Fraunhofer

FRAUNHOFER-GESELLSCHAFT

MASCHINELLES LERNEN – KOMPETENZEN, ANWENDUNGEN UND FORSCHUNGSBEDARF

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS17019 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren.

Autorinnen und Autoren:

Inga Döbel | Fraunhofer IMW

Dr. Miriam Leis | Fraunhofer-Zentrale

Manuel Molina Vogelsang | Fraunhofer IMW

Dmitry Neustroev | Fraunhofer IMW

Dr. Henning Petzka | Fraunhofer IAIS

Dr. Stefan Rüping | Fraunhofer IAIS

Dr. Angelika Voss | Fraunhofer IAIS

Martin Wegele | Fraunhofer-Zentrale

Dr. Juliane Welz | Fraunhofer IMW

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

DLR Projektträger

Förderkennzeichen: 01IS17019

Datum: 29.03.2018

Inhaltsverzeichnis:

Executive Summary **4**
1 Maschinelles Lernen – Einordnung, Konzepte, Methoden und Grenzen **9**
1.0 Einführung: Aktuelle Perzeption des Maschinellen Lernens und der Künstlichen Intelligenz 9
1.1 Warum ist Maschinelles Lernen sinnvoll? 12
1.2 Geschichtlicher Hintergrund von ML und KI 14
1.3 Fachliche Einordnung des Maschinellen Lernens 16
1.4 Wichtige Lernstile des Maschinellen Lernens: Überwachtes, unüberwachtes und bestärkendes Lernen 25
1.5 Modelltypen und Algorithmen des Maschinellen Lernens 29
1.6 Die Renaissance Künstlicher Neuronaler Netze 36
1.7 Herausforderungen an die Güte und Qualität beim ML 47

2 Anforderungen an die ML-Forschung und aktuelle Forschungsthemen **55**
2.0 Einführung 55
2.1 Datenlage 57
2.2 Fähigkeiten 64
2.3 Akzeptanz, Sicherheit und Verlässlichkeit 80
Tabellarische Zusammenfassung der offenen Forschungsfragen 87
Anhang A: Glossar: ML-Fachbegriffe 89
Index: Kapitel 1 und 2 93

3 Kompetenzlandkarte Maschinelles Lernen: Publikationen, Forschungsförderung, Patente und Akteure **97**
3.1 Methodik und Forschungsdesign 97
3.2 Bibliometische Analyse der ML-Publikationen 103
3.3 Maschinelles Lernen in Deutschland – Publikationen 113
3.4 Analyse der ML-Projektförderung auf europäischer Ebene 121
3.5 ML-Patentanalyse 125
3.6 Maschinelles Lernen: Produkte, Märkte und Wirtschaftsakteure 130
Anhang B: Schlagworte für die Suchanfragen 150

4 Rahmenbedingungen für Maschinelles Lernen **155**
4.1 Aus- und Weiterbildung 155
4.2 Transfer in die Praxis 157
4.3 Datenverfügbarkeit und Governance 158
4.4 Rechtliche, ethische und soziale Gestaltung 159

5 Expertenkonsultation **164**
5.1 Einleitung 164
5.2 Maschinelles Lernen – Methoden und Lerndaten 165
5.3 Forschungsthemen und Forschungsbedarf 168
5.4 Empfehlungen für die Forschungspolitik 171
5.5 Kompetenzlandschaft in Deutschland 172

5.6 Sozioökonomische, rechtliche und politische Rahmenbedingungen 177
5.7 Ausblick 180

6 Fazit 181

Danksagungen 182

Anlagen: Verzeichnisse 184
Literaturverzeichnisse nach Kapiteln 184
Weiterführende Literatur zum Thema Maschinelles Lernen 198
Abbildungsverzeichnis 199
Tabellenverzeichnis 202

Executive Summary

Kaum ein anderes Forschungsfeld hat in letzter Zeit so viel Aufsehen erregt wie das Maschinelle Lernen (ML) mit den damit einhergehenden rasanten Fortschritten auf dem Gebiet der Künstlichen Intelligenz (KI).

Diese Publikation gibt eine kompakte Einführung in die wichtigsten Konzepte und Methoden des Maschinellen Lernens, einen Überblick über Herausforderungen und neue Forschungsfragen sowie eine Übersicht zu Akteuren, Anwendungsfeldern und sozioökonomischen Rahmenbedingungen der Forschung mit Fokus auf den Standort Deutschland. Die Basis hierfür ist das vom BMBF geförderte wissenschaftliche Projekt »Maschinelles Lernen – Kompetenzen, Anwendungen und Forschungsbedarf«, das vom Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS, dem Fraunhofer-Zentrum für Internationales Management und Wissensökonomie IMW sowie der Zentrale der Fraunhofer-Gesellschaft durchgeführt wurde. Neben der statistischen Auswertung von wissenschaftlichen Publikationen, Patenten und Projekten wurden Marktstudien und viele wissenschaftliche Publikationen analysiert sowie Interviews mit 18 Fachleuten für ML durchgeführt, ebenso wie ein Workshop mit 16 Fachleuten aus Wissenschaft, Wirtschaft und Politik.

In Expertenkreisen wird ML als Schlüsseltechnologie für moderne KI-Techniken gesehen, weshalb insbesondere im ökonomischen Kontext KI und ML oft synonym verwendet werden. Maschinelles Lernen und insbesondere das sogenannte Deep Learning (DL) eröffnen völlig neue Möglichkeiten in der automatischen Sprachverarbeitung, Bildanalyse, medizinischen Diagnostik, Prozesssteuerung und dem Kundenmanagement. Wirtschaftsmedien sprachen 2017 vom »Jahr der KI« und die Anwendungsmöglichkeiten werden mit der Fortführung der Digitalisierung weiter steigen.

Die wissenschaftliche ML-Forschung ist längst nicht ausgeschöpft und insbesondere Forschungsfragen zu ML mit extrem großen oder sehr kleinen Datenmengen, zur Kombination von ML mit physikalischem oder Expertenwissen, sowie Sicherheit und Transparenz von ML-Modellen sind hochaktuell und hochrelevant.

Statistiken zu Publikationen in wissenschaftlichen Fachzeitschriften zeigen, dass 60% aller Publikationen zu ML aus China, den USA, der EU und Indien kommen. China weist hier ein besonders hohes quantitatives Wachstum von jährlich 17,5% im Zeitraum 2006 bis 2016 auf, wobei sechs der zehn meistpublizierenden Hochschulen und Forschungseinrichtungen aus China sind. 51% der erfassten Publikationen können dem Anwendungsfeld der Bild- und Videoauswertung zugeordnet werden, gefolgt von 22% zur Sprachverarbeitung.

In Europa entfallen die meisten Publikationen auf Großbritannien, gefolgt von Deutschland. Innerhalb von Deutschland gibt es jedoch regionale Unterschiede. Die Bundesländer

mit sowohl der höchsten Publikations- als auch Patentdichte sind Baden-Württemberg, Bayern und Nordrhein-Westfalen. Bei den Publikationen zum Deep Learning kann ab 2013 weltweit ein merkbarer Anstieg verzeichnet werden. Davor war der Anteil vernachlässigbar gering, und auch 2016 ist er mit 2,6% in Fachzeitschriften und 6,8% in Konferenzbeiträgen geringer als erwartet.

