



Smart Data innovation challenges

Abschlussbericht zum Projekt SDI- C
Förderkennzeichen: 01IS19030A-G

Projektlaufzeit: 01. August 2019 - 31. Dezember 2022

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

umProjektpartner



Prof. Dr. Michael Beigl,
Karlsruher Institut für
Technologie (Koordinator)
michael.beigl@kit.edu



Prof. Dr. Paul Lukowicz
Deutsches
Forschungszentrum für
Künstliche Intelligenz



Dr. Elmar Dorner
SAP SE



Dr. Joachim Köhler
Fraunhofer-Institut für
Intelligente Analyse- und
Informationssysteme



Dirk Mayer
Software AG



Prof. Morris Riedel
Forschungszentrum Jülich



Felizitas Müller
IBM Deutschland
GmbH



Zum Bericht

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren vom [Karlsruher Institut für Technologie \(KIT\)](#), [Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme \(IAIS\)](#), [Deutsches Forschungszentrum für Künstliche Intelligenz GmbH \(DFKI\)](#), [Forschungszentrum Jülich GmbH](#), [SAP SE](#), [IBM Deutschland GmbH](#) und der [Software AG](#), welche Projektpartner innerhalb des Projektes waren, sowie der [SICOS BW GmbH](#), die das Projekt im Rahmen eines Unterauftrags unterstützt hat, sowie des [Fraunhofer IOSB](#), welches an mehreren Mikroprojekten beteiligt war.

Mit diesem Bericht legen die Projektpartner der Smart Data Innovation Challenges (SDI-C) auch formell ihren gemeinsamen Abschlussbericht vor. Der öffentliche gemeinsame Abschlussbericht folgt dem vorgegebenen Muster für Förderprojekte des Bundes: Der folgende Abschnitt entspricht Kurzbericht (Teil 1), die folgenden Teile beschreiben inhaltlichen Entwicklungen in einer eingehenden Darstellung (siehe Teil 2 gemäß Anlage 2 zu Nr. 8.2 NKBF 98). Wir haben trotzdem versucht den Bericht so leserlich wie möglich für ein breites Publikum zu halten.

Die Inhalte des Berichts werden auch auf der Webseite www.sdil.de veröffentlicht. Hier werden auch weitere Materialien und Informationen zu den 27 geförderten Mikroprojekten und weiteren vergangenen sowie laufenden Mikroprojekten bereitgestellt.

r

Als Sprecher des Smart Data Innovation Lab (SDIL) bin ich stolz, Ihnen mit diesem Bericht zum Abschluss der Smart Data Innovation Challenges einen vielfältigen Einblick in unseren „Datenreinraum“ geben zu können. Das Projekt, aber vor allem die vielen Mikroprojekte innerhalb des Projekts, haben gezeigt, wie wichtig gemeinsame Infrastrukturen für die Förderung datengetriebener Innovationen auf Basis industrieller Daten sein können. Seit der Konzeption des SDIL im Jahr 2014 haben wir eine rasante Entwicklung im Bereich der KI-Innovationen erlebt, die sich in den Mikroprojekten widerspiegelt. Die Smart Data Innovation Challenges haben Forscher und Unternehmen zusammengebracht und zu gemeinsamen Innovationen beigetragen. Gleichzeitig haben sie die Entwicklung einer gemeinsamen Plattform nachhaltig beeinflusst.

Unser Dank gilt dem Bundesministerium für Bildung und Forschung sowie dem Projektträger am DLR in Berlin, die das Projekt unterstützt und begleitet haben, dabei stets das langfristige Potenzial datengetriebener Innovationen im Blick hatten und die Forschung auf Basis industrieller Daten unterstützt haben und weiterhin unterstützen.



Prof. Michael Beigl

*Sprecher Smart
Data Innovation Lab*



Zum Bericht	2
1. Die Smart Data Innovation Challenges	6
2. Eingehende Darstellung der Ergebnisse	8
Herausforderungen	8
Projektziel 1: Erfolgreiche KI-Lösungen durch schnelle Evaluierung auf realen Daten	9
Projektziel 2: Attraktive Projektbedingungen für kooperative Forschung	9
Lösungsansatz 1: KI- Mikroprojekte und Mischbetrieb	9
Lösungsansatz 2: Weiterentwicklung einer geteilten KI-Infrastruktur	10
Voraussetzungen zu Beginn	10
KI-Mikroprojekte	11
Die Challenges und der Auswahlprozess	13
Mikroprojekte Call 1: Einreichungsfrist 26.09.2019	13
Risk Management 4.0	14
Cognitive Assistant:	16
E-Scooter Detektion	18
Personenstromerkennung	20
EvaRD	22
Auswertung von hochdimensionalen Messdaten	24
Mikroprojekte Call 2: Einreichungsfrist 08.03.2020	27
Transper	28
Smart Air Quality Sensor Calibration	30
Feature Indexing für Parfümflaschen durch Deep Learning	32
KI -Modelle für Smart Cities	34
Chatbot Mining	36
Mikroprojekte Call 3: Einreichungsfrist 25.10.2020	39
Solving Accounting Optimization Problems in the Cloud	40
BERTI-4.0	42
Cognitive Matchmaking bei Communities von Lernplattformen	44
WearNet	46
EpiNet	48
LatentAgel4Wafer	50
Mikroprojekte Call 4: Einreichungsfrist 12.07.2021	53
Grumo	54
Raumdaten-KI in der Agrarwirtschaft	56
Airformer	58
KI-Turbine	60
Mikroprojekte Call 5: Einreichungsfrist 20.02.2022	63
Hipwings	64
Traffic Counter	66
Rapid Urbanism KI Agent	68
SmartNesting@HPC	70
DeepSoilLearning	72
KD4RE	74

Weiterentwicklung der Plattform	76
Compute Infrastruktur Plattform	77
Business Technology Platform BTP Multichord SAP	80
IBM Cloud mit AutoML	84
Software AG IOT-Cloud für SDIL	86
Angebotsweiterentwicklung	89
Smart Data Days, Workshops und Hackathons	90
SDIL in der Presse	91
Notwendigkeit und Angemessenheit der Arbeiten im Projekt	93
Wichtige Entwicklungen außerhalb des Projektes	93
Informationen zum Projektumfang	93
Fazit und Ausblick	94
Wissenschaftliche Veröffentlichungen mit Projektbezug	95



1. Die Smart Data Innovation Challenges

Das Projekt „Smart Data Innovation Challenges“ wurde vom 01.08.2019 bis zum 31.12.2022 vom Bundesministerium für Bildung und Forschung (BMBF) gefördert (Förderkennzeichen [01IS19030A-G](#)). Das Projekt wurde in Kontext des Smart Data Innovation Labs (SDIL, www.sdil.de) durchgeführt.



Abbildung 1: Komponenten der Smart Data Innovation Challenges

Das SDIL (Smart Data Innovation Lab) ist eine Plattform, die die Zusammenarbeit zwischen Wirtschaft, öffentlicher Hand und Forschung im Bereich KI, Big Data und Smart Data-Technologien fördert. Es wurde von Unternehmen wie IBM Deutschland, SAP und der Software AG sowie Forschungseinrichtungen wie dem Deutschen Forschungszentrum für Künstliche Intelligenz, der Fraunhofer-Gesellschaft, dem Forschungszentrum Jülich und dem Karlsruher Institut für Technologie ins Leben gerufen. Das SDIL ermöglicht es seinen Nutzern, Forschung auf industriellen Daten voranzutreiben.

Mit dem Projekt „SDI-C“ sollte die bisherige Dynamik der SDIL-Forschung zu industrierelevanter datenbasierter KI weiter beschleunigt werden, indem die Anzahl der auf der Plattform umgesetzten Projekte weiter erhöht und damit auch eine Skalierung der Lösungen erreicht wird. Neben der Qualität der Forschung ist es für den Transfer von KI-Ansätzen in die Industrie wichtig, eine möglichst große Zahl von Referenzprojekten vorweisen zu können und damit zu demonstrieren, dass schnelle Co-Innovationen auf Daten möglich sind.

In dem vom BMBF über drei Jahre geförderten Projekt wurden insbesondere innovative KI-Mikroprojekte gefördert. Diese durch das DFKI, Forschungszentrum Jülich, die Fraunhofer Gesellschaft (insbesondere das IAIS aber auch das IOSB) sowie das KIT zusammen mit Unternehmen durchgeführten Mikroprojekte wurden mit Bereitstellung der gemeinsamen Infrastruktur, sowie mit bis zu 6 Personenmonaten bei den Forschungspartnern gefördert. Die Auswahl erfolgte kompetitiv in fünf Auswahlrunden. Die im Projekt geförderten, aber auch andere Projekte die auf dem SDIL durchgeführt wurden finden sich auf der Webseite des SDILs unter www.sdil.de.

Ziel des Projektes war es einen skalierbaren „Mischbetrieb“ verschiedenster Projektformen zu ermöglichen. Dies ist insbesondere durch die Neustrukturierung der Infrastruktur und der initialen Betriebskonzepte möglich geworden, welche im Projekt vorangetrieben wurde. So konnte das SDIL auch mit anderen Projekten auf regionaler bis internationaler Ebene kooperiert, wie zum Beispiel mit dem Kompetenzzentrum KI-Engineering oder dem EU-Hubs4Data Projekt, welches eine Föderation von datengetriebenen Innovationsräumen innerhalb der Big Data Value Association europaweit vorantreibt

Das Projekt ist im August 2019 gestartet und war ursprünglich für drei Jahre geplant, wurde jedoch für fünf weitere Monate mittelneutral bis zum Ende 2022 verlängert. Das Jahr 2021 markierte dabei einen wichtigen Meilenstein für das Projekt, da mit Unterstützung des BMBF die „On-Premise“-Infrastruktur am KIT in den BW-Unicluster 2.0 migriert werden konnte. Hierzu hat das BMBF ein Cluster mit 76 A100 GPUs gefördert, die seitdem zur Forschung auf industriellen Daten zur Verfügung stehen und essenziell dazu beigetragen haben, dass eine skalierbare und effiziente Integration von industrieorientierten KI-Projekten in den Rechenzentrumsbetrieb möglich wurde. Gleichzeitig wurden skalierende Cloud-Dienste von IBM, SAP und Software AG eingebunden.

Die Abschlussveranstaltung im Januar 2023 in dem die Ergebnisse der letzten Mikroprojekte sowie die Ergebnisse der Plattformumstrukturierung der Öffentlichkeit vorgestellt wurden markierte das endgültige Ende des Verbundprojektes. Die gemeinsamen Ergebnisse werden jedoch nahtlos innerhalb des SDIL weiterverwertet. Weiterhin profitieren die beteiligten Unternehmen von der Verfügbarkeit passgenauer KI-Innovationen, die innerhalb der Projekte entwickelt wurden, und die sofort mit Abschluss des Mikroprojektes in Produkte, Dienste und Prozesse transferiert werden können.

Zusammenfassend sind wichtigsten Ergebnisse, die hierdurch während der Projektlaufzeit innerhalb des SDILs erreicht werden konnten:

- Die erfolgreiche Ausschreibung und Organisation von 27 Mikroprojekten innerhalb von fünf Ausschreibungen zu den Themenbereichen:
 - Hochfrequente Datenströme und Zeitreihen
 - Optimierung von und mit KI
 - Technologien zur Bewältigung der Pandemie und KI-Leuchttürme
 - KI auf räumlichen Daten oder Graphen und Datensparsames Lernen
 - Effiziente Nutzung von GPU-Clustern und KI-Model- Lebenszyklus
- Der Transfer vielfältiger Forschungsergebnisse innerhalb einzelner Mikroprojekte und die schnell-getaktete Verwertung von Forschungsergebnissen
- Die erfolgreiche Umgestaltung der SDIL-Plattform hin zu einem nachhaltigen Betriebsmodell innerhalb bestehender Cloud- und HPC-Systeme:
 - Die Anbindung von Cloud-Diensten und die Integration in Prozesse am SDIL
 - Die Inbetriebnahme einer eigenen GPU-Partition innerhalb eines Forschungsclusters für KI-Anwendungen nur für die Forschung auf Industriellen Daten.

2. Eingehende Darstellung der Ergebnisse

Das große Asset des SDIL ist die kurzfristige Möglichkeit, schnell Projekte zu starten und durchzuführen und damit Innovationen anzustoßen. Das SDIL arbeitet dabei an der Schnittstelle zwischen Forschung und praktischer Anwendung. Ziel ist es, KI-Algorithmen und Ansätze aus der aktuellen Forschung auf bestehenden Daten der Industrie anzuwenden und umgekehrt KI-Algorithmen basierend auf den Anforderungen der Industrie bzw. der industriellen Datensätze zu erforschen.

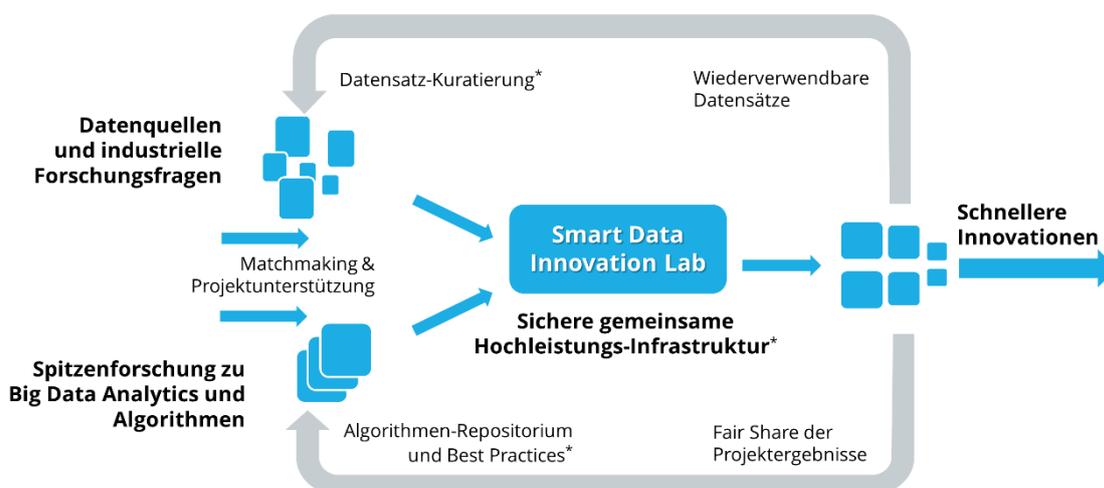


Abbildung 2: Smart Data Innovationszyklus

Herausforderungen

Die Smart Data Innovation Challenges setzen genau hier an und sollen diese Dynamik weiter beschleunigen. Neben der Qualität der Forschung ist es für den Transfer von KI-Ansätzen in die Industrie wichtig, über möglichst viele Referenzprojekte zu verfügen und zeigen zu können, dass eine Umsetzung zu geringen Kosten (insbesondere Zeitkosten, d.h. schnell auf Basis bestehender Infrastruktur) möglich ist.

Das Konzept der Smart Data Innovation Challenges erweitert das Smart Data Innovation Lab um eine wichtige Komponente. Damit kann die Forderung nach schnellerer, innovativer Forschung niederschwellig erreicht werden. Anhand drängender prototypischer Fragestellungen aus der Industrie sollen sich die besten datengetriebenen KI-Ansätze bewähren. Umgekehrt sollen Forscher motiviert werden, ihre Ansätze des maschinellen Lernens und der Big Data Analytics auf reale Datensätze anzuwenden.

Die KI-Mikroprojekte wurden zu sowohl von Industrie und Forschung getragen, wobei der jeweilige Forschungspartner für die Erforschung und Anwendung neuer Algorithmen zu 100 % finanziert wurden und damit auch zum Stand der Forschung beiträgt. Der Anwendungspartner,

der Daten und/oder Systeme beisteuert, profitiert direkt von der prototypischen Umsetzung und der durch das Projekt geförderten Sichtbarkeit. Der Beitrag der Industrie bestand in der Bereitstellung von Daten und personellen Ressourcen für gemeinsame Projekte. Beide Seiten profitieren gleichzeitig von der Infrastruktur, die im Projekt weiterentwickelt wurde.

Projektziel 1: Erfolgreiche KI-Lösungen durch schnelle Evaluierung auf realen Daten

Ein wichtiger Aspekt war die Weiterentwicklung der Plattform. Gerade für die im Kern an SDI-C beteiligten Industriepartner (IBM, SAP, SAG) ist die Evaluierung der Nutzbarkeit ihrer technischen Systeme für forschungsnahe Innovationen interessant.

Auch der Betrieb einer KI-Plattform auf Basis existierender HPC-Infrastruktur (wie z.B. am KIT) stellt eine Herausforderung dar, die erst durch möglichst verschiedene reale Projekte begreifbar wird.

Projektziel 2: Attraktive Projektbedingungen für kooperative Forschung

Umgekehrt kann nur eine erfolgreiche Evaluierung von Forschungsergebnissen an realen Daten zu industriellen Innovationen führen. KI-Lösungen sind dann Katalysatoren für Innovationen, wenn sie auch praktisch erprobt werden.

Das SDIL möchte die Forschung an realen Datenquellen fördern. Es hatte sich im Vorfeld des Projektes gezeigt, dass die Nutzung einer gemeinsamen, hochskalierbaren Plattform nur ein Baustein für erfolgreiche Projekte ist.

Lösungsansatz 1: KI- Mikroprojekte und Mischbetrieb

Aufgrund der kurzen Laufzeit haben „Mikroprojekte“ mit einer Laufzeit von 3-8 Monaten einen klar definierten Fokus und schaffen sowohl schnellere Innovation als auch schnelles Feedback für eine Weiterentwicklung von KI-Lösungen. In 5 Ausschreibungsrunden wurden daher im Projekt jeweils durchschnittlich 6 Mikroprojekte ausgewählt, die sich mit Problemstellungen von Unternehmen beschäftigten und an den ausgewählten Forschungseinrichtungen (KIT, DFKI, Fraunhofer, FZJ) mit jeweils 6 Personenmonaten Aufwand unterstützt wurden.

Die Auswahl der Projekte erfolgte durch den Lenkungskreis des Smart Data Innovation Lab, der dem BMBF jeweils eine Liste von Mikroprojekten zur Förderung vorschlug. Die Auswahl im Peer-Review-Verfahren erfolgte nach klar definierten wettbewerblichen und wissenschaftlichen Kriterien in Form eines offenen Ideenwettbewerbs innerhalb der beteiligten Forschungseinrichtungen.



Auch weiterhin sollen solche Projektformen die Infrastruktur des SDIL nutzen können, (diese wie auch bisher die strategischen Ziele des SDIL (den Transfer von datengetriebener KI von Wissenschaft zu Wirtschaft) voranbringen. Hierzu planen wir die Nutzung der Restkapazitäten aus, die natürlicherweise entstehen. Ziel ist es im Projekt ein nachhaltiges Priorisierungsmodell (Freemium-Nutzung) zu etablieren, dass die Nutzung für Projekte, die für den Betrieb aufkommen (einschließlich der im Vorhaben geförderten Mikroprojekte) zwar bevorzugt, aber für den nicht-kommerziellen Betrieb Restkapazitäten frei zur Verfügung stellt.

Lösungsansatz 2: Weiterentwicklung einer geteilten KI-Infrastruktur

Das KIT übernahm während der Laufzeit den Betrieb der Infrastruktur, fördert eine öffentlichkeitswirksame Ausschreibung und Veröffentlichung von Ergebnissen, übernimmt rechtliche und organisatorische Hilfestellungen und leistet wichtige technologisch orientierte Begleitforschung. Cloud-basierte Infrastruktur (siehe unten) wurde von den beteiligten Industriepartnern projektspezifisch angebunden.

Schulungen zur optimalen Nutzung der Technologien wurden als Webinare und Hackathons veranstaltet und von interessierten Nutzern im Vorfeld und während der Ausführung von Mikroprojekten genutzt.

Plattformpartner direkt auf eigene virtuelle Umgebungen zugreifen und neue Software direkt im SDIL erproben können. Hierzu sind früh im Projektvorhaben entsprechende Sicherheitsstandards entwickeln, die den hohen Vertraulichkeitsstandards des SDIL gerecht werden. Parallel wurden Datenbereitstellungsstrategien entsprechend den flexibleren Virtualisierungs- und Betriebskonzepten angepasst.

Voraussetzungen zu Beginn

"Data Engineering" mittels KI, beziehungsweise auch datenzentrisches Maschinelles Lernen wird in Zukunft ein entscheidender Faktor sein. Um aus dieser Entwicklung Wettbewerbsvorteile für den Standort Deutschland und Europa ziehen zu können, haben Partner aus Industrie und Wissenschaft 2014 das "Smart Data Innovation Lab" (SDIL) gegründet, das am Karlsruher Institut für Technologie (KIT) betrieben wird. Seit der Gründung bis zum Start der Smart Data Innovation Challenges wurde die Plattform kontinuierlich um neue Hard- und Softwarelösungen erweitert und von anwendungsnahen Forschungsprojekten genutzt. Allen Projekten gemeinsam ist die Aufbereitung, Verarbeitung und Analyse großer Datenmengen mit Methoden der KI auf leistungsfähigen Rechnersystemen. Hierfür steht das SDIL als „Datenreinraum“ bereit, dass auch technische, organisatorisch und rechtliche Sicherheit für die Verarbeitung sensibler Daten liefert.



Abbildung 3: On-Premise System am SDIL zu Projekt beginn

Im Vergleich zu Laboren in Forschungsdisziplinen wie Chemie oder Physik gibt es in der industriellen Datenanalyse noch keine Tradition von Standardprozessen oder -werkzeugen, die gerade beim Einsatz automatisierter Systeme wichtig sind. Das vom BMBF bis Mai 2019 geförderte Forschungsprojekt Smart Data Innovation-Prozesse, -Werkzeuge und -Betriebskonzepte (SDI-X) hat hierfür die Grundlagen geschaffen.

Die SDIL-Plattform war zu Projektbeginn aus grundverschiedenen hochskalierbaren Hard- und Softwarelösungen verschiedener Hersteller für KI-getriebene Big Data Analytics aufgebaut. Im Laufe der Zeit wurde sie durch einheitliche replizierbare Best Practices, Werkzeuge und Datenspeicher ergänzt, die rechtliche Rahmenbedingungen, technische Betriebskonzepte und ausführbare Templates für Analytics Workflows mit KI-Werkzeugen abdecken. Die Heterogenität hatte tendenziell hohe Betriebskosten als Folge. Auch stieß das Projekt auf ein Problem: Aufgrund der Nutzungsbedingungen war das SDIL eigentlich nur für nicht-kommerzielle Projekte interessant und gleichzeitig waren Kosten kaum auf einzelne Projekte abbildbar.

KI-Mikroprojekte

Das SDIL bietet die Möglichkeit für schnelle Transferprojekte im Bereich der datengetriebenen Innovation. Die üblichen Zeiten für den Aufbau einer eigenen Infrastruktur und die Etablierung organisatorischer und technischer Prozesse entfallen. Ein Zeitraum von zwei Wochen von der Idee bis zum Projektstart ist realistisch.

Vorwettbewerbliche Forschung an Daten mit neuen Algorithmen ist aber immer noch risikoreich. Umgekehrt gibt es für etablierte Verfahren oft etablierte Cloud-Infrastrukturen mit attraktiven Preismodellen. Die Idee der KI-Mikroprojekte war es daher, ein Förderinstrument zu schaffen, das abseits von klassischen Förderinstrumenten wie KMU-Innovativ des BMBF oder dem Zentralen Innovationsinstrument Mittelstand des BMWK insbesondere kleinen

Unternehmen die Möglichkeit bietet, Technologien zu erproben. Allerdings mit einer deutlich kürzeren Laufzeit und einer „Fail-Fast“-Strategie (auch wenn die im Folgenden vorgestellten Projektbeispiele überwiegend erfolgreich sind, zeigen sie oft auch die Grenzen der verfügbaren Daten). Die geförderten Mikroprojekte erhöhen somit die Chance auf einen verwertbaren Forschungstransfer. Gerade die Investition in das Thema Künstliche Intelligenz hat für Unternehmen einen hohen Stellenwert, auch wenn viele deutsche Unternehmen hier noch zögern.

Die Tatsache, dass nur der akademische Partner gefördert wird, erhöht die Geschwindigkeit der Bewilligung. Die Aussicht auf eine schnelle Verwertbarkeit macht es aber für viele Unternehmen interessant, auch bei einem Forschungsrisiko Eigenleistungen zu erbringen. Damit schließt das SDI-C Projekt eine Lücke in der Förderlandschaft auf nationaler Ebene im Bereich des Forschungstransfers für KI, der sonst meist nur durch Instrumente auf europäischer Ebene (sog. Cascade Funding Projekte) oder regionaler Ebene (Innovationsgutscheine) adressiert wird. Der Overhead des SDI-C Mikroprojektes ist deutlich geringer, was viele Unternehmen dazu bewogen hat, sich mit ihren Problemstellungen einzubringen. Die Möglichkeit, mit Unternehmen zusammenzuarbeiten und reale Daten zu nutzen, stärkt umgekehrt die Forschung.

Durch eine konstante Anzahl von inhaltlich unterschiedlichen, aber organisatorisch vergleichbaren Mikroprojekten hat das Projekt eine Grundauslastung (5-6 gleichzeitig laufende Projekte) für das SDIL erzeugt und den administrativen Aufwand erstmals auf die Mikroprojekte abbildbar gemacht. Damit wurde indirekt auch die Möglichkeit eröffnet, Kosten für zusätzliche Einzelprojekte abzuschätzen. Dieser Verstärkungseffekt führte dazu, dass z.B. auch einzelne EU-geförderte Mikroprojekte parallel durchgeführt werden konnten. Darüber hinaus war es während der Projektphase weiterhin möglich, gänzlich nichtkommerzielle Projekte wie z.B. Masterarbeiten ohne Förderung am SDIL mit Zugang zur Infrastruktur zu unterstützen.

Unterschiedliche Projekte basieren auf unterschiedlichen Daten, die unterschiedliche KI-Ansätze erfordern. Ziel einer erfolgreichen Dateninnovationsplattform muss es sein, möglichst viele Ansätze zur Verfügung zu stellen, diese erfolgreich an realen Problemstellungen zu testen und so auch eine rasche Rückkopplung zwischen Praxis und Forschung zu schaffen. Die Förderung möglichst vielfältiger Mikroprojekte war daher auch für die Weiterentwicklung des SDIL essenziell.

Die Challenges und der Auswahlprozess

Das Projekt war in fünf Ausschreibungsrunden (Calls) gegliedert, die den Challenges entsprachen. Nur die innovativsten Projekte wurden unterstützt. Die Ausschreibung jeder Projektrunde definierte ein bis zwei Fokusthemen, um Synergie- und Austauscheffekte zu verstärken und auch bestimmte Aspekte der Weiterentwicklung der Plattform zu ermöglichen und zu evaluieren. Die Weiterentwicklung war agil angelegt, daher wurden zusätzlich themenoffene Ausschreibungen durchgeführt, um auch den Bedarf an neuen Themen und Technologien abzubilden und in das Projekt zu integrieren. Die Themen wurden jeweils kurz vor der Ausschreibung gemeinsam festgelegt, um auch auf Entwicklungen wie die Pandemie reagieren zu können.

