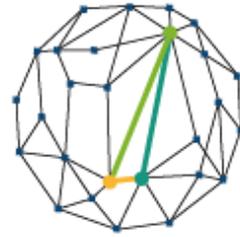




BERICHT

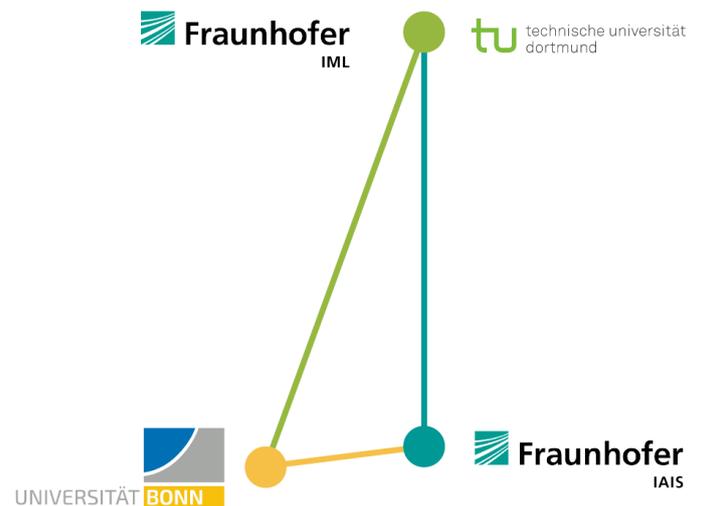
08/2020



ML2R
Kompetenzzentrum
Maschinelles Lernen
Rhein-Ruhr

Ein zusammenfassender Bericht über die bisherigen Aktivitäten des Kompetenzzentrums Maschinelles Lernen Rhein-Ruhr (ML2R)

Das ML2R verbindet Pionier-Institutionen der Forschung in Deutschland:



Inhaltsverzeichnis

Inhalt

Vorwort	1
Forschung	2
Transfer	11
Sichtbarkeit	14
Kontaktinformationen	18
Zentrumsinformationen	18

Das ML2R informiert



Vorwort

Maschinelles Lernen spielt eine zentrale Rolle auf dem Gebiet der Künstlichen Intelligenz (KI). Alan Turing, der als Gründer der KI gilt, erklärte bereits früh, dass Computer eher unterrichtet als programmiert werden sollten. In seinem berühmten Aufsatz „Computing Machinery and Intelligence“ schrieb er 1950, dass ein Lehrer einer Maschine „...will often be very largely ignorant of quite what is going on inside, although he may still be able to some extent to predict his pupil's behavior. ... Most of the programmes which we can put into the machine will result in its doing something that we cannot make sense (if at all), or which we regard as completely random behaviour.“¹ Diese frühe Aussage berührt die Frage der Vertrauenswürdigkeit, die nun die Forschung und Entwicklung der KI vorantreibt. Wir untersuchen das Maschinelle Lernen, um seine Modelle zu verstehen. Wir beweisen (statistische) Grenzen der Korrektheit, Fairness, Robustheit und des Ressourcenverbrauchs von Algorithmen und ihrer Implementierungen. Dies ist immer noch eine Herausforderung, da es über die abstrakte Ebene der statistischen Theorie hinausgeht und die Ebene der Algorithmen und, darüber hinaus, die Ebene ihrer Implementierungen auf verschiedenen Rechnerarchitekturen untersucht. Daher gibt es experimentelle Methoden, welche ebenfalls das Ziel der Vertrauenswürdigkeit anstreben: Benchmark-Tests und interaktive Untersuchungen oder Erklärungen von gelernten Modellen. Das ML2R legt seinen Fokus auf menschenorientiertes Maschinelles Lernen, arbeitet an Garantien für Klassen von Lernmodellen, z.B. probabilistische graphische Modelle, und untersucht die theoretischen und empirischen Grenzen des Ressourcenverbrauchs auf unterschiedlicher Hardware, sogar auf Quantencomputern.

Als Kompetenzzentrum engagieren wir uns für den Aufbau des Maschinellen Lernens in Forschung, Lehre und Anwendung. Wir laden Kolleginnen und Kollegen zu einem Forschungsaufenthalt bei uns ein - im Jahr der Covid-19-Pandemie scheiterte dies zu unserem großen Bedauern. Wir bieten eine Online-Sommerschule mit einem Hackathon an und haben eine Vielzahl von Unternehmen über ihre spezifischen Anwendungsmöglichkeiten des Maschinellen Lernens informiert. Die Vernetzung des ML2R mit den anderen Kompetenzzentren und geförderten deutschen Projekten zum Maschinellen Lernen sowie mit den französischen Kompetenzzentren zur Künstlichen Intelligenz unterstützt den Aufbau einer Gemeinschaft für Maschinelles Lernen.

Dieser Bericht strebt an, Sie als Leser anzuregen, mehr über die Forschung des ML2R zu erfahren, an der Sommerschule teilzunehmen, unseren Campus als Gastforscher zu besuchen oder von einer engeren Zusammenarbeit im Sinne eines Enterprise Innovation Campus zu profitieren, bei dem gemeinsame Teams aus Mitarbeitern von Unternehmen und Forschern des Kompetenzzentrums an spezifischen praktischen Problemen arbeiten.

In jedem Fall wünschen wir Ihnen viel Vergnügen und bleiben Sie gesund!

Prof. Dr. Katharina Morik
Sprecherin des ML2R
17. August 2020

Prof. Dr. Stefan Wrobel
Sprecher des ML2R
17. August 2020

¹ Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460. <https://academic.oup.com/mind/article/LIX/236/433/986238>

Das ML2R informiert



Forschung

DIE PARTNER

Das Kompetenzzentrum Maschinelles Lernen Rhein-Ruhr (ML2R) vereint vier führende Forschungsinstitutionen im Bereich des Maschinellen Lernens (ML): die Technische Universität Dortmund (TU Dortmund), das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS, das Fraunhofer-Institut für Materialfluss und Logistik IML und die Rheinische Friedrich-Wilhelms-Universität Bonn (Universität Bonn). Sie arbeiten mit dem gemeinsamen Ziel Spitzenforschung zu etablieren, den wissenschaftlichen Nachwuchs zu fördern und den Technologietransfer in Unternehmen zu stärken.

Technische Universität Dortmund (TU Dortmund). Die Fakultät für Informatik an der TU Dortmund ist eine der bundesweit größten und forschungsstärksten Informatikfakultäten. Neben dem Kompetenzzentrum Maschinelles Lernen Rhein-Ruhr ist auch der DFG-geförderte Sonderforschungsbereich 876 „Verfügbarkeit von Information durch Analyse unter Ressourcenbeschränkung“ an der Fakultät angesiedelt. Die Sprecherin des ML2R, Prof. Dr. Katharina Morik, hat die Professur am Lehrstuhl für Künstliche Intelligenz inne und ist Leiterin des Sonderforschungsbereiches 876.

Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme (Fraunhofer IAIS). Als Teil der größten Organisation für anwendungsorientierte Forschung in Europa ist das Fraunhofer IAIS eines der führenden Wissenschaftsinstitute auf den Gebieten Künstliche Intelligenz (KI), Maschinelles Lernen und Big Data in Deutschland und Europa. Das Institut unterstützt Unternehmen bei der Optimierung von Produkten, Dienstleistungen, Prozessen und Strukturen sowie bei der Entwicklung neuer digitaler Geschäftsmodelle. Der Sprecher des ML2R, Prof. Dr. Stefan Wrobel, ist Institutsleiter des Fraunhofer IAIS und gleichzeitig leitender Professor des Bereichs Informationssysteme und Künstliche Intelligenz am Institut für Informatik der Universität Bonn.

Fraunhofer-Institut für Materialfluss und Logistik (Fraunhofer IML). Das Fraunhofer IML gilt als erste Adresse in der ganzheitlichen Logistikforschung und arbeitet auf allen Feldern der inner- und außerbetrieblichen Logistik. Das Institut erarbeitet sowohl Problemlösungen zur unmittelbaren Nutzung für Unternehmen sowie Vorlauforschung. Der Vertreter des Fraunhofer IML beim ML2R, Prof. Dr. Dr. h. c. Michael ten Hompel, ist Leiter des Instituts und Inhaber des Lehrstuhls für Förder- und Lagerwesen an der Technischen Universität Dortmund.

Rheinische Friedrich-Wilhelms-Universität Bonn. Das Institut für Informatik an der Universität Bonn war eines der bundesweit ersten universitären Institute für Informatik. Die Forschungsgebiete am Institut gliedern sich in vier Bereiche: Algorithmen, Graphics-Vision-Audio, Informations- und Kommunikationsmanagement sowie Intelligente Informationssysteme und Künstliche Intelligenz. Der Vertreter der Universität Bonn beim ML2R, Prof. Dr. Christian Bauckhage, ist leitender Professor der Forschungsgruppe Machine Learning and Artificial Intelligence.

