Computercluster auf Basis eines Hadoop-Frameworks.

Projektplan

Erstes Semester: Einarbeitung in die Messsoftware und das Themengebiet. Sammeln notwendiger Informationen für weitere Arbeitsschritte. Die Erstellung eines Forschungsexposés ist Prüfungsleistung.

Zweites Semester: Entwurf einer Methode zur Generierung von geeigneten Trainingsdaten. Erstellung von Algorithmen zur Analyse von Messreihen. Recherche zu relevanten Arbeiten im Themenumfeld der Druckmessung von Personen mit diabetischem Fuß-Syndrom.

Drittes Semester: Weiterentwicklung der Algorithmen. Anwendung und Umsetzung von Methoden des maschinellen Lernens, insbesondere Neuronale Netze oder Gaußsche Prozessmodelle.

Viertes Semester: Entwicklung und Evaluation der ML-Verfahren zur Analyse und Darstellung von Messreihen aus der Fußdruck-Messung. Erzeugung von Prognosen auf Basis der Modelle. Die Modelle sollen in der Lage sein, Bewegungsmuster zu erkennen und auf Überbelastungen hinweisen.

Eignungskriterien

Zwingend:

- Grundkenntnisse in der Informatik
- Programmierkenntnisse in Python oder Matlab
- Teamfähigkeit und Interesse an wissenschaftlicher Arbeit
- Selbstständige Arbeitsweise

Optional:

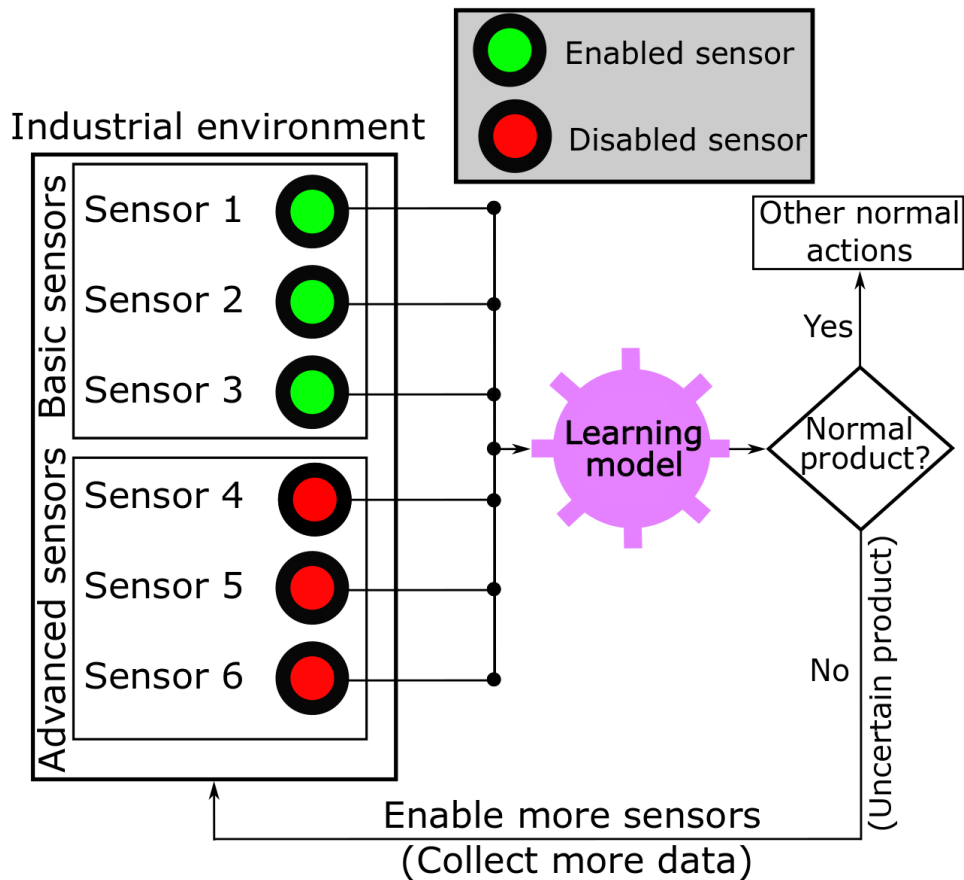
- Kenntnisse in Bereichen des maschinellen Lernens
- Kenntnisse über entsprechende Bibliotheken und Pakete für Data Science in Python.