Bereits vor Beginn der Förderphase hatte das SDIL den Lenkungskreis als Gremium etabliert, das gemeinsam über den Zugang zur Plattform entscheidet. Im Steering Committee sind sowohl akademische als auch industrielle Partner vertreten. Alle Partner kennen die Möglichkeiten der Plattform und bringen unterschiedliche Kompetenzen ein, die auch für die Begutachtung wichtig sind. Da insbesondere die akademischen Partner sowohl als Antragsteller als auch als Gutachter auftreten, war der jeweilige akademische Partner nicht an der Begutachtung seiner eigenen Anträge beteiligt. Dies hatte den Nebeneffekt, dass die industrielle und die akademische Seite mit jeweils drei Personen an der Begutachtung der Mikroprojektanträge teilnahmen. Die Diskussion wurde durch jeweils eine schriftliche Stellungnahme von mindestens einem akademischen und einem industriellen Partner vorbereitet. Die Entscheidung über eine zur Förderung vorgeschlagene Liste von 5-7 Projekten erfolgte nach intensiver Diskussion immer einstimmig.

Mikroprojekte Call 1: Einreichungsfrist 26.09.2019

Die Themen der ersten Ausschreibungsrunde ergaben sich aus dem ursprünglichen Projektantrag. Im Vordergrund standen „Hochfrequente Datenströme und Zeitreihen“ sowie Mikroprojekte mit Pilotcharakter. Die Projektauswahl fand am 07.10. statt. Wie erwartet wurden bereits 2 Monate nach Projektstart 6 Projekte ausgewählt, die unmittelbar starten können.



Bildquelle: Stephen Dawson auf Unsplash

Risk Management 4.0

Maschinelles Lernen zur vorausschauenden kausalen Analyse von Risikofaktoren über Investmentportfolien

Laufzeit: 01.11.2019 – 30.04.2020

Projektpartner

Karlsruher Institut für Technologie

Ployplearn Ravivanpong, (ployplearn.ravivanpong@kit.edu)

Dr. Till Riedel (till.riedel@kit.edu)

DWS Investment GmbH

Dr. Pascal Stock (pascal.stock@dws.com),

Die Kernidee des SDI-C Mikroprojektes "Risk Management 4.0" war die explorative Analyse hinsichtlich der Möglichkeit des maschinellen Lernens zur vorausschauenden kausalen Analyse von Risikofaktoren über Investmentportfolien hinweg. Solche automatisierten vorausschauenden Risikoanalysen fragt auch die Bundesanstalt für Finanzdienstleistungen (BaFin) in ihrem Konsultationspapier vom 16.07.2018 nach. Die Regulatorik auf der Ebene der EU und der BRD zur Anwendung des maschinellen Lernens von Finanzinstituten zielt darauf diese Anwendungen transparent und nachvollziehbar zu gestalten. Folglich sind die wissenschaftlichen Erkenntnisse der Erklärbarkeit der Kausalität des maschinellen Lernens mit Produktionsdaten eines Finanzinstituts von übergeordnetem Interesse für das Finanzsystem.

Projektziele

Die Forschungskoooperation beinhaltet neben der explorativen Testphase zur deskriptiven Analyse der Risikodaten und der Potentialanalyse darauf aufbauend mehrere Teilprojekte. Insbesondere werden folgenden Punkte unter dem Thema „Risk Identification, Warning and Data Optimization“, das unter anderem aus den zukünftigen regulatorischen Notwendigkeiten abgeleitet ist, im Rahmen des SDI-C Projekts untersucht: die ex-ante Analyse der kausalen Zusammenhänge und darauf basierenden Vorhersage von Risikotreibern von Fonds und Wertpapieren, sowie der ex-post Analyse der Risikotreiber, die zu Risikogrenzüberschreitungen führten.

Ziel ist es ein automatisiertes Frühwarnsystem zur Identifikation wesentlicher Risikofaktoren und der kausalen Analyse ihrer Auswirkungen über mehrere Wertpapiere und Investmentvermögen hinweg zu entwickeln. Dieses Frühwarnsystem soll Risikomanager bei der täglichen Detailanalyse und Kommunikation der Risikotreiber mit den Portfoliomanagern, Kunden und dem Regulator unterstützen.

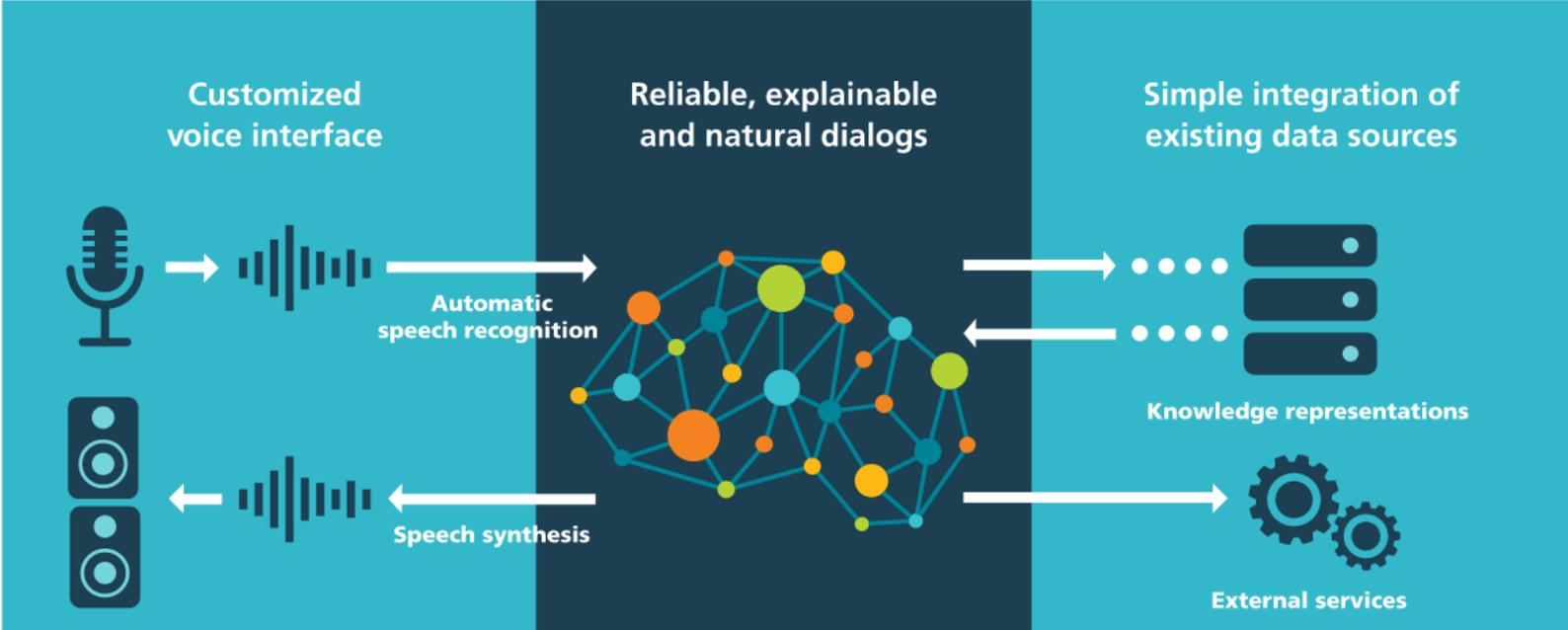
Projektergebnisse

Der wissenschaftliche Beitrag ist die Erklärbarkeit der Kausalität zwischen Risikofaktoren, die die Risiken in Investmentvermögen treiben. Aufgrund der zukünftigen regulatorischen Anforderungen sind diese Erkenntnisse für das Finanzsystem von übergeordneter Bedeutung. Zudem unterscheiden sich strukturierte Finanzdaten erheblich von den Produktionsdaten der Industrie, so dass die Erklärbarkeit der Kausalität die wesentliche Herausforderung zur Schaffung von regulatorisch notwendiger Transparenz ist. Vorgelagert ist die Erforschung der automatisierten Optimierung der großen Datenmengen in der Finanzindustrie

Die Daten umfassen für die erste Phase der explorativen Datenanalyse für die Machbarkeitsstudie, die dem eigentlichen Forschungsvorhaben vorgeschaltet ist, 167 tägliche Risikokennzahlen von 50 Investmentvermögen im Zeitraum Januar 2017 bis September 2018. Diese Daten sollen mit weiteren Investmentvermögen über einen längeren Zeitraum schrittweise erweitert werden. Zudem werden die Risikokennzahlen mit den Marktdaten der in den Investmentvermögen enthaltenen Wertpapieren von Thomson Reuters ergänzt. Die schrittweise Ergänzung der Daten dient der Erweiterung der Cluster-Analysen und des Trainierens des maschinellen Lernens. Auch werden in der Praxis weitere Datenquellen angebunden und verknüpft, weil dies aufgrund der regulatorischen Vorschriften, wie BCBS 239, für Finanzdienstleister verpflichtend ist.

Nächste Schritte

Der Verwertungsplan sieht vor, dass das KIT und die DWS Investment GmbH die Forschungsergebnisse für erste Anwendungen teilen und sowohl gemeinsame als auch individuell intern, weiterentwickeln für darauf aufbauenden Anwendungen, die die Erklärbarkeit der Kausalität zwischen Risikotreibern in Finanzdaten bedingen. Gemeinsame wissenschaftliche Publikationen sind vorgesehen.



Bildquelle: Fraunhofer SPEAKER Plattform

Cognitive Assistant:

Service-Assistent für Steuerberater, Wirtschaftsprüfer und Rechtsanwälte bei DATEV

Projektzeitraum

01.11.2019 – 30.06.2020

Projektpartner

Fraunhofer IAIS

Dr. Joachim Köhler (joachim.koehler@iais.fraunhofer.de)
 Prof. Dr. Jens Lehmann (Jens.Lehmann@iais.fraunhofer.de)

DATEV eG

Prof. Dr. Andreas Both (andreas.both@datev.de)

Die Kernidee des SDI-C Mikroprojektes "Cognitive Assistant" war es, eine Technologie zur dynamischen Erkennung von sogenannten "Intents" für Dialogsysteme zu entwickeln, welche "online", d.h. in dem Fall während des Betriebs einer Service-Hotline, ermittelt werden können. So kann ein Nutzer zum Beispiel melden, dass ein bestimmter Service nicht mehr funktioniert, vermehrt Fragen zu aktuellen Themen (z.B. Stichtage für die Einreichung von Dokumenten) gestellt werden oder sich juristische Grundlagen ändern (z.B. Erlassung von neuen Richtlinien).

Projektziele

Innerhalb des Projektes wurde der „Intent“ (Absicht) eines Nutzers mittels Natural-Language-Understanding-Prozessen abgeleitet und daraus mögliche auslösbare Aktionen bestimmt. Dies erfordert sowohl die Generalisierung von eingegebenen und erkannten Mustern als auch die semi-automatische Ableitung der „Intents“. Kann die Entwicklung dieser Technologie erfolgreich durchgeführt werden, so ist eine dynamische Erweiterung des Dialogsystems möglich, bei dem sowohl neue „Intents“ als auch dafür extrahierte Trainingsdaten dem für das Dialogsystem zuständigen Data Scientist vorgeschlagen werden. Ist die zunächst auf eine Domäne bezogene Analyse erfolgreich, kann die Technologie auf andere Domänen erweitert und so langfristig dynamische Märkte mit semi-automatischen Chatbots erschlossen werden.

Projektergebnisse

Zunächst wurde ein Datensatz erstellt und bereitgestellt, um die *Maschine Learning* Aufgabe zu lösen. Bei dem zu verwendenden Datensatz handelt es sich um Daten der öffentlich zugänglichen Info-DB, welche durch die 800 Mitarbeiter bei DATEV aus den Bereichen Service, Beratung und technisches Consulting bereitgestellt wurden. Die Daten wurden auf abgesicherten Servern gespeichert und für die weitere Verarbeitung im Kontext des Projektpartners DATEV bereitgestellt. Mit diesen Daten wurden DATEV und anwendungsspezifische „Intents“ mittels NLP-Verfahren trainiert, optimiert und evaluiert. Diese optimierte Intenterkennung wurde in eine Chatbot-Engine durch die Projektpartner integriert. Dabei kam erste Versionen der Fraunhofer SPEAKER-Plattform in einer on-premise Variante zum Einsatz.

Nächste Schritte und mögliche Verwertung

Eine mittel- und langfristige Verwertung ist durch die Integration der Technologie als Modul in der SPEAKER-Plattform gesichert. Dazu wird die Methoden zum Training des Intenterkenners als Modul in der Fraunhofer SPEAKER-Plattform integriert und mittelfristig Nutzern der Plattform bereitgestellt.

Referenzen

SPEAKER Plattform: <https://www.speaker.fraunhofer.de/>



Bildquelle: Martin Katler auf Unsplash

E-Scooter Detektion

Leichtes kamerabasiertes Verkehrserkennungssystem mit optimiertem neuronalem Netzwerk für hohe Zielerkennungsleistung auf leistungsschwachen Geräten

Projektzeitraum

01.11.2019 – 30.04.2020

Projektpartner

Karlsruher Institut für Technologie

Yexu Zhou (zhou@kit.edu)

Dr. Till Riedel (riedel@kit.edu)

Novel Sense GmbH

Dr. Markus Scholz (scholz@novelsense.com)

Die NovelSense UG befasst sich mit der Entwicklung innovativer Sensoranwendungen für die Mobilität der Zukunft. Innerhalb des SDI-C Mikroprojektes „E-Scooter Detektion“ sollen insbesondere neue Technologien für die infrastruktureitige Erfassung von Verkehrsdaten entwickelt und erprobt werden. Das Ziel ist die Entwicklung von Hardware und Softwareprodukten für Verkehrsplaner und Infrastrukturentwickler. Im vorliegenden Vorhaben sollen KI-Methoden zum Einsatz kommen, um eine kontinuierliche Erkennung und Zählung bestimmter Verkehrsteilnehmer mit einer Kamera zu ermöglichen. Ein hohes

Projektziele

Entwickelt wird ein akkubetriebenes und kamerabasiertes Verkehrszählgerät, welches sich durch seine kleine Baugröße, geringen Energieverbrauch und flexible Anbringungsmöglichkeit auszeichnet. Zum Einsatz kommt hier eine Kamera, welche in einem speziell entwickelten Gehäuse seitlich der Fahrbahn angebracht werden kann. Die Erkennung der Verkehrsteilnehmer erfolgt auf Basis der Kameradaten im Gerät mit einem hardwarebeschleunigten KI-Coprozessor mittels eines Neuronales Netzes. Der KI-Coprozessor erlaubt die Klassifikation und Detektion auf Echtzeit-Videodaten. Tracking und Geschwindigkeitsberechnung der erkannten Verkehrsteilnehmer wird anschließend durchgeführt. Fokus des Vorhabens im Rahmen der Smart Data Innovation Challenge ist die Realisierung der Erkennung von E-Scooter, die in Städten wie Berlin, Hamburg oder München bereits massenhaft eingesetzt werden. Hierbei soll sowohl die Durchfahrtzählung wie auch die Richtungserkennung realisiert werden.

Projektergebnisse

Die Detektion von E-Scootern wurde mittels Transfer-Learning mit Merkmals-Extraktion Anteilen vortrainierten Netzen sowie dem Fein-Tuning der Gewichte neu hinzugefügter Klassifikationsschichten erreicht. Die Herausforderung liegt in der Optimierung des eingesetzten Netzes, um unter Einhaltung der Ressourcenbeschränkungen des Hardware KI-Coprozessors den Verlust an Detektionsgenauigkeit zu minimieren, so sollen beispielsweise auch aktuelle Methoden zur Reduktion der Netzquantisierung untersucht werden.

Auf dem Markt für temporär installierbare Verkehrszähler werden aktuell so gut wie keine kamerabasierten Geräte eingesetzt. Die zum Einsatz kommenden Technologien umfassen hauptsächlich Induktion- oder Pneumatik-basierte Technologien, die einige Nachteile bzgl. Genauigkeit und Einsatzbereich aufweisen. Dies bezieht sich vor allem auf die Unterscheidung verschiedener Verkehrsteilnehmer. Ein besonderer Vorteil der vorgeschlagenen technischen Lösung ist dabei die Datenschutzkonforme Erhebung der Daten. Darüber hinaus wird die Erkennung von E-Scootern durch andere Anbieter am Markt, unabhängig von der verwendeten Technologie, derzeit noch nicht angeboten. Die für Training, Test und Validierung des neuronalen Netzes benötigten Daten sind aus frei verfügbaren Datensätzen sowie aus durch NovelSense erhobene Datensätze zusammengestellt und können durch Folgeprojekte kostenfrei verwendet werden.

Nächste Schritte

Kamerabasierte Verkehrserkennungssysteme bieten zahlreiche Vorteile gegenüber anderen Arten von Verkehrszählern, wie beispielsweise die Fähigkeit, eine breitere Palette von Spezies zu erkennen. Allerdings sind die aktuellen Bilderkennungsmodelle für die Zielerkennung oft zu groß, um auf leichten Geräten (Microcontroller) eingesetzt zu werden. Dadurch sind kamerabasierte Verkehrserkennungssysteme kostspielig und energieintensiv.





Bildquelle: Ryoji Iwata auf Unsplash

Personenstromerkennung

Vorhersagen von Personenströmen, wenn nur kurzfristige Trainingsdaten zur Verfügung stehen

Projektzeitraum

01.11.2019 – 30.04.2020

Projektpartner

Deutsches Forschungszentrum für künstliche Intelligenz (DFKI) GmbH
Dipl.-Inf. Gernot Bahle (Gernot.Bahle@dfki.de)

SIS Software GmbH

Dr. Tobias Franke (t.franke@sis-software.de)

Der Projektpartner SIS ist seit vielen Jahren im Bereich der Personenstromanalyse bei Großveranstaltungen kommerziell aktiv. Hierbei werden z.B. mithilfe von Smartphone Apps Bewegungsdaten der Besucher der Veranstaltung aufgezeichnet und in Echtzeit ausgewertet, um z.B. Rettungskräfte schnellstmöglich mit Informationen zu überfüllten Bereichen des Geländes zu versorgen. Innerhalb des SDI-C Mikroprojektes wurde erforscht werden, inwiefern Vorhersagen von Personenströmen möglich sind, wenn nur kurzfristige Trainingsdaten – anstelle von Daten aus mehreren vorherigen Iterationen der Veranstaltung – zur Verfügung

Projektziele

Mit ihrer Technik zur Erfassung des Verhaltens von Personenströmen mithilfe von Smartphones ist die SIS Software GmbH aktuell marktführend. Der nächste natürliche Schritt in dieser Thematik ist die Vorhersage von zukünftigen Bewegungen der Menschenmengen innerhalb eines kurzfristigen Zeitraumes. Der bisherige Stand der Technik benötigt hier jedoch entweder umfangreiche Referenzdaten, um ein valides Modell der Bewegungsströme zu erstellen oder ist schlicht und ergreifend nicht präzise genug, um kommerziellen Anforderungen Stand zu halten. Im Projekt soll basierend auf dem Datensatz eine neuartige Modellierung von Personenströmen erforscht werden, bei der simple Verhaltensmuster (z.B. das Verhalten nach dem Durchqueren eines Tores) anhand von aktuellen Trainingsdaten erlernt werden und diese Muster zu einem größeren Modell fusioniert werden. Ziel ist es, ein Modell zu erstellen, dass bereits nach einigen Stunden "Beobachtungsdauer" der Veranstaltung valide kurzfristige Vorhersagen für das Verhalten der Personenströme ermöglicht

Projektergebnisse

Für das Projekt wurde zwei Datensätze von den Veranstaltungen Rock am Ring aus den Jahren 2018 und 2019 (jeweils ca. 20.000 Teilnehmer über 3 Tage), sowie ein Datensatz von einer Großveranstaltung in Zürich (ca. 17.000 Teilnehmer) bereitgestellt. Der Datensatz aus Zürich kann auch zukünftig für Folgeprojekte weiterverwendet werden.

Im Rahmen des Projektes wurde die Bereinigung der Datenbasis (z.B. Trajektoriengenerierung aus GPS-Location-Fixes), das Mapping auf ein geeignetes geographisches Modell (im Fall von Zürich: Open Street Map Graph) und die Modellierung der eigentlichen Bewegungsinformation durchgeführt. Dazu wurden die Trajektorien verschiedenen Clustern zugeordnet, die jeweils einem Bewegungsprofil entsprechen. Eine agentenbasierte Simulation projiziert darauf aufbauend die Bewegungen von Personen in die Zukunft. Die zu prognostizierenden Personenströme ergeben sich aus der Aggregation der Einzelbewegungen.

Zur Auswertung wird der Veranstaltungsraum für die beschriebene Algorithmik in Zellen unterteilt. Die Validierung erfolgt durch den Vergleich der simulierten (zukünftigen) Personenzahl mit der tatsächlichen (aus den Daten). Prognosen können für einen Zeitraum von 10-20 Minuten erstellt werden. Je nach betrachtetem Gebiet (Randbereich, Kernbereich, Engstellen) haben die Vorhersagen eine unterschiedliche Qualität und Abweichungen zwischen 5% und 30% (Zellen mit wenigen Daten weisen die höchsten Abweichungen auf). Als Ergebnis konnte gezeigt werden, dass der Algorithmus vor allem für den Einsatz in hochfrequentierten Bereichen geeignet ist.

Nächste Schritte

Die Projektergebnisse sollen im Rahmen einer hochwertigen Veröffentlichung publiziert werden. Zudem sollen die Ergebnisse direkt in kommerzielle Produkte des Projektpartners einfließen. Dies wird direkt zu einer Erhöhung der Sicherheit bei Großveranstaltungen führen, da kritische Situationen (z.B. Verstopfungen an Toren und Durchgängen) frühzeitiger erkannt werden könnten als dies bisher der Fall ist.



Bildquelle Thomas Griesbeck auf Unsplash

EvaRD

Evaluation von Verfahren zur automatischen Rekonfiguration von Transportdrohnen

Projektzeitraum

01.11.2019 – 30.04.2020

Projektpartner

Fraunhofer IOSB-INA

Kaja Balzereit (kaja.balzereit@iosb-ina.fraunhofer.de)

Viafly GmbH

Die Fähigkeit, autonom Entscheidungen zu treffen und danach zu agieren ist ein Kernthema der Künstlichen Intelligenz. Dennoch sind Transportdrohnen heutzutage nicht in der Lage, Fehler in ihrem System selbständig zu erkennen und zu beheben. Dies führt zu Abstürzen von Drohnen und kann hohe Sach- und Personenschäden zur Folge haben. Um die funktionale Sicherheit von Transportdrohnen zu gewährleisten, stellt der assoziierte KMU-Partner Viafly dem Fraunhofer IOSB-INA Daten über fehlerfreie und fehlerhafte Drohnenflüge zur Verfügung. Das Fraunhofer IOSB-INA verwendet sein Know-How im Bereich der automatischen Fehlererkennung und -behandlung, um die Daten zu analysieren. Dabei sollen Fehler frühzeitig erkannt werden und rechtzeitig Gegenmaßnahmen eingeleitet werden. Die Ergebnisse dieser

Projektziele

Transportdrohnen werden in extremen Umgebungen eingesetzt und unterliegen daher Umwelteinflüssen wie zum Beispiel Temperaturschwankungen, Seitenwinden und Ähnliches, die die Dynamik der Drohne beeinflussen. Unruhige Flüge bis hin zu Abstürzen sind die Folge dieser Störungen.

Ziel dieses Projektes ist es, mithilfe von Methoden der modellbasierten Fehleranalyse und Rekonfiguration, einem Teilgebiet der künstlichen Intelligenz, die Auswirkungen dieser Störungen auf die Transportdrohne zu erkennen und geeignete Gegenmaßnahmen zu bestimmen, um den Flug der Drohne zu stabilisieren und Abstürze zu vermeiden. Zum Beispiel soll ein ausfallender Rotor der Drohne durch eine Anpassung der Motorleistungen an die anderen Rotoren abgefangen werden.

Projektergebnisse

Der wissenschaftliche Beitrag ist die Evaluation eines Rekonfigurationsverfahrens für Transportdrohnen. Dafür wurde ein Konzept entwickelt, das Transportdrohnen in die Lage versetzt, strukturelle und unvorhergesehene Fehler zu händeln. Fehler werden durch die Abweichungen von gelernten Modellen detektiert und an eine Rekonfigurationseinheit übergeben. Diese Rekonfigurationseinheit basiert auf logischen Kalkülen, die die Anforderungen für einen stabilen Flug beschreiben.

Der Ansatz wurde auf einer interaktiven Simulation eines Quadcopters simuliert. Es wurde ein Fehlerfall durch einen Motorausfall, der durch andere verfügbare Motoren kompensiert werden muss, in das System injiziert. Es konnte gezeigt werden, dass durch die Rekonfiguration der Absturz des Quadcopters verlangsamt wird.

Nächste Schritte

Das Fraunhofer IOSB-INA baut mit dem Projekt seine Kompetenzen im Bereich der Künstlichen Intelligenz für industrielle Anwendungen, insbesondere dem stetig wachsenden Markt der Drohnen, weiter aus. Diese Erfahrungen werden an Unternehmen, insbesondere KMU, transferiert.

Veröffentlichungen / Referenzen

Balzereit, K., Fullen, M., & Niggemann, O. (2020). A Concept for the Automated Reconfiguration of Quadcopters. In [LWDA \(pp. 180-191\)](#)



Bildquelle: Dieffenbacher

Auswertung von hochdimensionalen Messdaten

Auswertung von hochdimensionalen Messdaten aus der industriellen Fertigung für Prozessüberwachung und Qualitätsvorhersage

Projektzeitraum

01.11.2019 – 30.06.2020

Projektpartner

Fraunhofer IOSB

Dr. rer. nat. Constanze Hasterok (constanze.hasterok@iosb.fraunhofer.de)

Dr.-Ing. Julius Pfrommer (julius.pfrommer@iosb.fraunhofer.de)

Dr.-Ing. Thomas Usländer (thomas.uslaender@iosb.fraunhofer.de)

Dieffenbacher GmbH Maschinen- und Anlagenbau

Jürgen Woll (juergen.woll@dieffenbacher.de)

Stillstandszeiten von Maschinen und Anlagen in der Fertigung sind teuer und sollen möglichst vermieden werden. Verfahren der Prozessüberwachung eignen sich dafür, Abweichungen vom regulären Betrieb der Anlage zu erkennen. Dadurch können Störungen frühzeitig erkannt und behoben werden. Die Herausforderung des SDI-C Mikroprojektes „Auswertung von hochdimensionalen Messdaten“ bei der Identifizierung der Position der Störung im Prozess.

Projektziele

Durch Qualitätsvorhersagen anhand der eingestellten Prozessparameter soll die Produktion von Ausschuss vermieden. Oft kann die Qualität des hergestellten Produktes nur in aufwändigen Labormessungen bestimmt werden, sodass diese Information zeitverzögert vorliegt. Eine sofortige Vorhersage nach Messung der Prozessparameter spart hier wertvolle Zeit und vermeidet Ausschuss.

Der Beitrag dieses Projekts ist die Auswertung der Daten, um die oben genannten Problemstellungen zu adressieren. Mit dem ungeförderten Projektpartner Dieffenbacher wurden zunächst ML-Verfahren auf ihre Eignung hinsichtlich der Anforderungen im Kontext der Prozessüberwachung und Qualitätsvorhersage getestet und validiert. Geeignete Verfahren wurden ausgewählt, wobei prototypisch deren Nutzen in der Anwendung gezeigt wurde.

