Künstliche Intelligenz – Die dritte Welle

Ute Schmid[1] Voker Tresp[2] Matthias Bethge[3] Kristian Kersting[4] Rainer Stiefelhagen[5]

Abstract: Aktuelle Forschungsarbeiten aus dem Bereich Künstlichen Intelligenz werden vorgestellt. Dabei werden drei Perspektiven auf das Gebiet Maschinelles Lernen präsentiert, die über rein datenintensive Blackbox-Verfahren hinausgehen: Es werden Methoden vorgestellt, mit denen Erklärungen für die Entscheidungen von KI-Systemen generiert werden, aktuelle neurowissenschaftlich Ansätze zum maschinellen Sehen gezeigt und eine Möglichkeit Vorwissen in den Prozess des machinellen Lernens einzubringen aufgezeigt.

Keywords: Erklärbare KI; Wissensgraphen; Computer Sehen

1 Einführung

Das Thema Künstliche Intelligenz erfährt seit einigen Jahren sehr viel Aufmerksamkeit. Insbesondere besteht großes Interesse, KI-Technologie für viele Anwendungsbereiche – von Industrie 4.0 über Medizin bis hin zu Bildung – zu erschließen. Zunehmend zeigt sich, dass hier datenintensive Black-Box-Ansätze des maschinellen Lernens alleine nicht geeignet sind. Für einen robsten und transparenten Ensatz zu ermöglichen, müssen Klassifikationsentscheidungen adaptiv und kontextsensitiv sein, sich an menschliche Anforderungen anpassen und sollten vorhandenes breichsspezifisches Wissen berücksichtigen können. Drei Perspektiven auf entsprechende Weiterentwicklungen von Methoden des maschinellen Lernens werden im Folgenden präsentiert.

Die kurzen Texte beruhen auf teilweise überarbeiteten Kurzfassungen der drei eingeladenen Vorträge von Ute Schmid, Matthias Bethge und Volker Tresp im Rahmen einer von Kristian Kersting und Rainer Stiefelhagen organisierten Session zum Thema Künstliche Intelligenz.

[1] Universität Bamberg, Kognitive Systeme, ute.schmid@uni-bamberg.de
[2] Ludwig Maximilian Universität München & Distinguished Research Scientist at Siemens AG, Corporate Technology, volker.tresp@siemens.com
[3] Universität Tübingen & Amazon Scholar, Computational Neuroscience & Machine Learning, matthias.bethge@bethgelab.org
[4] Technische Universität Darmstadt, Künstliche Intelligenz und Maschinelles Lernen, kersting@cs.tu-darmstadt.de
[5] Karlsruhe Institute of Technology, Computer Vision for Human-Computer Interaction, rainer.stiefelhagen@kit.edu

doi:10.18420/inf2020_08

2 Die Dritte Welle der KI – Vom rein datengetriebenem Blackbox Lernen zu interaktiven und erklärbaren Ansätzen

Maschinelles Lernen wird als eine der wichtigsten Zukunftstechnologien für viele Bereiche der Wirtschaft und der Gesellschaft angesehen. Insbesondere Erfolge von Ansätzen der tiefen neuronalen Netzen auf Bilddaten versprechen, dass Modelle für komplexe Entscheidungsszenarien direkt aus Rohdaten gelernt werden können. Zunehmend zeigt sich allerdings, dass rein datengetriebene Ansätze in vielen Bereichen nicht umsetzbar sind: Zum einen können die hohen Anforderungen an die Menge und die Qualität an Daten, die hier für benötigt werden, häufig nicht oder nur mit sehr hohem Aufwand generiert werden. Zum anderen sind Entscheidungen von Blackbox-Modellen in vielen Bereichen rechtlich und ethisch unzulässig. Entsprechend wird aktuell die sogenannte *3rd Wave of AI* ausgerufen, nach der nun Erklärbarkeit und Parterschaftlichkeit die Ansätze des rein datengetreibenen maschinellen Lernens ablösen [TK19]. Maschinelles Lernen bietet eine Fülle verschiedener Ansätze. Je nach Problembereich können häufig auch direkt interpretierbare Ansätze eingesetzt werden [Mu18; Ru19].

Erklärungen können in verschiedenen Modalitäten erfolgen – insbesondere visuell und verbal. Für komplexe Entscheidungen sind visuelle Erklärungen häufig nicht ausdrucksstark genug. Beispielsweise kann ein Hervorheben desjenigen Bereiches im Bild eines Gewebeschnitts, auf dem die Entscheidung eines gelernten Modells maßgeblich beruht, häufig nur als Plausibilitätscheck dienen – etwa, dass tatsächlich der Bereich, in dem sich Tumorgewebe befindet, beachtet wird. Verbale Erklärungen können dagegen auf relevante Merkmalsausprägungen (‘Das Tumorgewebe hat einen Durchmesser größer 2 mm’), auf

Abb. 1: Beispielhafte Umsetzung eines Ansatzes zum erklärenden interaktiven maschinellen Lernen im Bereich Medizin.

quantifierte Aussagen (‘Alle identifizierten Metastasen sind kleiner als 1 mm’) und auf Relationen (‘Das Tumorgewebe berührt das Fettgewebe’) verweisen [SF20]. In einer beispielhaften Umsetzung für die digitale Gewebepathologie (siehe Abb.1) wird demonstriert, wie verbale Erklärungen erzeugt und von Domänenexperten korrigiert werden können. Durch interaktives Lernen wird es dadurch möglich, gelernte Modelle durch Einbringen von Expertenwissen inkrementell zu verbessern.