Erwerbbar Kompetenzen

- Erfahrung und Kompetenz zur Generierung und Beurteilung geeigneter Trainingsdaten
- Erstellen von Analyseergebnissen
- Erfahrung in der Anwendung von Methoden aus dem Deep Learning
- Wissenschaftliches Arbeiten, Schreiben und Präsentieren

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: 20 Adaptive Sensoraktivierung mit aktivem/maschinellern Lernen für IoT-gestützte Qualitätssicherung



Projektübersicht

Anzahl Studierende	1-2
Art	Studienprojekt
Projektverantwortung	Dr. Alaa Othman Prof. Dr.-Ing. Wolfram Schenck
Projektkontext	Projekt innerhalb des Center for Applied Data Science Gütersloh (CfADS) mit hochschulinternen Partnern.

Abstrakt

Efficiently managing the huge amount of data generated by IoT sensors in industrial environments is challenging. In this proposal, we explore the integration of active learning technique to optimize sensor usage by intelligently enabling and disabling sensors based on data patterns indicating uncertainties. In doing so, our approach ensures thorough inspection of products while conserving network resources. This intelligent and adaptive strategy reduces unnecessary energy consumption in battery-powered IoT environments, such as factories, where wearables and sensors play a critical role. Furthermore, the proposed strategy provides a strategic balance between network congestion, energy conservation, and quality assurance, marking a significant advancement in IoT sensor network management for industrial applications.

Kurzbeschreibung

In the rapidly evolving landscape of industrial environments, the proliferation of IoT sensors has led to an unprecedented generation of data. However, this surge in data comes with a significant challenge - efficient management and utilization. The extensive use of sensors to monitor various parameters, processes, and conditions in factory environments results in a tremendous volume of data streams. This presents challenges in terms of data transmission, network bandwidth utilization and, most importantly, the power consumption of battery-powered IoT devices.

To address these challenges, we present an adaptive sensor activation system infused with active learning algorithms. The core motivation is to harness the power of data-driven decision making to optimize sensor utilization. Rather than indiscriminately activating all sensors in real time, our system intelligently identifies patterns in product data that indicate uncertainties or potential errors. By selectively activating additional sensors based on these predictions, we achieve the dual goal of ensuring a comprehensive inspection of products and reducing the strain on network resources. This not only allows for more in-depth analysis when needed, but also conserves energy-a critical consideration in IoT environments with battery-powered devices.

In practice, our adaptive sensor activation system finds its ideal application in environments where IoT sensors and wearables play a central role, such as modern factories. In our IoT factory in Gütersloh, consider a scenario where Robotinos and other IoT devices monitor and track various processes. Through active learning, the system can identify anomalies or uncertainties in the monitored data streams. In response, it strategically activates additional sensors for closer inspection, ensuring that network congestion is minimized, energy is conserved, and product quality is rigorously maintained. This paradigm shift in sensor network management holds great promise for improving the efficiency, sustainability, and overall performance of industrial IoT environments.

Aufgabenstellung

The student(s) will develop and implement an adaptive sensor activation system for industrial IoT environments. This system will integrate active learning algorithms to intelligently activate sensors based on patterns identified in product data, specifically targeting uncertainties and potential errors. The goal is to optimize sensor usage and address the challenges posed by large data streams in large-scale industrial environments. Students will work on refining and implementing the algorithm, ensuring that additional sensors are selectively activated, minimizing network congestion, and conserving energy

in environments with battery-powered devices. The practical application of this system will be demonstrated in our IoT factory, improving both data efficiency and product quality through intelligent sensor activation.

Bezug zum Thema Data Science

The proposed research aligns seamlessly with core data science principles, integrating active learning algorithms into industrial IoT environments. This aligns with the research master's curriculum, which emphasizes advanced techniques for pattern recognition, predictive analytics, and data-driven decision making. The project provides an opportunity for students to explore and apply cutting-edge data science methods in a real-world context, fostering a deep understanding of how these techniques contribute to the optimization of sensor networks and resource utilization in industrial environments.