Ergebnisse

Im ersten Schritt wurden hochdimensionale Daten in enger Kooperation mit Prozessexperten aufbereitet, um deren Verwertbarkeit für maschinelle Lernverfahren und den Informationsgehalt zu sichern. Hierbei wurden Features in geeigneter Weise aggregiert und so die Dimensionen des Datensets reduziert.

Im zweiten Schritt wurden die aufbereiteten Daten für die kontinuierliche Überwachung des Herstellungsprozesses genutzt, wobei Verfahren des Maschinellen Lernens (ML) darauf trainiert wurden, reguläre Betriebszustände der Anlagen zu erkennen. Abweichungen vom regulären Betrieb können so durch den Algorithmus entdeckt und dem Bediener mitgeteilt werden.

Im dritten Schritt wurden zusätzlich zu den Prozessdaten, im Labor gemessene Daten der Produktqualität mit einbezogen. Dabei konnte gezeigt werden, dass die Prozessdaten einen signifikanten Informationsgehalt über die zu erwartende Qualität beinhalten, sodass Qualitätsvorhersagen von den Prozessdaten abgeleitet werden können.

Nächste Schritte

Die Verwertung der wissenschaftlichen Ergebnisse erfolgt durch das Fraunhofer IOSB. Verfahren und Algorithmen kommen in weiteren Forschungs- und Industrieprojekten zum Einsatz. Zusätzlich haben der Projektpartner Dieffenbacher und das Fraunhofer IOSB eine gemeinsame Patentierung der erarbeiteten Ideen und Ergebnisse angemeldet. Dieffenbacher ist im Bereich Maschinen- und Anlagenbau tätig und stellt Pressensysteme, sowie komplette Produktionsanlagen für die Holz-, Automobil-, Luftfahrt- und Recyclingindustrie her. Dieffenbacher beschäftigt über 1.600 Mitarbeiter und ist weltweit an 16 Produktions-, Service- bzw. Vertriebsstandorten tätig.

Veröffentlichungen

Pfrommer, J.; Hasterok, C.; Rehak J.; Woll, J; Störner, P.: Überwachung der Produktion von Werkstoffplatten, insbesondere Holzwerkstoffplatten, insbesondere unter Verwendung einer selbstorganisierenden Karte, Patentanmeldung DE102021206044.7, 2021, WO2022263458A1

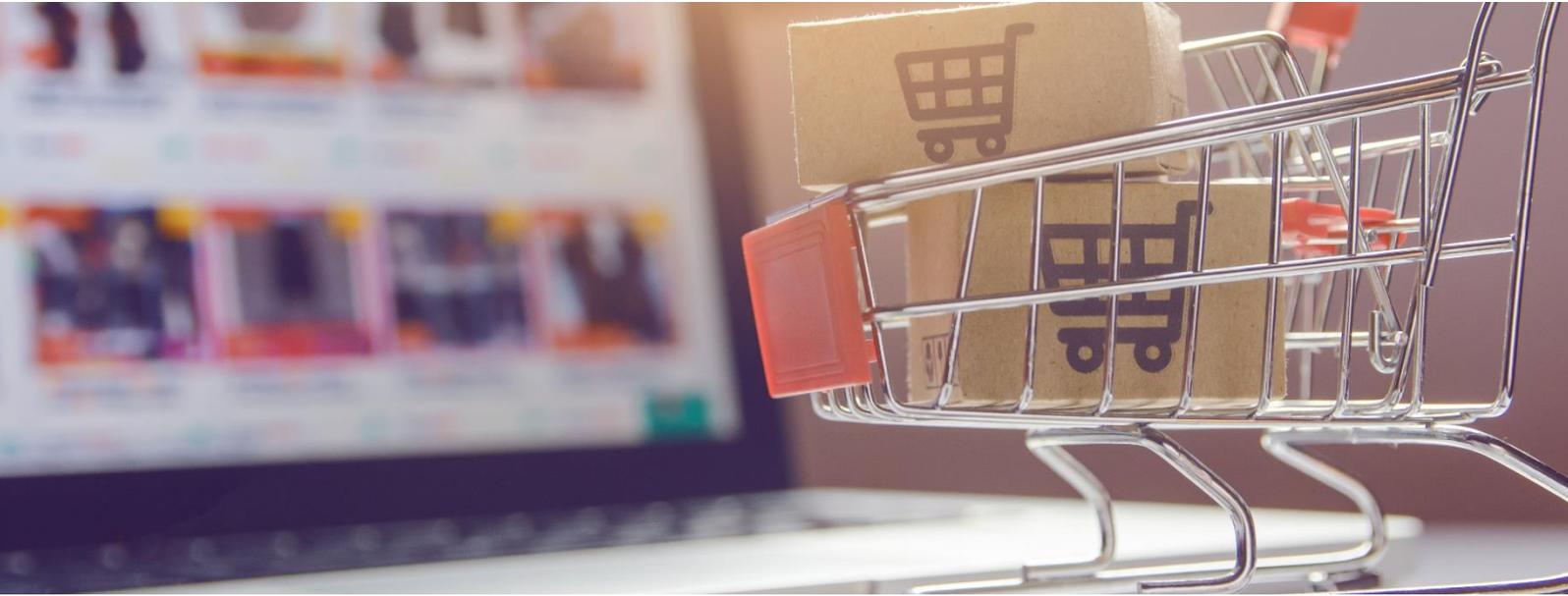


Mikroprojekte Call 2: Einreichungsfrist 08.03.2020

Nachdem die erste Projektrunde angelaufen ist und teilweise schon erste Ergebnisse sichtbar sind, wurde bereits die zweite Ausschreibungsrunde gestartet. Das Schwerpunktthema dieser Ausschreibungsrunde lautete „KI-Mikroprojekte zur Optimierung mit oder durch KI“. Der Ausschreibung war eine Diskussion wichtiger Trends vorausgegangen. Insbesondere die Themen AutoML und Hyperparametrisierung sollten in der Ausschreibung berücksichtigt werden.

Die Auswahl Sitzung fand am 30.03.2020 statt und war erstmals auch von der Ausrufung der Pandemie am 11.03.2020 geprägt, auch wenn zu diesem Zeitpunkt niemand die Auswirkungen der Pandemie auf die Mikroprojekte absehen konnte. Bereits zu diesem Zeitpunkt war jedoch klar, dass auf Anregung mehrerer Partner unabhängig von der Ausschreibung auch Rechenressourcen für Projekte im Zusammenhang mit der Pandemie zur Verfügung gestellt werden würden. Eine weitere Entwicklung zu diesem Zeitpunkt war, dass der Bedarf an KI-Rechenressourcen stetig zunahm. Zum Zeitpunkt des Auswahlgremiums war der damals einzige von IBM gespendete GPU-Rechenknoten bereits voll ausgelastet und die Zuteilung der Ressourcen eine Herausforderung. Etwas, das zu Beginn des Projekts noch nicht absehbar war.

Es wurden wieder 6 Anträge nach dem bewährten Verfahren ausgewählt, wobei ein Projekt aufgrund von neuen Priorisierungen beim Industriepartner als Folge der Pandemie nicht mehr zustande kam. Zwei weitere Projekte mussten deutlich verspätet gestartet werden, die meisten Projekte starteten zu diesem Zeitpunkt nicht wie geplant.



Bildquelle: NaMaKuKi auf vecteezy

Transper

Transparente Personalisierung im E-Commerce

Projektzeitraum

01.07.2020-31.01.2021

Projektpartner

Karlsruher Institut für Technologie

Dr. Michael Färber (michael.farber@kit.edu)

Dr. Anna Nguyen

econda GmbH

Dr. Philipp Sorg

Im SDIL-Mikroprojekt "TransPer: Transparente Personalisierung im E-Commerce" werden Aspekte der Kausalität, Robustheit und Unsicherheit von KI-Anwendungen im industriellen Umfeld betrachtet. Spezifisch fokussiert sich das Projekt darauf, wie Produktempfehlungen im eCommerce-Bereich transparenter gestaltet werden können, um eine bessere Kundenzufriedenheit zu erreichen und die Einhaltung rechtlicher Vorgaben zu gewährleisten.

Projektziele

Für viele mittelständische Unternehmen ist der Online-Shop das zentrale Kommunikationsmittel zu Kunden. Produktvorschläge sind dabei eine gängige Form der Personalisierung, die zur Steigerung des Umsatzes und der Kundenzufriedenheit führen. Die Vorschlagssysteme, wie sie für Online-Shops heutzutage eingesetzt werden, basieren in der Regel auf maschinellen Lernverfahren, bei denen anhand der gespeicherten Historie von Nutzern und ihrem Verhalten Empfehlungen für einzelne Nutzer abgeleitet werden (collaborative filtering). Methodisch kommen insbesondere tiefe neuronale Netzwerkarchitekturen zum Einsatz. Diese haben jedoch den signifikanten Nachteil, dass sie per se nicht transparent sind und es für den Endanwender nicht möglich ist, zu erkennen, warum ein neurales Netz eine bestimmte Entscheidung getroffen hat. Gemäß den Leitlinien der Europäischen Union und der Deutschen Bundesregierung hinsichtlich des Einsatzes von künstlicher Intelligenz werden in Zukunft jedoch Transparenz- und Erklärbarkeitskomponenten hinsichtlich der Personalisierung notwendig. In diesem Projekt wird daher ein Framework zur Erklärbarkeit von Produktvorschlägen am Beispiel des Personalisierungsservice von econda konzipiert, implementiert und evaluiert. Um die Einsatzmöglichkeiten von Methoden zur Erklärbarkeit anhand mehrerer Use Cases zu evaluieren, werden Modelle und Daten von econda-Kunden genutzt.

Projektergebnisse

Transper verwendet neuartige, auf schichtweiser Relevanzausbreitung basierende Erklärungsmaße und kann mit heterogenen Daten und komplexen neuronalen Netzwerkarchitekturen umgehen, wie z. B. Kombinationen mehrerer neuronaler Netzwerke in einer größeren Architektur. Wir wenden unseren Rahmen auf zwei reale Online-Shops an und bewerten ihn. Wir zeigen, dass die von TransPer bereitgestellten Erklärungen dazu beitragen, (i) die Vorhersagequalität zu verstehen, (ii) neue Ideen zur Verbesserung des neuronalen Netzes zu finden, (iii) den Online-Shops zu helfen, ihre Kunden zu verstehen, und (iv) rechtliche Anforderungen wie die von GDPR zu erfüllen.

Nächste Schritte

Die Ergebnisse aus dem Projekt bereiten für die econda GmbH die Basis, um Transparenzmechanismen bei Bedarf in Online-Shops und in Zukunft Aspekte von transparenter Personalisierung weiter zu betrachten, um insbesondere auf zukünftige regulative Anforderungen reagieren zu können.

Veröffentlichungen / Referenzen

Anna Nguyen, Franz Krause, Daniel Hagenmayer, Michael Färber
Quantifying Explanations of Neural Networks in E-Commerce Based on LRP
Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD'21), Seiten: 251-267, Springer, Juli, 2021



Bildquelle Uni Augsburg, smartaq.net

Smart Air Quality Sensor Calibration

KI-gestützte Interpolation zur entfernten Kalibrierung von Feinstaub-Sensoren

Projektzeitraum

01.07.2020-31.01.2020

Projektpartner

Karlsruher Institut für Technologie

Dr. Paul Tremper (paul.tremper@kit.edu)

Dr. Till Riedel (riedel@kit.edu)

GRIMM AEROSOL TECHNIK

Volker Ziegler

Umweltsensoren müssen regelmäßig kalibriert werden, um baubedingte und äußere Einflüsse korrekt in der Interpretation der Rohdaten zu berücksichtigen und entsprechend zu korrigieren. Low-Cost-Sensoren werden bevorzugt in Citizen-Science-Anwendungen eingesetzt, wobei eine langwierige Kalibrierung mit einem teuren Referenzgerät in der Regel der kostengünstigen und einfachen Konzeption vieler Citizen-Science-Anwendungen widerspricht. Im SDI.-C Mikroprojekt „Smart Calibration“ sollen entfernte Kalibrierungsmethoden auf Basis KI-gestützter Interpolation erforscht werden.

Projektziele

Im vom KIT TECO geleiteten Projekt "SmartAQnet" wurde ein offener Datensatz von Luftqualitätsdaten erstellt, der in seiner Zusammensetzung einzigartig ist: Über den Zeitraum von fast drei Jahren wurde ein engmaschiges, heterogenes Messnetzwerk in der Stadt Augsburg aufgebaut, das Daten aller Qualitätsgrade von Low-Cost Sensoren bis hin zu offiziellen Messstationen des Umweltbundesamtes enthält. Ein Messnetzwerk, wie es im SmartAQnet-Projekt aufgebaut wurde, bietet hier die Möglichkeit, durch seine räumliche wie zeitliche Dichte an Referenzstationen Sensoren vollautomatisiert mobil zu kalibrieren. Auf diesem einzigartigen Datensatz soll nun gezeigt werden, dass eine solche automatisierte, mobile Kalibrierung, gestützt durch ein engmaschiges Netz von Referenzstationen, qualitativ verwertbare Ergebnisse liefern kann. Das erlaubt SmartCities ihre Luftqualität mit kostengünstigen Low-Cost-Sensoren engmaschig zu überwachen, indem diese durch permanenten Live-Kontakt zu einem darunterliegenden, deutlich dünneren Netz qualitativ hochwertiger Referenzstationen stehen

Projektergebnisse

Im Projekt wurden Algorithmen entwickelt, die mithilfe von Maschinellem Lernen die Sensoren automatisch kalibrieren, ohne dass eine Kollokation notwendig ist. Aufgrund des komplexen, nichtlinearen Einflusses von Umwelteinflüssen sind KI-Methoden hier hervorragend geeignet, um stabile Ergebnisse zu erzielen. Um Projekt letztendlich KI-gestützte Interpolationstechniken untersucht. Es wurden verschiedene Interpolationstechniken verglichen, die auf zeitlich-räumlichem maschinellem Lernen basieren. Unter Verwendung des Datensatzes wird die Exposition auf einer Route durch die Stadt geschätzt und sie mit mobilen Messungen verglichen. Die Ergebnisse zeigen, dass verschiedene auf maschinellem Lernen basierende Interpolationsmethoden zwar recht unterschiedliche Ergebnisse liefern, die Validierung von auf maschinellem Lernen basierenden Ansätzen aber immer noch eine Herausforderung darstellt.

Nächste Schritte

Die indirekte Kalibrierung von Sensoren stellt noch immer eine nichtgelöste Herausforderung dar. Das KIT betrachtet das Problem weiterhin im Rahmen des Helmholtz HEPTA Projektes. Die experimentellen Low-Cost GRIMM EDM 80 Geräte der werden mittlerweile von einem Betreiberkonsortium weiterbetrieben, welche weiterhin kostengünstige Kalibrierungsmethoden im Feld testet.

Veröffentlichungen / Referenzen

Tremper, Paul, Till Riedel, and Matthias Budde. "Spatial interpolation of air quality data with multidimensional Gaussian processes." *INFORMATIK 2021* (2021)

Lücking, Markus, Niclas Kannengießer, Maurice Kilgus, Till Riedel, Michael Beigl, Ali Sunyaev, and Wilhelm Stork. "The merits of a decentralized pollution-monitoring system based on distributed ledger technology." *IEEE access* 8 (2020)



Bildquelle: Ulysse Pointcheval auf Unsplash

Feature Indexing für Parfumflaschen durch Deep Learning

Automatisierte Verbesserung der Parfümerie Pieper Produkt-Datenbank

Projektzeitraum

01.12.2020-31.05.2021

Projektpartner

Forschungszentrum Jülich

Prof. Dr. Ing. Morris Riedel

Stadt-Parfümerie Pieper GmbH

Rene Zinta, Referent Geschäftsführung

Die Kernidee des Forschungsprojektes „Feature Indexing für Parfümflaschen durch Deep Learning (DL) am Beispiel Parfümerie Pieper“ war eine automatisierte Verbesserung der Produkt-Datenbanken des Unternehmens Parfümerie Pieper. Der aktuelle Trend im Parfümgeschäft geht in die Richtung der Nutzung von Duftflakons in sehr speziellen Formen. Dabei tritt für den Kunden der Markenname nur noch im Hintergrund auf, viel wichtiger scheint eine Identifikation mit der spannenden Form und den individuellen Ausprägungen der Duftflakons.

Projektziele

Die Forschungsk Kooperation beinhaltet neben der Analyse der Bilddaten und Erprobung innovativer DL-Modelle, auch der Aufbau von zwei spezifischen Modulen damit Pieper die Indexierung der Datenbanken automatisiert und vor allem ohne viel technische Kenntnisse durchführen kann. Das erste Modul „Colour Detection Module“ erzeugt aus den Bildern der Parfümflaschen eine Liste der dominanten Farben. Filialkunden von Pieper nennen oft Farben in Verbindung mit Formen von Parfümflaschen unbekannter Hersteller. Das zweite Modul „Shape Detection Module“ nutzt DL zur Erfassung von „Formen-Features“ der Bilder von Parfümflaschen und vortrainierte DL-Netzwerke.

Projektergebnisse

Der erste wissenschaftliche Beitrag ist der gelungene Aufbau des „Colour Detection Module“ das verschiedene automatisierte Schritte enthält, um aus einem Bild einer Parfümflasche die nennenswerten Farben zu generieren. Da die Indexierung dazu helfen soll das Menschen ihr Parfum finden, mussten vor allem „Red, Green, and Blue (RGB)-Formate“ in „Hue, Saturation, and Lightness (HLS)-Formate“ konvertiert werden, die wesentlich besser verständlich für Menschen sind. Der zweite wissenschaftliche Beitrag ist das Modul „Shape Detection Module“ Ein vortrainiertes DL-Netzwerk auf [ImageNet](#) Daten hat bereits eine hohe Anzahl von semantischen Labels (bspw. Lampe, Teddy, Rotwein Dekanter, etc.), die hier benutzt werden. Daher wird ein DL-Modell verwendet das sich RESNET-50 nennt und auf ImageNet Daten trainiert wurde. Dieses vortrainierte (engl. pre-trained) DL-Netzwerk wird dann mit Parfümflaschen weiter trainiert (engl. Transfer-Learning). Dazu waren SDIL Computing Ressourcen notwendig da RESNET-50 Modelle mit bspw. 25.583.592 trainierbaren Parametern trainiert wurden.

Durch Einsatz dieser beiden Module konnte die Datenbank der Parfümerie Pieper mit neuen „Farben und Formen Features“ erfolgreich automatisiert angereichert werden.

Nächste Schritte

Umgesetzt in einem Prototyp, sollen die Projektergebnisse der Parfümerie Pieper helfen Kundenanfragen wie „ich suche ein Goldbaren Parfum“ schnell und gezielt zu antworten. Vor dem Projekt enthielt die Datenbank nur Informationen wie Preis, Artikelnummern, oder Beschreibungen des Duftes wie „blumig“, aber eher nur sehr begrenzte Angaben der Duftflakons.

Das Forschungszentrum verwertet die Projektergebnisse dieses Kooperationsprojektes bereits in dem EU-Projekt „Research on AI- and Simulation-Based Engineering at Exascale“ ([CoE RAISE](#)). Die Verwertung der Projektergebnisse ist auch in das [ON4OFF Projekt](#) eingeflossen, wo dann mittels dieser Ansätze auch spezifische Recommender und Services erstellt wurden.



Bildquelle: Apichet Chakreeyarut auf vecteezy.com

KI -Modelle für Smart Cities

Vorhersagen am Beispiel Einzelhandel und Smart Parking in Dortmund

Projektzeitraum

01.12.2020-31.05.2021

Projektpartner

Forschungszentrum Jülich

Prof. Dr. Ing. Morris Riedel

Osudio Deutschland

Stefan Herold, Geschäftsführer

Das SDI-C Mikroprojekt "KI-Modelle für Smart Cities am Beispiel Einzelhandel und Smart Parking in Dortmund" wurde stark von der COVID-19-Pandemie beeinflusst. Die vorgesehenen Datenlieferanten konnten aufgrund fehlender Veranstaltungen und geschlossener Einzelhandelsgeschäfte ihre Daten nicht zur Verfügung stellen. Dennoch wurde beschlossen, das Projekt fortzusetzen und ähnliche Ziele mit vorhandenen Daten der Stadt Dortmund zu verfolgen. Ein "Datensvisualisierungs-Dashboard" sowie Vorhersagemodelle auf Basis von Autoregressionsmethoden wurden entwickelt und befinden sich bereits in der Nutzungsphase.

Projektziele

Die Forschungsk Kooperation umfasst neben der Analyse vorhandener Datenquellen und der Erprobung verschiedener Prognosemodelle vor allem die Entwicklung eines flexiblen und erweiterbaren Datenvisualisierungs-Dashboards. Dabei sollen in erster Linie Daten der Stadt Dortmund zum Test verwendet werden, das Dashboard selbst soll aber später auch für andere Städte nutzbar sein. Darüber hinaus sollte das Dashboard vor allem auch für „Nicht-IT- oder KI-Experten“ nutzbar sein.

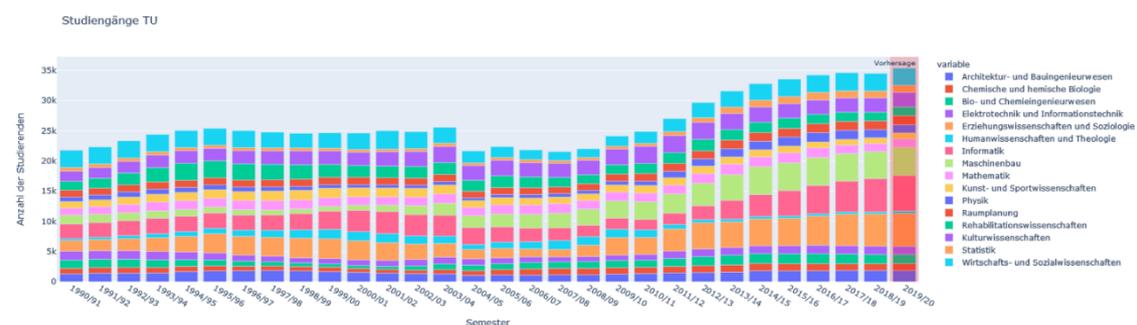
Projektergebnisse

Der erste wissenschaftliche Beitrag bestand in der Analyse vorhandener Datensätze für das Dashboard, einschließlich Echtzeitdaten der Stadt Dortmund und der Verknüpfung von Ereignis-, Verkehrs- und Verkaufsdaten. COVID-19 hatte einen starken Einfluss auf das Projekt, was zu einer Änderung der Ziele und Ergebnisse führte. Die Verfügbarkeit von öffentlichen Daten und Datenschutzprobleme schränkten langfristige Prognosen und den Einsatz komplexer maschineller Lernverfahren (ML) ein. Stattdessen wurden statistische Datenmethoden verwendet, um Vorhersagemodelle zu erstellen, z.B. die Anzahl der Studenten pro Semester, Autopendler, Radfahrer pro Stunde in Dortmund, Arbeitslosenzahlen und Flugbewegungen am Flughafen Dortmund. Dabei kam vor allem die Methode der "Autoregression" zum Einsatz, die einfache Vorhersagen auf Zeitreihendaten liefert.

Das Ergebnis ist ein benutzerfreundliches "Datenvisualisierungs-Dashboard", das die Autoregressionsmodelle und Daten flexibel darstellt. Die Projektpartner können nun viel einfacher die Daten einsehen und die Vorhersagemodelle betrachten. Trotz der COVID-19-Krise wurde das Projekt von allen Beteiligten als Erfolg gewertet.

Nächste Schritte

Das Forschungszentrum verwertet die Ergebnisse dieses Kooperationsprojektes bereits im CASA Healthcare Lab im Rahmen der Forschung zum Acute Respiratory Distress Syndrome (ARDS) und anderen medizinischen Zeitreihendaten. Die Verwertung der Projektergebnisse ist auch in das Projekt ON4OFF eingeflossen, wo weitere Zeitreihen im Bereich des stationären Einzelhandels analysiert wurden. Die DoData GmbH sowie die Osudio GmbH (jetzt SQLI Deutschland GmbH) verwerten die Ergebnisse in Studien und prüfen die Integration in zukünftige Dienstleistungen, auch möglicherweise über die Grenzen Dortmunds hinaus.





Bildquelle: Khunkorn Laowisit auf vecteezy.com

Chatbot Mining

Automatisierte Analyse und Optimierung von Kunden-Service Chatbots

Projektzeitraum

01.07.2020-31.01.2020

Projektpartner

Karlsruher Institut für Technologie

Ulrich Gwenuch (ulrich.gwenuch@kit.edu)

hsag Heidelberger Services

Daniel Schloss

Im Energiewirtschaftsbereich setzen viele kleine Energieversorger wie Stadtwerke Chatbots ein, um Kundenfragen zu beantworten und einfache Probleme zu lösen. Allerdings haben mittelständische Unternehmen Schwierigkeiten, die Chatbot-Daten zu analysieren und den Kundenservice zu optimieren. Das SDI-C Mikroprojekt „Chatbot Mining“ zielt darauf ab, eine automatisierte Alternative zur manuellen Analyse dieser Daten zu entwickeln, um Optimierungspotenziale zu identifizieren. Dabei werden Methoden des Maschinellen Lernens eingesetzt, um die Dialoge zwischen Chatbots und Kunden zu analysieren und die Ergebnisse

Projektziele

In der Energiewirtschaft setzen mittlerweile auch viele kleine Energieversorger wie Stadtwerke Chatbots ein, um häufige Fragen von Kunden zu beantworten und einfache Probleme automatisiert zu lösen. Erfahrungen zeigen jedoch, dass es insb. für mittelständische Unternehmen schwierig ist, die Daten aus dem Betrieb des Chatbots zu analysieren und für eine Optimierung ihres Kundenservices zu nutzen. Das Mikroprojekt zielt darauf ab, eine automatisierte Alternative zur manuellen Analyse dieser Daten aufzuzeigen, um die enormen Potenziale dieser Daten für die Optimierung ihres Chatbots und ihres Kundenservices allgemein zu heben.

Im Projekt soll aus dem Betrieb mehrerer Kundenservice-Chatbots Optimierungspotenziale für die Kundenschnittstelle identifiziert werden. Hierzu kommen Methoden des Maschinellen Lernens zum Einsatz, um die bereits existierenden Dialoge zwischen Chatbots und Kunden von Stadtwerken zu analysieren. Im zweiten Schritt ist die operative Umsetzung in konkreten Systemen beim Partner hsg Heidelberg Services AG geplant.

Projektergebnisse

Aus wissenschaftlicher Sicht stellt das vorgeschlagene Konzept einen interessanten Beitrag zur datengetriebenen Optimierung von Chatbots dar. Insbesondere der Ansatz, durch eine Emotionsanalyse von Dialogprotokollen eines Chatbots einen optimalen Zeitpunkt für die Übergabe/Weiterleitung des Kunden an einen menschlichen Kundenservice-MitarbeiterInnen vorherzusagen, ist neuartig und innovativ.

Der Ansatz wurde in einem Conversation-Mining-System implementiert, das zur visuellen Analyse von Kunden-Chatbot-Konversationen auf Prozessebene verwendet werden kann. Die Ergebnisse von vier Fokusgruppen-Evaluierungen im Anschluss an das Mikroprojekt zeigen, dass das System Chatbot-Entwicklern und -Betreibern helfen kann, Ansatzpunkte für Chatbot-Verbesserungen zu identifizieren. Die Forschung trägt damit neues Designwissen für Conversation-Mining-Systeme bei.