DAS TEAM

Die in der gemeinschaftlichen Zusammenarbeit erzeugten Synergien haben das Kompetenzzentrum Maschinelles Lernen Rhein-Ruhr seit seinem Start im Jahr 2018 als international und bundesweit sichtbares Forschungszentrum etabliert. Sein Prestige fußt hierbei auf der exzellenten Forschungsleistung des kontinuierlich gewachsenen Wissenschaftlerstabs. Zurzeit umfasst das Team des Kompetenzzentrums 32 wissenschaftliche Mitarbeiter. Im Dezember 2019 gab das Bundesministerium für Bildung und Forschung (BMBF) einem Antrag des ML2R um **Mittelaufstockung** statt. Das Kompetenzzentrum wird nun mit **zusätzlichen 8 Millionen Euro** bis Ende 2022 gefördert. So ist es dem ML2R möglich, das Team um **bis zu 25 neue wissenschaftliche Mitarbeitende** aufzustocken.

Das ML2R informiert

ASSOZIIERTE PARTNER

Zusammen mit seinen assoziierten Partnern treibt das ML2R **interdisziplinäre** und **institutionenübergreifende Forschungsprojekte** voran. Im Rahmen dieser Kooperationen erhalten die Partner Zugriff auf die Expertise und Ressourcen des Kompetenzzentrums, zum Beispiel Algorithmen. Prof. Dr. Dr. Wolfgang Rhode, Experte auf dem Gebiet der Astroteilchenphysik an der TU Dortmund, ergänzt das ML2R um anspruchsvolle Daten mit tiefgreifenden Forschungsfragen. Mit seiner Spezialisierung auf autonome intelligente Systeme bringt Prof. Dr. Sven Behnke von der Universität Bonn seine Expertise in kognitiver Robotik und Computer Vision ein. Prof. Dr. Jürgen Gall von der Universität Bonn wird mit seinen Arbeiten in den Bereichen Handlungserkennung, der Schätzung menschlicher Körperposen sowie Objektdetektion und -segmentierung die Forschungsanstrengungen des ML2R unterstützen.

FORSCHUNGSSCHWERPUNKTE

Die gemeinsamen Forschungsbestrebungen innerhalb des Kompetenzzentrums basieren auf der Erforschung und dem Aufbau von **modularen ML-Anwendungen**, bei welchen Systeme aus einzelnen Komponenten aufgebaut und zu komplexen Architekturen verknüpft werden. Dieser Ansatz umfasst die besonderen Aspekte von **menschenorientierten maschinellen Lernverfahren**, **ML unter Ressourcenbeschränkung** und **mit komplexem Wissen**. Die Forschenden des ML2R widmen sich zudem Fragestellungen rund um die Themen **vertrauenswürdiger maschineller Lernverfahren** sowie dem **Maschinellen Lernen auf Quantencomputern**.

AUSGEWÄHLTE FORSCHUNGSERGEBNISSE

Erklärbare und vertrauenswürdige KI

Die Akzeptanz maschineller Lernverfahren ist eine wichtige Grundlage für die zukünftige Verbreitung und Nutzung von ML-Anwendungen in der industriellen Produktion und in Hochrisiko-Einsatzbereichen. Die Ergebnisse und Prozesse des Maschinellen Lernens müssen **interpretierbar** und **erklärbar** sein. Dies stellt heute ein prominentes Thema auf fast allen aktuellen KI- oder ML-Konferenzen dar. Die wegweisende Doktorarbeit von Dr. Stefan Rüping zu „Learning interpretable models“ an der TU Dortmund im Jahr 2006² untersuchte die drei Dimensionen Verständlichkeit, Genauigkeit und Effizienz mit Beiträgen auf den Ebenen der Optimierung der Interpretierbarkeit eines Lernalgorithmus mit und ohne Kenntnis seiner Internen (White Box- und Black Box-Ansatz). Dieses Thema wird in den heutigen Standorten von ML2R aufgegriffen.

Ein kürzlich publizierter Beitrag deckt erklärbare latente Faktoren in Daten auf, die das menschliche Verständnis erleichtern. Rekurrente Neuronale Netze filtern Prototypen innerhalb von minimal einschließenden Bällen aus den Daten. Die entsprechende Arbeit von Prof. Dr. Christian Bauckhage, Dr. Tiansi Dong und Rafet Sifa wurde auf der International Conference on Artificial Neural Networks 2019 (ICANN 2019) mit einem der beiden Best Paper Awards ausgezeichnet.³

² Rüping, S. (2006). *Learning interpretable models* (Doctoral dissertation). TU Dortmund University, Dortmund, Germany. https://eldorado.tu-dortmund.de/bitstream/2003/23008/1/dissertation_rueping.pdf

³ Bauckhage, C., Sifa, R., Dong, T. (2019). Prototypes Within Minimum Enclosing Balls. *ICANN 2019*, S. 365-376. Cham, Switzerland: Springer. https://link.springer.com/chapter/10.1007/978-3-030-30493-5_36

Das ML2R informiert

Es wurde oft behauptet, dass durch die Induktive Logische Programmierung (ILP) erlernte Modelle besonders leicht zu verstehen sind. Das kürzlich erschienene Paper „Mining Tree Patterns with Partially Injective Homomorphisms“ überbrückt die Lücke zwischen ILP und Graph Mining-Verfahren.⁴ Es baut auf dem Erfolg des vorangegangenen Beitrages über probabilistische häufige Teilbäume für Graphenklassifizierung und -abruf auf.⁵ Hierbei beschränkte ein einfaches probabilistisches Verfahren die Mustersprache auf Bäume und lockerte die Anforderungen an die Vollständigkeit des Mining-Algorithmus sowie an die Korrektheit des Mustervergleichsoperators, indem Transaktions- und Abfragegraphen durch kleine Zufallsstichproben ihrer überspannenden Bäume ersetzt werden. Auf diese Weise können chemische Graphen-Datensätze effizient bearbeitet werden, selbst solche mit 42.687 Verbindungen, von denen 39.337 verbunden sind (der NCI-HIV-Datensatz), oder der ZINC-Datensatz mit 8.946.757 (8.946.755 verbundenen) unmarkierten, so genannten 'Lead-Like'-Molekülen aus der Zink-Datenbank mit kaufbaren chemischen Verbindungen. Die theoretische Grundlage, empirischen Tests und verständlichen Muster ermöglichen eine effektive Unterstützung der Wissenschaft durch Maschinelles Lernen.

Im Rahmen des Netzwerkes KI.NRW hat das Fraunhofer IAIS-Team unter der Leitung von Dr. Maximilian Poretschkin ein White Paper zum Thema „Trustworthy Use of Artificial Intelligence – Priorities from a Philosophical, Ethical, Legal, and Technological Viewpoint as a Basis for Certification of Artificial Intelligence“ erarbeitet, das einen Bezugsrahmen für die weitere Arbeit des Kompetenzzentrums bietet.⁶

Im Rahmen des ML2R wurde gemeinsam mit Behrooz Omidvar-Tehrani und Jean-Michel Renders auf der Extended Database Conference 2020 ein Workshop organisiert, dessen Schwerpunkt auf der **gesamten Pipeline des Maschinellen Lernens** lag.⁷ Aufgrund der Covid-19-Pandemie ging der Workshop selbst online aber das Panel wurde zu einem gemeinsamen Paper.⁸ ML2R wird seine Arbeit zur Optimierung von gesamten ML-Pipelines und zur Erfassung und Strukturierung von Qualitäts- und Ressourcenverbrauchsgarantien des Maschinellen Lernens fortsetzen.