Verfügbare Ressourcen

- **Access to a fully-equipped IoT factory:** The project benefits from the availability of a comprehensive IoT factory that facilitates the creation of simulations, data collection, and real-world testing. This resource allows students to work in an authentic industrial environment, enhancing the practical applicability of their research.
- **Expert Oversight:** The project is supported by knowledgeable and experienced supervisors who are well-versed in data science techniques, including active learning, as well as IoT and industrial applications. Students can tap into this expertise for guidance, feedback, and mentorship throughout the research process, ensuring the quality and relevance of their work.
- **Collaborative spaces:** The CfADS group at FHBielefeld provides collaborative spaces where students can engage with peers, share insights, and leverage additional resources. These hubs foster an environment that encourages interdisciplinary collaboration and promotes a holistic approach to problem solving.
- **No additional hardware required:** The project leverages existing hardware resources within the IoT factory, eliminating the need for additional investment. Students can seamlessly integrate their research into the existing infrastructure, streamlining the implementation of adaptive sensor activation using active learning algorithms.

Projektplan

First Semester: Project Setup and Exploration

Become familiar with the IoT factory infrastructure and data collection processes. Develop a clear understanding of industrial processes and sensor networks. Begin a literature review on active learning in IoT environments. Define initial requirements for the adaptive sensor activation system. Create a Research Exposé.

Second Semester: Initial Development and Prototyping

Develop a prototype of active learning algorithms in a simulated environment. Implement basic sensor activation mechanisms based on active learning. Start integrating the prototype with the IoT factory for initial testing. Investigate potential challenges and refine the system design. Publish a Technical Report about the research area.

Third Semester: System Refinement and Integration

Enhance the active learning algorithms for more sophisticated sensor activation. Integrate the system with real-time data from the IoT factory. Evaluate the performance of adaptive sensor activation in a controlled environment. Begin to explore energy saving aspects for battery powered devices. Publish a Paper about the results on a conference.

Fourth Semester: Optimization and Final Evaluation

Optimize the system for efficiency, minimizing unnecessary sensor activations. Fine-tune active learning algorithms based on feedback and performance metrics. Conduct extensive testing in various scenarios within the IoT factory. Summarize results, compare with existing methods, and provide recommendations for further improvements. Master's thesis and colloquium

Eignungskriterien

Mandatory: Strong programming skills, particularly in languages suitable for IoT applications (e.g., Python).

Optional:

- Familiarity with machine learning concepts and algorithms.
- Understanding of IoT systems and sensor networks.

Erwerbbarer Kompetenzen

- Proficiency in implementing and using active learning methods, especially in the context of adaptive sensor activation.
- Hands-on experience with IoT systems, including working with sensor data, communication protocols, and network optimization.
- Competence in developing and integrating machine learning algorithms within an IoT environment.
- Knowledge and practical skills in designing and executing experiments for testing and optimizing adaptive sensor activation strategies.

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitel: 21 Entwicklung eines autonomen Buchhaltungssoftware Assistenten als autonomen Agenten auf Basis von Large Language Models



Projektübersicht

Anzahl Studierende	1
Art	Projekt mit externen Partnern
Projektverantwortung	Prof. Dr. Stefan Berlik
Projektkontext	Projekt in Zusammenarbeit mit der Firma Diamant Software GmbH

Abstrakt

In diesem Projekt wird ein autonomer Agent entwickelt, der als Kernkomponente ein Large Language Model (LLM) hat und in einer Buchhaltungsumgebung interagieren kann. Das Forschungsprojekt ist in dem Unternehmen Diamant Software GmbH durchzuführen.

Kurzbeschreibung

Das Unternehmen Diamant Software entwickelt die Rechnungswesen- und Controlling-Software Diamant. Die Software bietet u. a. Funktionen von Rechnungswesen und Controlling wie Finanzbuchhaltung, Rechnungseingangsmangement, Kosten- und Leistungsrechnung an. Das Unternehmen strebt danach, die intelligenteste Rechnungswesen-Software der Welt herzustellen. Dabei spielt die Anwendung der künstlichen Intelligenz (KI) eine unmittelbare und unverzichtbare Rolle. Um die KI-Funktionalitäten für die Software zu erforschen und weiterzuentwickeln, betreibt das Unternehmen ein eigenes KI-Kompetenzzentrum in Darmstadt. In diesem KI-Team ist das Projekt zu realisieren.