In unserer Patentrecherche als Indikator für die technologische Leistungsfähigkeit von Regionen und Einrichtungen entfallen die Hauptaktivitäten auf die USA, China und Südkorea. 73% aller im Zeitraum 2006 bis 2016 erfassten Patente stammen aus diesen Ländern, mit den Unternehmen Microsoft, Google, Amazon, Facebook, Samsung (Südkorea) und Huawei (China) an der Spitze. In Deutschland sind die patentstärksten Akteure Siemens AG, Robert Bosch GmbH, Deutsche Telekom AG, Daimler AG, BMW AG und SAP SE. Deutsche mittelständische Unternehmen mit 49 bis 249 Mitarbeitenden weisen vergleichsweise wenige Patentaktivitäten auf. Im Hinblick auf die Standorte von KI-Start-ups ist Berlin, wo über 50 Unternehmen ihren Sitz haben, nach London die zweitgrößte Region in Europa.

Um den Standort Deutschland international zu stärken, haben die konsultierten Fachleute ausdrücklich auf die Aus- und Weiterbildungssituation hingewiesen. Ihnen zufolge muss in Deutschland noch viel stärker ML-bezogen aus- und weitergebildet werden, nicht nur in der Informatik, sondern auch in den Anwendungsdisziplinen. Gleichzeitig sollten Aus- und Weiterbildungsangebote stärker interdisziplinär orientiert sein, um KI-basierte Kompetenzen in der beruflichen Breite aufzubauen. Zusätzlich müssen entsprechende Arbeitskräfte global angeworben werden, was angesichts des weltweiten Wettbewerbs um Talente sowie der zu erwartenden steigenden Nachfragen nach ML-basierten Produkten und Dienstleistungen eine Herausforderung darstellt.

Deutschland verfügt über eine gute wissenschaftliche Basis in ML. Für die Sicherung der Wettbewerbsfähigkeit ist ausschlaggebend, den Anwendungsbezug in der Forschung zu stärken und dies beispielsweise auch in öffentlichen Forschungsausschreibungen stärker einzufordern. Ferner wurde konstatiert, dass derzeitige Maßnahmen zur Unternehmensförderung eher junge Start-ups anstatt etablierte Traditionsunternehmen ansprechen. Im Hinblick auf die Förderung der Anwendung von ML in Deutschland liegt gerade hier viel Potenzial, insofern sollten KMU durch mehr Fachinformationen zum Einsatz und dem Nutzen von ML unterstützt werden.

Der Zugang zu hinreichend großen und qualitativ hochwertigen Datenbeständen wurde für den Erfolg und die Wettbewerbsfähigkeit Deutschlands als bislang ungelöste Herausforderung gesehen, insbesondere in der Medizin und der industriellen Produktion. Hier sind Governance-Strukturen, die den kontrollierbaren und sicheren Datenaustausch ermöglichen, sowie datenschutzrechtliche Bedingungen zu berücksichtigen oder anzupassen.

In Zukunft werden Maschinen zusehends entscheidungsrelevante Ergebnisse generieren. Hierzu ist es wichtig, auf der technologischen Seite die Sicherheit, Robustheit und hinreichende Nachvollziehbarkeit von automatisierten Entscheidungsprozessen zu gewährleisten. Gleichzeitig muss dafür gesorgt werden, dass ML-Anwendungen mit juristischen Fragen wie Haftung und Verantwortlichkeit für algorithmisch getroffene Entscheidungen vereinbar sind, was zudem auch technisch umsetzbar sein muss. Dies auszuformulieren und regulativ umzusetzen ist ein wichtiges und komplexes Anliegen, das einen inter- und transdisziplinären Einsatz erfordert. Für die weitere Verbreitung maschineller Lernverfahren in die Anwendung ist nicht zuletzt auch die gesellschaftliche Akzeptanz von zentraler Bedeutung. Hierfür ist eine breite öffentliche Diskussion und Einbindung verschiedener gesellschaftlicher Gruppen erforderlich.

KAPITEL 1

Maschinelles Lernen – Einordnung, Konzepte, Methoden und Grenzen

Dr. Miriam Leis | Fraunhofer Zentrale
Dr. Henning Petzka | Fraunhofer IAIS
Dr. Stefan Rüping | Fraunhofer IAIS
Dr. Angelika Voss | Fraunhofer IAIS

Inhaltsverzeichnis Kapitel 1

1 Maschinelles Lernen – Einordnung, Konzepte, Methoden und Grenzen 9

1.0 Einführung: Aktuelle Perzeption des Maschinellen Lernens und der Künstlichen Intelligenz 9

1.1 Warum ist Maschinelles Lernen sinnvoll? 12

1.2 Geschichtlicher Hintergrund von ML und KI 14

1.3 Fachliche Einordnung des Maschinellen Lernens 16
- 1.3.1 Stochastik und Bayessches Verfahren 16
- 1.3.2 Der Analogismus 18
- 1.3.3 Der Konnektionismus 19
- 1.3.4 Der Symbolismus 21
- 1.3.5 Der Bezug zwischen Big Data und Maschinellem Lernen 23

1.4 Wichtige Lernstile des Maschinellen Lernens: Überwachtes, unüberwachtes und bestärkendes Lernen 25
- 1.4.1 Überwachtes Lernen 25
- 1.4.2 Unüberwachtes Lernen 26
- 1.4.3 Semi-überwachtes Lernen 28
- 1.4.4 Bestärkendes Lernen und sequentielles Entscheiden 28

1.5 Modelltypen und Algorithmen des Maschinellen Lernens 29
- 1.5.1 Regressionsmodelle 29
- 1.5.2 Entscheidungsbäume 30
- 1.5.3 Cluster 32
- 1.5.4 Kernmethoden 33
- 1.5.5 Künstliche Neuronale Netze 34
- 1.5.6 Bayessche Modelle 35
- 1.5.7 Sequentielle Entscheidungsmodelle 35

1.6 Die Renaissance Künstlicher Neuronaler Netze 36
- 1.6.1 Funktionsweise von tiefen KNN 37
- 1.6.2 Lernen von Datenrepräsentationen in tiefen KNN 39
- 1.6.3 Neue Aufgaben für tiefe KNN 42
- 1.6.4 Typen von tiefen Neuronalen Netzen 43

1.7 Herausforderungen an die Güte und Qualität beim ML 47
- 1.7.1 Qualität der Daten 47
- 1.7.2 Overfit, Underfit und Generalisierbarkeit 48
- 1.7.3 Performanz und Kostenfunktion 49
- 1.7.4 Robustheit 51

1 Maschinelles Lernen – Einordnung, Konzepte, Methoden und Grenzen

1.0 Einführung: Aktuelle Perzeption des Maschinellen Lernens und der Künstlichen Intelligenz

In Expertenkreisen wird Maschinelles Lernen als Schlüsseltechnologie der Künstlichen Intelligenz (KI) verstanden. Künstliche Intelligenz (KI) ist ein Teilgebiet der Informatik mit dem Ziel, Maschinen zu befähigen, Aufgaben »intelligent« auszuführen. Dabei ist weder festgelegt, was »intelligent« bedeutet, noch welche zum Einsatz kommen. Nachdem die KI in der Vergangenheit mit anderen Techniken einige Rückschläge erlitten hatte, hat ihr das Maschinelle Lernen jetzt zu einer regelrechten Renaissance verholfen. Deshalb werden »Maschinelles Lernen« und »Künstliche Intelligenz« insbesondere im wirtschaftlichen Kontext oftmals vereinfacht synonym verwendet. Wirtschaftsmedien sprachen von »2017 als dem Jahr der KI«.[1,2] Inzwischen ist ein enormer globaler Wettbewerb rund um das Zukunftsfeld der ML-Technologien und KI-Anwendungen entfacht, der insbesondere zwischen den USA und China ausgetragen wird.