Menschenorientierte ML

Bei der Anwendung in der Praxis interagieren Systeme und Lösungen des Maschinellen Lernens typischerweise mit menschlichen Benutzern. Dies legt nahe, die Interaktion und das Feedback der Benutzer zu verwenden, um

⁴ Schulz, T. H., Horvarth, T., Welke, P., Wrobel, S. (2018). Mining Tree Patterns with Partially Injective Homomorphisms. *Proc. ECML PKDD*, pp. 585-601. http://publica.fraunhofer.de/eprints/urn_nbn_de_0011-n-5523328.pdf

⁵ Welke, P., Horváth, T., Wrobel, S. (2018). Probabilistic frequent subtrees for efficient graph classification and retrieval. *Machine Learning*, 107, 1847–1873. <https://link.springer.com/content/pdf/10.1007/s10994-017-5688-7.pdf>

⁶ Cremers, A. et al. (2019). *Trustworthy Use of Artificial Intelligence. Priorities from a Philosophical, Ethical, Legal, and Technological Viewpoint as a Basis for Certification of Artificial Intelligence*. Sankt Augustin, Germany: Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS. https://www.iais.fraunhofer.de/content/dam/iais/KINRW/Whitepaper_Thrustworthy_AI.pdf

⁷ Morik, K., Omidvar-Tehrani, B., Renders, J.-M. (2020). Explainability for Trustworthy ML Pipelines. *Proc. Workshop ETMLP*. <https://europe.naverlabs.com/etmlp/>

⁸ Giannotti, F., Kotthaus, H., Morik, K., Piatkowski, N., Schlunder, P. (2020). Explainability for Trustworthy ML Pipelines: A discussion. B. Omidvar-Tehrani/J.-M. Renders (Mod.), *Panel of the International Workshop on Explainability for Trustworthy ML Pipelines of the ETMLP*. https://drive.google.com/file/d/1J3Ng8c9kX_ixwHrCBPbjQoqgEx9R7b5X/view

Das ML2R informiert

die Fähigkeiten eines Systems kontinuierlich zu verbessern. Im Rahmen des ML2R untersuchen wir daher Ansätze für **interaktives Maschinelles Lernen**, das **Lernen aus implizitem und explizitem Benutzer-Feedback** sowie das **Lernen von interpretierbaren oder erklärbaren Modellen** (siehe oben).

Interaktives Lernen ermöglicht unter anderem das Angehen von Problemen, welche durch die Knappheit von gelabelten Trainingsdaten in bestimmten Anwendungsbereichen entstehen. Ein Beispiel hierfür ist das Problem der Relationsextraktion im Text Mining. ML2R-Wissenschaftler*innen konnten zeigen, dass eine Fernüberwachung in Kombination mit aktivem Lernen und Experten-Feedback während der Laufzeit neuronalen Netzen ermöglichte, die Ergebnisse der Relationsextraktion zu verbessern.⁹

Ein weiteres Anwendungsbeispiel ist das Social Media Mining für Community Policing und Notfallmaßnahmen. Hier entwickelte ML2R den ersten Prototyp eines Systems, das raumbezogene Verteilungen und in Social Media Posts ausgedrückte Meinungen analysiert. Da die auf Social-Media-Plattformen verwendete Sprache häufigen Änderungen unterliegt, integriert das System verschiedene Techniken des Maschinellen Lernens, insbesondere durch häufige Abfragen nach Benutzer-Feedback.¹⁰

Quantum Computing

Die **Quantum Machine Learning (QML)**-Forschung innerhalb des ML2R lässt sich in zwei Hauptbereiche unterteilen: **Hardware** und **Algorithmen**. Im Bereich der QML-Hardware werden erste Studien auf einem Adiabatischen Quantencomputer (AQC), genauer gesagt auf dem D-Wave 2000Q-System der NASA, durchgeführt. Das Fraunhofer-Center für Quanten-Computing (14 Fraunhofer-Institute sind bereits beteiligt) betreibt kooperativ zusammen mit IBM ein IBM Q System One.

Das **adiabatische Theorem der Quantenmechanik** besagt, dass durch die graduelle Veränderung der Hamiltonian eines Systems das System in seinem n-ten Eigenzustand verbleibt (vorausgesetzt, wir initialisieren das System in diesem Zustand). Adiabatische Quantencomputer stützen sich auf dieses Prinzip zur Lösung **quadratisch binärer unbeschränkter Optimierungsprobleme** (Quadratic Binary Unconstrained Optimization Problems, **QUBOs**). AQCs beginnen mit der Lösung eines eingebauten trivialen QUBO-Problems H und wandeln H allmählich in eine vom Benutzer spezifizierte QUBO-Instanz H^* um - wenn die Größe der Änderungen klein genug war, wird der Zustand des Systems schließlich ein Optimierer von H^* sein (postuliert durch das adiabatische Theorem). Eine QUBO-Instanz kann durch ihre Koeffizientenmatrix oder, äquivalent, als Parameter eines Ising-Modells dargestellt werden. Letzteres macht adiabatisches Quanten-Computing aus Sicht des Maschinellen Lernens (ML) interessant, da viele sehr rechenintensive ML-Probleme als Ising-Modell (re-)formuliert werden können, unter anderem: Maximum-a-posteriori-Schätzung in probabilistischen Modellen wie Markov Random Fields, Boltzmann-Maschinen und Hopfield-Netzen.

Unsere initiale Studie zeigt, dass viele relevante probabilistische Modelle nicht durch AQCs aus der realen Welt optimiert werden können. Jenseits der 64 Dimensionen lässt die Hardware-Architektur voll besetzte Koeffizienten-Matrizen nicht zu, was groß angelegte ML-Modelle ausschließt. Motiviert durch diese Tatsache haben

⁹ Adilova, L., Giesselbach, S., Rüping, S. (2018). *Making Efficient Use of a Domain Expert's Time in Relation Extraction*. <https://arxiv.org/pdf/1807.04687.pdf>

¹⁰ Kirsch, B., Giesselbach, S., Knodt, D., Rüping, S. (2018). Robust End-User-Driven Social Media Monitoring for Law Enforcement and Emergency Monitoring. In G. Leventakis/M. Haberfeld (eds.), *Community-Oriented Policing and Technological Innovations* (S. 29-36). Cham, Switzerland: Springer. <https://library.open.org/bitstream/handle/20.500.12657/27840/1002165.pdf?sequence=1#page=41>

Das ML2R informiert

ML2R-Wissenschaftler*innen einen eigenen quanteninspirierten Hardware-Löser entwickelt, der dichte Koeffizientenmatrizen ermöglicht.¹¹ Dieser Solver ist eine **FPGA**-basierte Hardware-Implementierung eines evolutionären Algorithmus mit Mutationsraten-Anpassung, der speziell für die QUBO-Zielfunktion entwickelt wurde. Eine Studie über handelsübliche FPGA-Hardware zeigt, dass das System 2048-dimensionale QUBO-Instanzen in wenigen Sekunden lösen kann und dabei weniger als 10 W verbraucht. Allein die Kühlung des 2000Q-Systems von D-Wave verbraucht 25 kW.¹²

Parallel dazu untersuchen wir Quanten-Gate-Computer. Während AQCs spezialisierte QUBO-Löser sind, ermöglichen gate-basierte Quantencomputer dem Benutzer die direkte Manipulation von Quantenbits mit einer benutzerspezifischen Quantenschaltung. Erste Studien an verschiedenen IBM-Q-System-Geräten (von 5 bis 53 Quantenbits) zeigen, dass die Anwendbarkeit durch die Quantenbit-Konnektivität des Systems und der unterstützten Menge von Quanten-Gattern eingeschränkt ist. Das heißt, nicht jede theoretisch gültige Quantenschaltung kann von realen Quanten-Gate-Maschinen ausgeführt werden. Im QML-Kontext versuchen wir, diese Einschränkung zu überwinden, indem wir eine randomisierte Suchheuristik entwickeln, welche Quantenschaltungen für die Klassifizierung lernt. Ausgehend von einem Datensatz eines Klassifizierungsproblems gibt die Methode eine Quantenschaltung aus, die Daten klassifizieren kann, indem sie nur solche Quantengatter verwendet, die physikalisch realisierbar sind.

Im Bereich der QML-Algorithmen untersuchen wir die Neuformulierung bekannter ML-Probleme zu Ising- oder QUBO-Formen, wodurch sie für AQC und unseren FPGA-basierten Löser zugänglich werden. Diese Forschungsanstrebungen umfassen Clustering-Probleme¹³, Max-Sum Dispersion¹⁴, die im Zusammenhang mit Internetsuchen oder Textzusammenfassungen auftritt, sowie Markov Random Fields¹⁵, und eine Version der Support-Vektor-Maschine.¹⁶ In einem zweiten Forschungszweig untersuchen wir die Leistung von Machine-Learning-Modellen, die bereits in QUBO-Form vorliegen, wie Boltzmann-Maschinen.¹⁷ Hierbei liegt der Fokus auf der bedingten Unabhängigkeitsstruktur des Modells. Das langfristige Ziel dieser Arbeit ist eine flache Einbettung eines Modells in die physikalische Quantenbit-Konnektivität, da die Tiefe von Quantenschaltungen ein hohes Rauschen in die Berechnung induziert.

Der Arbeitskreis "Physik, moderne Informationstechnologie und Künstliche Intelligenz" der Deutschen Physikalischen Gesellschaft (DPG) organisierte seine Herbsttagung 2019 rund um das Thema Quantentechnologie.

¹¹ Piatkowski, N., Mücke, S. (2020). *System zur evolutionären Optimierung pseudoboolescher Polynome*. Munich, Germany: Deutsches Patent- und Markenamt [Patent: DE 10 2020 206 127.0].