Autonome Agenten sind intelligente Systeme, die autonom handeln und auf komplexe Aufgabenstellungen reagieren können. Autonome Agenten mit LLM nutzen LLMs als zentrale Komponente, um Informationen zu verarbeiten und Entscheidungen zu treffen. Ein LLM, wie beispielsweise GPT (Generative Pre-trained Transformer) von OpenAI, ist ein leistungsstarkes Sprachmodell, das auf einem neuronalen Netzwerk basiert und ein tieferes Verständnis von natürlicher Sprache aufweist. Autonome Agenten, die auf LLMs aufbauen, können komplexe Probleme in der natürlichen Sprachverarbeitung angehen. Sie sind in der Lage, Texte zu verstehen, zu generieren und in Dialogen zu interagieren, basierend auf den Informationen und Mustern, die sie aus großen Textdatensätzen gelernt haben.

Durch den Einsatz von autonomen Agenten im Buchhaltungsbereich lassen sich zahlreiche zeitintensive Aufgaben übernehmen, wodurch Mitarbeitende in der Lage sind, ihre Ressourcen effizienter zu nutzen und sich verstärkt auf andere wichtige Aufgaben zu konzentrieren. Dies führt zu einer Optimierung der Arbeitsabläufe und einer Steigerung der Produktivität im gesamten Buchhaltungsprozess.

Aufgabenstellung

Der/die Studierende wird im Rahmen dieses Projekts einen autonomen Agenten konzipieren und entwickeln, der ein LLM als zentrale Ressource nutzt. Der besondere Schwerpunkt liegt dabei auf der Implementierung von Funktionalitäten, die eine nahtlose Interaktion dieses autonomen Agenten in einer Buchhaltungsumgebung ermöglichen. Diese zu entwickelnde Buchhaltungsumgebung soll sowohl bestimmte Datenschnittstellen, beispielweise zu der Buchhaltungssoftware Diamant Software, als auch Tools und Funktionalitäten haben, die Interaktionsmöglichkeiten zwischen dem autonomen Agenten und Buchhaltungsmitarbeitenden erlauben. Dem Agenten können APIs zu verschiedenen Anwendungen wie Diamant Software und Microsoft Teams zur Verfügung gestellt werden. Eine alternative Möglichkeit besteht darin, dass die Diamant Software selbst als Umgebung genutzt wird und über den Browser angesprochen wird. Das Ziel besteht darin, dass der entwickelte autonome Agent in der Lage ist, vielfältige komplexe Aufgaben aus dem Bereich des Rechnungswesens eigenständig und effizient zu bewältigen. Beispielweise kann der Agent eine Buchungsaufgabe von einem menschlichen Nutzer in natürlicher Sprache verstehen und eigenständig die Buchung in der Diamant Software durchführen. Abbildung 1 zeigt den verallgemeinerten Interaktionsprozess zwischen dem Nutzer, dem Agenten und der Umgebung.

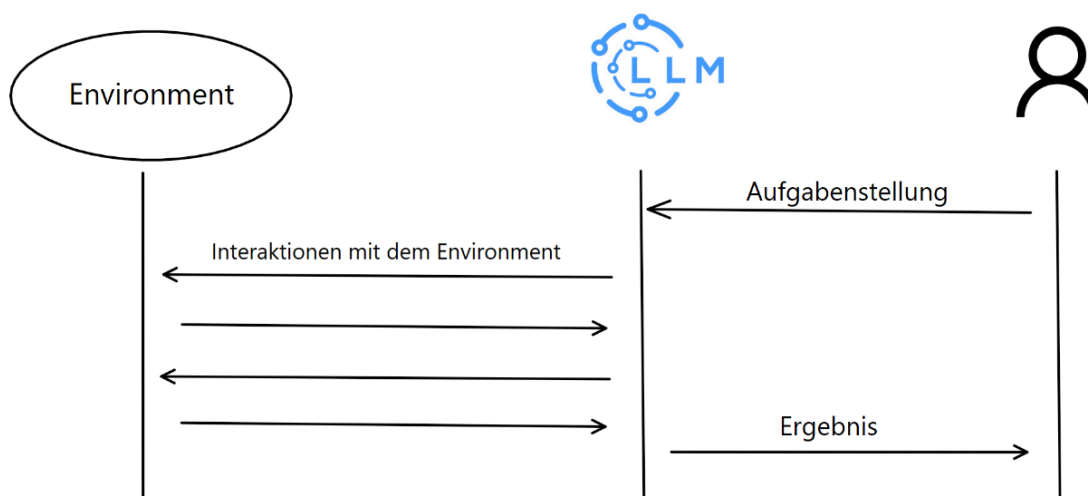


Abbildung 1: Interaktionen zwischen dem Nutzer, dem autonomen Agenten und dem Environment