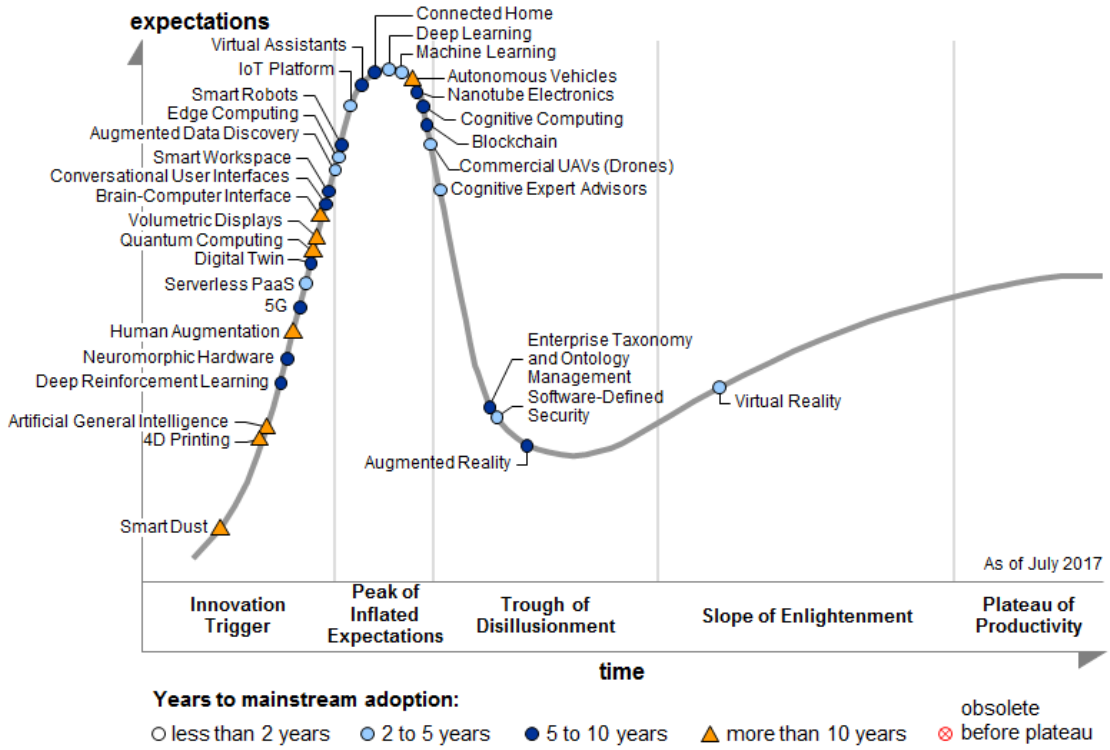
Maschinelles Lernen (ML) bezweckt die Generierung von »Wissen« aus »Erfahrung«, indem Lernalgorithmen aus Beispielen ein komplexes Modell entwickeln. Das Modell, und damit die automatisch erworbene Wissensrepräsentation, kann anschließend auf neue, potenziell unbekannte Daten derselben Art angewendet werden. Immer wenn Prozesse zu kompliziert sind, um sie analytisch zu beschreiben, aber genügend viele Beispieldaten – etwa Sensordaten, Bilder oder Texte – verfügbar sind, bietet sich Maschinelles Lernen an. Mit den gelernten Modellen können Vorhersagen getroffen oder Empfehlungen und Entscheidungen generiert werden ganz ohne im Vorhinein festgelegte Regeln oder Berechnungsvorschriften.

ML-Anwendungen oder »lernende Maschinen« sind nicht nur auf physische Geräte und Roboter beschränkt, sondern können auch rein digitale Anwendungen in IT-Systemen sein, wie verschiedene Arten von »Robos« und Bots, zum Beispiel Chatbots, Social Bots, Gamebots oder Robo-Player, Robo-Advisors oder Robo-Journalisten. ML-Techniken und KI-Anwendungen sind dabei, sämtliche Branchen und Lebensbereiche nachhaltig zu beeinflussen.

[1] Fortune 2016
[2] Forbes 2017

2016 stand »Maschinelles Lernen« bereits auf dem Scheitelpunkt des »Gartner Hype Cycle«.[3] 2017 kam dort das »Tiefe Lernen« (Englisch: Deep Learning) in mehrschichtigen Künstlichen Neuronalen Netzen hinzu, das beeindruckende Erfolge insbesondere im Bild- und Sprachverstehen zeigt. Themen auf der Spitze der Hype-Kurve haben primär medienbedingt inflationäre Erwartungen hervorgerufen und werden in der nächsten Zeit in eine wissenschaftlich fundierte Ernüchterungsphase eintreten. Das bedeutet jedoch nicht zwangsläufig, wie gelegentlich angenommen, dass das Thema an Bedeutung verlieren wird. Im Gegenteil: in der Phase nach den inflationären Erwartungen stellt sich heraus, welche Ideen und Anwendungen wirklich realisierbar sind und welchen wirklichen Nutzen sie für die Gesellschaft, Wirtschaft und Technologielandschaft haben werden.

Abbildung 1: Gartner Hype Cycle für Emerging Technologies 2017[4]



ML hat insbesondere in den letzten Jahren eine Reihe technischer Anwendungen ermöglicht, die viele Menschen als »intelligent« bezeichnen. Beispiele sind Maschinen, die sinnvoll auf natürliche Sprache reagieren, Gesichter und Objekte erkennen, passgenaue Vorschläge (z.B. zu Musikstücken oder Waren) anbieten oder automatisch Strukturen in un-

[3] Gartner 2016
[4] Gartner 2017

übersichtlichen Datensätzen ausfindig machen. Solche Anwendungen werden auch als »kognitive Maschinen«, »kognitive Systeme«, »Cognitive Computing« bezeichnet.