¹² Mücke, S., Piatkowski, N., Morik, K. (2019). Hardware Acceleration of Machine Learning Beyond Linear Algebra. *ECML PKDD Workshops*, 1, S. 342-347. https://link.springer.com/chapter/10.1007/978-3-030-43823-4_29

¹³ Mücke, S., Piatkowski, N., Morik, K. (2019). Hardware Acceleration of Machine Learning Beyond Linear Algebra. *PKDD/ECML Workshops*, 1, S. 342-347. https://link.springer.com/chapter/10.1007/978-3-030-43823-4_29; Mücke, S., Piatkowski, N., Morik, K. (2019). Learning Bit by Bit: Extracting the Essence of Machine Learning. *Proc. LWDA*, S. 144-155. http://ceur-ws.org/Vol-2454/paper_51.pdf; Bauckhage, C., Ojeda, C., Sifa, R., Wrobel, S. (2018). Adiabatic Quantum Computing for Kernel k=2 Means Clustering. *Proc. LWDA*. <http://ceur-ws.org/Vol-2191/paper3.pdf>; Bauckhage, C., Piatkowski, N., Sifa, R., Hecker, D., Wrobel, S. (2019). A QUBO Formulation of the k-Medoids Problem. *Proc. LWDA*, S. 54-63. http://ceur-ws.org/Vol-2454/paper_39.pdf

¹⁴ Bauckhage, C., Sifa, R., Wrobel, S. (2020). Adiabatic Quantum Computing for Max-Sum Diversification. *Proc. SIAM International Conference on Data Mining*, S. 343-351. <https://epubs.siam.org/doi/pdf/10.1137/1.9781611976236.39>

¹⁵ Mücke, S., Piatkowski, N., Morik, K. (2019). Hardware Acceleration of Machine Learning Beyond Linear Algebra. *ECML PKDD Workshops*, 1, S. 342-347. https://link.springer.com/chapter/10.1007/978-3-030-43823-4_29

¹⁶ Mücke, S., Piatkowski, N., Morik, K. (2019). Learning Bit by Bit: Extracting the Essence of Machine Learning. *Proc. LWDA*, S. 144-155. http://ceur-ws.org/Vol-2454/paper_51.pdf

¹⁷ Piatkowski, N. (2019). Hyper-Parameter-Free Generative Modelling with Deep Boltzmann Trees. *ECML PKDD*, 2, pp. 415-431. <https://ecmlpkdd2019.org/downloads/paper/637.pdf>

Das ML2R informiert

Katharina Morik, die im Vorstand der derzeit rund 900 Mitglieder zählenden Arbeitsgruppe sitzt, hielt auf der Herbsttagung in Freiburg einen Vortrag über Maschinelles Lernen und knüpfte Kontakte zu internationalen Kollegen, die an ML und Quanten-Computing arbeiten.

ML mit limitierten Ressourcen

Die Ressourcen Laufzeit und Speicher werden bereits seit langer Zeit untersucht, zumeist in der O-Notation der Worst-Case-Komplexität, manchmal aber auch im Worst-Case der Ausführung. Angesichts der Heterogenität moderner Rechnerarchitekturen und der Herausforderung durch große Datenmengen und Datenströme im Internet der Dinge steht die Forschung im Bereich des Maschinellen Lernens vor großen Herausforderungen. Insbesondere **FPGAs** eignen sich gut für Anwendungen im Internet der Dinge, erfordern aber gleichzeitig eine Neuimplementierung bekannter ML-Algorithmen. Der Ressourcenbedarf mehrerer Implementierungen für einen Algorithmus muss charakterisiert werden. Beispielsweise bietet selbst der bekannte Entscheidungsbaum-Algorithmus inzwischen neuartige Forschungsfragen: Welche Implementierung ist für welche Rechenarchitektur und Anwendungsanforderungen am besten geeignet?¹⁸ Deep Learning-Modelle können auch auf FPGAs ausgeführt werden - es existieren bereits mehrere Implementierungen. Im Rahmen einer kürzlich erschienenen Arbeit des ML2R zusammen mit dem Sonderforschungsbereich 876 und seinem Astrophysik-Projekt wurde ein Ende-zu-Ende-Lernen aus Bildströmen eines Teleskops mit Hilfe von Deep Learning erfolgreich entwickelt und das gelernte Modell auf einem FPGA direkt am Teleskop implementiert.¹⁹

Forscher am Fraunhofer IML setzen in aktuellen KMU-Projekten einen leistungsfähigen Supercomputer ein, um Modelle für die Bilderkennung zu trainieren. Für mobile Roboter werden energieeffiziente Prozessoren eingesetzt, um Bilddaten von den Robotern in die Cloud zu senden, wo Berechnungen durchgeführt werden. Falls Skalierungs- und Datenschutzerwägungen dieses Verfahren verbieten, sind zwei Ansätze möglich. Zum einen kann automatisch ein für das System optimierter Programmcode generiert werden, der unabhängig von gängigen Softwarepaketen und damit wesentlich effizienter ist.²⁰ Zum anderen können die für die Problemlösung konzipierten Modelle, wie z.B. das Zählen von Objekten, spezialisiert werden, so dass keine unnötigen Ergebnisse berechnet werden müssen.²¹ Beide Ansätze zusammen ermöglichen den Einsatz der Modelle direkt auf den ressourcenbeschränkten Systemen.

Umgekehrt können wir uns die ML-Algorithmen ansehen und erkennen, wie wir das Modell so verändern können, dass weniger Ressourcen verbraucht werden. Bei probabilistischen graphischen Modellen konnte gezeigt werden, dass die Verwendung von ganzzahligen Variablen während des gesamten Lernprozesses und der Inferenz Ressourcen spart, ohne zu viel Genauigkeit zu verlieren. Das komplexe PGM könnte sogar auf Geräten mit extrem niedrigem Stromverbrauch ausgeführt werden!²²

¹⁸ Buschjäger, S., Morik, K. (2020). Decision Tree and Random Forest Implementations for Fast Filtering of Sensor Data. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 65(1), 209-222. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7962153>

¹⁹ Buschjäger, S., Pfahler, L., Buss, J., Morik, K., Rhode, W. (2020). On-Site Gamma-Hadron Separation with Deep Learning on FPGAs. *ECML PKDD* (accepted, to be published), Ghent, Belgium.

²⁰ Urbann, O. et al. (2020). *A C Code Generator for Fast Inference and Simple Deployment of Convolutional Neural Networks on Resource Constrained Systems*. <https://arxiv.org/pdf/2001.05572.pdf>

²¹ Urbann, O., Stenzel, J. (2019). A Convolutional Neural Network that Self-Contained Counts. *Journal of Image and Graphics*, 7(4), 112-116. <http://www.joig.org/uploadfile/2019/1113/20191113120905251.pdf>

²² Piatkowski, N., Lee, S., Morik, K. (2016). Integer undirected graphical models for resource-constrained systems. *Neurocomputing*, 173, 9-23. <https://reader.elsevier.com/reader/sd/pii/S0925231215010449?token=6EBE8492BD3AD0DBFB42198F5513C0A7213285115EC85DED41508C43F1E8B5452CAFB78F52FA92303BF731010EC1EC72>

Das ML2R informiert

Probabilistisches Gap Filling auf räumlich-zeitlichen Datensätzen

Räumlich-zeitliche Datensätze ermöglichen die gleichzeitige Untersuchung von verschiedenen Orten über die Zeit. In vielen Fällen sind solche Daten jedoch unvollständig, was die Analyse erheblich erschwert. **Satellitenbilder** zum Beispiel liefern wichtige Hinweise zu Vorgängen auf der Erdoberfläche und können exemplarisch Auswirkungen des Klimawandels sowie Prozesse der Urbanisierung und agrarischen Landflächennutzung abbilden. Allerdings stellen Verdeckungen der Erdoberfläche durch Wolken ein zentrales Problem bei der Verarbeitung dar. ML2R-Wissenschaftler*innen der TU Dortmund haben deshalb an einem Verfahren gearbeitet, um Lücken in beliebigen räumlich-zeitlichen Datensätzen zu füllen. Im Gegensatz zu bisherigen Ansätzen kommt die Methode ohne Hilfsinformationen aus und benötigt keine spezifischen Dateneigenschaften. Die Lücken werden durch ein wahrscheinlichkeitsbasiertes Modell aufgefüllt, welches Vorhersagen unter Berücksichtigung der teilbeobachteten räumlich-zeitlichen Nachbarschaft macht. Ihre Methode wendeten die Wissenschaftler*innen erfolgreich auf Satellitenbilder an, wo sie die durch Wolken verdeckten Teile der Oberfläche rekonstruieren konnten.²³