Nächste Schritte

Auf Basis der initialen Ergebnisse des Projektes hat das KIT zusammen mit hsg ein visuelles Analysetool für Chatbots prototypisch erstellt, welche es ermöglicht verschiedenste Verläufe zu analysieren. Die ersten Ergebnisse wurden nach dem Projekt bereits veröffentlicht.

Veröffentlichungen / Referenzen

Schloß, Daniel, Ulrich Gnewuch, and Alexander Maedche. "Towards Designing a Conversation Mining System for Customer Service Chatbots." (2022).

Schloß, Daniel, Juan DG Espitia, and Ulrich Gnewuch. "Designing a Conversation Mining System for Customer Service Chatbots." (2023).



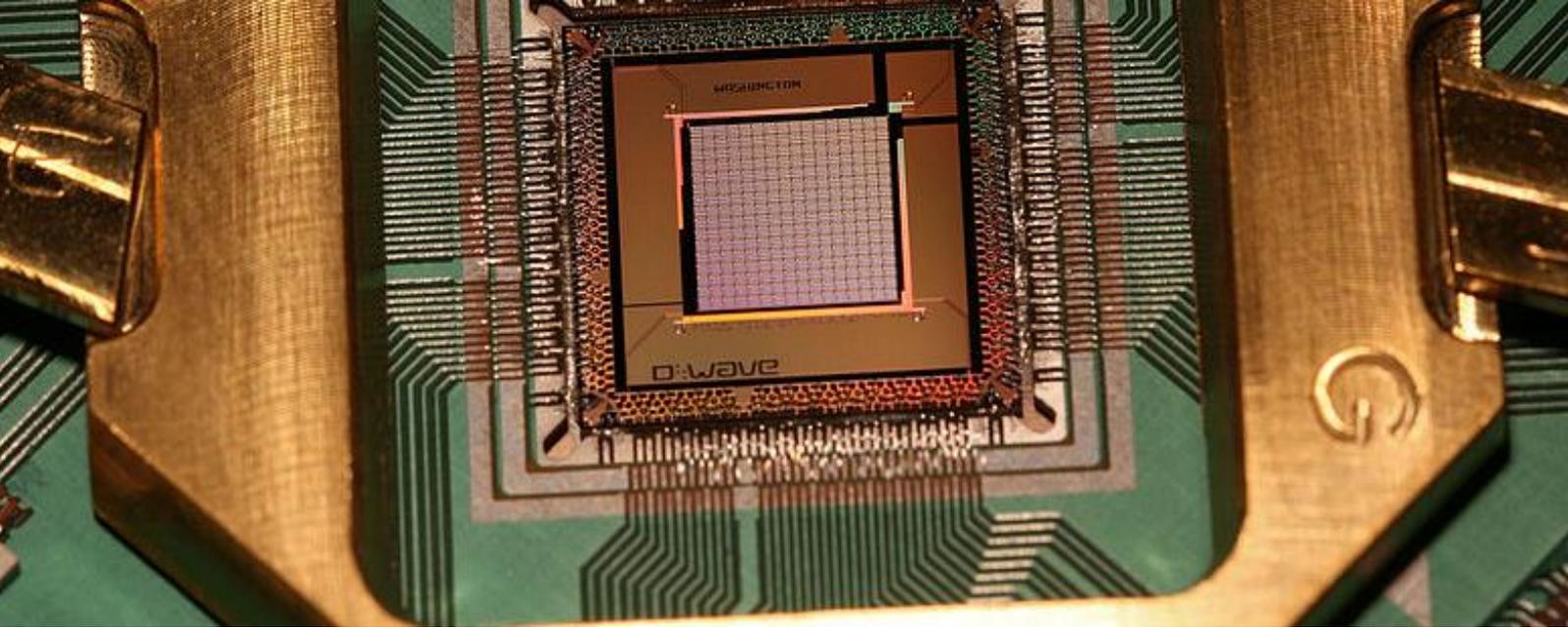
Mikroprojekte Call 3: Einreichungsfrist 25.10.2020

Die Auswahl Sitzung fand am 09.11.2020 im Rahmen der ersten gemeinsamen Smart Data Open Days und unter dem Eindruck der Präsentation der Ergebnisse der ersten Projekte statt. Die Veranstaltung wie die Sitzung fanden online statt.

Schon bei der Auswahl der Projekte zur Auswahlrunde zuvor war klar, dass Projekte zur Unterstützung der Bewältigung der Pandemie im Zentrum der dritten Runde von Mikroprojekten stehen müssten. Mittlerweile war jedoch auch dem Enthusiasmus einer schnellen technischen Bewältigung ein Realitätssinn gewichen. So wurden auch deutlich weniger Themenvorschläge hier eingereicht als erwartet. Interessanterweise beschäftigen sich die ausgewählten Projekte zu dem Thema mit wichtigen Themen, die vor der Pandemie wahrscheinlich niemand so stark in den Fokus gerückt hätte: den Homeschooling sowie der Datenbeschaffung während Epidemien. Das Thema Chipproduktion erreichte erst während der Laufzeit seine Relevanz im Kontext der Pandemie. Andere Leuchtturmprojekte beschäftigten sich mit großen Sprachmodellen, Foundational Models für Wearables und Anwendungen von Quantencomputing. Der Wunsch nach richtungsweisenden Projekten mit Leuchtturm-Charakter wurde in der Ausschreibung auch zuvor nochmals hervorgehoben.

Insgesamt wurden wieder 6 Projektanträge ausgewählt. Die zugehörigen Mikroprojekte konnten alle erfolgreich abgeschlossen werden. Die Aktualität der Themen zeigt in dieser Ausschreibungsrunde nochmals das Potential von Mikroprojektausschreibungen.

Insgesamt wurde auf Grund der organisatorischen Folgen der Pandemie, die Projektlaufzeit bei den meisten Projekten deutlich überschritten, was jedoch keine sonstigen Folgen hatte. Die Ergebnisse wurden mit großer Resonanz auf dem Smart Data Open Day 2021 präsentiert, der erstmalig von der SICOS BW GmbH im Unterauftrag des KIT organisiert wurde



Bildquelle: Mwjohnson0, D Wave TwoX auf Wikimedia Commons (CC-BY-SA 4.0)

Solving Accounting Optimization Problems in the Cloud

Nutzung von Quantum-Algorithmen für Konsistenzprüfung in Finanzberichten

Projektzeitraum

01.02.2021 – 31.12.2021

Projektpartner

Fraunhofer IAIS

Dr. Rafet Sifa (rafet.sifa@iais.fraunhofer.de)

PricewaterhouseCoopers GmbH Wirtschaftsprüfungsgesellschaft

Dr. Bernd Kliem (bernd.kliem@pwc.com)

Viele Anwendungen im automatisierten Prüfwesen sowie in der Analyse und Konsistenzprüfung von Finanzdokumenten können teilweise als das Teilsummenproblem formuliert werden: Gegeben eine Menge von Zahlen und eine Zielsumme, finde die Teilmenge von Zahlen, deren Summe der Zielsumme entspricht. Das Problem ist NP-schwer, und daher sind klassische Lösungsalgorithmen in vielen realen Anwendungen nicht praktikabel.

Wir behandeln das Problem als ein QUBO (quadratic unconstrained binary optimization) Problem und zeigen, wie Gradientenabstieg in Hopfield-Netzwerken zuverlässig Lösungen für sowohl künstliche als auch reale Daten findet. Wir skizzieren, wie dieser Algorithmus durch adiabatische Quantencomputer (Quantenannealer) und spezialisierte Hardware für digitales Annealing angewendet werden kann, und führen Experimente auf Quantenannealing-Hardware durch.

Projektziele

Unter dem Projektthema „Solving Accounting Optimization Problems in the Cloud“, untersuchen wir die Anwendbarkeit von Quantum-Computing-Methoden auf die Problemstellung der Konsistenzprüfung in Tabellen in Finanzberichten. Wir formulieren das Problem als QUBO und untersuchen verschiedene Lösungsansätze:

1. Optimierung von Hopfield-Netzwerken auf GPUs
2. Optimierung von Hopfield-Netzwerken auf FPGAs
3. Direkte Lösung auf Quantum-Annealing Hardware

Wir untersuchen dabei sowohl künstlich erzeugte Test-Daten als auch Daten aus realen Finanzberichten.

Projektergebnisse

Im Laufe dieses Projekts haben wir untersucht, wie das Teilsummenproblem eine wichtige Rolle bei der Automatisierung des Finanzprüfungsprozesses spielt und wie das Teilsummenproblem als eine bekannte Problemarchitektur neu formuliert werden kann, die durch die Anwendung von Gradientenabstieg auf die Energielandschaft von Hopfield-Netzwerken gelöst werden kann. Wir haben festgestellt, dass der vorgeschlagene Algorithmus zuverlässig korrekte Summenstrukturen für künstliche und reale Daten findet.

Wir haben einen Überblick über die Möglichkeiten adiabatischer Quantencomputer für die vorgeschlagene Aufgabe und ihre aktuellen Einschränkungen gegeben. Wir haben die Leistungsfähigkeit von Quantenannealern für das Teilsummenproblem bewertet und festgestellt, dass der Algorithmus für Probleme mit kleinem Wertebereich zuverlässig korrekte Lösungen findet. Mit der zunehmenden Leistungsfähigkeit von Quantenhardware wird der Einsatz von Quantencomputern für reale Anwendungen im Finanzbereich in naher Zukunft eine realistische Möglichkeit sein.

Nächste Schritte

Wir untersuchen die Integration des Algorithmus in bestehende intelligente Prüfungssoftware und planen weitere Experimente auf leistungstärkerer Quantum-Hardware.

Veröffentlichungen / Referenzen

David Biesner, Thore Gerlach, Bernd Kliem, Christian Bauckhage, Rafet Sifa:
Solving Subset Sum Problems using Quantum Inspired Optimization Algorithms with Applications in Auditing and Financial Data Analysis. ICMLA 2022:



Bildquelle Bosch

BERTI-4.0

Pre-trained Language Model für Conversational Agents im Kontext von Industrie 4.0

Projektzeitraum

01.1.2021 – 31.12.2021

Projektpartner

Fraunhofer IAIS

Dr. Joachim Köhler (joachim.koehler@iais.fraunhofer.de)

Dr. Diego Collarana (diego.collarana.vargas@iais.fraunhofer.de)

Bosch Research

Dr. Irlan Grangel-Gonzales

Die Nutzung natürlicher Benutzerschnittstellen über Sprachassistenten ist dabei entscheidend. Fortschritte im Natural Language Processing (NLP) haben das Verstehen natürlicher Sprache vorangetrieben, doch die direkte Anwendung auf Industrie 4.0-Einstellungen zeigt oft unbefriedigende Ergebnisse. Um dieses Problem zu überwinden, wurde im SDI-C Mikroprojekt „BERTI-4.0“ entwickelt - ein spezialisiertes BERT-Modell für Industrie 4.0-Texte. Das Modell wurde durch Vortraining auf allgemeinen Korpora und anschließendes Feintuning auf I4.0-Domänen-Korpora entwickelt und erfolgreich für Named Entity Recognition (NER) und Question Answering (QA) eingesetzt.

Projektziele

Das Hauptziel der Industrie 4.0 ist die Schaffung intelligenter Fabriken, in denen Cyber-Physikalische-Systeme und Menschen zusammenarbeiten. Die Bereitstellung natürlicher Benutzerschnittstellen über Sprachassistenten ist eine entscheidende Voraussetzung für die Verwirklichung dieser Vision. Mit den Fortschritten im Natural Language Processing (NLP) hat das Verstehen natürlicher Sprache an Popularität gewonnen, und Deep Learning hat den Einsatz effektiver Sprachmodelle gefördert. Die direkte Anwendung von NLP-Fortschritten auf Industrie 4.0 Settings führt jedoch oft zu unbefriedigenden Ergebnissen, da sich die Wortverteilung von Korpora aus dem öffentlichen Bereich zu I4.0-Korpora verändert hat.

Der Gesamtprozess des Vortrainings und der Fine-Tunings von BERTI-4.0 ist in Abbildung 1 dargestellt. Zunächst wurden BERTI-4.0 mit Gewichten aus dem Standard-BERT-Modell initialisiert, das auf allgemeinen Korpora vortrainiert wurde. Dann wird BERTI-4.0 noch einmal auf I4.0-Domänen-Korpora (Standards, z.B. RAMI, OPC UA und wissenschaftliche Textartikel zu I4.0) vortrainiert. Um die Wirksamkeit unseres Ansatzes zu zeigen, wird BERTI-4.0 anhand von der Entwicklung von zwei I4.0-Sprachassistentenkomponenten, genauer Named Entity Recognition (NER) und Question Answering (QA), verfeinert und evaluiert. Es wurden Pre-Training-Strategien mit verschiedenen Kombinationen und Größen von allgemeinen Domänenkorpora und I4.0-Korpora getestet. Außerdem wurden am Ende des Projekts REST APIs für die Partner bereitgestellt, um die Modelle service-orientiert zu nutzen.

Projektergebnisse

In dem Projekt wurde eine Datenpipeline entwickelt, um die Daten des Projektpartners zur weiteren Datenanalyse nutzbar zu machen. Die Erfassung von Daten über Industrie 4.0-Standards und -Richtlinien erforderte eine PDF-Lesekomponente, die kompliziert zu entwickeln war. Dazu wurden entsprechende PDF-Analysekomponenten entwickelt und angewendet. Das Projekt war in der Lage, ein neues BERT-Modell zu trainieren, das auf Text für Industrie 4.0 spezialisiert ist. So konnte es Antworten liefern und Entitäten aus diesem Bereich erkennen.

Nächste Schritte

Das Fraunhofer IAIS konnte seine Kompetenzen im Bereich der Künstlichen Intelligenz für Industrie 4.0-Anwendungen weiter ausbauen. Diese Erfahrungen sollen zukünftig an Unternehmen, insbesondere KMU, weitergegeben werden. Darüber hinaus wird das Fraunhofer IAIS seine wissenschaftliche Reputation durch die Veröffentlichung der Ergebnisse und der Auswertung weiter steigern. Die zukünftige Zusammenarbeit mit der Robert Bosch GmbH wird den Entwicklungsprozess von Conversational Agents und Dokumentenanalyse in der Industrie 4.0 weiter vorantreiben. Darüber hinaus kann die Robert Bosch GmbH ihre Expertise im Bereich Conversational AI weiter ausbauen und sich so von anderen Unternehmen am Markt differenzieren. Darüber hinaus kann in Folgeprojekten der Einsatz von Large Language Models, d.h. GPT-3, untersucht werden.



Bildquelle: Compare Fibre auf Unsplash

Cognitive Matchmaking bei Communities von Lernplattformen

Nutzung maschinellen Lernens zur Zuordnung von Fragestellungen und Experten in Lernplattformen

Projektzeitraum

01.1.2021 – 31.12.2021

Projektpartner

Fraunhofer IAIS

Dr. Rafet Sifa (rafet.sifa@iais.fraunhofer.de)

Daniel Jung Academy

Daniel Jung (dj@danieljung.education)

Community-Plattformen zur Beantwortung von Fragen wie Stackoverflow gehören zu den beliebtesten interaktiven Umgebungen im Internet, in denen Einzelpersonen Wissen austauschen können. Das Ermitteln von Experten, die Fragen beantworten könnten, ist eine der größten Herausforderungen dieser Plattformen. Im Rahmen des SDI-C Mikroprojektes „Cognitive Matchmaking“ hat das Fraunhofer IAIS für die Daniel Jung Academy ein Vorschlagssystem basierend auf state-of-the-art Methoden gebaut, welches Communityfragen entsprechenden Experten zuweist.

Projektziele

Unter dem Projektthema „Cognitive Matchmaking bei Communities von Lernplattformen“, untersuchen wir die Anwendbarkeit von Recommender Systemen zum Matching von Communityfragen und Experten in Lernplattformen. Bestehende Systeme stehen vor der Herausforderung tausende von Experten, Nutzern mit Millionen von Fragen zu matchen, welches einen potenziellen Anwendungsfall für large-scale Recommender Systeme darstellt.

Im speziellen werden zwei verschiedene Recommender Systeme für die Zuordnung von Fragen mit Experten untersucht:

1. Topic Modeling basierte Recommender Systeme (LDA-Rec)
2. Transformer-Embedding basierte Recommender Systeme (SBERT-Rec)

Projektergebnisse

Wir schlagen zwei verschiedene Ansätze für Empfehlungssysteme vor, nämlich LDA-Rec und SBERTRec. LDA-Rec nutzt das probabilistische Topic Modeling von LDA zur Repräsentation von Fragen und Benutzer innerhalb eines Topic-Raums. Die Unähnlichkeit von Frage- und Benutzerpaaren wird über die Kullback-Leibler-Divergenz geschätzt. Das SBERT-Rec Empfehlungssystem besteht aus einem SBERT-Modell zur Repräsentation von Fragen/Benutzern und einem anschließenden Cosinus Moduls zur Schätzung der Ähnlichkeit. LDA-Rec unterscheidet sich grundlegend von SBERT-Rec in der Art und Weise, wie die Repräsentationen geschätzt werden. Das LDA-Modell geht von einer geschlossenen Welt aus und lernt granulare Themen innerhalb eines Korpus, während das SBERT-Modell Weltwissen einbezieht, wodurch es weniger auf ein bestimmtes Thema fokussiert ist. Außerdem liefert LDA interpretierbare korpusinterne Themen, die wertvolle Einblicke, in die auf den Plattformen diskutierten, Themen geben.

Unsere Ergebnisse zeigen, dass SBERT-Rec in beiden Datensätzen LDA-Rec übertrifft bzgl. des durchschnittlichen Rank Scores. Während SBERT-Rec in einem Open-World-Szenario ohne Vorannahmen über die zugrunde liegenden Themen des Korpus besser abschneidet, findet LDA-Rec einzigartige Themen innerhalb eines bestimmten Closed-World-Korpus

Nächste Schritte

Im Hinblick auf zukünftige Forschung wäre es interessant, das autoregressive Sprachmodell Generative Pre-trained Transformer 3 (GPT-3) zu verwenden, das noch mehr Weltwissen einbezieht. Wir werden auch die Auswirkungen von Tags auf die Darstellung der Fragen untersuchen.

Veröffentlichungen / Referenzen

Stenzel, R., Lübbering, M., Ulusay, B., Uedelhoven, D., & Sifa, R. (2022). Matching Experts to Questions: A Comparison of Recommender Systems, LWDA 2022: 57-66



Bildquelle Forster Rohner AG auf Wikimedia commons (CC-BY-SA-3.0)

Matching Experts to Questions: A Comparison of Recommender Systems. LWDA 2022: 57-66

WearNet

WearNet DeepLearning Toolkit für Wearable Bewegungsanalysen

Projektzeitraum

01.1.2021 – 30.09.2021

Projektpartner

Deutsches Forschungszentrum für künstliche Intelligenz (DFKI) GmbH

Prof. Dr. Paul Lukowicz (Paul.Lukowicz@dfki.de)

norm—projects

Pascal Stüsser

Das Mikroprojekt "WearNet" wurde mit dem Ziel durchgeführt, modulare Deep Learning Systeme für Smart Clothing zu entwickeln. Diese Systeme sollten es Industriepartnern ermöglichen, Smart Textilien ohne großen Entwicklungsaufwand herzustellen. Dabei lag der Fokus auf der Modularität der Hardware und der dazugehörigen Software. Das Ziel bestand darin, innovative Deep Learning Modelle zu entwickeln, die mithilfe großer Mengen an annotierten Sensordaten trainiert werden, um vorab trainierte und leicht anpassbare Modelle zu erstellen. Dies ermöglichte es der Firma "Funktion" kundenspezifische Lösungen schnell zu*

Projektziele

Intelligente Kleidung, bei der die Elektronik in der textilen Struktur integriert ist, wird seit langem als zukunftssträftig gesehen. Trotz des großen Interesses bleiben intelligente Textilien derzeit allerdings ein Nischenprodukt. Dies ist sehr stark der Tatsache geschuldet, dass die meisten bisherigen Entwicklungen, vertikale, eng definierte Spezialanwendungen sind, die nicht von der "economies of scale" profitieren können.

Im Bereich der Hardware gibt es daher derzeit Bemühungen, breit einsetzbare Plattformen. Das Ziel des Projektes besteht darin, durch Training geeigneter Netze auf sehr große Datenmenge für das modulare Sensorsystem der Firma „Funktion*“ ein Toolkit bereitzustellen, das auf einer Vielzahl von typischen Problemen vor-trainiert ist und mit einer kleinen Menge von Trainingsdaten an verschiedene Anwendungen angepasst werden kann. Im Bereich der sensorbasierten Analyse von Aktivitäten und Bewegungen gibt es derzeit keine solchen allgemeinen Toolkits, welches im Projekt nun realisiert werden soll

Projektergebnisse

Im Rahmen des Mikroprojekts wurden innovative Deep Learning Modelle auf Basis der Transformer-Architektur entwickelt. Diese Modelle wurden mit Hilfe großer Mengen öffentlich verfügbarer, annotierter Sensordaten auf dem SDIL-Cluster trainiert. Durch die Berücksichtigung physikalischer Randbedingungen und den Einsatz von Adversarial Learning-Konzepten wurde ein hohes Maß an Unabhängigkeit von konkreten Nutzern und Sensoren erreicht. Die entwickelten Modelle erwiesen sich als äußerst leistungsfähig und übertrafen bestehende Systeme bei Weitem.

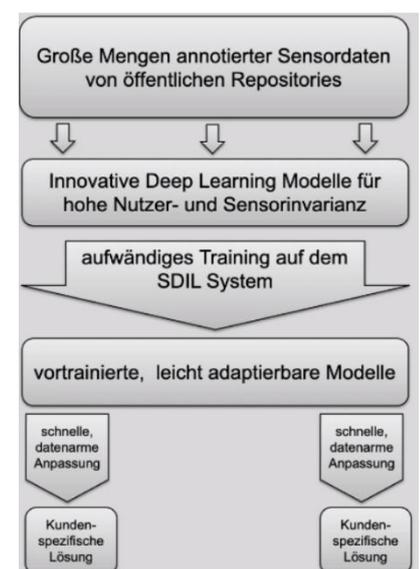
Die entwickelten Deep Learning Modelle bilden eine solide Grundlage für die modularen Smart Clothing Systeme von "Funktion*". Mit Hilfe der Ergebnisse kann die Integration der Modelle in die bestehende Hardware erfolgen, um ein voll funktionsfähiges System zu schaffen. Dabei war es wichtig, die Modularität beizubehalten, um den Kunden eine einfache Anpassung ihrer Smart Textilien zu ermöglichen.

Nächste Schritte

Nun müssen umfangreiche Tests und Validierungen durchgeführt werden, um sicherzustellen, dass die entwickelten Modelle den hohen Anforderungen an Leistung, Genauigkeit und Flexibilität gerecht werden. Die entwickelten Deep Learning Modelle stellen einen entscheidenden Wettbewerbsvorteil für die Firma "Funktion*" dar, da sie eine schnelle und maßgeschneiderte Entwicklung von kundenspezifischen Lösungen ermöglichen.

Veröffentlichungen / Referenzen

Fortes Rey, Vitor, Sungho Suh, and Paul Lukowicz. "Learning from the Best: Contrastive Representations Learning Across Sensor Locations for Wearable Activity Recognition." Proceedings of the 2022 ACM International Symposium on Wearable Computers. 2022.



Ben je in de twee weken voor je test benaderd door de GGD in verband met contactonderzoek? *

Nee, ik ben niet door de GGD benaderd in verband met contactonderzoek

Ja, ik ben wel door de GGD benaderd in verband met contactonderzoek

Bildquelle Coneno

EpiNet

Automatische Datenqualitätsverbesserung für partizipative Umfragesysteme unter Einsatz von Natural Language Processing

Projektzeitraum

01.1.2021 – 30.09.2021

Projektpartner

Deutsches Forschungszentrum für künstliche Intelligenz (DFKI) GmbH

Peter Hevesi (peter.hevesi@dfki.de)

Coneno GmbH

Marco Hirsch

Im SDI-C Mikroprojekt „EpiNet“ wird im Bereich epidemiologische Datenerfassung eine Software-lösung dafür genutzt werden, Validität und Zuverlässigkeit in der Erstellung und Kontrolle von Umfragen automatisiert zu verbessern. Diese KI-basierte Assistenzlösung soll mit Hilfe von Methoden aus dem Natural Language Processing, Domainexperten (Epidemiologen) befähigen, für die allgemeine Bevölkerung verständlichere und intuitivere Fragebogen zu erstellen

Projektziele

Das Mikroprojekt "EpiNet" hatte das Ziel, Methoden aus dem Natural Language Processing (NLP) zu untersuchen und einzusetzen, um die Datenqualität von auf Umfragen basierenden partizipativen Systemen zu verbessern. Konkret sollte erforscht werden, wie moderne Sprachtechnologien, insbesondere Transformer-Modelle, die Qualität von Fragen in Umfragebögen optimieren können. Die Untersuchung sollte zeigen, wie diese Methoden effektiv eingesetzt werden können, um bessere Fragen zu stellen und somit die Datenqualität in partizipativen Studien zu erhöhen.

Projektergebnisse

Im Rahmen des Mikroprojekts wurden verschiedene Aspekte sinnvoller Assistenzfunktionen untersucht, die zur Verbesserung der Datenqualität beitragen können. Dazu gehörten:

- a) Analyse der Verbreitung einzelner Wörter im Fragetext,
- b) Zuordnung von Fragen zu spezifischen Themen wie Gesundheit oder Politik,
- c) Klassifizierung, ob eine Frage gesellschaftlich sensitive Themen berührt, und
- d) automatische Erkennung der geeigneten Form des Antwortbereichs basierend auf dem Fragetext.