Bezug zum Thema Data Science

Die Entwicklung von autonomen Agenten ist ein aktuelles und essenzielles Thema im Bereich der künstlichen Intelligenz (KI) bzw. Data Science. Insbesondere die Integration von LLMs in autonomen Agenten stellt einen Fortschritt in Data Science dar. In Verbindung mit LLM sind autonome Agenten in der Lage, nicht nur auf vorhandene Daten zuzugreifen, sondern auch kontextbezogene und generative Fähigkeiten zu entwickeln.

Verfügbare Ressourcen

- Ansprechpartner im KI-Team des Unternehmens Diamant Software GmbH
- Verschiedene LLMs, z. B. von OpenAI
- Diamant Software
- Hardware für die Programmierung

Projektplan

Erstes Semester: Formulierung des Forschungsexposees, Einarbeitung in autonomen Agenten und LLMs, Konkretisierung des Forschungsvorhabens.

Zweites Semester: Entwicklung der Buchhaltungsumgebung für den autonomen Agenten, Entwicklung eines simplen Agenten, Auswertung erster Ergebnisse.

Drittes Semester: Entwicklung eines komplexen autonomen Agenten. Auswertung weiterer Ergebnisse.

Viertes Semester: Implementierung und Vergleich von weiteren Verfahren, Fine-Tuning, Finale Evaluierung.

Eignungskriterien

Zwingend:

- Programmierkenntnisse
- Gute Englischkenntnisse
- Grundkenntnisse in der Buchhaltung

Optional:

- Programmierkenntnisse in Python
- Erfahrung mit Diamant Software
- Grundlegende Kenntnisse in autonomen Agenten

Erwerbbarer Kompetenzen

- Entwicklung eines komplexen autonomen Agenten
- Anwendung von LLMs
- Wissenschaftliches Schreiben

Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: 22 Maschinelle Intelligenz für die Erkennung von Anomalien, sowie die Prädiktion von Interaktionen, anhand von Bewegungsinformationen im Smart Home



Projektübersicht

Anzahl Studierende	1
Art	Projekt mit externen Partnern
Projektverantwortung	Prof. Dr.-Ing. Thorsten Jungeblut
Projektkontext	Projekt in Zusammenarbeit mit dem KogniHome - Technikunterstütztes Wohnen für Menschen e.V. in Bielefeld; Datensätze und Hardwareressourcen werden gestellt; Anstellung als wissenschaftliche Hilfskraft ist möglich; Über die Projektbearbeitung besteht die Möglichkeit der engen Zusammenarbeit mit Vereinsmitgliedern des KogniHome e.V., u.a. Steinel, Hettich, Achelos, HUM Systems, C&S GmbH uvm.

Abstrakt

Ziel des Projektes ist die Mitentwicklung eines Systems, das mittels KI-Verfahren das Nutzungsverhalten der Bewohner*innen eines Smart Homes erlernt, um anschließend, geeignete Handlungsempfehlungen bereitzustellen oder, bei der Erkennung von Anomalien, geeignete Maßnahmen einzuleiten. Wissenschaftliche Herausforderung des Projekts ist die Anwendung und

Evaluation von maschinellen Lernverfahren zum Lernen von Interaktionsmustern basierend auf den gelieferten Daten der Minimalsensorik, sowie der Prädiktion zukünftiger Interaktion und der Anomalieerkennung.