Kombination von Expertenwissen und Neuronalen Netzen zur Modellierung von Umweltauswirkungen in der Landwirtschaft

Das Bestreben, bessere Strategien für die Auswahl und den Anbau von Nutzpflanzen zu entwerfen, bildet seit jeher einen zentralen Kern der Forschung in den Bereichen Landwirtschaft und mathematische Modellierung. Hierbei befasst sich eine Vielzahl von Studien mit der **Beziehung zwischen Umweltbedingungen und Pflanzenwachstum** - insbesondere mit den Auswirkungen bestimmter Umweltbelastungen (z.B. Hitze und Dürre) auf Nutzpflanzen und ihre Ernte. Im vergangenen Jahr haben ML2R Expert*innen zusammen mit Kolleg*innen des Exzellenzclusters PhenoRob an der Universität Bonn erfolgreich einige dieser Fragen mit Hilfe der Künstlichen Intelligenz untersucht. Sie agierten hierbei an der Schnittstelle zwischen mathematischer Modellierung und Techniken aus der Analyse von Big Data und Deep Learning. Ihre vorgeschlagene Lösung stützt sich einerseits auf Wissen aus der klassischen Agrarforschung in Form empfindlicher Modelle, die explizite Umwelt- und Pflanzenmerkmale beschreiben (Auswirkungen der Photosynthese, der Transpiration von Nutzpflanzen und der Hydrodynamik des Bodens). Zum anderen wurden diese spezifischen Domänenkenntnisse genutzt, um Convolutional Neural Networks (CNN), mit denen geeignete Datenrepräsentationen und verborgene Verbindungen zwischen Umwelt- und Pflanzenmerkmalen aufgedeckt werden sollen, weiter zu lenken und zu regulieren. Der darauf aufbauende Ansatz des Maschinellen Lernens **kombiniert das Wissen von Experten mit der Flexibilität und Ausdruckskraft großer Daten und neuronaler Netze**. Dieser Forschungsansatz liefert neuartige Algorithmen zur Analyse von Umweltbelastungen und zur Identifizierung der widerstandsfähigsten Pflanzensorten (Hybride). Der Beitrag wurde 2019 im Rahmen der jährlichen "Crop Challenge in Analytics" des weltweit tätigen Landwirtschaftskonzerns Syngenta eingereicht und zum Sieger des Wettbewerbs gekürt.²⁴

AUSZEICHNUNGEN

Die dem Kompetenzzentrum vorstehenden Persönlichkeiten vereinen wissenschaftliche Expertise und Exzellenz mit einem weitreichenden Netzwerk wissenschaftlicher und industrieller Kontakte. Ihre Leistungen im Feld maschineller Lernverfahren wurden vielfach ausgezeichnet.

²³ Fischer, R., Piatkowski, N., Pelletier, C. Webb, G., Petitjean, F., Morik, K. (2020). No Cloud on the Horizon: Probabilistic Gap Filling in Satellite Image Series. *DSAA* (accepted, to be published), Sydney, Australia.

²⁴ Weitere Informationen erhalten Sie unter: <https://www.syngenta-us.com/thrive/research/winners-2019-crop-challenge.html>

Das ML2R informiert

Die Sprecherin des ML2R, **Prof. Dr. Katharina Morik**, wurde im Oktober 2019 durch die Gesellschaft für Informatik e.V. (GI) mit dem **GI-Fellowship** ausgezeichnet. Katharina Morik wurde hierbei als Pionierin des Maschinellen Lernens gewürdigt, da sie das Forschungsgebiet auch auf internationaler Ebene maßgeblich vorangebracht hat. Insbesondere ihre Zusammenarbeit mit Yves Kodratoff von der Université Paris-Sud in den frühen Jahren des Faches hat den Grundstein für die europäische Forschungscommunity des Maschinellen Lernens gelegt. Die so entstandene Tagung ECML PKDD umfasst nun alle europäischen Länder und hatte im letzten Jahr 800 Teilnehmerinnen und Teilnehmer. Zudem wurden die Interdisziplinarität und hohe Anwendungsrelevanz ihrer wissenschaftlichen Arbeit hervorgehoben.

Prof. Dr. Stefan Wrobel, Sprecher des ML2R, wurde im Juli 2019 durch die Gesellschaft für Informatik e.V. (GI) als **einer von „zehn prägenden Köpfen“ der deutschen Geschichte der Künstlichen Intelligenz** ausgezeichnet. Im vom Bundesministerium für Bildung und Forschung (BMBF) ausgerufenen „Wissenschaftsjahr 2019 – Künstliche Intelligenz“ entschied sich die Jury für zehn herausragende Wissenschaftlerinnen und Wissenschaftler, die KI in Deutschland maßgeblich prägen. Prof. Dr. Stefan Wrobel wurde als Experte im Bereich des Maschinellen Lernens gekürt.

Prof. Dr. Dr. h. c. Michael ten Hompel, Vertreter des Kompetenzzentrumspartners Fraunhofer IML, gilt als einer der Väter des Internet der Dinge und veröffentlichte zahlreiche Fachbücher und über vierhundert Artikel rund um Logistik und IT. Da er die Digitalisierung des Logistikstandortes Ruhr wegbereitend vorangetrieben hat, wurde Michael ten Hompel im Dezember 2018 die **Ehrendauszeichnung „Bürger des Ruhrgebiets“** verliehen. Im April 2019 wurde er zudem mit dem **„HERMES Verkehrs.Logistik.Preis 2019“** für seine herausragenden Leistungen in der internationalen Logistik geehrt.

Prof. Dr. Christian Bauckhage, Vertreter des Kompetenzzentrumspartners Universität Bonn, wurde gemeinsam mit den Forschenden Rafet Sifa (Fraunhofer IAIS) und Dr. Tiansi Dong (Bonn-Aachen International Center for Information Technology) mit dem **Best Paper Award** auf der International Conference on Artificial Neural Networks (ICANN 2019) ausgezeichnet. In dem institutionsübergreifenden Projekt unter der Flagge des ML2R widmeten sich die Forscher den „Prototypes within Minimum Enclosing Balls“. Sie stellten hierbei ein Verfahren vor, das aus einer gegebenen Datenmenge diejenigen charakteristischen Datenpunkte herausfiltert, die für menschliche Nutzer besonders gut zu interpretieren sind (»explainable latent factors«).

WISSENSCHAFTLICHE NACHWUCHSFÖRDERUNG

Das Kompetenzzentrum Maschinelles Lernen Rhein-Ruhr hat sich der Förderung des wissenschaftlichen Nachwuchses im Schlüsselbereich maschineller Lernverfahren verschrieben. An den im Kompetenzverbund beteiligten Institutionen hat die Förderung und Vermittlung von Fachwissen an den aufstrebenden Nachwuchs lange Tradition. Das ML2R bietet Masterabsolvent*innen die Möglichkeit einen Abschluss mit Doktorgrad im Forschungsbereich maschineller Lernverfahren zu erwerben. Exzellenten Wissenschaftler*innen wird zudem die Möglichkeit einer Beschäftigung als PostDoc geboten.

Die herausragenden Leistungen des ML2R-Nachwuchses zeigten sich exemplarisch in der Auszeichnung der Dissertationsarbeit von ML2R-Forscher Dr. Nico Piatkowski mit einem **Dissertationspreis** der TU Dortmund. Er schloss seine Arbeit mit dem Titel „Exponential Families on Resource-Constrained Systems“ mit Auszeichnung (summa cum laude) ab. Es ist die erste umfassende Untersuchung von Ressourcenbedarfen und Qualitätsgrenzen für probabilistische grafische Modelle. Die Arbeit bietet in einem ersten Schritt eine gründliche und anleitende Übersicht über Exponentialfamilien, bevor sie die neuartigen Lernmodelle von räumlich-zeitlichen Modellen, ganzzahligen Zufallsfeldern und eine innovative Approximation der Partitionsfunktion vorstellt. Im Anschluss an seine Promotion nahm Nico Piatkowski eine Position als PostDoc am Kompetenzzentrum an.

Das ML2R informiert

AUSSICHTEN

Das ML2R strebt an, **internationale Forschungsk Kooperationen** voranzutreiben. Ausgehend vom weiten Netzwerk bestehender internationaler Kontakte in der Forschungsgemeinschaft des Maschinellen Lernens ist daher geplant, gemeinsame Forschungsprojekte und Publikationen umzusetzen. Sollten Sie oder Ihr wissenschaftliches Team an der Umsetzung einer gemeinschaftlichen Forschungsk Kooperation, etwa im Rahmen von Publikationen, Workshops, Panels etc., interessiert sein, so kontaktieren Sie uns gerne.