ML-basierte Maschinen werden zukünftig in vielen Bereichen zusehends Entscheidungen selbstständig treffen können. Das ruft aus unterschiedlichen Perspektiven neue rechtliche Fragestellungen auf den Plan, beispielsweise zur Haftung bei Schäden und Mängeln, zur Verantwortung von Inhalten und Urheberrechtsfragen, zur Transparenz von Entscheidungen, zum Daten- und Verbraucherschutz oder zur Frage, inwieweit den Entscheidungen von solchen Maschinen Folge zu leisten ist. Die zentrale ethische Herausforderung ist es, die Maschinen so zu gestalten, dass sie mit unseren Gesellschafts-, Rechts- und Wertevorstellungen kompatibel sind. Diese gesellschaftliche Debatte muss jetzt beginnen.[5]

Das Verständnis von Künstlicher Intelligenz hat sich über die Jahrzehnte hinweg gewandelt. Dies wird auch in den medial hervorgehobenen Höhepunkten der KI-Entwicklung reflektiert: 1996 wurde der IBM-Schachcomputer »Deep Blue« gefeiert, als er den Schachweltmeister Kasparow im Schachspiel durch umfängliche Suche und viel Rechenleistung besiegte. 2011 erlangte der IBM-Computer »Watson« einen symbolträchtigen Sieg gegen menschliche Mitstreiter im Quiz-Spiel »Jeopardy«, wo insbesondere die maschinellen Fähigkeiten zur natürlichen Sprachverarbeitung im Vordergrund standen. 2017, zwanzig Jahre nach dem Sieg von »Deep Blue« im Schachspiel, besiegte der Computer »AlphaGo« der Google-Tochterfirma Deep Mind, Ke Jie, den weltrangbesten Spieler im viel komplexeren und schwer vorhersagbaren Go-Spiel. Das ging nur, indem der Computer von sich aus gelernt hat, Situationen und Züge zu bewerten und die vielversprechendsten im Spiel gegen sich selbst auszutesten.

Die oftmals erstaunlich scheinenden Leistungen ML-basierter KI-Systeme implizieren jedoch nicht, dass die Maschine irgendein Verständnis oder gar »Bewusstsein« davon hat, was die Daten bedeuten, die sie verarbeitet, warum und in welchem Kontext sie das tut und was die Daten bedeuten, auch wenn es für einen Menschen unter Umständen so aussehen könnte, als ob die Maschine wirklich »denken« würde.

Ebenso häufen sich in populären Medien und fiktiven Filmen fehlgeleitete Vorstellungen von »superintelligenter« und »menschenähnlicher« KI mit eigenen Intentionen, Motiven und Bewusstsein. Das führt dazu, dass in der Öffentlichkeit oftmals über Dinge diskutiert wird, die jenseits von langfristigen Realisierungsmöglichkeiten liegen.

Als Zukunftsvision der KI sehen einige Forscher, wie der »Deep Mind« Gründer Demis Hassabis, eine sogenannte »Artificial General Intelligence«[6,7], die nicht nur konkrete Aufgaben in einem eingeschränkten Gebiet löst, sondern allgemeine kognitive Leistungen mit

[5] Bitkom 2017
[6] Technology Review 2017a
[7] Technology Review 2017b

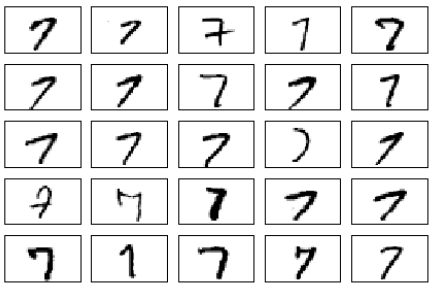
menschenähnlicher Flexibilität erbringen könnte. Zwar möchte man gerne ML-basierte Systeme ohne großen Aufwand auf ähnliche, aber dennoch andersartige Aufgaben effizient umtrainieren – hierdurch könnte beispielsweise ein Service-Roboter schneller neue Aktionen erlernen und in einem breiteren Spektrum eingesetzt werden – aber das hat noch lange nichts mit »genereller KI« und menschlicher Flexibilität zu tun.

Für realistische Zwecke sind Debatten und Bestrebungen zu einer künstlichen Universal- oder gar Superintelligenz innerhalb absehbarer Zeithorizonte wenig sinnvoll. Die existierenden KI-Systeme sind mit großem Aufwand konstruiert und auf enge Problemstellungen trainiert. Die Forschung beschäftigt sich vorrangig damit, den Trainingsaufwand zu reduzieren, Sicherheit, Robustheit und Transparenz zu verbessern, die Modelle an neue Aufgaben anzupassen und die Kompetenzen von Mensch und Maschinen zweckmäßig zu kombinieren.

1.1 Warum ist Maschinelles Lernen sinnvoll?

Es gibt viele Aufgaben, die für Computer zu variabel sind, als dass sie sinnvoll und effektiv durch konkrete, im Voraus festgesetzte Berechnungsvorschriften gelöst werden könnten. Immer wenn Prozesse zu kompliziert sind, um sie analytisch zu beschreiben und es viele Beispiele gibt, sind ML-Techniken die beste Wahl. Möchte man mit einem Computer handgeschriebene Ziffern erkennen, etwa um die handschriftlichen Adressen auf Briefen automatisch den entsprechenden Stadtteilen zuzuordnen, so ist es schwer, systematisch zu beschreiben, wie alle denkbaren handschriftlichen Varianten zum Beispiel einer »7« der Ziffer »Sieben« zugeordnet werden könnten.

Abbildung 2: Beispielhafte Variationen handschriftlicher »7«[8]



Das Maschinelle Lernen bietet hier eine effektive Alternative: Die Maschine generiert das gewünschte Ergebnis nicht durch im Vornhinein festgelegte Regeln (»so wird vorgegangen, um eine Ziffer Sieben zu erkennen«), sondern es wird ein Lernalgorithmus bereit gestellt, der aus vielen Beispielen Regelmäßigkeiten extrahiert.

[8] LeCun/Cortes/Burges 2015

Der Algorithmus erhält viele handschriftliche Varianten der verschiedenen Ziffern zusammen mit der korrekten Zuordnung und baut ein Modell mit differenzierenden Merkmalen auf. Das kann er auch auf neue, bis dahin unbekannte handschriftliche Zeichen anwenden, um sie der passendsten Ziffer zuzuordnen. So ist es möglich, einen praxistauglichen, automatischen Postleitzahlenleser zu entwickeln.

Ein Modell ist eine Abstraktion der Wirklichkeit. Beim Maschinellen Lernen erzeugt der Lernalgorithmus ein Modell, das Beispieldaten generalisiert, so dass es anschließend auch auf neue Daten angewendet werden kann.

Je mehr Beispiel- bzw. Trainingsdaten der Lernalgorithmus erhält, umso mehr kann er sein Modell verbessern und die Fehlerquote verringern. Insbesondere kann man auch noch im Betrieb der PLZ-Leser weiter falsch und richtig zugeordnete Ziffern sammeln, damit der Lernalgorithmus das Modell noch weiter verbessern kann. Die Maschine lernt so ständig weiter.

Da eine endliche Anzahl von Beispielen die Gesamtheit aller denkbar möglichen Varianten natürlich unvollständig beschreibt, ist jedes gelernte Modell zwangsläufig mit Unsicherheit behaftet. Deshalb wird oftmals eine Wahrscheinlichkeitseinschätzung dafür mitgeliefert, wie sicher die Maschine den Ausgabewert einschätzt. Die Ziffer in Abbildung 2 Abbildung 3, Reihe 1 und Spalte 5 könnte mit gewisser Wahrscheinlichkeit auch eine Zwei sein, dann wäre die präferierte Ausgabe eine »7«, aber als zweite Präferenz auch eine »2« möglich.