Kurzbeschreibung

Die Digitalisierung dringt in immer weitere Lebensbereiche vor, und die damit verbundene Vernetzung von vielfältigen Komponenten des täglichen Umfelds führt zu einer Situation, in der ein stetig höherer Anspruch an die alltägliche Technik gestellt wird. Die Systeme sollen smarter werden, automatisiert und eigenständig handeln. Im Idealfall antizipiert die Technik die Bedürfnisse des Menschen und eine Korrektur durch diesen ist nicht mehr nötig. Bedarfsgerechte Mensch-Technik-Interaktion erfordert die Anpassung des Intelligenten Technischen Systems (ITS) an den Nutzungskontext und nicht umgekehrt. Voraussetzung für eine intuitive Interaktion ist daher zuallererst die sichere Erkennung des Nutzungskontextes, d.h. wo befindet sich der Nutzer und welche Handlung führt er gerade durch. Das Wissen über regelmäßig auftretende Interaktionsmuster ermöglicht es dem ITS zukünftige Interaktionen vorherzusagen und Assistenzfunktionen prädiktiv zu steuern. Bildgebende Sensorik (z.B. Kameras, hochauflösende Time-Of-Flight-Sensoren) ermöglichen durch Personen-, Objekt-, Gesten- oder gar Gesichtserkennung eine leistungsfähige Erkennung des Handlungskontextes, bringen aber die Problematik der Erfassung personenbezogener Daten mit sich. Gerade im privaten häuslichen Umfeld, aber auch in Bürogebäuden oder der Produktion kann dies aus datenschutzrechtlichen Gründen unerwünscht sein.

Minimalsensorik, wie z.B. Bewegungs- oder Präsenzmelder erzeugen nicht unmittelbar personenbezogene Daten, sondern erfassen nur punktuelle Informationen über die Anwesenheit von Personen oder Objekten in einem räumlich eingeschränkten Bereich. Für eine umfassende Erfassung des Nutzungskontextes verspricht die Kombination einer Vielzahl von einfachen Sensoren eine ausreichende Erfassung des Nutzungskontextes bei gleichzeitiger Wahrung der Privatsphäre. Ist für eine Assistenzfunktion komplexere Sensorik notwendig (z.B. für eine Sprach- oder Gestenerkennung), dann braucht sie nur (und auch nur genau dann) aktiviert zu werden, wenn die Nutzung des Assistenzsystems zumindest absehbar ist. Die kontinuierliche Erfassung des Nutzungskontextes ermöglicht aber auch das Lernen von regelmäßigen Handlungsmustern. Aus diesen gelernten Handlungsmustern kann das ITS eine zukünftige zu erwartende Interaktion ableiten und Empfehlungen für die Aktivierung von Assistenzfunktionen geben bzw. diese vorbereiten (Prädiktive Steuerung). Darüber hinaus lassen sich aber auch Abweichungen vom potenziellen Regelzustand (Anomalien) erkennen und geeignet darauf reagieren.

Die Ableitung von Handlungsempfehlungen durch das ITS kann in einfachster Implementierung regelbasiert erfolgen (z.B. „Wenn der Bewegungsmelder im Flur aktiviert wird, dann soll das Licht angeschaltet werden“). In Systemen mit einer hohen Anzahl an Sensoren ist dieses aber sehr aufwendig, nicht flexibel anpassbar und schlecht auf wachsende Umgebungen skalierbar. Darüber hinaus liefern verschiedene Sensoren unterschiedlich abstrakte Informationen (z.B. lokale Bewegung (PIR, Ultraschall) oder Bewegung über mehrere Räume hinweg (HF/Mikrowellen), binär „an/aus“ oder Entfernungen) oder gar komplexere Informationen kamerabasierter Systeme (z.B. Anzahl erkannter Personen/Haustiere/Objekte (Staubsaugerroboter/Transportplattform)). Daher gilt es die unterschiedlichen Informationen zu abstrahieren, ggf. zu anonymisieren und in einem ganzheitlichen Ansatz zum Lernen des Nutzungskontextes, zum Abschätzen zukünftiger Interaktion und zum Erkennen von Anomalien zu nutzen. Wissenschaftliche Herausforderung des Projekts ist daher die Anwendung und Evaluation von maschinellen Lernverfahren zum Lernen von Interaktionsmustern basierend auf den gelieferten Daten der Minimalsensorik, sowie der Prädiktion zukünftiger Interaktion und der Anomalieerkennung.

Aufgabenstellung

Der/die Studierende soll in diesem Projekt ein ITS mitentwickeln, das mittels maschinellen Lernverfahren das Nutzungsverhalten der/des Bewohner/s eines Smart Homes erlernt, um anschließend, anhand der Echtzeitdaten, geeignete Handlungsempfehlungen bereitzustellen oder, bei der Erkennung von Anomalien, geeignete Maßnahmen einzuleiten. Die Ausarbeitung des konkreten Anwendungsszenarios, zusammen mit dem Team des KogniHome e.V., ist Teil des Projekts.