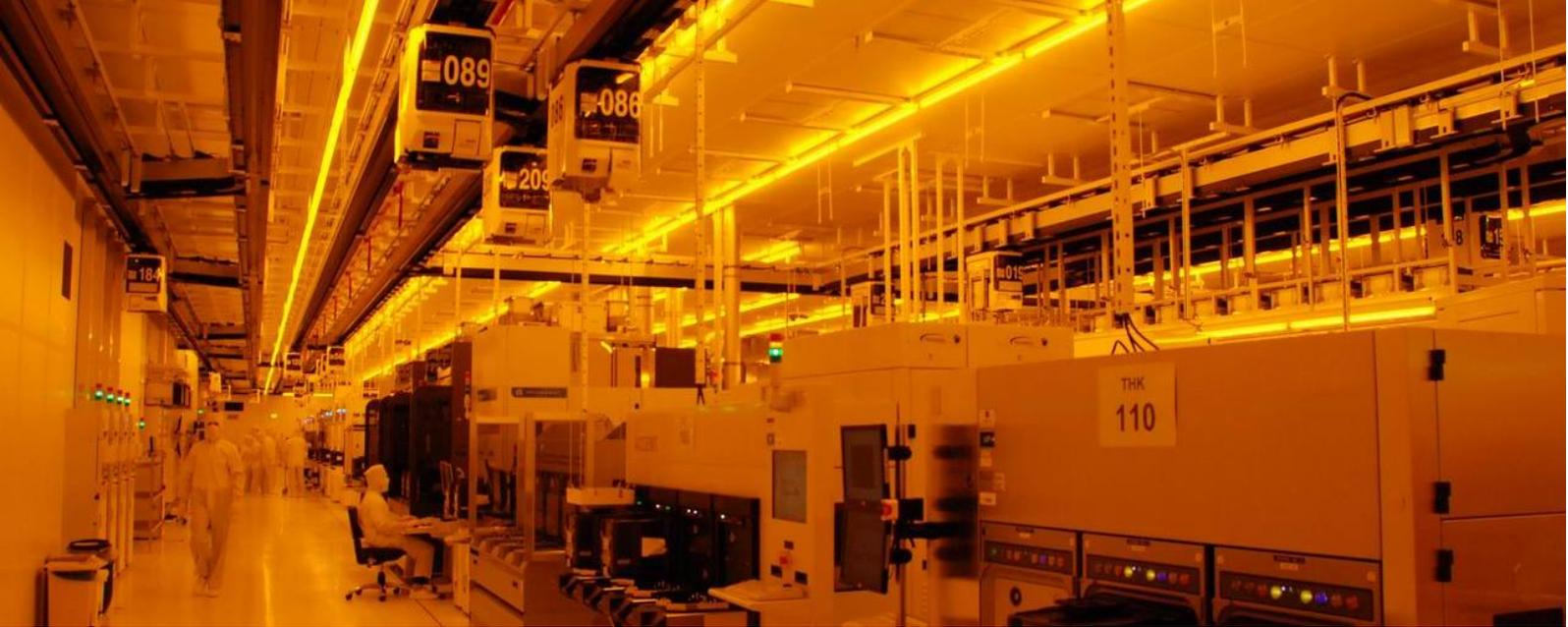
Es wurde festgestellt, dass die Verwendung von spezialisierten Transformer-Modellen, insbesondere des sogenannten "BERT"-Models, vielversprechende Ansätze für die Datenanalyse und Frageoptimierung bietet. Diese Modelle ermöglichen es, Ansätze des Transfer Learning zu nutzen, um die Gewichtung der Modelle auch mit kleineren Datensätzen feinzustimmen.

Die Methoden wurden auf Testdatensätzen evaluiert, wobei vielversprechende Ergebnisse erzielt wurden. Die Erkenntnisse und verwendeten Methoden wurden in einer umfassenden Masterarbeit dokumentiert und zusammengefasst.

Nächste Schritte

Um die erzielten Ergebnisse in der Praxis anzuwenden und zu erproben, wurden ausgewählte Methoden, insbesondere diejenigen für die Word-Verbreitungsanalyse, in enger Zusammenarbeit mit dem Industriepartner in eine Serveranwendung umgesetzt. Die Implementierung nutzt gängige Technologien und Schnittstellen, um sie einfach zu erweitern und in den Praxiseinsatz zu bringen.

Als nächster Schritt ist es wichtig, die Serveranwendung einem Praxistest zu unterziehen, um die Effektivität und Leistungsfähigkeit der Methoden in realen Umfrageumgebungen zu überprüfen. Dabei sollten potenzielle Anpassungen und Verbesserungen berücksichtigt werden, um eine optimale Datenqualität zu gewährleisten.



Bildquelle GlobalFoundries

LatentAgel4Wafer

Merkmalsräume zur Beschreibung von Zuständen in der Wafer-Produktion

Projektzeitraum

01.3.2021 – 31.12.2021

Projektpartner

Karlsruher Institut für Technologie

Dr. Till Riedel (riedel@kit.edu)

Ployplearn Ravivanpong, (Ployplearn.ravivanpoing@kit.edu)

GlobalFoundries

Dr. Alexander Lajn

Bisherige Prognosen über Alterung und mögliche Ausfälle basieren im Wesentlichen auf Erfahrungswerten erfahrener Prozessingenieure. Statistische und KI-gestützte Predictive Maintenance stoßen hier an ihre Grenzen (curse of dimensionality), da für die vorhandenen kontinuierlichen Prozessdaten keine Beschreibungen der zustandsabhängigen Effekte vorliegen. Ziel des SDI-C Mikrorprojektes „LatentAgel4Wafer“ war es daher, signifikante latente Variablen zu erlernen, um diese Effekte quantifizierbar zu machen.

Projektziele

Die Fab 1 von GLOBALFOUNDRIES in Dresden ist Leuchtturm der verbliebenen Mikroelektronikproduktion in Deutschland. Ein Wafer wird in über 1000 Arbeitsschritten über mehrere Monate gefertigt. Ein einzelner Fehler im Produktionsprozess reicht, um die funktionale Ausbeute (yield) dramatisch zu reduzieren. Bei einem solchen über viele Jahre gereiften Prozess ist die Anwendung von KI zur Optimierung eine Herausforderung. Im Erfolgsfall hat das Projekt auch in dieser Hinsicht Leuchtturmcharakter für ähnliche Anwendungen in anderen gereiften High-Tech-Bereichen.

Sowohl Alterung als auch potenzielle Ausfälle können (Stand heute) nur bis zu einem bestimmten Grad durch erfahrene Prozessingenieure auf Grund von Erfahrungswerten prognostiziert werden. Sie sind mit klassischen Werkzeugen der statistischen und KI-gestützten Predictive Maintenance nicht signifikant zu erfassen (curse of dimensionality). Typische Prozessdaten sind kontinuierlich erfasste Drücke, Gasflüsse, Beschleunigungsspannungen. Außer einzelner Grenzwerte, die nur den Fehlerfall beschreiben, existieren bis heute keine Beschreibungen zustandsabhängiger Effekte für diese Daten. Daher ist es das Ziel des Projektes signifikante latente Variablen zu lernen, die diese Effekte quantifizierbar machen.

Projektergebnisse

Dazu wurden im Projekt mehrere neuartige Lernarchitekturen entwickelt, um den Merkmalsraum auf Basis von Mustern in den Prozessdaten effizient zu "komprimieren". Dieser latente Merkmalsraum sollte schließlich den aktuellen Zustand einzelner Komponenten beschreiben, um Wartungsintervalle anzupassen und frühzeitige Diagnosen für systematische Probleme aufzuzeigen. Die Ansätze mittels Convolutional Autoencoder Modellen zeigten erste interessante Ergebnisse. Die entworfenen Long Short Term Memory Modelle konvergierten jedoch nicht. Eine Herausforderung stellte die Kodierung der teilweise unterabgetasteten Zeitreihen dar. Ein überraschender erster Erfolg bei der Erkennung von Anomalien wurde durch die Anwendung statistischer Analysen auf die als Baseline untersuchten Zeitreihen erzielt.

Eine wichtige Erkenntnis war, dass die effiziente Vorverarbeitung großer Sensor- und Logdaten (im Umfang von mehreren hundert GByte mittels Dask auf dem HTCondor-Cluster des SDIL) eine wichtige Grundvoraussetzung für die Exploration möglicher Merkmalsräume ist und sowohl noch viel Rechenzeit als auch Entwicklungszeit in Anspruch nimmt.

Nächste Schritte

Aufgrund des hohen Reifegrades der Prozesse in der Halbleiterindustrie geht es wirtschaftlich nur sekundär um eine Steigerung der Fertigungsqualität, sondern darum, die gleiche Qualität mit besser angepassten Maßnahmen zu erreichen. Dies kann die hohen Total Cost of Ownership für die Fertigungsanlage deutlich reduzieren. GlobalFoundries will dazu eigene Data Science Workflows zur Prozessüberwachung mit KI aufbauen.

Basierend auf den Ergebnissen entwickelt das KIT verschiedene Methoden im Bereich der skalierbaren Zeitreihenanalyse weiter und integriert diese in verschiedene KI-Werkzeuge.



Mikroprojekte Call 4: Einreichungsfrist 12.07.2021

Nachdem die Ausschreibungsrunde zuvor keinen klaren technischen KI-Fokus hatte, wurden für die vierte Ausschreibungsrunde mehrere Schwerpunkte vorgeschlagen, zum einen sollte „KI auf räumlichen Daten oder Graphen“ gefördert werden zum anderen wurde das Thema „Datensparsames Lernen“, insbesondere unter Einsatz großer Rechenressourcen als prioritär angesehen.

Das Auswahlgremium fand am 26. Juli 2021 statt. Parallel hatte das KIT bereits einen Aufstockungsantrag gestellt, um gefördert durch das Ministerium einen eigenen Rechencluster für das SDIL beschaffen zu können, nachdem die Ressourcen deutlich überlastet waren und auch das Angebot nicht mehr dem Bedarf vollständig entsprach. Die Hoffnung, die sich zum Teil bestätigen konnte, war, dass die ausgewählten Projekte bereits die neuen Rechenressourcen zumindest in der Endphase nutzen können.

Ausgewählt wurden ursprünglich 6 Projekte, wobei durch die ungleiche Verteilung der Projekte auf die Partner eine Imbalance an Fördermitteln auftrat, die zum Teil zu spät erkannt wurde. In der Folge der Verzögerungen konnte ein Projekt schließlich nicht durchgeführt werden. Bei einem weiteren Projekt scheiterte die vertragliche Einigung zwischen den Partnern, so dass zuletzt nur vier Projekte erfolgreich abgeschlossen werden konnten.

Abgeschlossen werden konnten drei Projekte, die sich im weitesten Sinne mit geolokalisierten Sensordaten beschäftigen. Sowie ein Projekt (mit Verzögerung) welches aus dem Kontext des KI-Engineering Kompetenzzentrums kommt und nur gefördert die Cloudangebote im SDIL genutzt hat.



Bildquelle: Botaurus auf wikimedia commons

Grumo

Datengetriebene Modelle zur Vorhersage von Grundwasserständen

Projektzeitraum

01.10.2021-01.06.2022

Projektpartner

Karlsruher Institut für Technologie

Dr. Tanja Liesch (tanja.liesch@kit.edu)

DISY GmbH

Dr. Andreas Abecker (andreas.abecker@disy.net)

Im SDI-C Mikroprojekt „Grumo“ werden datengetriebene Modelle zur Vorhersage von Grundwasserständen, die bisher nur punktuell für einzelne Messstellen entwickelt wurden, in die Fläche übertragen. Dies ist ein neuartiger Ansatz, der auch auf andere Domänen übertragbar ist. Die Ergebnisse sind für zahlreiche Zielgruppen wie Behörden (Wasser- und Umweltbehörden), Wasserversorger, Landwirtschaft, Forstwirtschaft und die Öffentlichkeit von Interesse. Eine entsprechende Aufbereitung und Darstellung z.B. mittels Webanwendungen oder Analysesystemen wie Cadenza der DISY GmbH.

Projektziele

Grundwasser bildet in vielen Regionen die Grundlage der Trinkwasserversorgung. Darüber hinaus versorgt oberflächennahes Grundwasser Pflanzen, bildet wertvolle Feuchtbiotope und speist Bäche und Flüsse. Eine dynamische, flächenhafte Modellierung und Vorhersage von Grundwasserständen ist für ein nachhaltiges Wasserressourcenmanagement von großer Bedeutung. Die bisherige Praxis beschränkt sich weitgehend auf die reine punktuelle Beobachtung von Grundwasserständen, Projektziel war die flächenhafte Modellierung von Grundwasserständen mittels neuartiger KI-Lernarchitekturen.

Als Inputdaten dienten sequenzielle flächenhafte meteorologische Inputparameter (z.B. Niederschlag aus Radardaten), räumliche Daten (wie Landnutzung, Beschaffenheit des Bodens) und Daten aus Netzwerken (Wasserstände/ Abflüsse in Oberflächengewässern), die auf den Zusammenhang mit punktuell gemessenen Zeitreihen zahlreicher Grundwasser-Messstellen trainiert werden. Fokusgebiet war dabei zunächst der baden-württembergische Teil des Oberrheingrabens, der in seinem Porengrundwasserleiter das größte und wichtigste Grundwasservorkommen Süddeutschlands beherbergt. Darüber hinaus wird die Übertragbarkeit auf ganz Baden-Württemberg im Rahmen des Projekts evaluiert.

Projektergebnisse

Für die Vorhersage vieler Messstellen gleichzeitig müssen die Eingaben für das Modell charakterisierbar sein. Dafür werden statische Parameter wie der Durchlässigkeitsbeiwert (kf-Wert), die Charakterisierung der Landnutzung, die Beschreibung des Aquifers und die Clustering-Features aus vorherigen Arbeiten genutzt.

Die Grundwasserstände werden mit Hilfe von Graph Convolutional Networks (GCNs) modelliert. GCNs sind ähnlich den bekannten Convolutional Neural Networks (CNNs) und ermöglichen die Verarbeitung von Informationen auf Graphen, die Knoten und Kanten repräsentieren. Der Graph wird anhand der geografischen Koordinaten oder mit Hilfe von multidimensionaler Skalierung erstellt. Das Temporal Graph Convolutional Network (T-GCN) kombiniert GCN und GRU, um räumliche und zeitliche Abhängigkeiten zu modellieren. Es ermöglicht eine autoregressive Modellierung der Grundwasserstände unter Einbeziehung der Grundwasserstände der Nachbarknoten. Für die Fusion externer Variablen wie dem Niederschlag kann ein Datafusion-Ansatz genutzt werden. Für das primäre Fokusgebiet, den baden-württembergischen Teil des Oberrheingrabens in Südwestdeutschland, war ein Datensatz von ca. 600 Grundwassermessstellen verfügbar. Es wurden verschiedene Modellansätze getestet: T-GCN (Temporal Graph Convolution Network), A3T-GCN (Attention Temporal Graph Convolutional Network) und MTGNN (Multivariate Time Series Graph Neural Network).

Nächste Schritte

Die prinzipielle Eignung des Ansatzes zur Modellierung der Grundwasserhöhen konnte gezeigt werden (auch wenn die Datenfusion mit meteorologischen Messwerten keine Verbesserungen brachte). Die Ergebnisse bieten DISY die Möglichkeit Aufbereitung und Darstellung z.B. mittels Webanwendungen oder Analysesystemen wie Cadenza.



Bildquelle: Vecteezy

Raumdaten-KI in der Agrarwirtschaft

Verbesserung von KI Modellen basierend auf räumlichen Daten in der Agrarwirtschaft

Projektzeitraum

01.10.2021-01.06.2022

Projektpartner

Forschungszentrum Jülich

Prof. Dr. – Ing. Morris Riedel,

Heliopas GmbH

Ingmar Wolff

Benno Ommerborn Avino

Die Kernidee des Forschungskooperationsprojektes „Raumdate-KI in der Agrarwirtschaft“ war die Unterstützung des KMU Heliopas, das bereits einige erste KI-Ansätze realisiert hatte. Die Kernexpertise von Heliopas liegt jedoch im Bereich der Bodenfeuchte mit Anwendungen in der Agrarwirtschaft (z.B. WaterFox App) und benötigte daher SDIL-Experten im Bereich des maschinellen Lernens (ML) gepaart mit Know-how der Recheninfrastruktur, um z.B. von Hyperparameter-Optimierungstools für ML-Modelle zu profitieren. Um vor allem auch innovative Ansätze des Deep Learning (DL) im Projekt nutzen zu können, war die Einbindung weiterer Datenquellen notwendig und daher ebenfalls Bestandteil des Forschungskooperationsprojektes.

Projektziele

Die Forschungskoooperation umfasst neben der explorativen Datenanalyse und der Erprobung innovativer Deep Learning (DL) Modelle sowie der Hyperparameteroptimierung vor allem den Aufbau einer effizienten und qualitativen „Preprocessing Pipeline“ zur Integration weiterer relevanter Datenquellen über die Zeit. Dazu sollten neue Datenquellen identifiziert und innovative neue ML- und DL-Modelle getestet werden, um die Vorhersagen im Bereich der Bodenfeuchte für das KMU Heliopas zu verbessern.

Projektergebnisse

Der erste wissenschaftliche Beitrag ist die erfolgreiche Entwicklung der „Preprocessing Pipeline“, die Messdaten von verschiedenen Messstationen und Satelliten verwendet. Dabei handelt es sich um Zeitreihen mit unterschiedlichen Messintervallen. In den ersten Schritten der Pipeline wurden die Daten von unnötigen Stationen bereinigt (engl. remove stations) und die Qualität der Daten aufbereitet (engl. clean labels). Anschließend wurden Lücken in den Zeitreihen durch verschiedene „Resampling“-Techniken geschlossen und mit Hilfe eines „Anchor-Features“ gefiltert. Im Transformationsschritt wurden frequenzbasierte Daten in eine Frequenzform transformiert (engl. data transformation). Zuletzt wurden die Daten in Zeitfenster aufgeteilt (engl. split time windows). Unterschiedlich große Zeitfenster wurden analysiert und ausgewählt, um diese im Hyperparameter Tuning final verwenden zu können.

Die Optimierung sowie der Einsatz innovativer DL-Modelle sind weitere wissenschaftliche Beiträge zur „Preprocessing Pipeline“. Die Projektergebnisse haben dem KMU Heliopas geholfen. Ein Zitat von Benno Avino (Heliopas CTO & Mitbegründer): „Das SDI-C Projekt hat es uns ermöglicht, uns auf die Verarbeitung der Satellitenbilder zu konzentrieren und die Modellierung und das Training in fähige Hände zu geben.... Darüber hinaus hat das Hyperparameter-Tuning zu erheblichen Verbesserungen unseres Modells geführt.“

Ein letzter Teil der Projektergebnisse war jedoch auch sehr wichtig: Die Integration neuer, qualitativ hochwertiger Datenquellen erfordert Fachwissen auf dem Gebiet der Bodenfeuchte. Das Projekt hat deutlich gemacht, dass mehr Fachwissen auf dem Gebiet der Bodenfeuchte wünschenswert wäre.

Nächste Schritte

Die Forschungsergebnisse werden bereits in anderen Projekten genutzt, wie dem EU-Projekt "Research on AI- and Simulation-Based Engineering at Exascale" ([CoE RAISE](#)) und im FZJ CASA Healthcare Lab für die Forschung des Acute Respiratory Distress Syndrome (ARDS). Die Firma Heliopas plant die Integration der Ergebnisse in ihre Produkte.



Bildquelle: Gemeinde Kirchheim b. München

Airformer

Transformer Basierte Repräsentation Spatio-Temporale Physikalischer
Randbedingungen in Deep Learning Systemen für Luft-Qualität Vorhersagen

Projektzeitraum

01.10.2021-01.06.2022

Projektpartner

Deutsches Forschungszentrum für künstliche Intelligenz (DFKI) GmbH

Prof. Dr. Paul Lukowicz (Paul.Lukowicz@dfki.de)

Gemeinde Kirchheim b. München

Tobias Schock

Das SDI-C Mikroprojekt „Airformer“ beschäftigt sich mit räumlich und zeitlich verteilten Luftqualitätsdaten, wobei die Analyse auf Zusammenhängen in mehrdimensionalen, multimodalen (Luftqualität, Topologie, Verkehrsdaten) Datenräumen stattfindet. Mit der Adaption des Transformer-Ansatzes mit Einbettung physikalischer Randbedingungen im Aufmerksamkeitsmechanismus wurde hier eine neue Lernmethode angestrebt.

Projektziele

Das Projekt "Airformer" hatte zum Ziel, räumlich und zeitlich verteilte Luftqualitätsdaten zu analysieren, um Vorhersagen und "what-if Szenarien" zu erstellen. Dabei fand die Analyse in mehrdimensionalen und multimodalen Datenräumen statt, die verschiedene Datenquellen wie Luftqualität, Topologie, Verkehrsdaten, Geräuschanalyse, Wetter und Wirtschaftsaktivität integrierten. Das Hauptaugenmerk lag auf der Anwendung von Transformer-basierten Deep Learning Ansätzen zur Vorhersage und Analyse der Luftqualitätsentwicklung in Smart Cities.

Projektergebnisse

Dazu wurden Transformer-basierte Deep

Learning Ansätze für die Vorhersage und Analyse der Luftqualitätsentwicklung in Smart Cities adaptiert. Die wichtigsten Innovationen sind:

- Die Verwendung des Transformer Konzepts an sich, das bisher in diesem Bereich nicht verwendet wurde (derzeit gibt es zwar erste Ansätze, die das Attention Konzept verwenden, jedoch nur im Zusammenhang mit LSTMs).
- Die Einbettung physikalischer und semantischer Randbedingungen in den Aufmerksamkeitsmechanismus des Transformators im Sinne des Konzepts der Physically Informed Neural Networks
- Ein Transfer-Learning-Ansatz, bei dem die Einbettung als allgemeine Repräsentation, der für die Luftqualität in Ballungsräumen relevanten Zusammenhänge aus großen Mengen öffentlich verfügbarer Daten vortrainiert wird, um dann eine schnelle und datensparsame Anpassung an neue Umgebungen zu ermöglichen.

Die Modelle wurden am SDIL-Cluster auf großen existierenden Luftqualitätsdatensätzen entwickelt und vortrainiert und dann mit historischen Daten aus Kirchheim an den konkreten Anwendungsfall angepasst.

Nächste Schritte:

Die entwickelten Modelle sollen den Projektbeteiligten in Kirchheim zur Verfügung gestellt werden, insbesondere der Firma HawaDawa, um sie für den weiteren Einsatz in der Stadt und anderen Projekten nutzen zu können.

Als nächster Schritt wird die Anwendung der Modelle auf die gesammelten Daten im Rahmen des Smartheim Projektes in Kirchheim erfolgen. Diese Daten umfassen Informationen von insgesamt 14 Luftqualitätssensoren sowie MAC-Sensoren zur Verkehrsverfolgung, Geräuschsensoren als breit angelegte Aktivitätsinformation und die Stadttopologie, einschließlich der Verteilung von Gewerbeaktivitäten.

Um die Effizienz und Qualität der Modelle kontinuierlich zu verbessern, wird auch die Erweiterung und Verbesserung der Sensoren und Daten in Betracht gezogen. Dadurch können die Modelle aktualisiert und auch anderen Folgeprojekten zur Verfügung gestellt werden.



Bildquelle: Siempelkamp

KI-Turbine

KI-basierte Betriebsüberwachung von Turbinen

Projektzeitraum

01.05.2022-01.09.2022

Projektpartner

Fraunhofer IOSB

Dr. Christian Kühnert (christian.kuehnert@iosb.fraunhofer.de)

Siempelkamp NIS Ingenieurgesellschaft mbH

Martin Allin (martin.allin@siempelkamp-nis.com)

Das SDI-C Projekt „KI-Turbine“ wurde im Rahmen des KI-Engineering-Kompetenzzentrums CC-King als sogenannter Transfer-Check konzipiert. Kern war die Fragestellung, wie ML-Modelle in den produktiven Einsatz gebracht werden können. Daher hat das Fraunhofer IOSB mit dem Smart Data Innovation Lab zusammengearbeitet, über es den Zugriff auf eine cloud-basierte kollaborative Arbeitsumgebung erhielt. Hierüber konnten die Modelle entwickelt und später für den Betrieb ausgebracht werden.

Projektziele

Die Firma Siempelkamp betreut große Dampf- und Gasturbosätze, für welche über viele Jahre ein Monitoringkonzept basierend auf hochaufgelöster Schwingungserfassung entwickelt wurde. Aktuell werden mit Hilfe von softwareseitig definierten Toleranzbändern Abweichungen einzelner Messgrößen vom Normalbetrieb erkannt. Durch eine detaillierte Auswertung der Daten von Prozessexperten kann dann eine Anlagendiagnose erstellt und Schaden an Maschinen frühzeitig diagnostiziert oder verhindert werden. In Zukunft soll die Überwachung von Turbinen jedoch verstärkt rechnergestützt erfolgen.

Die Kernidee des Projektes war es, ein KI-basiertes Assistenzsystem zu entwickeln, welches das Normalverhalten der Gasturbosätze erlernt und bei auftretenden Anomalien eine Rückmeldung an den Experten gibt. ML-Modelle wurden eingesetzt um automatisiert das normale Anlageverhalten zu lernen und anormales Verhalten präzise zu erkennen. Wichtig hierbei ist, dass die ML-Modelle möglichst automatisiert und robust gelernt und in das Assistenzsystem integriert werden können. Das KI-basierte Assistenzsystem sollte zudem skalierbar sein, damit ML-Modelle von neuen Anlagen mit geringem Aufwand eingebunden werden können.

Projektergebnisse

Ziel des Einsatzes von AutoML war es, die Arbeitszeit für einen Data Science Experten zur Erstellung einer ML-Pipeline zu reduzieren. Da im Projekt unterschiedliche Turbinen von Siempelkamp betrachtet wurden, bedeutet dies, dass auch mehrere ML-Modelle trainiert und überwacht werden mussten. Durch den Einsatz von AutoML-Verfahren entfiel eine unter Umständen aufwändige manuelle Parametrierung der Modelle. Somit konnten die ML-Modelle schneller für die einzelnen Turbinen in Betrieb genommen werden.

Nächste Schritte

TransferCheck unter Unterstützung ermöglichte der Firma Siempelkamp NIS Ingenieurgesellschaft, einen Einstieg in den Einsatz von KI in Ihrem Tätigkeitfeld zu finden. Dabei hat das Projekt die Grundlagen gelegt die nötig sind, um die KI in den realen Praxisbetrieb zu überführen: von der statistischen Datenauswertung bis hin zur Anwendung des ML-Modells in der Cloud.

Die Kooperation mit dem Kompetenzzentrum CC-King mit dem SDIL soll zukünftig intensiviert werden, um so integrierte Transferleistungen anbieten zu können

Veröffentlichungen / Referenzen

Interview Christian Kühnert zu CC-King Transfer-Check: <https://www.ki-engineering.eu/de/praxisbeispiele/transferchecks/ueberwachung-dampf-gasturbosaeetze-cloud-siempelkamp.html>



Mikroprojekte Call 5: Einreichungsfrist 20.02.2022

Innerhalb der letzten Ausschreibungsrunde stand nun die vollkommen neu aufgestellte SDIL-Plattform zur Verfügung. Ziel war es daher insbesondere die effektive Nutzung der GPU-Ressourcen in den Vordergrund der Ausschreibung zu rücken. Daneben wurde das Cloudangebot, was innerhalb der Projektlaufzeit entwickelt wurde nochmal in den Vordergrund gerückt. Zudem war das Projekt zu diesem Zeitpunkt bereits mittelneutral verlängert, so dass die Projekte erst mit Abschluss des Jahres fertiggestellt werden mussten, wobei ein Ende im Oktober angestrebt wurde. Die spätere Frist ermöglichte es die Einrichtung des Clusters abzuschließen.

Ausgewählt wurden 6 höchst unterschiedliche Projekte, die die technischen Möglichkeiten ganz verschieden nutzen von Transformermodellen bis hin zu Reinforcement-Learning. Alle sechs Projekte konnten erfolgreich durchgeführt werden und erfolgreich die neuen Möglichkeiten nutzen. Die Ergebnisse wurden zusammen mit den Ergebnissen des 4. Calls im Januar 2023 auf dem „Smart Data Open Day 2022“ präsentiert.