Abbildung 3: Automatische Ziffernerkennung mit ML[9]

Es ist zu beachten, dass die Qualität der maschinellen Ausgaben maßgeblich von den verwendeten Trainingsdaten abhängt, ebenso wie von dem Feedback, falls solches während des maschinellen Lernprozesses an die Maschine zurückgespielt wird. Werden der Maschine falsche Beispiele gegeben oder fehlen einschlägige Beispiele, lernt sie auch nicht das Richtige.

1.2 Geschichtlicher Hintergrund von ML und KI

Maschinelles Lernen kann auf eine recht lange Geschichte zurückblicken und ist aus Methoden der Statistik und KI hervorgegangen. Angeregt durch das Verständnis verteilter neuronaler Prozesse im Gehirn entstanden bereits in den späten 1940er Jahren erste Konzepte von Künstlichen Neuronalen Netzen (KNN) und fanden zehn Jahre später erste Implementierungen. Ende der 1960er haben zwei bekannte KI-Wissenschaftler, Minsky und Papert, gezeigt, dass ein einziges Neuron die elementare Entweder-oder-Logik nicht lernen kann und größere Neuronale Netze mit wenigen lokalen Vernetzungen in ihrer Ausdrucksfähigkeit beschränkt sind. Dies führte in den 1970er Jahren zur Stagnation der KI-Forschung, insbesondere an KNN, und dem zum sogenannten ersten »KI-Winter«.

In den 1980er Jahren konzentrierte sich die Forschung auf symbolische Expertensysteme. Ihre Wissensbasis bestand aus manuell eingegebenen logischen Regeln, die sich auf ma-

[9] Ryerson University 2017

nuell selektierte Merkmale oder ebenfalls manuell konstruierte Objekthierarchien bezogen. Solche Wissensrepräsentationen bezeichnet man als »symbolisches Wissen«. Es stellte sich aber heraus, dass ein konsistenter Ausbau größerer Wissensbasen immer schwerer wurde. Man erkannte, dass praktisch niemals alle denkbaren Vorbedingungen für eine Aktion explizit angegeben werden können. Zudem traten Probleme im Umgang mit neuen Informationen auf, die bereits eingegebenem Wissen widersprechen. Das führte Ende der 1980er zum zweiten »KI-Winter«.

Mitte der 1980er wurden Neuronale Netze zwar durch die Back-Propagation-Methode wieder interessant. Für praktische Anwendungen stellten sich ab 1995 aber andere Lernmethoden, insbesondere Stützvektormaschinen, als handhabbarer heraus.

Erst um die Jahrtausendwende ermöglichten Fortschritte in den Computertechnologien und das Aufkommen von »Big Data« das Lernen von sehr komplexen, sogenannten »tiefen« Künstlichen Neuronalen Netzen, auch als »Deep Learning« bekannt. Damit begann der Erfolg der heutigen KI.

Tabelle 1: Überblick zu ausgewählten Meilensteinen im Einsatz von Maschinellem Lernen

heute	ML-basierte Systeme sind inzwischen in der Lage, ▪ radiologische Bilder so gut wie Mediziner zu analysieren ▪ automatisch unklare Bilder zu vervollständigen ▪ selbst KI-Software zu schreiben und zu trainieren ▪ Börsengeschäfte anhand eigener Prognosen selbstständig durchzuführen ▪ in komplexen Spielen wie Go und Poker gegen Menschen zu gewinnen ▪ sich selbst Wissen, Spiele und Strategien beizubringen
2017	**KI** (Alpha Go) **gewinnt** im **Go-Spiel** gegen den weltrangersten Spieler Ke Jie
2011	**KI gewinnt** im **Quiz-Spiel** (IBM Watson)
2010er	Bedeutende **Erfolge** mit **Deep Learning** (v.a. in der Sprachverarbeitung, Objekterkennung, Mustererkennung, Bioinformatik)
2000er	**Popularitätsgewinn des ML:** **Revival der Neuronalen Netze** (Big Data und schnelle Computer); Verbreitung der **Kernel-Methoden** des ML
1996	**KI gewinnt im Schach gegen den Weltmeister Kasparow** (IBM Deep Blue)
1990er	Durchbruch: **Stützvektormaschinen (SVM)**
1985-1995	Stagnation der Forschung und Entwicklung: **Aufgabe der Expertensysteme**

	(zu hohe Komplexität und langsame Computer)
1980er	Praktische Anwendung der »Back Propagation«-Methode für ML und KNN; Forschung an Expertensystemen
1980er	Boom der **humanoiden Robotik** (Japan)
1974-1980	Stagnation der FuE: **Scheitern Neuronaler Netze** (zu langsame Computer)
1960er	Entwicklung: Bayessche Netze, probabilistisches ML und semantische Netze
1950er	Pionierarbeiten im **Maschinellen Lernen** (ML), Begriffsprägung der **Künstlichen Intelligenz** (KI)
1940er	Theorie der »**Künstlichen Neuronalen Netze**« (KNN)

1.3 Fachliche Einordnung des Maschinellen Lernens

Ganz allgemein verfolgen die schließende bzw. induktive Statistik und das Maschinelle Lernen ähnliche Ziele, nämlich aus neuen Daten möglichst treffende Vorhersagen zu generieren. Während in der Statistik und besonders im Zweig der Stochastik für jedes Datum der errechnete Wahrscheinlichkeitswert von zentralem Interesse ist, wird dieser bei den meisten ML-Methoden lediglich indirekt genutzt. Die Wahrscheinlichkeiten können das ausschlaggebende Kriterium zur Ergebnisgenerierung sein, der exakte Wahrscheinlichkeitswert ist hierfür jedoch meist weniger wichtig. Fast alle Modelle, die ML-Verfahren aus Beispielen erzeugen, sind letztendlich statistische Modelle. Die Statistik bildet deshalb ein Fundament für die Theorie des Maschinellen Lernens. Statistische Methoden motivieren Modelle des Maschinellen Lernens und Erkenntnisse aus der Statistik helfen, maschinelle Lernverfahren aus Sicht der Wahrscheinlichkeitstheorie zu verstehen.

Maschinelles Lernen und Künstliche Intelligenz sind von verschiedenen »Denkschulen« beeinflusst worden, die jeweils unterschiedliche Ideen in den Fokus stellen[10]. Den folgenden vier kommt die größte Bedeutung zu, da sie die Höhen und Tiefen des ML und der KI maßgeblich geprägt haben.

1.3.1 Stochastik und Bayessche Verfahren

Bei stochastischen Methoden interessiert man sich für die Bestimmung ganzer Wahrscheinlichkeitsverteilungen. Auf Basis der vorliegenden Daten soll nicht nur für jedes wei-

[10] Domingos 2016

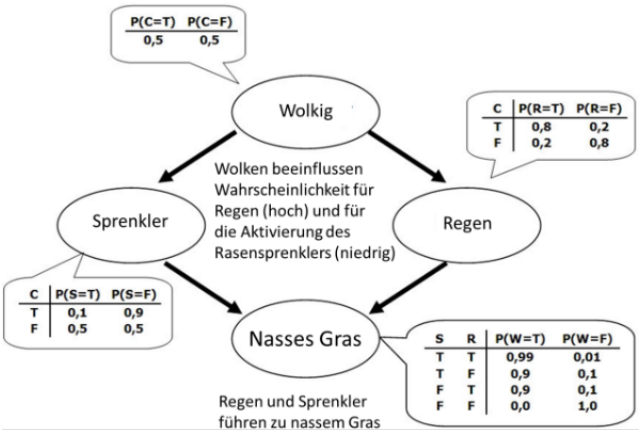
tere, neue Beispiel ein Merkmal vorhergesagt werden, sondern man versucht, zusätzlich eine Wahrscheinlichkeitsaussage zu treffen. Im Vordergrund der stochastischen Verfahren stehen die Modellierung und der kontrollierte Umgang mit Unsicherheiten auf Basis von inhaltlichen und problemspezifischen Überlegungen.