Gemeinsam mit dem Sonderforschungsbereich 876 „Verfügbarkeit von Information durch Analyse unter Ressourcenbeschränkung“ wird das ML2R im Zeitraum des 31.08. - 04.09.2020 eine **internationale Sommerschule** veranstalten. Die virtuelle Veranstaltung wird Expertinnen und Experten aus den Forschungsfeldern der Datenanalyse (Maschinelles Lernen, Data Mining, Statistik) und der Eingebetteten Systeme (Cyber-Physikalische Systeme) zusammenführen. In ihren Vorträgen widmen sie sich der Ressourcenbeschränkung von Geräten im Rahmen von Datenanalysen. Die Sommerschule wird den Teilnehmenden zudem die Möglichkeit geben, ihre Forschung vorzustellen und sich im Austausch darüber untereinander zu vernetzen. Begleitet wird die Sommerschule von einem Hackathon in Form einer Kaggle Challenge, in welcher die Teilnehmenden ihr Wissen über Maschinelles Lernen und Cyber-Physikalische Systeme in einem realen Szenario testen können. Weitere Informationen zur Sommerschule erhalten Sie unter:

<https://www-ai.cs.tu-dortmund.de/summer-school-2020/>

Das ML2R informiert



Transfer

UNTERNEHMENSKOOPERATIONEN

Das ML2R ermöglicht Unternehmen im Rahmen einer Kooperation den Zugang zu Strategien, Wissen und Daten maschineller Lernverfahren (ML). Anhand von Daten, die durch die Unternehmen zur Verfügung gestellt werden, können **Potenziale Maschinellen Lernens analysiert** und die **Umsetzung von Anwendungen in der Praxis unterstützt** werden. Bisherige Anwendungsinteressenten des ML2R bilden ein breites Wirtschaftsspektrum ab, welches exemplarisch die Bereiche Logistik und Industrie 4.0, Telekommunikationsanbieter und das Compliance-Gebiet umfasst. Cybersicherheit, Anomalieerkennung, Qualitätsprognosen in Prozessen zur Ressourceneinsparung, realzeitliche Entscheidungsunterstützung sowie die Erklärbarkeit beziehungsweise Transparenz von Maschinellern konnten als mögliche Anwendungsbereiche in den Unternehmen identifiziert werden. Der Bedarf einer Vernetzung und Zusammenarbeit von exzellenter Forschung und innovativem Unternehmertum zeigte sich in 42 Gesprächen, welche von ML2R-Mitarbeiter*innen mit Unternehmen geführt wurden. So fanden unter anderem Gespräche mit den Firmen KraussMaffei, Deutsche Telekom AG/Telekom Security, Wilo SE und targens statt.

Das Team des ML2R unterstützt Unternehmen nicht nur im Rahmen von Kooperationen bei der Einführung von Maschinellern in die unternehmenseigenen Prozesse und Produkte, sondern hilft auch beim **Auf- und Ausbau der notwendigen Kompetenz**. Das noch in der Planung befindliche Seminarangebot ist hierbei nur ein Baustein. Angedacht sind darüber hinaus auch Co-Working Formate, bei denen Unternehmen gemeinsam mit ML2R-Expertinnen und Experten an innovativen Anwendungen des Maschinellen Lernens arbeiten. Dabei spielt der Aufbau von Know-How in den Unternehmen und der direkte Transfer von Forschung in die Praxis eine entscheidende Rolle.

Die **Generierung von C Quelltext**²⁵ aus trainierten Modellen zur Integration in firmeneigenen Frameworks ist nicht nur für kleine und mittelständische deutsche Unternehmen interessant. Im ML2R wird ein bereits verfügbarer Code Generator als eigenständige Lösung weiterentwickelt. Gleichzeitig ist das Fraunhofer IML an **Googles Open Source Projekt "Intermediate Representation Execution Environment" (IREE)** beteiligt, welches eine C Code Generierung des Fraunhofer IML integriert. Innerhalb des IREE werden zudem weitere Möglichkeiten zur einfachen Bereitstellung und Anwendung evaluiert.

KION STIFTUNGSPROFESSUR

Durch eine verstärkte Kooperation mit Industriepartnern treibt das ML2R eine anwendungsorientierte Erforschung des Maschinellen Lernens voran und ermöglicht den Transfer erstklassiger Forschung in einem Schlüsselgebiet der Künstlichen Intelligenz in die Industrie. Die Kooperationen erweisen sich hierbei von beiderseitigem Nutzen und ermöglichen es Forschern realpraktische Bedürfnisse der hiesigen Wirtschaft in ihrer Arbeit zu adressieren.

Die KION GROUP AG hat die **Stiftungsprofessur** „Maschinelles Lernen für industrielle Anwendungen“ an der TU Dortmund ins Leben gerufen. Ein entsprechender Einstellungsprozess wird zurzeit durchgeführt. Der ausgewählte Wissenschaftler wird sich in die Forschungsaktivitäten des ML2R einbringen und fördert somit den Ausbau des Kompetenzzentrums. Zudem wird durch die Einrichtung der Stiftungsprofessur ein Beitrag zur Grundfinanzierung der angestrebten Verstärkung des ML2R geleistet.

²⁵ Fraunhofer IML hat einen Open Source Neural Network Code Generator (NNCG) veröffentlicht (Abzurufen unter: <https://github.com/iml130/nncg>).

Das ML2R informiert

FLAGSHIP-ANWENDUNGEN

Die Flagships bilden den Rahmen für die anwendungsorientierte Evaluierung der im Kompetenzzentrum erforschten Methoden in konkreten Szenarien. Die Generalisierbarkeit der Ergebnisse wird durch die Auswahl der Anwendungsgebiete, welche sich durch technische und branchenspezifische Charakteristika unterscheiden, gegeben. Flagship-Projekte des ML2R zentrieren sich zurzeit auf die Bereiche Industrie und Logistik 4.0 sowie Cognitive Process Automation.

An dem Fraunhofer IML und der TU Dortmund existiert ein wissenschaftliches Versuchsfeld, das auf Szenarien der **Industrie 4.0** und **Logistik 4.0** basiert. Demonstratoren bilden spezifische Logistiksubsysteme, an denen sich die Wirkung der fortschreitenden Autonomisierung und der neuartigen Methoden des Maschinellen Lernens am stärksten zeigen werden. Der Fokus der hierauf aufbauenden Demonstratoren liegt sowohl auf der Ebene der logistischen Prozesse beziehungsweise Anlagensteuerung als auch auf der menschlichen Aktivität. Sie zeigen somit, in welcher Weise sich der Betrieb und die Optimierung logistischer Systeme durch die Anwendung des Maschinellen Lernens weiterentwickeln werden. Ein im Rahmen des ML2R entwickelter Demonstrator ermöglicht die Objekterkennung mithilfe tiefer neuronaler Netze bei Raten von ca. 30 Hz und 0,5 W Energieverbrauch. In Forschungsprojekten zur Indoor-Lokalisierung für intralogistische Prozesse werden zudem Lokalisierungsansätze durch autonome Systeme, anhand standardisierter RGB Kameras sowie über externe, nicht-optische Sensorik entwickelt.

Am Fraunhofer IML ermöglicht die "**ai arena**" den Zugang zu neuartiger praxis- und industrienaher Hardware. Der Forschungsschwerpunkt liegt sowohl auf der Entwicklung von Code-Generatoren als auch auf der KI-autonomen Schwarmrobotik mit Mensch-Technik-Interaktion. Letztere zielt darauf ab, Schwärme von Robotern und Drohnen unter ständiger Optimierung erfolgreich zu koordinieren und zu navigieren. Mit Hilfe des Reinforcement Learning haben Wissenschaftler die Simulation virtueller Welten mit einzelnen Robotern getestet. Diese mussten durch einen Raum navigieren, welcher mit Hindernissen versehen war. Das Verfahren soll nun in der realen Welt auf einen Schwarm von realen Robotern angewendet werden.

ai arena am Fraunhofer IML:



Darüber hinaus haben das Fraunhofer IML und der Lehrstuhl für Materialfluss- und Logistikmanagement der TU Dortmund das "**PhyNetLab**" eingerichtet. Bestehend aus cyber-physikalischen Objekten (PhyNodes) stellt das PhyNetLab anhand realistischer Maßstäbe ein Testfeld für Lagergebäude dar. Es greift somit den Bedarf an Flexibilität und Skalierbarkeit für neu entstehende Lager sowie die aus der Multidisziplinarität in den Bereichen Logistik, eingebettete Systeme und drahtlose Kommunikation resultierende Komplexität solcher Systeme auf.