Bildquelle Vintus

Hipwings

Deep Learning Ansätze für intelligente ergonomische Sitzmöbel

Projektzeitraum

01.10.2021-01.06.2022

Projektpartner

Deutsches Forschungszentrum für künstliche Intelligenz (DFKI) GmbH

Prof. Dr. Paul Lukowicz (Paul.Lukowicz@dfki.de)

Vintus GmbH

Dr. Johannes Wappenschmidt

Das SDI-C Mikroprojekt „Hipwings“ zielt darauf ab Deep-Learning-Ansätze für die Vorhersage und Analyse von Benutzerhaltung und -aktivitäten auf einem intelligenten Stuhl zu entwickeln. Ziel ist die Entwicklung von transformer-basierten Erkennungssystemen, die Informationen aus dem Bewegungssensor und das Motordrehmoment der Sitzsteuerung kombinieren, um die Haltung und Aktivität des Benutzers in Echtzeit zu erkennen. Die Informationen werden mit dem Wissen von Ergonomieexperten kombiniert, um die Körperhaltung zu beeinflussen, sowohl in Form von aktiven Formänderungen durch den im Stuhl eingebauten Motor als auch durch Benachrichtigungen über eine Smartphone-App. Methodisch werden Transformer-Decoder-Encoder Methoden für sensorbasierte Erkennung menschlicher Aktivitäten (Human Activity Recognition, HAR) weiterentwickelt, insbesondere im Hinblick auf die Einbindung physikalischer und biomechanischer Randbedingungen.

Projektziel:

Das Hauptziel des Projekts "Hipwings" bestand darin, Deep Learning Architekturen für die sensorbasierte Erkennung menschlicher Aktivitäten zu verbessern, was ein bedeutendes und ungelöstes Forschungsproblem im Bereich der Künstlichen Intelligenz darstellt.

Projektergebnisse:

Im Rahmen dieses Projekts wurden innovative Deep-Learning-Ansätze entwickelt, um die Vorhersage und Analyse von Benutzerhaltung und -aktivitäten auf einem intelligenten Stuhl zu ermöglichen. Die wichtigsten Ergebnisse waren:

- Entwicklung von Transformer-basierten Erkennungssystemen: Diese Systeme bauten auf den umfangreichen Vorarbeiten des Antragstellers im Bereich der sensorbasierten Aktivitätserkennung mit Transformer-Architekturen auf. Sie kombinierten Informationen aus dem Bewegungssensor und dem Motordrehmoment der Sitzsteuerung, um die Haltung und Aktivität des Benutzers in Echtzeit zu erkennen.
- Kombination von Benutzeraktivitätsinformationen mit ergonomischem Fachwissen: Die erfassten Aktivitätsdaten des Benutzers wurden mit dem Wissen von Ergonomie-Experten verknüpft, um die Körperhaltung zu beeinflussen. Dies geschah durch aktive Formänderungen des Stuhls mittels der eingebauten Motoren und Benachrichtigungen über das Internet oder eine Smartphone-App.
- Der Datensatz des intelligenten Stuhls "Hipwings" von Vintus wurde für das Projekt zur Verfügung gestellt. Der Stuhl verfügt über Bewegungssensoren, um die Bewegungen des Nutzers zu erkennen, sowie Motoren, die die Form des Stuhls in Echtzeit anpassen und so auf die Körperhaltung einwirken.

Methodisch wurden moderne Transformer-Decoder-Encoder-Methoden für die sensorbasierte Erkennung menschlicher Aktivitäten (Human Activity Recognition, HAR) weiterentwickelt, insbesondere durch die Integration physikalischer und biomechanischer Randbedingungen, die von Ergonomieexperten vorgegeben wurden.

Nächste Schritte

Die Firma Vintus evaluiert die erzielten Ergebnisse, um smartere Bürostühle zu entwickeln. Angesichts des gesundheitlichen Risikos von Bewegungsmangel im Büro oder Homeoffice ermöglichen die smarten Bürostühle "Hipwings" automatisch Bewegung während der Büroarbeit. Durch die Sensorik erkennen die Stühle die Bedürfnisse der Nutzerinnen und Nutzer in verschiedenen Arbeitssituationen und aktivieren motorisch bewegte Sitzteile entsprechend. Dies ermöglicht konzentriertes Arbeiten und Bewegung zur gleichen Zeit, wodurch die Arbeitszeit für zusätzliche Bewegung im Alltag genutzt werden kann.



Bildquelle: Millertime83 auf Wikimedia Commons (CC-BY-SA 3.0)

Traffic Counter

Automated lightweight efficient object tracking model optimization.

Projektzeitraum

01.04.2022 – 30.09.2022

Projektpartner

Karlsruher Institut für Technologie

Yexu Zhou (zhou@kit.edu)

Dr. Till Riedel (riedel@kit.edu)

Novel Sense UG

Dr. Markus Scholz (scholz@novelsense.com)

Das SDI-C Mikroprojekt „Traffic Counter“ konzentriert sich auf die Entwicklung eines leichten, stromsparenden Verkehrszählers auf der Grundlage von IoT-Geräten. Im Vergleich zu bestehenden kamerabasierten Verkehrszählern ist es kostengünstiger und einfacher zu installieren. Um die begrenzten Ressourcen optimal zu nutzen und die unvorhersehbaren Leistungseinbußen zu vermeiden, die mit der Quantisierung einhergehen, wird ein Framework für die automatische quantisierte neuronale Optimierung entwickelt. Das Framework sucht automatisch in dem definierten Suchraum und erreicht schließlich den besten Kompromiss zwischen Genauigkeit und Effizienz. Dieser Suchraum besteht aus quantisierten Operatoren. Das bedeutet, dass das endgültige Modell ohne weitere Quantifizierung direkt auf der Hardware eingesetzt werden kann.

Projektziele

Kamera-basierte Verkehrszähler verwenden sowohl Objekterkennungs- als auch Objektverfolgungstechniken. Im Vergleich zu Modellen, die die Erkennung und Verfolgung separate durchführen, zeigen Modelle, die gleichzeitig Objekte erkennen und verfolgen, in der Regel eine überlegene Leistung, da sie kontextuelle Informationen im Video nutzen. Solche Modelle sind jedoch oft in ihrer Struktur komplex und rechenintensiv, was ihre Bereitstellung auf leichtgewichtigen Geräten erschwert. Darüber hinaus können Quantisierung und Pruning die Leistung dieser Modelle beeinträchtigen. Um diese Herausforderungen zu bewältigen, zielt dieses Projekt darauf ab, Pruning und Quantisierung in ein hardwarebewusstes neuronales Architektursuch-Framework zu integrieren, um automatisch leichtgewichtige Modelle zu optimieren, die gleichzeitig Objekte erkennen und verfolgen.

Projektergebnisse

Wir haben ein Framework für die gleichzeitige Objekterkennung und -verfolgung entwickelt, das die hardwarebewusste neuronale Architektursuche (NAS, bzw. Neural Architecture Search) benutzt. Während des Modelltrainings, dieses Framework balanciert intelligent die Modellleistung und die Inferenzzeit, indem es dynamisch die Größe des Faltungskernels auswählt und Pruning-Techniken anwendet. Das resultierende Modell kann die beste Leistung innerhalb der Inferenzzeitbeschränkungen zeigen. Dadurch entfällt die Notwendigkeit, die Modellstruktur manuell anzupassen und zu optimieren, und die Effizienz der Modellstrukturoptimierung kann erheblich verbessert werden. Darüber hinaus sind alle Faltungsoperatoren im Framework 8-Bit-quantisiert, was eine direkte Hardwarebereitstellung ohne Leistungsverluste durch die Post-Quantisierung gewährleistet.

Es wurden umfangreiche Experimente mit vorbereiteten Daten durchgeführt und ein Vergleich mit anderen State-of-the-Art-Modellen durchgeführt. Das optimierte Modell zeigte nicht nur eine überlegene Tracking- und Zählleistung, sondern auch eine geringere Größe (nur 3 MB) und schnellere Inferenzzeit. Diese Ergebnisse bestätigen die Wirksamkeit und Praktikabilität unseres Frameworks und machen es zu einer vielversprechenden Lösung für realweltliche Anwendungen.

Nächste Schritte

Die Ergebnisse sind wichtig für die technologische Weiterentwicklung der Abakus Hardware von Novelsense. Durch die Architektursuche ist es insbesondere möglich einen Hardware-Software-Co-Design-Ansatz zu wählen und Hardware und KI-Modell, aber auch das Erkennungsproblem vor Ort genau aufeinander abzustimmen.

Eine Herausforderung für die weitere Verkleinerung von Modellen ist es ausreichend spezifische Trainingsdaten zu bekommen. Gerade Neural Architecture Search Ansätze brauchen ausreichend Trainingsdaten. Hier arbeitet Novelsense gerade weiter mit dem SDIL zusammen, um den Ansatz auf synthetische Daten zu adaptieren.

Das KIT verwertet die NAS-Ansätze innerhalb mehrerer Projekte unter anderem zur Optimierung von Wearables.



Bildquelle Martin Sanchez auf Unsplash

Rapid Urbanism KI Agent

Integration ethischer und sozialer Überlegungen in ML-basierte Planung

Projektzeitraum

01.06.2022 – 31.11.2022

Projektpartner

Deutsches Forschungszentrum für künstliche Intelligenz (DFKI) GmbH

Peter Hevesi (Peter.hevesi@dfki.de)

IAUAI gGmbH

Matthias Nohn, LF MPP Dipl.-Ing. (Geschäftsführer)

Methodisch beschäftigte sich das Mikroprojekt mit der Frage, wie menschliches Expertenwissen in ML-Verfahren eingebettet werden kann, und insbesondere wie ethische und soziale Überlegungen in Systeme integriert werden können („Responsible AI“). Die konkrete Anwendung ist die Unterstützung einer nachhaltigen, sozial verträglichen Stadtplanung. Dabei soll die existierende manuelle Planungsplattform „Rapid-Urbanism-Explorer“ des Industriepartners IAUAI GmbH um eine KI-Komponente erweitert werden, die automatisch Planungsvarianten vorschlägt

Projektziele

Das Mikroprojekt "Rapid Urbanism" hatte das Ziel, menschliches Expertenwissen in ML-Verfahren zu integrieren, insbesondere die Integration ethischer und sozialer Überlegungen in Systeme, die relevante Vorgänge in der realen Welt beeinflussen („Responsible AI“). Das Projekt adressierte ein wichtiges und aktuelles Forschungsproblem und beabsichtigte, eine KI-Komponente in die bestehende manuelle Planungsplattform "Rapid-Urbanism-Explorer" des Industriepartners IAUI gGmbH (<http://rapidurbanism.com/explorer-ai/>) einzubetten. Diese KI-Komponente sollte automatisch Planungsvarianten vorschlagen, die sowohl die lokalen Gegebenheiten als auch sozioökonomische Ziele optimal berücksichtigen. Die Generierung solcher Vorschläge sollte durch das Lernen aus einer großen Menge öffentlich verfügbarer Daten (Open Street Map, Daten öffentlicher Verwaltungen) erfolgen und durch Experteninput in Richtung sozioökonomischer Nachhaltigkeit verfeinert werden.

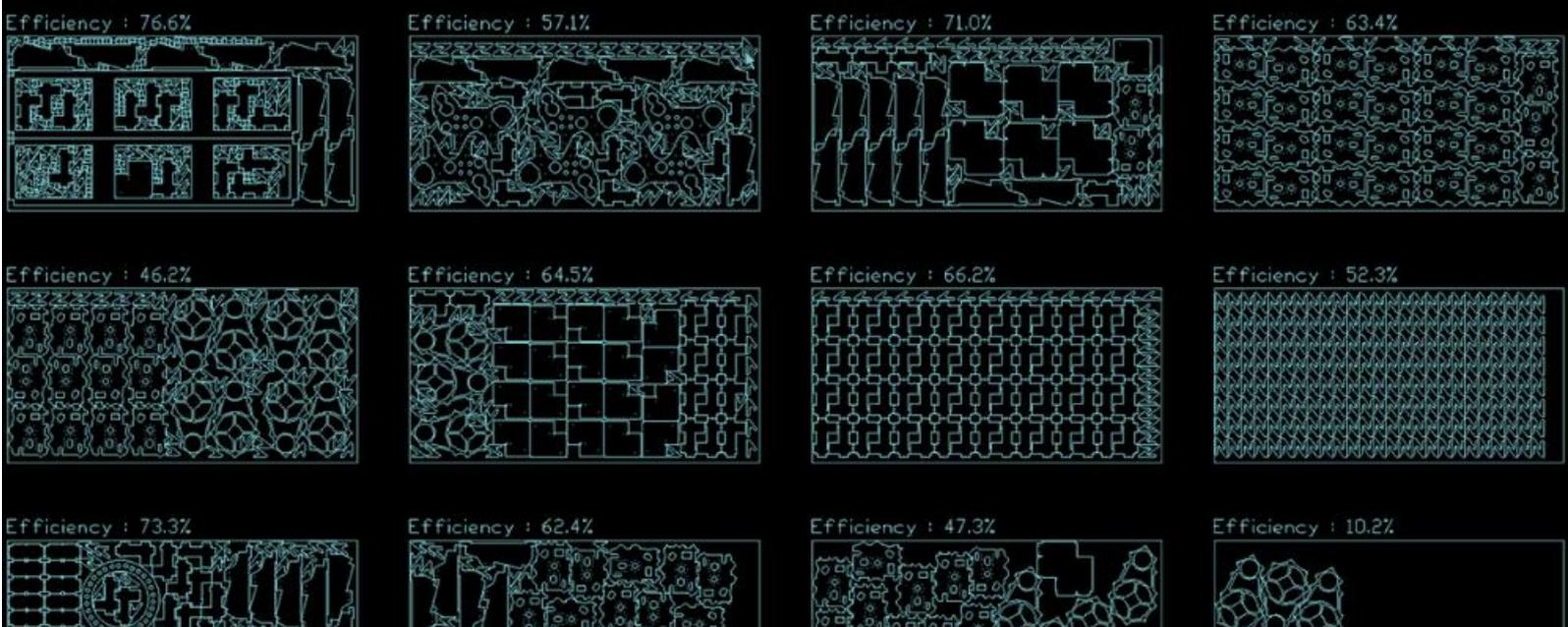
Projektergebnisse

Um das Problem der automatischen Generierung von Grundstückszuordnungen und deren Nutzungstypen (z. B. Gewerbegebiete, Wohnflächen, Erholungsflächen) zu lösen, wurden drei Ansätze verfolgt:

- Repräsentation des Wesens gültiger/guter Portionierungen: Es wurde untersucht, wie Wissen über die Zusammenhänge zwischen verschiedenen Objekten in gültigen und guten Zuordnungen in den Attention Mechanismus von Transformer-Modellen integriert werden kann. Zudem wurden Graph Neural Network Architekturen als induktiver Bias verwendet, um die Repräsentation der Portionierungen zu kodieren. Ziel war es, trotz der hohen Komplexität den Lernaufwand zu begrenzen und Transparenz zu gewährleisten.
- Optimierung bezüglich sozioökonomischer Verträglichkeit: Die Repräsentation und der darauf basierende Decoder wurden im Hinblick auf von Experten vorgegebene sozioökonomische Aspekte optimiert. Dabei wurden Deep Reinforcement Learning Ansätze adaptiert, die dem System ermöglichen, durch interaktive Zusammenarbeit mit den Experten inkrementell besser zu werden. Dies war besonders herausfordernd, da solche Kriterien teilweise noch in der Entwicklung sind und daher nur begrenzte Daten für klassisches Training vorhanden waren.
- Integration in die Rapid-Urbanism-Explorer Plattform: Die entwickelten Ansätze wurden erfolgreich in die Plattform der Firma IAUI gGmbH integriert, um die automatische Erstellung von Planungsvarianten zu unterstützen.

Nächste Schritte

Die erzielten Projektergebnisse können nun dazu verwendet werden, die bestehende manuelle Planungsplattform "Rapid-Urbanism-Explorer" um eine KI-Komponente zu erweitern, die eine nachhaltige und sozial verträgliche Stadtplanung unterstützt. Durch die Einbettung von Expertenwissen und ethischen Überlegungen in die KI-gestützte Planung können optimale Lösungen für die lokale Umgebung und die sozioökonomischen Ziele erzielt werden. Der Einsatz von Multi-GPU/Multi-Node-Systemen ermöglicht dabei effizientes Training und schnellere Iterationszyklen, um die optimale Konfiguration zu finden.



Bildquelle jetcam.net via wikimedia commons (CC-BY-SA 4.0)

SmartNesting@HPC

Optimierung der Clusternutzung für das RL von Nesting Problemen

Projektzeitraum

01.09.2022 – 31.12.2022

Projektpartner

Karlsruher Institut für Technologie (KIT)

Yiran Huang (yiran.huang@kit.edu)

TRUMPF Werkzeugmaschinen SE + Co. KG

Dr. Jens Ottnad

Kirolos Abdou

Im SDI-C Mikroprojekt „SmartNesting@HPC“ ging es darum einen von TRUMPF entwickelten auf Reinforcement Learning basierten Nesting-Algorithmus effizient auf ein HPC-System zu übertragen. Beim Nesting (Schachteln) geht darum, möglichst viele Teile aus einer Blechtafel herauszuschneiden und dabei Material zu sparen. Es gibt zwar schon Software für das Nesting, diese kann aber nur die Form der Teile berücksichtigen, weitere wichtige Faktoren lässt es außen vor. Deshalb schachteln die meisten Mitarbeiter von Hand. Würde ein Computer alle Faktoren berücksichtigen würde dies aktuell zu lange dauern da er schlichtweg zu wenig Rechenleistung hat, um die perfekte Nesting-Lösung auf Knopfdruck zu liefern. Hier können Reinforcement Learning Ansätze helfen schnellerer Optimierungsheuristiken zu finden.

Projektziele

TRUMPF arbeitet momentan daran dieses klassische Optimierungsproblem mit unterschiedlichen Herangehensweisen zu beschleunigen. Ein vielversprechender Ansatz ist die Formulierung als Reinforcement Learning Problem. Ziel des Projektes ist es die Nutzung skalierbarer Architekturen zu ermöglichen. Dabei sollen verschiedene Möglichkeiten der Kombination von KI-Beschleunigern HPC-Systemen und Cloud-Ressourcen erprobt werden.

Nesting ist ein sowohl in der Forschung als auch in der Wirtschaft hochrelevantes kombinatorisches Optimierungs-Problem. Ähnlich zu anderen Problemklassen hat es das Potential durch den Einsatz von KI radikal verändert zu werden. Die „Vorhersage“ einer Optimalen Schachtelung wäre im Gegensatz zur Kombinatorischen Optimierung eine enorme Ersparnis. Aufgrund der komplexen geometrischen Randbedingungen und wenn, wie im Fall von TRUMPF, weitere Optimierungsziele hinzukommen, schwer zu definieren und hat einen hohen Forschungsanspruch. Schon heute werden für Nesting HPC-Systeme eingesetzt. Die Nutzung von KI-Beschleunigern hat hier eine hohe Relevanz zur Verbesserung des Stands der Technik.

Projektergebnisse

In einer explorativen Datenanalyse haben wir verschiedene Parameter und Einstellungen mit verschiedenen Modellen getestet. Wir haben festgestellt, dass die bestehenden Algorithmen auf zwei Arten optimiert werden können: (i) durch die Verwendung eines komplexeren Modells: Das derzeit verwendete einfache Modell schränkt die Fähigkeit des Algorithmus zur Extraktion von Attributen ein. Darüber hinaus bedeutet die geringe Nachfrage nach Grafikprozessoren in den einfachen Modellen, dass eine große Anzahl von Grafikprozessoren ungenutzt bleibt und nicht effektiv genutzt wird. (ii) durch die Verwendung anderer RL-Algorithmen. Der Vorteil des derzeitigen Algorithmus sind die geringen Kosten für die Broadcasting des Netzes. Wenn sich das Modell ändert, sollte der RL-Algorithmus entsprechend angepasst werden, um die Kosten für die Übertragung zu senken und den Einsatz von GPUs zu erhöhen. Darüber hinaus liefert uns der Ansatz der Kontrollvariablen die besten Hyperparameter für die aktuellen Ressourcenbeschränkungen.

Nächste Schritte

Je nachdem, ob die Architektur des Algorithmus geändert werden kann, haben wir die folgenden zwei Vorschläge. Wenn es möglich ist, die Architektur des Algorithmus zu ändern, sammeln wir mehr Informationen über die SmartNesting-Umgebung und entwickeln gezielt einen vernünftigen und umfassenden Algorithmus neu. Wenn es nicht möglich ist, den Algorithmus zu ändern, empfehlen wir, experimentelle Daten unter Verwendung der oben beschriebenen Parameter zu sammeln und schlagen vor, dass das KIT Ihre zukünftige Algorithmus-Entwicklung durch die Bereitstellung von Daten aus experimentellen Läufen unterstützt.



Bildquelle: Dylan de Jonge auf Unsplash

DeepSoilLearning

Effiziente Nutzung von Multi-GPU-Clustern zur Hyperparameteroptimierung von Deep Learning (DL) Modellen unter Verwendung innovativer Bodenfeuchtemessdaten

Projektzeitraum

01.06.2022 – 31.12.2022

Projektpartner

Forschungszentrum Jülich

Dr. Carsten Montzka

Prof. Dr. – Ing. Morris Riedel

Heliopas GmbH

Ingmar Wolff

Benno Ommerborn Avino

Das Forschungskooperationsprojekt „Effiziente Nutzung von Multi-GPU-Clustern zur Hyperparameteroptimierung von Deep Learning (DL) Modellen unter Verwendung innovativer Bodenfeuchtemessdaten“ baut auf den Ergebnissen des erfolgreichen Forschungskooperationsprojektes „Verbesserung von KI-Modellen auf Basis von raumbezogenen Daten in der Agrarwirtschaft“ auf. Die Kernidee des neuen Projektes ist die erfolgreich aufgebaute „Preprocessing Pipeline“ und KI-Modelle vor allem durch die Nutzung neuer innovativer Datenquellen zu verbessern (engl. fine-tuning). Das SDIL-Expertenteam für Recheninfrastruktur, Hyperparameteroptimierung und KI-Modellierung wurde daher um einen etablierten Wissenschaftler im Bereich Bodenfeuchte erweitert (Bodenfeuchte-Experte).

Projektziele

Die Forschungsk Kooperation umfasst neben der weiteren Integration innovativer Datenquellen und der Verbesserung bestehender DL-Modelle sowie der Hyperparameteroptimierung vor allem das Fine-Tuning der „Preprocessing Pipeline“. Bei der Auswahl geeigneter Datenquellen wurde die wissenschaftliche Expertise von Heliopas im Bereich der Bodenfeuchte durch einen etablierten Senior Wissenschaftler des FZJ ergänzt. Die verbesserte Quantität und Qualität der ausgewählten Datensätze hat das Potential, die Genauigkeit der DL-Modelle für KMU weiter zu verbessern.

Projektergebnisse

Die bereits etablierte „Preprocessing Pipeline“ aus dem ersten Projekt wurde in diesem Projekt mit deutlich besseren qualitativen und mehr aussagekräftigen quantitativen Datenquellen noch effektiver ausgestaltet. Basierend auf der effektiveren „Preprocessing Pipeline“ wurden bestehende, aber auch neue DL-Modelle wie Long-Short Term Memory (LSTM), Gated Recurrent Units (GRUs) und Convolutional Neural Networks (CNNs) trainiert. Dabei wurden insbesondere die Vorteile von Multi-GPU-Clustern genutzt, um den Trainingsprozess deutlich zu beschleunigen. Die verbesserte Datenqualität sowie die höhere Anzahl an Daten führten zu einer stetigen Verbesserung der Genauigkeit der DL-Modelle. Abgerundet wurden diese Projektergebnisse durch eine weitere Anwendung der Hyperparameteroptimierung mit dem Ray Tune Tool.

Insgesamt haben die Projektergebnisse dem KMU Heliopas erneut geholfen. Benno Avino (CTO & Mitbegründer von Heliopas): „Durch eine gründliche Datensatzexploration und Featureanalyse haben die Domänenexperten und Forscher ein völlig neues Verständnis des Problems und seiner Eigenschaften ermöglicht“.

Nächste Schritte

Wie bereits bei der Verwertung des ursprünglichen Heliopas-Projektes verwertet das FZJ die Ergebnisse dieses Verbundprojektes im Rahmen des EU-Projektes "Research on AI- and Simulation-Based Engineering at Exascale" (CoE RAISE) im Bereich der Hyperparameteroptimierung und innovativer DL-Modelle (z.B. LSTM, GRU) weiter. Auch die Nutzung der verbesserten „Preprocessing Pipeline“ ist bereits in der Praxis umgesetzt und wird z.B. im FZJ-Center for Advanced Simulation and Analytics (CASA) Simulation and Datalab (SDL) Healthcare im Rahmen der Erforschung des Acute Respiratory Distress Syndrome (ARDS) und anderer Atemwegserkrankungen (z.B. Covid-19, Pneumonie, etc.) eingesetzt.

Die Firma Heliopas wertet die Projektergebnisse aus, um sie in ihre Produkte zu integrieren. Die Verwertung der Projektergebnisse beinhaltet eine weitere intensive Zusammenarbeit zwischen dem Antragsteller aus dem FZJ und den KMU, die mit hoher Wahrscheinlichkeit zu einer langfristigen und strategischen Zusammenarbeit führen wird. Für den Aufbau dieser vertieften Zusammenarbeit war das Projekt unverzichtbar.



Bildquelle Pop Nukoonrat auf vercteezy

KD4RE

Erstellung generativer, wissensgetriebener Dialogsysteme unter Nutzung großer Transformer-Sprachmodelle

Projektzeitraum

01.06.2022-30.11.2022

Projektpartner

Karlsruher Institut für Technologie

Dr. Michael Färber (michael.farber@kit.edu)

Aleph Alpha GmbH

Jonas Andrulis (jonas.andrulis@aleph-alpha.de)

Das Ziel dieses Projekts war es, Unternehmen dabei zu unterstützen, große Datenmengen in Textform effektiv zu analysieren, indem ein tieferes Kontextverständnis ermöglicht wird. Oft verstecken sich wertvolle Informationen in komplexen Zusammenhängen, die nicht leicht durch herkömmliche Stichwortsuchen erkannt werden können. Das Projekt zielte darauf ab, diese Herausforderung mit Hilfe großer Transformers-Sprachmodelle zu lösen und die Nutzung von Daten und großen Sprachmodellen durch kleinere Unternehmen zu erleichtern.

Projektziele

Große Sprachmodelle haben zu enormen Fortschritten im KI-Bereich geführt und erlauben mit relativ geringem Aufwand neuartige, KI-basierte Anwendungen für Textzusammenfassung und -erstellung, oder Dialogsysteme zu erstellen. Die bisherigen, großvolumigen KI-Sprachmodelle wurden auf zentralen Infrastrukturen von Hyperscalern trainiert. Aleph Alpha entwickelt aktuell Große Sprachmodelle, welche jedoch neben technologischen Chancen auch eine Reihe von Risiken und Problemen bergen, wie zum Beispiel die mangelnde Integration von strukturiertem Wissen, um domänenspezifische Anwendungen abdecken zu können

Konkret wurden im Projekt sogenannte generative Dialogsysteme evaluiert, die einerseits sehr natürlich kommunizieren können (wie ein Sprachmodell) und andererseits Mehrwert bieten, indem sie Fakten aus konkreten Anwendungsfällen und Domänen integrieren (aus den Wissensgraphen). Damit können Unternehmen ihre bisherige, interne Datenbasis mittels natürlicher Sprache durchsuchbar machen.