Bayes Theorem ist eine mathematische Formel für die Bestimmung einer bedingten Wahrscheinlichkeit, also der Bedingung, dass ein Ereignis auftritt, wenn ein anderes Ereignis eingetreten ist. Mit dem Theorem kann Vorwissen in die Bestimmung von Wahrscheinlichkeiten eingebracht werden.

Bei Bayesschen Verfahren, benannt nach dem Statistiker Thomas Bayes, geht es im Kontext von ML darum, mit Hilfe von Bayes Theorem das jeweils wahrscheinlichste Modell auf Basis der vorliegenden Datenlage zur Beschreibung und Vorhersage zu generieren. Im Fokus steht das Schlussfolgern über zukünftige Ereignisse unter Unsicherheit durch Einbeziehung von Vorannahmen. Wahrscheinlichkeitsangaben können kontinuierlich aktualisiert werden, sobald neue Informationen, zum Beispiel Belege, die für einen bestimmten Sachverhalt sprechen, eintreffen.

Ein Bayessches Netz ist ein Mechanismus zur automatischen Anwendung des Satz von Bayes, um die Wahrscheinlichkeit von abhängigen Variablen aus damit zusammenhängenden Beobachtungen zu berechnen. Aus Vorwissen, ob zum Beispiel eine Person raucht oder nicht, und ob das Wetter warm oder kalt ist, kann man so die wahrscheinlichste Diagnose, zum Beispiel Lungenerkrankung oder Erkältung, herleiten. Beobachtbare und davon abhängige Größen werden als Zufallsvariable aufgefasst und durch Knotenpunkte in einem Netz repräsentiert. Kausale Beziehungen oder andere Abhängigkeiten werden durch Pfeile ausgedrückt. Die Knoten haben Tabellen zur Berechnung der Wahrscheinlichkeiten ihrer Werte aus den Werten ihrer Vorgängerknoten. Diese bedingten Wahrscheinlichkeiten können von Parametern abhängen, die aus den Daten gelernt werden.

Abbildung 4: Bayessches Netz[11]



Wenn größere Datenmengen nur schwer zu beschaffen sind, aber gleichzeitig Vorwissen vorliegt, das man gerne berücksichtigen möchte, sind Bayessche Verfahren oft vorteilhaft. Außerdem sind Bayessche Verfahren wichtig für die Theorie des Maschinellen Lernens, da die Beschreibung der Unsicherheiten mit Hilfe der Wahrscheinlichkeitstheorie mathematisch fundierte Begründungen erlaubt, warum bestimmte Modelle gute Ergebnisse versprechen.

1.3.2 Der Analogismus

In analogistischen Verfahren steht die Annahme im Mittelpunkt, dass Objekte, die bezüglich bestimmter Merkmale große Ähnlichkeiten aufzeigen, folglich einer gemeinsamen Klasse angehören. Darauf aufbauend werden Schlussfolgerungen auf Grund von Ähnlichkeiten gemacht. Dafür muss ein geeignetes Maß bestimmt werden, das eine gewünschte Ähnlichkeit gut repräsentiert. Oft geschieht das durch die Bestimmung von Merkmalen oder einer Repräsentation der Beispiele in einem mehrdimensionalen Zahlenraum, in dem die Ähnlichkeit durch die Distanz der Punkte dargestellt wird. Ausgehend von der Repräsentation kann man versuchen, die Beispiele in sogenannte Cluster zu gruppieren.

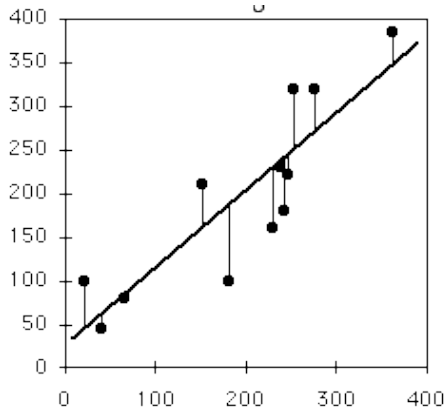
Bei den meisten ML-Verfahren geht es darum, jedem Beispiel als Label einen Wert zuweisen. Diese werden im Analogismus in Abhängigkeit von ähnlichen Beispielen gewählt, für die bereits Label vorliegen.

Bei einer Regression geht es darum, Merkmale in Abhängigkeit zu setzen. Zum Beispiel will man den Wert eines Hauses auf Basis von Merkmalen wie Größe, Wohngegend, Baujahr usw. schätzen. Nach analogistischem Denken führen ähnliche Merkmalswerte zu einem vergleichbaren Preis. Stehen einige Vergleichswerte zur Verfügung, so legt man

[11] Goodman/Tenenbaum 2016

dazwischen eine passende Gerade und erhält eine Schätzfunktion für den Preis, die man auch auf neue Beispiele anwenden kann.

Abbildung 5: Lineare Regression[12]



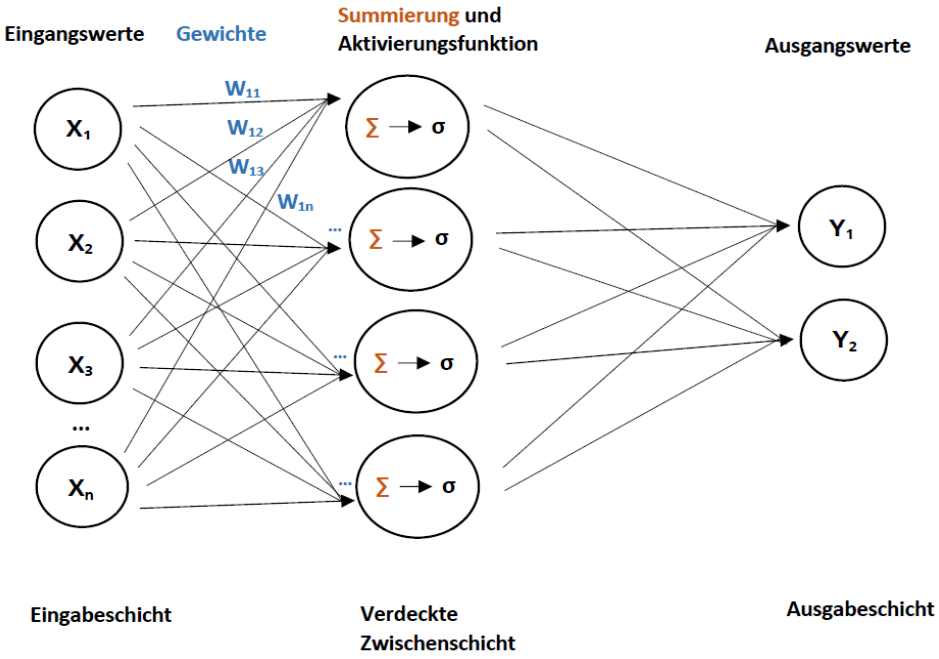
Auch hinter den Empfehlungssystemen von Verkaufsplattformen oder Videoportalen steckt der analogistische Gedanke: Kunden mit ähnlicher Produkthistorie haben ähnliche Interessen. So können Vorschläge gemacht werden, die in vielen Fällen auf Nutzer sinnvoll angepasst sind.