Das ML2R informiert

Im Flagship **Cognitive Process Automation** sind aktuell am Fraunhofer IAIS verschiedene Demonstratoren in der Vorbereitung. Sie extrahieren Informationen aus Rechnungsbelegen, Geschäftsberichten und Gerichtsurteilen mittels Natural Language Processing (NLP) und Maschinellern zur Entitätenerkennung und Anonymisierung. Zusätzlich unterstützen die Systeme menschliche Prüfer bei der Analyse und Annotation von Textbausteinen und Tabellen. Darüber hinaus, umfassen die Flagship Demonstratoren ein KI-unterstütztes Tool zur automatischen Abarbeitung von Checklisten im Hinblick auf die Beantwortung diverser Fragen oder die Erfüllung unterschiedlicher Voraussetzungen, z.B. die Vollständigkeit der obligatorischen Daten, und zum Konsistenzcheck des Dokumentinhaltes in unterschiedlichen Tabellen und Textpassagen sowie in unterschiedlichen Dokumenten.

Die Demonstratoren zur Belegerkennung und Analyse von Gerichtsurteilen integrieren dabei Methoden aus dem Statistischen Relationalen Lernen, die es erlauben, Vorwissen von Expert*innen in Form von Regeln auszuwerten und zu gewichten. Dies ermöglicht auch schon mit kleineren Trainingsmengen gute Ergebnisse zu erzielen und ist besonders hilfreich bei der Auswertung semi-strukturierter Daten, in diesem Fall Dokumenten mit Layout-Informationen.

AUSSICHTEN

Das ML2R ist bestrebt Kooperationen mit Partnern aus Industrie und Wirtschaft zu vertiefen. Basierend auf den vielen Gesprächen und Vorträgen in Unternehmen, sind **verstärkte Kooperationen mit Industriepartnern** in Form von Forschungsprojekten und gemeinsamen öffentlichkeitswirksamen Maßnahmen geplant. Auch die gemeinsame Teilnahme an themenspezifischen Konferenzen ist vorgesehen. Sollten Sie und Ihr Unternehmen an einer Kooperation mit dem ML2R, etwa im Rahmen eines Forschungsprojektes in Ihrem Unternehmen oder der gemeinsamen Teilnahme an Konferenzen, interessiert sein, so kontaktieren Sie uns gerne.

Zusätzlich sind **Seminarangebote** und **Schulungen** zu den ML2R Forschungsschwerpunkten in der Planung. Insbesondere die neuen Themen Quantum ML und Vertrauenswürdiges ML werden hierbei ins Zentrum gerückt. Aber auch anwendungsnahe Themen aus den beiden Flagships sind angedacht.

Das ML2R informiert



Sichtbarkeit

VERNETZUNG

Das Kompetenzzentrum selbst fußt auf der exzellenten Grundlagenforschung an den ML2R-Standorten - das sind der Sonderforschungsbereich 876 zur ressourcenlimitierten Datenanalyse und das Max-Planck-Institut für Cybersecurity am Standort Dortmund sowie die Exzellenzcluster "Robotik und Phänotypisierung für Nachhaltige Nutzpflanzenproduktion" und "Matter and Light for Quantum Computing" am Standort Bonn. Das Umfeld der Standorte bietet ein Ökosystem aus Forschung, Transfer und Innovation, wie die nachfolgenden Abbildungen zeigen.

Das "Centrum für Entrepreneurship & Transfer" (CET) der TU Dortmund fördert und unterstützt Existenzgründungen. Das Technologie Zentrum Dortmund beherbergt mehr als 300 kleine und mittelständische Unternehmen mit rund 10 000 Beschäftigten. Das Leibniz-Institut für Analytische Wissenschaften (ISAS) an der TU Dortmund treibt die chemische und physikalische Messtechnik voran, unter anderem die Erkennung von Viren durch ein innovatives Fernrohr.²⁶ "BAuA" ist die Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Gemeinsam mit der Fraunhofer Gesellschaft und weiteren Institutionen trägt die Universität Bonn das Bonn-Aachen International Center for Information Technology (b-it). "b-it" bietet drei internationale Master-Studiengänge in Informatik und Angewandter Informatik an und umfasst neun Forschungsgruppen - alle auf einem international exzellenten Niveau.

Das Fraunhofer IAIS mit Sitz in Sankt Augustin bei Bonn beherbergt unter anderem einen der Standorte von "AI4EU", das zentrale Leuchtturm-Projekt der Europäischen Union im Bereich Künstliche Intelligenz. Das Fraunhofer IAIS ist auch Teil der "International Data Spaces Association". Dieses bietet seinen mehr als 100 teilnehmenden Unternehmen und Institutionen verschiedener Branchen und Größen aus 20 Ländern, darunter mehrere Fortune-500-Unternehmen, eine vom Fraunhofer IAIS entwickelte virtuelle Architektur, die einen sicheren Datentransfer ermöglicht.

Das *innovative Umfeld der ML2R-Standorte:*



²⁶ Shpacovitch, V. et al. (2017). Application of the PAMONO-Sensor for Quantification of Microvesicles and Determination of Nano-Particle Size Distribution. *Sensors*, 17(2), 244. <https://www.mdpi.com/1424-8220/17/2/244/htm>

Das ML2R informiert

Die **Rhein-Ruhr Region** bietet Innovation und Tradition: international anerkannte Forschung zu Maschinellen Lernen (ML) und Big Data einerseits und ein herausragender Wirtschaftsraum mit über 20 DAX- und MDAX-Unternehmen andererseits.

Das ML2R profitiert in seiner Arbeit von dem ihm umspannenden Netzwerk. So beraten hochrangige wissenschaftliche und wirtschaftliche Akteure die Mitarbeiter*innen des Kompetenzzentrums entlang ihrer spezifischen Expertise im Rahmen einer Beiratstätigkeit. Das **Steering Board**, welches sich aus Wissenschaftlern zusammensetzt und der **Industriebeirat**, dem wirtschaftliche Beiräte angehören, gestalten die Forschungsagenda aus wissenschaftlicher sowie Anwendersicht mit und ermöglichen so eine externe Qualitätssicherung.

KOORDINIERUNG

Im September 2018 beauftragte Bundesforschungsministerin Anja Karliczek die Sprecherin des Kompetenzzentrums ML2R, Prof. Dr. Katharina Morik, mit der **Koordinierung der deutschen Kompetenzzentren** für Künstliche Intelligenz (KI). Das Netzwerk deutscher KI-Kompetenzzentren ist seither gewachsen und umfasst nun sechs Knotenpunkte: das Kompetenzzentrum Maschinelles Lernen Rhein-Ruhr (ML2R), das Berlin Institute for the Foundations of Learning and Data (BIFOLD), das Munich Center for Machine Learning (MCML), das Tübingen AI Center - Competence Center for Machine Learning, das Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig (ScaDS.AI) und das Deutsche Forschungszentrum für Künstliche Intelligenz (DFKI). Die **Geschäftsstelle für Koordinierung** am Standort der TU Dortmund leitet die inhaltliche Vernetzung und Abstimmung der bundesweiten KI-Kompetenzzentren.

Die Geschäftsstelle unter Federführung von Prof. Dr. Katharina Morik verantwortet die Abstimmung sowie den inhaltlichen und thematischen Austausch **zwischen dem Netzwerk deutscher KI-Kompetenzzentren und dem französischen National AI Research Network**. Die Zusammenarbeit zwischen Deutschland und Frankreich auf dem Forschungsgebiet der Künstlichen Intelligenz wurde somit erfolgreich intensiviert. Prof. Dr. Katharina Morik steht hierzu in kontinuierlicher Verständigung mit Prof. Dr. Bertrand Braunschweig, dem Koordinator des nationalen französischen Programms zur KI-Forschung am französischen Institut Inria. Gemeinsam mit ihm treibt sie die Konkretisierung der deutsch-französischen Kollaboration weiter voran – die gemeinsamen Konsultationen umfassten eine binationale Absichtserklärung sowie kollaborative Publikationsausschreibungen. Deutschland und Frankreich unterzeichneten die gemeinsame Absichtserklärung, welche als rahmensetzende Vereinbarung die Kooperationen der Länder im Forschungsfeld der Künstlichen Intelligenz bestärkt. Der Aufruf zur Einreichung von Vorschlägen für binationale Kooperationen wird demnächst veröffentlicht.

VERANSTALTUNGEN

Global Forum on AI for Humanity. Vom 28. bis 30. Oktober 2019 fand das Global Forum on AI for Humanity in Paris statt. Prof. Dr. Katharina Morik, Sprecherin des ML2R und Koordinatorin der deutschen KI-Kompetenzzentren, hat das Treffen gemeinsam mit Prof. Dr. Bertrand Braunschweig, Koordinator der französischen KI-Kompetenzzentren am Institut INRIA, und Prof. Dr. Malik Ghallab sowie namenhaften Experten aus Australien, Kanada, Deutschland, England, Frankreich, Japan, den Niederlanden und den USA organisiert. Das Treffen in Paris mit einer Ansprache durch den Gastgeber des Forums, der französische Präsident Emmanuel Macron, diente als formeller Startschuss für die Globale Partnerschaft für Künstliche Intelligenz (GPAI). Die zukünftige Agenda der GPAI-Arbeitsgruppen wurde formuliert und ein Buch ist in Vorbereitung.