Projektergebnisse

Das Projekt nutzte große Transformers-Sprachmodelle, die über mehr Hintergrundinformationen verfügen, um Informationen aus Texten zu extrahieren. Im Gegensatz zu herkömmlichen Verfahren, die tausende annotierte Beispiele benötigen, um akzeptable Ergebnisse zu erzielen, zeigte sich, dass diese großen Sprachmodelle mit deutlich weniger Daten auskommen. In einigen Fällen genügte sogar nur ein einziges annotiertes Beispiel, um relevante Informationen zu extrahieren. Solche Sprachmodelle können jedoch auf herkömmlichen Rechnern aufgrund ihrer Größe, oft hunderte von Milliarden von Parametern, nicht ausgeführt werden.

Um dieses Problem zu bewältigen, ging das Projekt einen innovativen Weg. Es wurde ein deutlich kleineres Modell verwendet, dem mithilfe des großen Sprachmodells beigebracht wurde, wie man relevante Informationen in Texten findet. Auf diese Weise kann das kleinere Modell die gewünschten Informationen extrahieren, ohne auf leistungsstärkeren Rechnern oder speziellen Infrastrukturen zu laufen. Dies ermöglicht Unternehmen, selbst von einem Laptop aus relevanten Informationen effizient zu analysieren und bisher ungenutztes Wissen zugänglich zu machen.

Durch die Anwendung dieser Methode können Unternehmen, insbesondere kleinere Unternehmen, ihre Daten effektiver nutzen, ohne teure und langwierige Prozesse der Datensammlung und Annotation durch Experten. Die Kombination von großen und kleinen Sprachmodellen ermöglicht es, maßgeschneiderte Modelle für jeden Anwendungsfall zu erstellen, wodurch die Flexibilität und Effizienz der Datenanalyse erheblich gesteigert wird.

Nächste Schritte

Insgesamt stellt das Projekt einen vielversprechenden Weg dar, um die Potenziale von großen Sprachmodellen für die effektive Datenanalyse in Unternehmen zu nutzen. Für Aleph Alpha stellt die Wissensextraktion aus Sprachmodellen insbesondere eine Möglichkeit dar dieses Wissen auch zu verifizieren.



Weiterentwicklung der Plattform

Die größte Herausforderung für den Betrieb einer unabhängigen Datenreinraums, die im Zentrum des SDIL steht und im Projekt adressiert wurde, bleibt die Nachhaltigkeit, da ein rein geförderter Betrieb schwierig ist, sobald die Förderung aus verschiedenen Quellen kommt. Noch schwieriger, aber wünschenswert, ist eine gemeinsame Finanzierung durch industrielle Drittmittel. Aufgrund der schwierigen Bedingungen für einen nahtlosen Übergang zwischen Forschung und kommerziellem Betrieb (EU-Beihilferahmen, steuerliche und haftungsrechtliche Risiken) ist es eine der großen Herausforderungen für Innovationsräume wie das SDIL, sowohl eine gesunde Balance als auch eine organisatorisch tragfähige Basis für einen Mischbetrieb zu finden, der sowohl gemeinsam geförderte oder von den Partnern getragene Verbundforschungsprojekte als auch Auftragsforschung und Transferprojekte

Gerade der Mischbetrieb mit industriellen Projekten und „reinen“ Forschungsprojekten ist immer dann herausfordernd, wenn vertrauliche Unternehmens-Daten im Spiel sind und gleichzeitig neuste KI-Lösungen getestet werden sollen. Es hatte sich im Vorprojekt gezeigt, dass der Betrieb verschiedenster Lösungen möglich ist, jedoch gerade KI-Lösungen und Bedarfe sich schnell weiterentwickeln. Herausfordernde Mikroprojekte sollen daher genutzt werden, um diese Entwicklung zu beschleunigen.



Compute Infrastruktur Plattform

Aufbau eines SDIL GPU-Clusters als Partition im Uni BW Cluster 2.0

SDIL-Infrastrukturpartner

Karlsruher Institut für Technologie (KIT)

Simon Raffener (simon.raffener@kit.edu)

In SDI-C zur Verfügung gestelltes Angebot:

- *GPU-Partition mit 76 A100 GPUs*
- *Jupyter Lab Hub Umgebung auf HPC-System*
- *Slurm HPC Scheduler und Accounting*
- *Large Scale Storage*
- *Gitlab mit CI/CD Integration des HPC-Systems*

Ausgangssituation

Als Compute Infrastruktur Plattform wurde im Rahmen des Smart Data Innovation Labs und des SDI-C Projektes bisher eine Plattform bestehend aus verschiedenen Komponenten angeboten. Diese umfasste neben den für die Berechnungen benötigten Knoten, beispielsweise einem IBM AC922 GPU-Knoten basierend auf der POWER9 Architektur, auch Knoten für die notwendigen administrativen und operativen Dienste der Plattform.

Zur Ermöglichung eines effizienteren Betriebs einer Compute Plattform für SDI-C, sowie auch für eine Vereinfachung der Auswahl der zu Nutzenden Ressourcen für insbesondere neue Nutzende, wurde im Jahr 2021 eine neue Plattform in Betrieb genommen.

Ergebnisse/Technische Neuerungen

Die neue Plattform wird synergistisch mit dem bwUniCluster 2.0 am KIT betrieben. Durch dieses Betriebskonzept wird ein stabilerer Betrieb mit einer höheren Verfügbarkeit ermöglicht sowie möglichst viele Synergien gehoben, beispielsweise bei der Ermöglichung neuerer Zugangsmöglichkeiten und aktueller Softwaremodule.

Das auf jedem Server bzw. Knoten installierte Betriebssystem ist Red Hat Enterprise Linux (RHEL) 8.x. Zusätzlich zu diesem Betriebssystem wurde eine Reihe von (quelloffenen) Softwarekomponenten wie Slurm installiert.

Login-Knoten

Der Zugang zu den Ressourcen der SDIL-Plattform wird den Nutzenden über zwei primäre Zugangswege zur Verfügung gestellt. Zum einen besteht die Möglichkeit für einen klassischen HPC-Zugang über dedizierte Login-Knoten und unter Nutzung des Job-Scheduling Systems. Dieser ist besonders für stark ressourcenintensive Rechenvorhaben nötig, bei denen Ressourcen entweder über einen längeren Zeitraum oder eine größere Anzahl von Ressourcen und Rechenknoten bzw. GPUs benötigt werden. Diese Knoten werden gemeinsam mit dem bwUniCluster 2.0 genutzt und können für interaktive Logins, Dateiverwaltung, Softwareentwicklung und interaktives Pre- und Postprocessing verwendet werden. Vier Knoten sind als Login-Knoten vorgesehen.

Zum anderen wird ein interaktiver Zugang über die im Bereich der Data Analytics und des Maschinellen Lernens verbreitete Lösung via JupyterLabs ermöglicht. Diese ermöglicht einen graphischen Zugriff auf die Compute Plattform mittels eines Webbrowsers. Die Nutzenden können darüber hinaus im Rahmen der Compute Plattform auf containerisierte Umgebungen zurückgreifen, die beispielsweise bestimmte Versionen der gängigen Python Softwaremodule für die Datenanalyse enthalten. Diese Container können dabei auch von externen Quellen stammen, so bietet beispielsweise der GPU-Hersteller Nvidia optimierte Container für die Ausführung bestimmter Anwendungen auf Nvidia GPUs an.

Rechenknoten

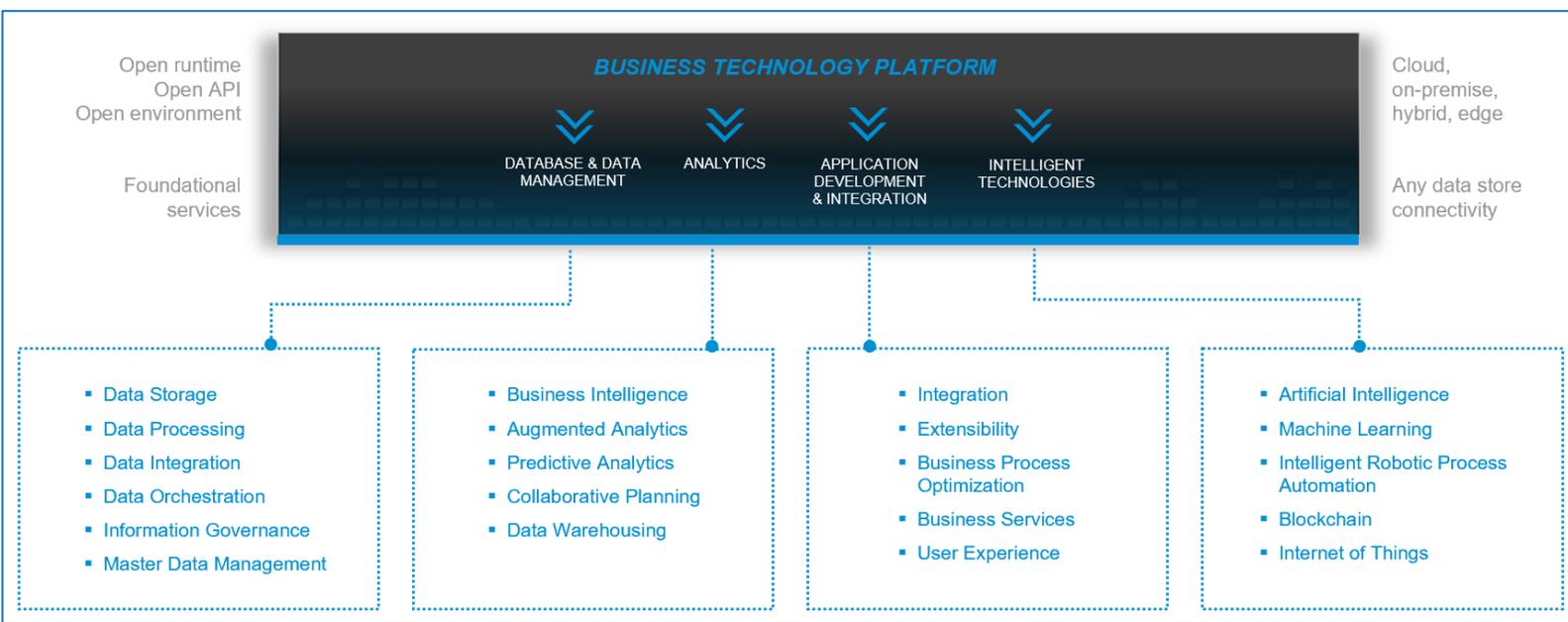
Die SDIL-Knoten sind für Berechnungen vorgesehen. Diese Knoten sind für die Nutzenden nicht direkt zugänglich, stattdessen müssen die Berechnungen an ein so genanntes Batch-System

übergeben werden. Das Batch-System verwaltet alle Rechenknoten und führt die in der Warteschlange stehenden Aufträge je nach ihrer Priorität aus, sobald die erforderlichen Ressourcen verfügbar sind. Ein einziger Auftrag kann Hunderte von Rechenknoten und mehrere Tausend CPU-Kerne auf einmal nutzen.

Insgesamt verfügt die Plattform über 19 Rechenknoten mit 76 Beschleunigern vom Typ NVIDIA A100-40. Mit insgesamt 38 CPU-Sockeln und 912 CPU-Kernen ergibt sich ein Primärspeicher von 7 TB, die lokale Festplatte umfasst 38 TB. Weiterhin ist das System mit einem InfiniBand HDR als Interconnect ausgestattet.

Ausblick

Das KIT wird die Compute Plattform weiterhin im Rahmen des Smart Data Innovation Labs, sowie des Folgeantrags Smart Data Innovation Services, angeboten. Insbesondere ist angedacht Verbesserungen bei den Zugangsmöglichkeiten sowie Möglichkeiten zur effizienteren Nutzung der verfügbaren Beschleuniger zu evaluieren.



Bildquelle SAP

Business Technology Platform | BTP | Multichord | SAP

Ein „Prüfstand“ für Forscher Algorithmen unter realen Bedingungen

SDIL-Infrastrukturpartner

SAP

Dr. Elmar Dorner (elmar.dorner@sap.com)

In SDI-C zur Verfügung gestelltes Angebot:

- *SAP Business Technology Platform Cloud Services (SAP BTP)*
- *SAP Data Intelligence Cloud*
- *SAP HANA Cloud*
- *SAP Data Warehouse Cloud*

Ausgangssituation und Ziele

Als ein Marktführer für Geschäftssoftware unterstützt SAP unterschiedlichste Unternehmen jeder Größe und Branche dabei, ihre Ziele bestmöglich zu erreichen. Die SAP-Technologien für maschinelles Lernen, das Internet der Dinge und fortschrittliche Analyseverfahren helfen den Kunden auf dem Weg zu einem intelligenten Unternehmen. SAP vereinfacht Technologie für Unternehmen, damit sie SAP- Software nach ihren eigenen Vorstellungen einfach und reibungslos nutzen können.

Gerade die anwendungsnahe Weiterentwicklung und Erprobung von KI-basierten Diensten ist daher ein wichtiger Faktor, für deren zukünftige breite Einsetzbarkeit, insbesondere innerhalb der von SAP HANA getriebenen Cloud-Lösungen. Der Ausbau langfristiger akademischer Partnerschaften und kollaborativer Ökosysteme rund um den Themenbereich KI, auch über das Projekt hinaus, steht somit ebenfalls im Fokus.

Die SAP Business Technology Platform kann ein integriertes Angebot geschaffen werden, das aus den vier Technologieportfolios, Datenbank- und Datenmanagements, Analytics, Anwendungsentwicklung und -integration sowie intelligente Technologien besteht. Auf Basis der SAP BTP ermöglicht SAP einen Prüfstand, SAP Edge AI Research Workbench, für KI-Forschung, Inbetriebnahme und Verfeinerung der Modelle im KI-Lebenszyklus.

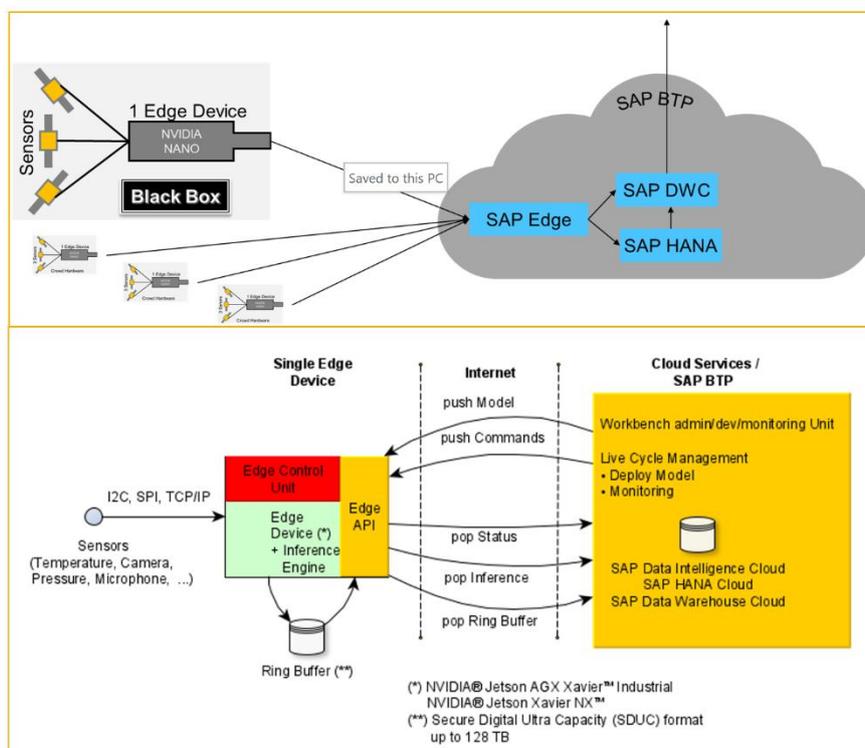


Abbildung 4: SAP Edge AI Research Workbench – Overview

Grundvoraussetzung ist ein vorhandenes TensorFlow / Keras AI-Modell und die entsprechenden Sensor-APIs zur Speisung der Inferenz-Engine. Das gewählte Setup ermöglicht Forschern die Out-of-the-Box-Anwendung.

"Data Engineering" mittels KI wird zu einem entscheidenden Wettbewerbsfaktor für den Standort Deutschland und Europa. Vor diesem Hintergrund unterstützt die SAP das "Smart Data Innovation Lab" (SDIL) am Karlsruher Institut für Technologie (KIT). Die

Mit dem „Prüfstand“ können Forscher befähigt werden, entwickelte Algorithmen unter realen Bedingungen anzuwenden, um deren Leistungsfähigkeit zu prüfen und die Optimierung zu unterstützen. Die lokale Vorverarbeitung der Daten ermöglicht breite Anwendungsszenarien, wie

- Überwachung von Werkzeugmaschinen – Verschleiß-, Alterungsvorhersage
- Biodiversitätsforschung – Insekten/Vogelmonitoring
- Landwirtschaft – Parameterüberwachung Boden

Weiterentwicklung der technischen Infrastruktur und der Angebotsmöglichkeiten für das Durchführen von KI-Mikroprojekten am SDIL war somit die zentrale Zielstellung im Projekt.

Ergebnisse

Die Aktivitäten im Projekt zielten darauf ab, Akademiker*innen die Möglichkeit zu bieten, Daten- und KI-Projekte zukünftig im Rahmen von SDIL/SDIC-Mikroprojekten mit dem "SAP Cloud Stack" durchzuführen. Um diese Zielstellung zu erreichen, wurde parallel in drei Handlungssträngen gearbeitet. 1. Mehrwertbetrachtungen zu Cloud-/AI-Services Erweiterung 2. Betreibermodell für die Cloud Services im Rahmen der SDIL/SDIC-Plattform 3. Technisches Cloud Service Enablement

Zu 1) Im Rahmen der Mehrwertbetrachtungen wurden in einer Reihe von Workshops und Gesprächen, grundsätzliche Anforderungen und Herausforderungen für datenbasierte Anwendungen und Dienste, als auch die Optimierung dieser mit oder durch den Einsatz von KI-Ansätzen, betrachtet und diskutiert. Wesentlichen Input für diese Betrachtungen lieferten zum einen der fachliche Austausch mit dem SDIC-Mikroprojekt „LatentAgeI4Wafer - AI-gestützte Vorhersage alterungsabhängiger Effekte auf Basis maschinell gelernter latenter Variablen in der Wafer-Produktion“ und das "Smart Air Quality Network" Projekt.

Zu 2) Die Bereitstellung des "SAP Cloud Stack" unter den Bedingungen öffentlicher Forschung erfordert ein nachhaltiges Bereitstellungskonzept, was Operative Kosten, Support, Enablement auch für die Forscher*innen nachhaltig bereitstellt. Unterschiedliche Ansätze wurden eruiert und ein erfolgversprechendes Modell für den universitären Kontext erarbeitet und umgesetzt. Das Modell berücksichtigt Erweiterbarkeit der Cloud-Dienste als auch Skalierung über einen unabhängigen, universitären Service-Dienstleister.

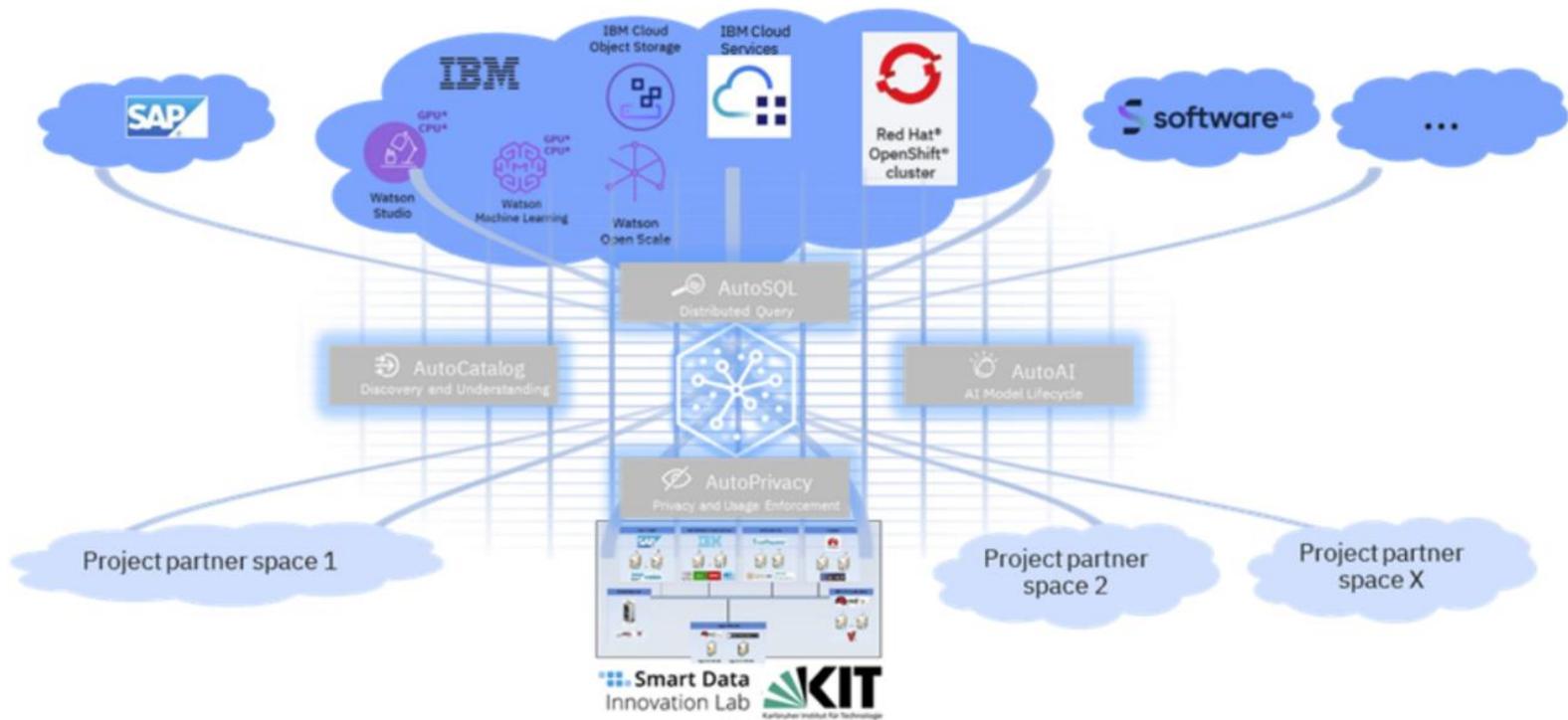
Zu 3) Die Verankerung der Fähigkeiten zum Einsatz der Cloud-Dienste im Forschungskontext und im Rahmen zukünftiger Mikroprojekte erfolgte über ein zugeschnittenes Enablement-Format mit dem Fokus auf die SAP Data Warehouse Cloud.

Aufbauend auf die Verankerung der Cloud-Dienste für den Forschungskontext erfolgte die Erweiterung des Serviceerweiterungsangebots mit dem Fokus, das Schließen der Transferlücke von KI-Forschung zur betrieblichen Anwendung zu erleichtern. Mit Hilfe des im Projekt zugeschnittenen Angebots der SAP Edge AI Workbench wird Anwendungsforschern die Möglichkeit geboten, entwickelte Algorithmen unter realen Bedingungen anzuwenden, um deren Leistungsfähigkeit zu prüfen und die Optimierung zu unterstützen.

Nächste Schritte

Ziel der SAP ist weiterhin die Forschung zu und den Transfer von KI in betriebliche Praxis zu fördern. Die Unterstützung von industrienahen Mikroprojekten beispielsweise durch Wissenstransfer und Enablement steht somit im Zentrum der geplanten Aktivitäten.

Der Fokus liegt auf der fachlich/technischen Beratung und Unterstützung der Projektbeteiligten. Dies beinhaltet beispielsweise die Erprobung und das Enablement SAP-bezogener Cloud-Dienste. Es kann jedoch, mikroprojektabhängig, Arbeiten zum Zuschnitt der einzusetzenden Cloud-Services und die (Weiter-)Entwicklung des Services-Angebotes, beinhalten.



Bildquelle: IBM

IBM Cloud mit AutoML

Ein Data Fabric Implementierung

SDIL-Infrastrukturpartner

IBM

Felizitas Müller (felizitas.mueller@ibm.com)

In SDI-C zur Verfügung gestelltes Angebot

- *IBM Cloud inkl. Cloudbasierter AutoML Dienste*
- *Beratung und Wissenstransfer*
- *Innovative Industrie4.0- und Nachhaltigkeits-Lösungen*
- *IBM Cloud User Trainings*

Ausgangssituation

Zu Projektbeginn 2019, hat IBM dem Smart Data Innovation Lab eine IBM Power9 mit den neuen AC922 Hardwareknoten zur Verfügung gestellt.

Insbesondere die Unterstützung von ML-Umgebungen auf Basis von Containertools standen hier den Nutzern zur Verfügung.

In dieser Zeit standen die folgenden Dienste zur Verfügung:

- IBM Power9 AC922
- IBM Cloud Pak for Data
- Power AI
- Maximo Visual Inspection

Ergebnisse

Im Laufe des Projektzeitraums hat IBM bei der Definition der Kriterien zur Vergabe der Mikroprojekte und Erstellung der Ausschreibung unterstützt und die entsprechende Infrastruktur dafür zur Verfügung gestellt.

Um die richtige Nutzung des IBM Power9 Systems innerhalb dieser Projekte zu gewährleisten, wurden entsprechende Workshops für Studierende organisiert, sowohl vor Ort als auch remote.

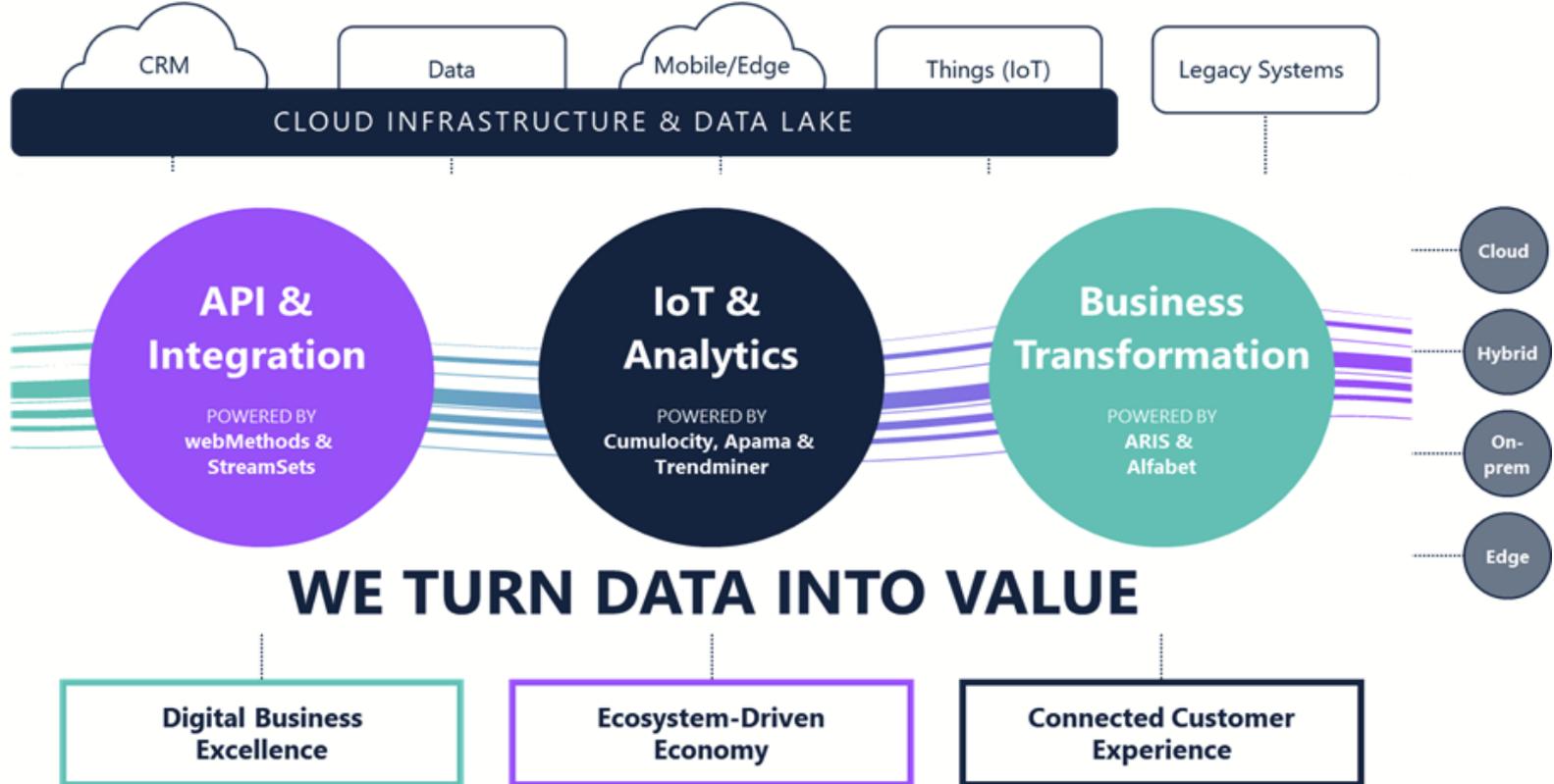
Des Weiteren wurde die Projektumgebung Anfang 2021 um die IBM Cloud und weitere Partner-Clouds ausgedehnt, ohne dass die Sicherheitsanforderungen komprimiert wurden. Die Projekte inkl. ihrer Daten konnten somit weiterhin untereinander gekoppelt werden und die Kontrolle über die Daten konnten zugesichert werden.