1.3.3 Der Konnektionismus

Der Konnektionismus sieht das Vorbild für ML in der Funktionsweise des Gehirns, wo Lernen durch Veränderung in der Stärke der Verbindungen zwischen einzelnen Nervenzellen oder Neuronen stattfindet. Dieser Prozess wird in vereinfachter Form für das Maschinelle Lernen nachgeahmt: In einer Datenstruktur werden »Knoten«, sogenannte künstliche Neuronen, schichtweise zu Künstlichen Neuronalen Netzen (KNN) verbunden, in denen Daten bzw. Signale weiter geleitet werden. An den Verbindungen liegen mathematische Gewichte, die mit den eintreffenden Signalstärken multipliziert und anschließend aufaddiert werden. In den Knoten bestimmt eine »Aktivierungsfunktion«, ob und in welcher Stärke das Signal weitergegeben wird. So wandern die Signale von der Eingabe- bis in die Ausgabeschicht. Da die Aktivierungsfunktionen nicht-linear gewählt werden können, kann ein KNN eine komplexe, nicht-lineare Funktion approximieren.

[12] McDonald 2017

Abbildung 6: KNN, eigene Darstellung



Künstliche Neuronale Netze verarbeiten Vektoren in einem mehrdimensionalen Zahlenraum, die von Schicht zu Schicht durch die Knoten transformiert werden, um auch komplexere, nichtlineare Funktionen zu approximieren.

Lernen in einem KNN bedeutet, aus dem Unterschied zwischen Ausgabewerten und richtigen Antworten rückwärts durch die Schichten Korrekturen für die Gewichte zu ermitteln. Als Signale verarbeiten Künstliche Neuronale Netze nur Ansammlungen von Zahlen, die, mathematisch gesprochen, als »Vektoren« transformiert werden. Da die Schichten hochgradig vernetzt sind, erschließt sich dem Betrachter die Bedeutung der inneren Knoten, der Gewichte und weitergeleiteten Wertekombinationen nicht ohne Weiteres

Deshalb spricht man hier auch von subsymbolischen Modellen. Dieser Umstand führt dazu, dass konnektionistische Ansätze kaum oder gar nicht für den Menschen nachvollziehbar sind. KNN eignen sich jedoch besonders gut, wenn extrem umfangreiche und hochdimensionale Daten verarbeitet werden müssen, wie bei der Bild- oder Sprachverarbeitung. Die Forschung an KNN führte in den 1960er Jahren zu ersten Anwendungen[13], kam dann aber fast zum Erliegen, als Beschränkungen in den damals benutzen Lernalgorithmen und Netzen aufgezeigt wurden[14]. Später bezeichnete man diese Phase als ersten »KI-Winter«.

Erst mit dem Aufkommen leistungsfähiger Computer, preiswerter Speichertechnologien und dem Anstieg der Datenmengen durch die Verbreitung der Internettechnologien um die Jahrtausendwende wuchs erneut das Interesse an Künstlichen Neuronalen Netzen. Mit vielen internen Schichten zeigen tiefe Neuronale Netze jetzt ihre großen Stärken in der automatischen Bilderkennung und natürlichen Sprachverarbeitung (Englisch: Natural Language Processing oder NLP).

1.3.4 Der Symbolismus

Ontologien sind in der Informatik sprachlich gefasste Beschreibungen von Konzepten und ihren Beziehungen.

Im Symbolismus werden Intelligenz- und Entscheidungsleistungen formalisiert, indem logikbasierte Beschreibungen von Wissen über Konzepte, ihre Eigenschaften und Beziehungen erstellt werden. Man bezeichnet solche Begriffssysteme als Ontologien.

Früher bildete man aus Fakten über Objekte Wissensbasen. Aus den Typen der Objekte und logischen Regeln konnte neues Wissen zielgerichtet hergeleitet werden. Man konnte zum Beispiel ausdrücken, dass Wasser nass ist, dass ein Ozean aus Wasser besteht, und,

[13] Rosenblatt 1958
[14] Wikipedia 2017a

dass nass wird, was im Ozean schwimmt. Aus einer Beobachtung, dass eine bestimmte Person im Ozean schwimmt, konnte dann gefolgert werden, dass sie nass sein muss[15].

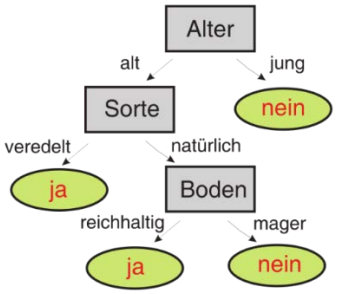
Expertensysteme waren mit ihrem sprachlich formulierten Wissen und den logischen Ableitungen gut nachvollziehbar, mussten aber mit großem Aufwand von Hand aufgebaut werden. Sie erlebten ihren Höhepunkt in den 1980er Jahren, bis klar wurde, dass sich größere Wissensbasen nur schwer konsistent erweitern ließen. Außerdem konnten praktisch nie alle denkbaren Vorbedingungen für die Anwendbarkeit der Regeln explizit angegeben werden. Es folgte der sogenannte zweite KI-Winter.

Vom Symbolismus ist im aktuellen Maschinellen Lernen kaum etwas übriggeblieben. Es gibt allerdings aus der Statistik motivierte Lernalgorithmen, die Merkmale wie Farben, Namen, Temperaturwerte oder Preisangaben interpretieren. Sie berücksichtigen aber nicht, dass dahinter Objekte mit Beziehungen oder sogar Regeln zur Ableitung neuen Wissens stehen könnten.

ML-Modelle, die symbolische Beschreibungen enthalten, sind recht gut nachvollziehbar. Hohe Akzeptanz finden sogenannte Entscheidungsbäume. Sie leiten die gesuchte Antwort, also etwa eine Kategorie oder Empfehlung, her, indem sie sukzessive Merkmale abfragen. Lernen in einem Entscheidungsbaum bedeutet, den Knoten von oben nach unten möglichst diskriminative Abfragen von Eigenschaften zuzuweisen. Je geschickter man in den Knoten abfragt, umso prägnanter und nachvollziehbarer wird der Baum. Das Resultat besteht aus verständlichen Regeln zur Lösung der Aufgabe.

Da die Nachvollziehbarkeit neben der Performanz eine wichtige Eigenschaft darstellt, ist es vorstellbar, dass symbolische Methoden in der Zukunft wieder eine wichtigere Rolle einnehmen werden.