Zur Veranschaulichung soll folgendes Szenario beschrieben werden: Sabine (74) steht jeden morgen zwischen 7 und 8 Uhr auf, geht auf die Toilette und macht sich danach einen Kaffee. In diesem Handlungsablauf sind drei Aktivitäten zu erkennen: Aufstehen, auf Toilette gehen, Kaffee kochen. In jeder dieser Aktivitäten kann es jetzt zu Handlungsempfehlungen bzw. Anomalien kommen. Beispielsweise könnte das ITS den Gang zur Toilette durch geeignete Beleuchtung vorbereiten, oder bereits die Kaffeemaschine anschalten. Gleichzeitig überwacht das ITS die Aktivitäten und erkennt Abweichungen vom gewohnten Verhalten. Stürzt Sabine z.B. auf dem Weg zur Toilette und steht nicht wieder eigenständig auf, so soll das ITS dies erkennen und z.B. den Notruf oder die Pflegestelle verständigen.

Bezug zum Thema Data Science

Die Evaluation und Anwendung von KI-/ML-Verfahren zur Zustandsüberwachung und Prädiktion sind ein Kernthema der Data Science und werden beispielsweise in den Modulen „Data Mining & Machine Learning“ sowie „Künstliche Intelligenz“ behandelt. Die Aufnahme hochaufgelöster Messdaten (z.B. Verbrauchsdaten) von einer Vielzahl an Sensoren in komplexen Wohnumgebungen stellt hohe Anforderungen an Organisation und Verarbeitung der Daten. Dieses ist Kern des Moduls „Big Data Architekturen“.

Verfügbare Ressourcen

- Informationen, die zur Erstellung des Szenarios benötigt werden (Systembeschreibung, logistische Abläufe, relevante Kennzahlen) werden vom KogniHome e.V. bereitgestellt
- Es besteht Zugriff auf die Forschungswohnung des KogniHome e.V.
- Über die Forschungswohnung des KogniHome e.V. stehen umfangreiche Testdatensätze zur Verfügung.
- Der Ansprechpartner im KogniHome e.V. wird über die Projektlaufzeit zur Verfügung stehen
- Benötigte Materialien werden vom KogniHome e.V. bereitgestellt.
- Hardware für das komplexere Machine Learning ist über das Data Science Lab, das CfADS, sowie dem KI-Rechencluster yourAI der FH Bielefeld verfügbar.

Projektplan

Erstes Semester: Erstellung eines Forschungsexposés als Prüfungsleistung. Einarbeitung in die Konzepte und Strukturen (IoT-Software, Smart Home Protokolle, Schnittstellen etc.) der Forschungswohnung des KogniHome e.V.

Zweites Semester: Erstellung des Systemkonzepts zur Anomalieerkennung zur prädiktiven Steuerung von Assistenzsystemen. Recherche zu relevanten Arbeiten im Themenfeld des Einsatzes von KI-Verfahren für das Lernen von Ereignissequenzen und für die Anomalieerkennung. Erstellung eines Papers, das einen Überblick über das jeweilige Forschungsgebiet gibt, als Prüfungsleistung.

Drittes Semester: Praktische Umsetzung verschiedenster maschinellen Lernverfahren im Bereich des unüberwachten und des (semi-)überwachten Lernens zur Optimierung einer automatisierten Interaktion von Mensch und Maschine, sowie deren Evaluation.

Viertes Semester: Masterarbeit und Kolloquium. Finale Evaluierung durch Vergleich der implementierten Strategien. Erstellung eines Papers mit ersten quantitativen Ergebnissen als Prüfungsleistung.

Eignungskriterien

Zwingend:

- Programmierkenntnisse (vzw. Python)
- Erfahrung mit dem Versions-Kontroll-System „git“

Optional:

- Erfahrung in der Elektronikentwicklung
- Erfahrung im Bereich Smart Home Technologien/IoT-Geräten
- Programmierung von Mikrocontrollern

Erwerbbar Kompetenzen

- Verfahren der künstlichen Intelligenz zur Problemlösung
- Sensornähe Informationsverarbeitung
- Prädiktive Assistenz und Anomalieerkennung (Übertragbarkeit auf industrielle Prozesse)
- Kompetenzen in der Arbeit im Team