Mit der Migration der IBM Angebote in die Cloud, wurden ab 2021 Schulungen zu den unterschiedlichen Angeboten durchgeführt, so dass Forscher, mittelständische Unternehmen und Studierende den Mehrwert für ihre Projekte nutzen konnten.

Nächste Schritte

Anschlussprojekte werden die Smart Data Innovation Services (SDI-S) innerhalb von KI-Mikroprojekten.

Ziel hierbei ist die Erprobung neuer KI-Verfahren, -Werkzeugen und -Methoden, die innerhalb von Diensten durch die Partner bereitgestellt werden. Ziel dieser Mikroprojekte ist eine kurze Zeitspanne zwischen Idee und Umsetzung von KI-Forschung mithilfe der Nutzung unserer Cloud Ressourcen.



Bildquelle: Software AG

Software AG IOT-Cloud für SDIL
Unterstützung der Wertschöpfung auf Daten

SDIL-Infrastrukturpartner

Software AG
Dirk Mayer (dirk.mayer@softwareag.com)

In SDI-C zur Verfügung gestelltes Angebot

- Software AG Complex Event Processing engine APAMA
- Software AG TERRACOTTA in Memory-DB
- Software AG IOT-Cloud mit Cumulocity und APAMA

Ausgangssituation und Ziele

Die Software AG ist der Softwarepionier der vernetzten Welt. Seit dem Jahr 1969 hat die Gesellschaft mehr als 10.000 Firmen und Organisationen dabei unterstützt, Menschen, Unternehmen, Systeme und Geräte durch Software zu verbinden. Mithilfe von Integration & API-Management, IoT & Analytics sowie Business & IT Transformation ebnet die Software AG den Weg zum vernetzten Unternehmen; ihre Produkte sind der Schlüssel für einen ungehinderten Datenfluss und eine reibungslose Zusammenarbeit. Das Unternehmen hat mehr als 5.000 Beschäftigte in mehr als 70 Ländern und einen Jahresumsatz von mehr als 830 Millionen Euro.

Die Software AG leitete die Forschungsarbeiten zu hochskalierbaren Streaming-Lösungen. Um die breite produktive Nutzbarkeit der Ergebnisse sicherzustellen, beteiligte sich die Software AG am Angebot geeigneter Schulungen für die Mikroprojekte.

Die Software AG hat sich an den Arbeiten zu BigData- und IOT-Lösungen und deren Portabilität auch auf Cloudlösungen, insbesondere der IOT-Plattformen beteiligt. Weiterhin stellt die Software AG hierzu geeignete Schulungen für das Smart Data Innovation Lab.

Zu Projektbeginn hat die Software AG dem Smart Data Innovation Lab eine komplette Instanz für die Software AG eigene weltweit führende Complex Event Processing (CEP) Software APAMA sowie die in Memory Datenbank TERRACOTTA zur Verfügung gestellt, auf durch die Partner und das SDIL zur Verfügung gestellter Hardware installiert und integriert. Insbesondere die Unterstützung von ML-Umgebungen mittels TERRACOTTA als auch die Echtzeitüberwachung und Streaming-Analyse auf Basis APAMA standen hier den Nutzern für Microprojekte zur Verfügung.

Ziel der Arbeit war es insbesondere KMUs und Forschungseinrichtungen diese Tools für ausgewählte Tests und prototypische Versuche zur Verfügung zu stellen sowie Erkenntnis für deren Weiterentwicklung zu sammeln. Ein weiteres Ziel war es, wie auch im Fall der anderen Software-Partner mehr externe Dienste transparent in die Workflows einzubinden und so Workflows über Rechenzentren hinweg zu gestalten, so sollten unter anderem auch Cloud-Dienste eingebunden werden.

Ergebnisse

Im Rahmen des Projektes wurden die erfolgreich beendeten Mikro-Projekte bis zum erfolgreichen Abschluss begleitet und kontinuierlich hinsichtlich Skalierbarkeit und Nachhaltigkeit des Projektergebnisses überprüft. Die in der anschließenden Analyse beobachteten und identifizierten Abläufe, techn. Voraussetzungen und Ansatzpunkte wurden erhoben und, soweit erforderlich, Maßnahmen und Implementierungen für die nächste Projektrunde definiert und umgesetzt. Es wurde gemeinsam mit dem KIT an der Integration einer Erweiterung des Angebots seitens der Software AG um Cumulocity IOT gearbeitet. Die Integration wurde erfolgreich abgeschlossen, in der 5. Ausschreibung für Mikroprojekte mit angeboten und erfolgreich genutzt. Die Projekte inkl. ihrer Daten konnten somit weiterhin



untereinander gekoppelt werden und die Kontrolle über die Daten konnten zugesichert werden.

Mit der Migration der Software AG Angebote in die Cloud, wurden ab 2021 Schulungen zu den unterschiedlichen Angeboten durchgeführt, so dass Forscher, mittelständische Unternehmen und Studierende den Mehrwert für ihre Projekte nutzen konnten.

Weiterhin wurde im Rahmen der jährlichen Smart Data Open Days das migrierte Software AG Angebot am und für das SDIL vorgestellt und in einer Diskussionsrunde mit den Partnern und interessierten Unternehmen und Forschungseinrichtungen erörtert.

Nächste Schritte

Die Software AG wird weiterhin das SDIL mit den dort entwickelten und zur Verfügung gestellten Services im Rahmen von Mikro-Forschungsprojekten sowie in der Schulung an diesen unterstützen.

Im Rahmen eines Folgeantrags wird die Software AG mit den Partnern und dem KIT versuchen weitere Cloudservices, z.B.: auf Basis der webMethods-Produktfamilie, im SDIL für die Forschung und Prototypenentwicklung zu entwickeln und implementieren. Interessenten, wie Forschungseinrichtungen in Zusammenarbeit mit KMUs könnten diese dann im Rahmen von Mikroprojekten für Test und Erprobung zur Verfügung gestellt werden.

Angebotsweiterentwicklung

Im Rahmen einer Projektaufstockung im Jahr 2021 wurde die SICOS BW GmbH im Unterauftrag des KITs eingebunden. Ziel war es die Angebotsentwicklung gerade gegenüber kleineren Unternehmen zu verbessern. Dabei baute die SICOS BW auf Ihrer Erfahrung bei der Vermittlung von HPC-Ressourcen sowie auf ihrem Engagement im Smart Data Solution Center BW auf.

Die SICOS BW GmbH hat in Zusammenarbeit mit dem Karlsruher Institut für Technologie (KIT) eine Angebotsliste für das Smart Data Innovation Lab (SDIL) und das Smart Data Science Center Baden-Württemberg (SDSC-BW) erstellt. Diese Liste bietet eine Vielzahl von Angeboten für kleine und mittelständische Unternehmen (KMUs) aus allen Teilen Deutschlands, die sich mit Data Analytics, Künstlicher Intelligenz und Maschinelles Lernen beschäftigen oder sich in diesen Bereichen weiterentwickeln möchten. Das Ziel ist es, den KMUs eine individuelle Auswahl anpassenden Angeboten zu bieten, die ihren Bedürfnissen und Anforderungen gerecht werden und somit einen Mehrwert für ihre Unternehmen darstellen.

Die Angebote sind entsprechend ihrer Laufzeit sortiert und in vier Kategorien unterteilt:

- Smart Data Consulting: Hier werden KMUs in Form von Erstberatung und Potentialanalyse unterstützt.
- Data Analytics & HPC Infrastruktur: Diese Kategorie bietet Angebote für Proof-of-Concepts und Transfer- oder Forschungsprojekte.
- KI-Werkzeuge & Algorithmen: KMUs erhalten hier Zugang zu verschiedenen KI-Tools und Algorithmen.
- Individuelle Weiterbildung: Die vierte Kategorie ermöglicht individuelle Weiterbildungsangebote für die Mitarbeiter der KMUs.

Die Vollumfänglichkeit der Angebote ist ein besonderes Alleinstellungsmerkmal (Unique Selling Proposition) des SDIL. Die Unternehmen können die Leistungen und den Zeitrahmen flexibel an ihre Bedürfnisse anpassen, sei es für einen einzigen Tag oder für mehrere Jahre.

Um den tatsächlichen Bedarf der KMUs zu ermitteln, führte die SICOS BW Stichprobenbefragungen durch. Dabei wurden verschiedene Faktoren berücksichtigt, wie die Verfügbarkeit finanzieller Mittel, die technischen Randbedingungen wie Breitbandanbindung und Reifegrad der IT-Infrastruktur sowie die Kompetenzen der Mitarbeiter und mögliche Weiterbildungsbedarfe. Auch die Erwartungen bezüglich des Returns on Investment (RoI) und die Bereitschaft für neue digitale Geschäftsmodelle wurden erfasst.

Eine Gap-Analyse wurde durchgeführt, um die bestehenden Angebote weiter zu optimieren. Dabei wurde sowohl der interne Blick auf das lückenlose Angebot am SDIL geworfen als auch externe Perspektiven berücksichtigt, um die Angebote an die Bedürfnisse der KMUs anzupassen und eventuelle Lücken zu schließen.

Im Rahmen der Öffentlichkeitsarbeit unterstützte die SICOS BW konzeptionell und bei der Implementierung, um die Angebote des SDIL und des SDSC-BW bekannt zu machen und

entsprechend zu veröffentlichen. Eine Projektwebseite des SDSC BW stellt eine Übersicht der Angebote dar, welche von der Erstberatung zur Potentialanalyse bis hin zum Transferprojekt reichen.

Abschließend erstellte die SICOS BW einen Umsetzungsplan für die Einführung und die produktive Bereitstellung der neuen Angebote. Dieser Plan umfasst zwei Phasen: eine für die Laufzeit des Projektes und eine zweite für die Zeit nach Projektende, um sicherzustellen, dass die Angebote auch langfristig erfolgreich und nachhaltig nutzbar sind.

Die Zusammenarbeit zwischen dem KIT und der SICOS BW hat somit eine umfassende Angebotsliste für das SDIL und das SDSC-BW geschaffen, die KMUs dabei unterstützt, die richtigen Angebote entsprechend ihrer individuellen Bedürfnisse auszuwählen und somit einen Mehrwert für ihre Unternehmen zu generieren.

Smart Data Days, Workshops und Hackathons

Trotz der Coronapandemie war das Projekt von kollaborativen Veranstaltungen geprägt. Diese fanden notgedrungen meist online statt. Es wurden verschiedene Formate ausprobiert, von der virtuellen Poster-Session mit Technologiedemonstrationen bis hin zu fast professionellen Fernsehproduktionen, um ein möglichst breites Publikum von Forschern und Unternehmen zu erreichen.

Im Bereich der Technologieschulungen fanden beispielsweise mehrere Workshops der IBM statt innerhalb derer das Angebot Hands-On vorgestellt wurde. Teilnehmer konnten direkt die Dienste testen. Die Workshops waren außerordentlich gut besucht. Ähnlich hat das KIT mehrere Einführungen in das Cluster veranstaltet.

Einen innovativen Weg hat die SAP mit Ihrem im Jahr 2021 veranstalteten Hackathon beschritten. Hier wurde die damals neu im SDIL integrierte Data Warehouse Cloud im Rahmen eines wettbewerbsartigen mehrtägigen Events zusammen mit Forschern auf die Anwendbarkeit für Forschungsdaten getestet. Gerade das direkte Feedback hat hier wichtige Impulse für die Weiterentwicklung gebracht.

Ursprünglich waren im Projekt „Roadshows“ zu den einzelnen Forschungspartnern geplant, die aber pandemiebedingt ausfallen mussten. Jedoch hat sich während der Projektzeit auch die Art der Zusammenarbeit stark verändert, so dass virtuelle Labs wie das SDIL beinahe natürlich online arbeiten. Ungeachtet dessen hat sich im Projekt doch gezeigt, dass der persönliche Kontakt, auch online ein wichtiger Faktor für die produktive Zusammenarbeit bleibt.

Insbesondere die Smart Data Open Days, welche jeweils jährlich die Ergebnisse der Mikroprojekte vorgestellt haben, wurden sehr positiv angenommen. Trotz der technischen Möglichkeiten, fehlte teilweise das Networking beim gemeinsamen Kaffee. Trotzdem wurde durch Diskussionen auf der virtuellen Bühne eine interaktive und anregende Atmosphäre ermöglicht. Nicht zuletzt der mit Unterstützung der SICOS veranstaltete „SDIL Open Day 2021“

war ein großer Erfolg. Die Videos der einzelnen Veranstaltungen sind weiterhin auf YouTube abrufbar:

<https://www.youtube.com/@smartdatainnovationlabsdil2842/>



SDIL in der Presse

In verschiedenen Fachmedien wurde über das SDIL im Zeitraum vom 16. August bis 31. Dezember 2021 berichtet. Dies erfolgte in Form von einer Pressemitteilung, Blogbeiträgen und Gastbeiträgen:

- Bigdata Insider, Dezember 2021 <https://www.bigdata-insider.de/zukunftstechnologien-mit-forschungs-partnern-vorantreiben-a-1082489/>
- SICOS BW, Dezember 2021 <https://www.sicos-bw.de/pressemitteilungen/sicos-bw-und-kit-spannende-fragestellungen-fuer-foerderprojekte-gesucht/>
- Datacenter Insider, November 2021 <https://www.datacenter-insider.de/das-kit-forschungs-cluster-fuer-industrieprojekte--bis-2022-kostenlos-a-1072831/>
- Nafems Online-Magazin, Oktober 2021 – Nr. 3/2021, 59. Ausgabe
- Big Data Insider, Oktober 2021 <https://www.bigdata-insider.de/kmu-erhalten-kostenlosen-zugriff-auf-kit-forschungscluster-a-1059472/>
- Mittelstands Café, September 2021 <https://www.mittelstandcafe.de/kit-forschungscluster-f-r-industrieprojekte-einsetzbar-1919877.html>
- Big Data Insider, September 2021 <https://www.bigdata-insider.de/kit-forschungscluster-unterstuetzt-industrieprojekte-a-1053950/>



- Digital Business Cloud, September 2021 <https://www.digitalbusiness-cloud.de/sdil-forschungscluster-von-kit-jetzt-auch-fuer-industrieprojekte-einsetzbar/>
- Digital Factory Journal, September 2021 <https://www.smart-production.de/digital-factory-journal/news-detailansicht/nsctrl/detail/News/kmu-profitieren-von-kit-forschungscluster>
- Ingenieurmagazin.com, September 2021 <https://www.ingenieurmagazin.com/it-elektrotechnik/kmu-data-analytics-sicos-bw/5393/>
- Industrial AI, August 2021 <https://www.ind-ai.net/newsarchiv/kit-forschungscluster-fuer-industrieprojekte-einsetzbar/>
- IT, Internet und Kommunikations-News, August 2021 <https://www.itiko.de/artikel/1919877/kit-forschungscluster-f-r-industrieprojekte-einsetzbar.html>
- IT & Production Online, August 2021 <https://www.it-production.com/news/maerkte-und-trends/karlsruher-institut-fuer-technologie/>
- Compamed Magazin, August 2021 https://www.compamed.de/de/Archiv/Big_Data_Hilfe_beim_Heben_von_Datensch%3%A4tzen
- Industrie 4.0 & IOT, August 2021 <https://www.i40-magazin.de/allgemein/karlsruher-institut-fuer-technologie/>

Folgende Gastbeiträge wurden in Folge der Pressemitteilung von weiteren Multiplikatoren aufgegriffen. Hierbei ging es insbesondere um den Veranstaltungshinweis zum Smart Data Open Day am 8. Dezember 2021:

- Wirtschaftsförderung Region Stuttgart, Dezember 2021 <https://ki-community.region-stuttgart.de/smart-data-open-day-2021/>
- Datacenter Insider, November 2021 <https://www.datacenter-insider.de/das-kit-forschungs-cluster-fuer-industrieprojekte--bis-2022-kostenlos-a-1072831/>
- IT, Internet und Kommunikations-News, November 2021 <https://www.itiko.de/artikel/1940290/smart-data-open-day-2021-innovationen-aus-daten-schaffen.html>
- PR-Inside, November 2021 <https://www.pr-inside.com/de/smart-data-open-day-2021-innovationen-aus-daten-schaffen-r4854554.htm>
- IT & Production Online, November 2021 <https://www.it-production.com/news/maerkte-und-trends/virtuelle-info-veranstaltung/>
- Industrie 4.0 & IOT, November 2021 <https://www.i40-magazin.de/industrie-4-0/einblicke-in-das-smart-data-innovation-lab/>
- Industry of Things, November 2021 <https://www.industry-of-things.de/online-veranstaltung-zeigt-wie-aus-daten-innovationen-werden-a-1077975/?cmp=beleg-mail>

SDIL in einer Studie zu Test- und Demonstrationsräumen, die vom dänischen Forschungsministerium in Auftrag gegeben wurde

<https://ufm.dk/aktuelt/nyheder/2022/ny-rapport-fra-icdk-munchen-test-og-demonstrationsfaciliteter-i-sydtyskland>

Notwendigkeit und Angemessenheit der Arbeiten im Projekt

Die Arbeiten im Projekt waren insbesondere auf die Mikroprojekte fokussiert, welche im Personaleinsatz gedeckelt waren. Damit wurden die zu Verfügung gestellten Personalressourcen ideal genutzt, um agil die Ergebnisse zu optimieren. Die Projekte haben dabei die Infrastruktur genutzt die innerhalb des Projektes aufgebaut wurden.

Wichtige Entwicklungen außerhalb des Projektes

Während der Projektlaufzeit hat sich die KI-Landschaft massiv weiterentwickelt. Zuletzt haben insbesondere die gehosteten Angebote von OpenAI (wie ChatGPT) und Huggingface zusammen mit Angeboten wie Google Colab die Zugänglichkeit von KI-Ressourcen für viele stark verbessert. Gerade die Forschung an neuen Modellen erfordert aber auch immer größere Rechenressourcen, damit die Forscher selbst Beiträge leisten können. Insbesondere die neu entstehenden grundlegenden Modelle können mit den typischerweise in Forschungsprojekten zur Verfügung stehenden Ressourcen gar nicht mehr gelernt werden. Das Tuning von Modellen, wie es bereits in Mikroprojekten durchgeführt wurde, wird hier immer wichtiger. Insbesondere bleibt das SDIL auch als On-Premise-Lösung für sensible Unternehmensdaten interessant.

Auch im Vergleich zur Preisentwicklung und dem stetig wachsenden Angebot von KI-Diensten in der Cloud ist die Compute-Infrastruktur am KIT für die adressierte Zielgruppe trotz steigender Strompreise und nicht zuletzt durch die Einbindung von Cloud-Diensten der Partner zu attraktiven Konditionen konkurrenzfähig. Positiv wirkt sich die Konsolidierung der Frameworks hin zu PyTorch und TensorFlow aus.

Auf europäischer Ebene mit den European Digital Innovation Hubs, aber auch auf nationaler Ebene mit verschiedenen Service- und Kompetenzzentren ist die Konkurrenz von Plattformen, die sich mit Transferleistungen (insbesondere auch im Bereich KI) insbesondere an KMU richten, gewachsen. Dennoch konnte sich das SDIL mit seinem Angebot bisher gut behaupten.

Auch die regulatorischen Rahmenbedingungen innerhalb der EU haben sich verändert. So kann sich das SDIL nicht problemlos als Datenintermediär positionieren, da der im September 2023 in Kraft tretende Data Governance Act hier das gleichzeitige Angebot von Dienstleistungen ausschließt. Andere Entwicklungen wie der Data Act könnten sich langfristig positiv auf das Projekt auswirken, da hier Bedingungen für den Zugang zu „IoT-Daten“ festgelegt werden.

Informationen zum Projektumfang

Das Projekt hatte einen Umfang von ca. 3,2 Mio. EUR Personalkosten, 27k EUR Sachkosten, 595k EUR Investitionen, 80k Unteraufträge und wurde vom BMBF mit ca. 3,6 Mio. EUR gefördert. Das gesamte Projekt hatte einen Umfang von ca. 325 Personenmonaten. Mehr als

die Hälfte (ca. 190 Personenmonate) entfiel auf die Mikroprojekte einschließlich des projektbezogenen Betriebs der Plattform. Der restliche Personalaufwand wurde für die Projektauswahl, die Weiterentwicklung der Plattform und Veranstaltungen eingesetzt.

Die größte Investition im Rahmen der Förderung war die Anschaffung des Rechenclusters im Wert von 500.000 EUR netto.

Aufgrund der Pandemie wurden deutlich weniger Reisen unternommen als ursprünglich geplant. Auch die Smart Data Open Days wurden online bzw. hybrid durchgeführt.

Fazit und Ausblick

Die Ergebnisse der Mikroprojekte wirken sich vor allem positiv auf die Zusammenarbeit zwischen Industrie und Wissenschaft aus. Die Industriepartner als Auftraggeber der Datenauswertung profitieren direkt von den Ergebnissen. Darüber hinaus eröffnen sich für die beteiligten Partner neue interessante Themen und potenzielle zukünftige Projektpartner in größeren Projekten. Die Zusammenarbeit zwischen Industrie und Wissenschaft wird gestärkt und ermöglicht eine verstärkte wirtschaftliche Nutzung von Datenanalysen.

Ein weitgehend unsichtbarer, aber wichtiger Teil des SDIL bleiben die Prozesse vom Vertragsabschluss über das On-boarding der Nutzer bis hin zur Präsentation der Ergebnisse. Hier ist die vertrauensvolle Zusammenarbeit aller Beteiligten, die sich auch auf Dritte erstreckt, ein wichtiger Faktor.

Auch wenn die Mikroprojekte aufgrund der begrenzten Laufzeit keine grundlegend neuen Forschungsergebnisse liefern können, werden dennoch interessante wissenschaftliche Transferergebnisse erzielt. Diese Ergebnisse können für andere Analysen und Projekte nützlich sein. Auch die Möglichkeit, zuvor entwickelte Algorithmen an neuen Datensätzen zu testen und zu erproben, trägt zur wissenschaftlichen Weiterentwicklung bei. Insgesamt zeigt sich, dass die Mikroprojekte sowohl für die Zusammenarbeit zwischen Industrie und Wissenschaft als auch für den wissenschaftlichen Erkenntnisgewinn von Bedeutung sind und die Grundlage für weitere Projekte und Forschungsarbeiten bilden. Auch für die Forschung war das SDI-C eine wichtige Ergänzung zu anderen Projekten. So werden in den nächsten Jahren mehrere Dissertationen abgeschlossen, die Aspekte der Mikroprojekte beinhalten.

Dies zeigt zusammenfassend deutlich, dass das SDIL eine nachhaltige wissenschaftliche und wirtschaftliche Relevanz besitzt und sich kontinuierlich weiterentwickelt. Die Anschlussfähigkeit des Projektes und des Smart Data Innovation Lab wird durch den Start des Projektes Smart Data Innovation Services unterstrichen. Das neue Projekt baut auf den Entwicklungen und Kontakten des SDI-C auf, um die Zusammenarbeit zwischen Forschung/Wissenschaft und Industrie weiter zu festigen und zu konkretisieren.

Wissenschaftliche Veröffentlichungen mit Projektbezug

Balzereit, K., Fullen, M., & Niggemann, O. (2020). A Concept for the Automated Reconfiguration of Quadcopters. In LWDA (pp. 180-191).

Biesner, D., Gerlach, T., Kliem, B., Bauckhage, C., & Sifa, R. (2022). Solving Subset Sum Problems using Quantum Inspired Optimization Algorithms with Applications in Auditing and Financial Data Analysis. ICMLA 2022.

Fortes Rey, V., Suh, S., & Lukowicz, P. (2022). Learning from the Best: Contrastive Representations Learning Across Sensor Locations for Wearable Activity Recognition. Proceedings of the 2022 ACM International Symposium on Wearable Computers.

Huang, Y., Zhou, Y., Hefenbrock, M., Riedel, T., Fang, L., & Beigl, M. (2022). Universal Distributional Decision-Based Black-Box Adversarial Attack with Reinforcement Learning. In International Conference on Neural Information Processing (pp. 206-215).: Springer

Huang, Y., Zhou, Y., Hefenbrock, M., Riedel, T., Fang, L., & Beigl, M. (2022). Automatic Feature Engineering Through Monte Carlo Tree Search. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 581-598). Springer

Nguyen, A., Krause, F., Hagenmayer, D., & Färber, M. (2021). Quantifying Explanations of Neural Networks in E-Commerce Based on LRP. ECML-PKDD'21, 251-267. Springer.

Petter, O., et al. (2019). Crowdsensing under recent mobile platform background service restrictions: A practical approach. Adjunct Proceedings of the UbiComp 2019

Schloß, D., Gnewuch, U., & Maedche, A. (2022). Towards Designing a Conversation Mining System for Customer Service Chatbots, ICIS 2022.

Schloß, D., Espitia, J. D. G., & Gnewuch, U. (2023). Designing a Conversation Mining System for Customer Service Chatbots, ECIS 2023.

Stenzel, R., Lübbering, M., Ulusay, B., Uedelhoven, D., & Sifa, R. (2022). Matching Experts to Questions: A Comparison of Recommender Systems, LWDA 2022: 57-66

Zhao, H., et al. (2022). Improving Human Activity Recognition Models by Learnable Sparse Wavelet Layer. ACM ISWC 2022,

Zhou, Y., et al. (2022). TinyHAR: A lightweight deep learning model designed for human activity recognition. ACM ISWC 2022.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

 **Smart Data**
Innovation Lab