3 Maschinelles Lernen mit Wissensgraphen

Maschinelles Lernen mit Wissensgraphen findet zunehmend Interesse, sowohl im akademischen als auch im industriellen Umfeld [Ni15]. Die Knoten in Wissensgraphen sind Konzepte (Entitäten, Klassen, Attribute,. . .), die anhand ihrer semantischen Eigenschaften und ihrer Beziehungen zueinander beschrieben werden. Wissensgraphen können über Ihre Adjazenztensoren dargestellt werden, aus denen Tensormodelle abgeleitet werden können, die es dann erlauben, neue Fakten abzuleiten [NTK11]. In unserem Vortrag haben wir aufgezeigt, wie maschinelles Lernen mit Wissensgraphen in industriellen Anwendungen [Hi18] (Abbildung 2) und zur Unterstützung klinischer Entscheidungen. verwendet werden kann. Wichtige Probleme, mit denen wir uns im klinischen Umfeld befassen, sind fehlende Daten, Erklärbarkeit und Bewertung von Ansätzen zur klinischen Entscheidungsunterstützung [Wu20] . Weitere Schwerpunkte unserer Arbeit sind Szenengraphen in der Bilderkennung [BMT17] und die Untersuchung von Bezügen zu Wahrnehmung und Gedächtnis [Tr15].

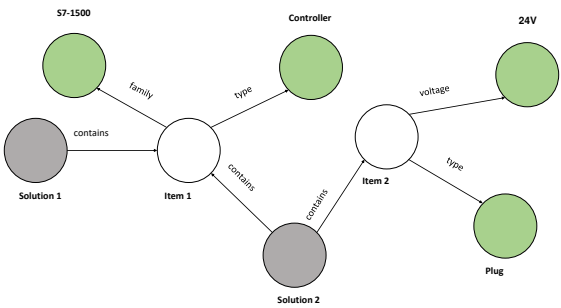

Abb. 2: Ein Wissensgraph beschreibt industrielle Komponenten, deren Eigenschaften und Bezüge zu Kundenlösungen.

4 Das sehen Maschinen aber anders

Maschinen werden immer besser darin, Wahrnehmungsaufgaben wie Objekt- oder Spracherkennung zu lösen. Die Eigenschaften, die von Maschinen verwendet werden, um zum Beispiel eine Katze von einem Hund zu unterscheiden, sind jedoch ganz anders als die Merkmale, die Menschen verwenden [Ge20]. Entsprechend vorsichtig sollte man sein, wenn man Maschinen kognitive Fähigkeiten, wie Objekterkennung oder Verstehen von Szenen zuschreibt. Psychologische Erkenntnisse und Untersuchungen können dazu beitragen, die Unterschiede zwischen menschlichem und maschinellem Sehen zu charakterisieren und zu verringern.

5 Abschließende Bewertung

Alle drei Beiträge haben den Fokus auf maschinellem Lernen als dem aktuell am meisten beachteten Bereich der Künstlichen Intelligenz Forschung. Jedoch zeigen die Beiträge deutlich auf, wo die aktuellen datenintensiven Ansätze ihre Grenzen haben. Die dritte Welle der KI Forschung erweitert den Blick über statistisches maschinelles Lernen hinaus hin zu hybriden Ansätzen, in denen klassische KI-Methoden aufgegriffen und weiterentwickelt werden. Zudem besteht neues Interesse an interdisziplinären Fragestellungen, insbesondere um Mensch-KI Interaktion partnerschaftlich zu gestalten.

Literatur

[BMT17] Baier, S.; Ma, Y.; Tresp, V.: Improving visual relationship detection using semantic modeling. In: ISWC. Springer, 2017.

[Ge20] Geirhos, R.; Jacobsen, J.-H.; Michaelis, C.; Zemel, R.; Brendel, W.; Bethge, M.; Wichmann, F. A.: Unintended cue learning: Lessons for deep learning from experimental psychology. Journal of Vision 20/11, S. 652–652, 2020.

[Hi18] Hildebrandt, M.; Sunder, S. S.; Mogoreanu, S.; Thon, I.; Tresp, V.; Runkler, T.: Configuration of industrial automation solutions using multi-relational recommender systems. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, S. 271–287, 2018.

[Mu18] Muggleton, S. H.; Schmid, U.; Zeller, C.; Tamaddoni-Nezhad, A.; Besold, T.: Ultra-Strong Machine Learning: comprehensibility of programs learned with ILP. Machine Learning 107/7, S. 1119–1140, 2018.

[Ni15] Nickel, M.; Murphy, K.; Tresp, V.; Gabrilovich, E.: A Review of Relational Machine Learning for Knowledge Graphs. Proceedings of the IEEE/, 2015.

[NTK11] Nickel, M.; Tresp, V.; Kriegel, H.-P.: A three-way model for collective learning on multi-relational data. In: Proceedings of the 28th International Conference on International Conference on Machine Learning. S. 809–816, 2011.

[Ru19] Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1/5, S. 206–215, 2019.

[SF20] Schmid, U.; Finzel, B.: Mutual Explanations for Cooperative Decision Making in Medicine. KI-Künstliche Intelligenz 34/2, S. 227–233, 2020.

[TK19] Teso, S.; Kersting, K.: Explanatory interactive machine learning. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. S. 239–245, 2019.

[Tr15] Tresp, V.; Esteban, C.; Yang, Y.; Baier, S.; Krompaß, D.: Learning with memory embeddings. NIPS 2015 Workshop on Nonparametric Methods for Large Scale Representation Learning/, 2015.

[Wu20] Wu, Z.; Yang, Y.; Ma, Y.; Liu, Y.; Zhao, R.; Moor, M.; Tresp, V.: Learning Individualized Treatment Rules with Estimated Translated Inverse Propensity Score. arXiv preprint arXiv:2007.01083/, 2020.