Thementag Künstliche Intelligenz. Prof. Dr. Katharina Morik organisierte den „Thementag Künstliche Intelligenz - Maschinelles Lernen als Schlüssel für die Künstliche Intelligenz“ der Nordrhein-Westfälischen Akademie der Wissenschaften und Künste, welcher am 29. Oktober 2019 in Düsseldorf stattfand. Im Rahmen von Vorträgen und einer Poster-Ausstellung wurden die Breite des Gebietes maschineller Lernverfahren sowie Anwendungsbeispiele aufgeführt. Der Sprecher des ML2R, Prof. Dr. Stefan Wrobel, hielt den Vortrag: „Künstliche Intelligenz und Maschinelles Lernen: Chancen und Herausforderungen“.

Das ML2R informiert

Treffen der deutschen und französischen Kompetenzzentren. Am 16. September 2019 kamen Vertreterinnen und Vertreter der sechs deutschen KI-Kompetenzzentren sowie der französischen KI-Institute am Rande der Konferenz ECML PKDD in Würzburg zusammen. Organisiert wurde das Treffen von Prof. Dr. Katharina Morik und Prof. Dr. Bertrand Braunschweig (INRIA), Koordinator der französischen KI-Institute. Die Wissenschaftler*innen präsentierten ihre aktuelle Forschung. Ziel des Treffens war die Konkretisierung eines virtuellen deutsch-französischen Zentrums zur Koordination der Kompetenzzentren beider Länder sowie die Erarbeitung einer gemeinsamen Erklärung zur verstärkten länderübergreifenden Zusammenarbeit.

Summer of Research. Eine Gelegenheit des gemeinsamen Austausches und Arbeitens erhielten die ML2R-Mitarbeiter*innen, welche an den Standorten Dortmund, Bonn und St. Augustin forschen, am 24. Juli 2019 beim Summer of Research in Sankt Augustin. Die Doktorand*innen, Postdocs und erfahrenen Wissenschaftler*innen widmeten sich insbesondere aktuellen Herausforderungen zum praktischen Einsatz des Maschinellen Lernens entlang ihrer verschiedenen Forschungsschwerpunkte. Es ist geplant, dass der Summer of Research dem ML2R jährlich ein Forum der Zusammenarbeit bieten wird.

Besuch von Forschungsministerin Karliczek. Bei ihrem Besuch des ML2R am 9. Juli 2019 in Dortmund erhielt Bundesforschungsministerin Anja Karliczek Einblicke in die Spitzenforschung des Kompetenzzentrums. In Begleitung von Journalist*innen begegnete sie Robotern, die KI und ML spielerisch begreifbar machen, entdeckte KI-Systeme, die gesprochene Sprache analysieren, Satellitenbilder verbessern und autonomes Fahren sicherer machen. Zudem summte über ihr ein Drohnenschwarm. Der Besuch gab den Teilnehmenden die Möglichkeit, die Exponate selbst auszuprobieren und sich über die Zukunftspläne des Kompetenzzentrums zu informieren.



Oliver Schaper / TU Dortmund



Oliver Schaper / TU Dortmund

Tagung von ML-Experten. Beginnend mit einer Eröffnungsveranstaltung im VIP-Club des Dortmunder Fußballvereins BVB 09 am Abend des 4. Juni 2019, lud das ML2R alle Forschenden der durch das Bundesministerium für Bildung und Forschung (BMBF) geförderten Machine Learning-Projekte nach Dortmund. Die über 100 Expert*innen aus mehr als 50 Projekten in Industrie, Wirtschaft und Wissenschaft nutzten die Gelegenheit, sich auszutauschen und untereinander zu vernetzen. In rund 30 Vorträgen und einer Posterausstellung wurden die vielfältigen Anwendungsbereiche des Maschinellen Lernens deutlich – von Logistik, Transport und Produktion, über Meteorologie, Biologie und ein breites medizinisches Anwendungsfeld bis hin zu Windkraftanlagen und Datensicherheit in Unternehmen.



Oliver Schaper / TU Dortmund



Oliver Schaper / TU Dortmund

Das ML2R informiert

Auftaktveranstaltung. Die Auftaktveranstaltung des ML2R am 23. Januar 2019 veranschaulichte den Dreiklang aus Grundlagenforschung, weitreichender Kompetenz und Wirtschaftlichkeit. Prof. Dr. Thorsten Joachims (Cornell University, USA) hielt einen Vortrag zum Thema "Learning from and for Interventions" und Prof. Dr. Michèle Sebag (Université Paris-Sud, Frankreich) präsentierte den Stand der Technik im Bereich "Selecting an algorithm (and hyper-parameters)". Natürlich wurde auch das wegweisende Kompetenzzentrum ML2R vorgestellt und Vertreter der Bundesregierung (Prof. Dr. Wolf-Dieter Lukas), der Landesregierung Nordrhein-Westfalens (Ministerin Isabel Pfeiffer-Poensgen) sowie die Rektorin der TU Dortmund, Prof. Dr. Dr. h. c. Ursula Gather, stellten seine gesellschaftliche Bedeutung heraus. Erfolgsgeschichten von ML-Anwendungen, z.B. bei T-Systems, Volkswagen AG und Pricewaterhouse Coopers, wurden durch eine Innovationsschau abgerundet, bei der die Firmen Comma Soft AG, Crosscan GmbH, Integrated Analytics GmbH, Old World Computing UG, Point 8 GmbH, QuinScape GmbH, RapidMiner GmbH, die Roberta-Initiative des Fraunhofer IAIS, sourcewerk GmbH und UNICEPTA Gesellschaft für Medienanalyse mbH das Ökosystem des Ruhrgebietes demonstrierten.



Oliver Schaper / TU Dortmund



Oliver Schaper / TU Dortmund

AUSSICHTEN

Das ML2R wird im Laufe des Jahres 2020 einen **virtuellen ML Showroom** einrichten, der Demonstratoren und Anwendungen des Maschinellen Lernens in einer virtuellen Umgebung erlebbar machen wird. Die Online-Plattform wird neben technologischen Demonstratoren auch den Zugang zu Onlinekursen und Lehrmaterial, kuratierten Daten und Algorithmen für die praktische Anwendung, sowie aktuellen Informationen zum Transferpotenzial von Ergebnissen der ML-Forschung in die Wirtschaft umfassen. Das ML2R wird so einem interessierten Publikum aus Wissenschaft, Wirtschaft und Gesellschaft einen offenen Zugang zu Wissen und Daten ermöglichen. Ziel ist es, gesellschaftliche und wirtschaftliche Vorbehalte gegenüber den Technologien Künstlicher Intelligenz abzubauen und kollaborierende Wissenschaftsbestrebungen zu stärken.

Im Zeichen des verstärkten Transfers von Ergebnissen der ML-Forschung in die Wirtschaft steht auch der vom ML2R konzipierte **Blog**, der aktuell in Vorbereitung ist. Im Blog werden die Expert*innen des ML2R über ihre Forschung berichten und so den Austausch über aktuelle Forschungsergebnisse, -themen und -anwendungsfragen, insbesondere auch mit Fachleuten aus Unternehmen, stärken.

Unter Federführung der Geschäftsstelle für Koordinierung am ML2R wird zudem ein **Newsletter** des bundesweiten Netzwerks **aller KI-Kompetenzzentren** aufgesetzt. Gegenstand des Newsletters, der sich an ein interessantes Publikum richtet, sind aktuelle Publikationen sowie Informationen aus den Kompetenzzentren.

Das ML2R informiert



Kontaktinformationen



Prof. Dr. Katharina Morik

Tel. 0231/755-5101

Fax 0231/755-5105

katharina.morik@tu-dortmund.de



Prof. Dr. Stefan Wrobel

Tel. 02241/14-3015 (Sokr. -3014)

Fax 02241/14-2436

stefan.wrobel@iais.fraunhofer.de

Zentrumsinformationen

Kompetenzzentrum für Maschinelles Lernen Rhein-Ruhr
Technische Universität Dortmund, Fakultät für Informatik
Otto-Hahn-Str. 12
44227 Dortmund

Kompetenzzentrum für Maschinelles Lernen Rhein-Ruhr
Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS
Schloss Birlinghoven
53757 Sankt Augustin



ML2R
Kompetenzzentrum
Maschinelles Lernen
Rhein-Ruhr