Abbildung 7: Entscheidungsbaum zur Vorhersage der Apfelernte[16]



[15] Cycorp 2017

[16] Grafik: André Flöter via Wikipedia, cc; https://de.wikipedia.org/wiki/Datei:Entscheidungsbaum.svg

Seit einigen Jahren finden Ontologien zur Verknüpfung verschiedener Datenbestände (Englisch: Linked Data) im semantischen Netz (Englisch: semantic web) und in sogenannten »Wissensgraphen« (Englisch: knowledge graph) neue Anwendung. Das größte semantische Netz aus öffentlichen Daten ist DBPedia, das automatisch aus Wikipedia extrahiert wird[17]. Google nutzt seit 2013 einen Wissensgraphen für die Infoboxen in seiner Suchmaschine, und der IBM Supercomputer »Watson«, der 2011 menschliche Teilnehmer im Quiz-Spiel »Jeopardy« besiegte, hat aus Texten extrahiertes Wissen in Wissensgraphen dargestellt.

Die Objekte und Relationen im Graphen werden eindeutig gekennzeichnet, um inhaltsvolle strukturierte Information in standardisierter Repräsentation zwischen Maschinen auszutauschen. Das ist ein wichtiger Beitrag zur Errichtung von intelligenten Maschinen, um den sich eine Initiative zur Errichtung internationaler Standards kümmert[18].

Googles Suchmaschine zeigt erste Anwendungen: Bei einer Anfrage nach der Hauptstadt Deutschlands wird die Antwort direkt ausgegeben, anstatt nur wie früher der Link zu einem Dokument, in dem die Antwort vermutlich zu finden ist. Die Suchmaschine hat in der Anfrage das Objekt »Deutschland« und die Relation »hat als Hauptstadt« richtig erkannt und muss innerhalb des Wissensgraphen nur vom »Deutschland«-Knoten der richtigen Verbindung folgen. Eine weitere Anwendung von Wissensgraphen verspricht eine Weiterentwicklung von textbasierten Suchmaschinen hin zu performanten Dialogsystemen, die Anfragen sinngemäß beantworten können.

Dazu müssen die Wissensgraphen möglichst vollständig sein. Einige Lernverfahren versuchen zu diesem Zweck, Relationen zwischen Knoten vorherzusagen, Knoten zu identifizieren, die dasselbe Objekt beschreiben, und ein Objekt einem Typ zuzuordnen.[19] Dabei kann es helfen, die potenziellen Relationen durch Hintergrundwissen einzuschränken: Ein Mensch kann nur Kind von anderen Menschen sein. Auch dieses Einbauen von Hintergrundwissen in Form von Ontologien ist standardisiert. Je komplexer die zugelassenen Beschränkungen sind, desto schwieriger wird jedoch das automatische Arbeiten mit dem Graphen.

1.3.5 Der Bezug zwischen Big Data und Maschinellem Lernen

Insbesondere die jüngsten Erfolge des Maschinellen Lernens können nicht losgelöst von anderen relevanten Entwicklungen und Kontexten gesehen werden, ohne die insbesondere die Künstlichen Neuronalen Netze nicht so populär geworden wären. Ausschlaggebend waren Fortschritte in der Computer- und Informationstechnologie: die steigende Prozessorgeschwindigkeit und Rechenleistung, die Verfügbarkeit und den Preisverfall für gro-

[17] http://wiki.dbpedia.org/
[18] W3C 2014
[19] Nickel et al. 2016

ße Datenspeicher, Fortschritte im Design effizienter Algorithmen, Programmiersprachen und Daten-Management-Systemen, sowie die Verbreitung der Internet-Technologien, die große Datenmengen für ML-Trainingszwecke verfügbar machte. Viele Methoden der »Big Data Analytics« sind Methoden des Maschinellen Lernens.

»Big Data« ist zunächst nur eine Bezeichnung für die immensen und stetig wachsenden Datenmengen, die durch die weltweite Verbreitung von Internettechnologien entstanden sind. »Big Data« zeichnet sich durch die sogenannten »5V« aus. Sie besagen, dass es sich hierbei um:

- extrem große Datenmengen handelt (volume),
- die in einer Vielfalt unterschiedlicher Datentypen – Bilder, Text, Sprache, Tabellen – vorliegen (variety)
- und sich schnell verändern können und deshalb auch schnell ausgewertet müssen (velocity).

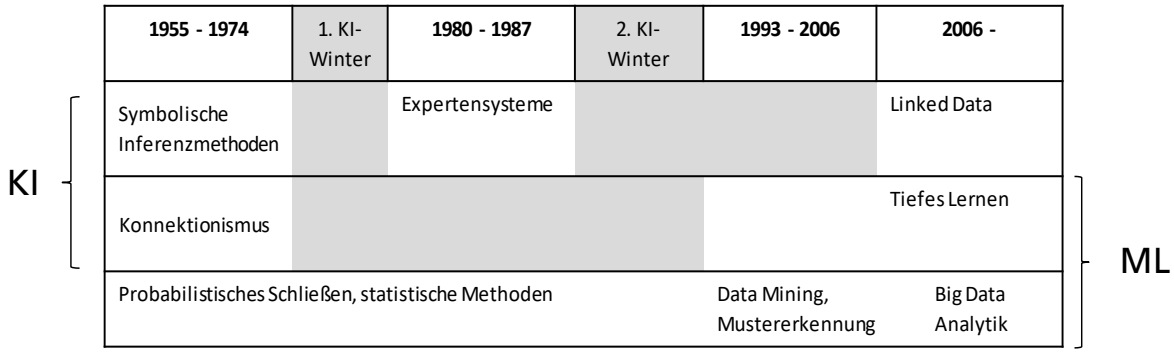
Zusätzlich wird die Anforderung gestellt, dass:

- die Daten glaubwürdig und guter Qualität sind (validity bzw. veracity)
- und die Auswertung einen Mehrwert liefert (value).

Mit Big-Data-Technologien ist es möglich geworden, diese Datenmengen ausfallsicher auf günstiger Hardware verteilt zu sammeln, zu speichern und zu verarbeiten. Daten gelangen oder entstehen in dedizierten Unternehmens-IT-Systemen und -Datenbanken, wo sie vor anderweitigem Zugriff gut geschützt sind. Der wahre Wert der Daten erschließt sich aber erst, wenn sie aus solchen abgeschotteten »Silos« extrahiert, kombiniert und intelligent analysiert werden können. So können neue Muster und Beziehungen erkannt werden, die auch zu besseren Prognosen führen. Hiermit ließen sich bspw. Stausituationen vorhersehen oder Produktions- und Logistikprozesse besser auf die jeweilige Nachfragesituation anpassen, um unnötige Fahrt- und Lagerkosten zu vermeiden. Intelligente Datenanalysen, in denen ML inzwischen eine bedeutende Rolle spielt, transformieren somit reines Big Data in »Smart Data«, also Daten, deren wertvoller Gehalt erschlossen wurde.

Der Erfolg hängt sehr stark von großen (Trainings-)Datenmengen ab, die wesentlich durch Big-Data-Technologie verfügbar wurden. Das gilt insbesondere für die Bild-, Audio- und Textdaten, die im Big-Data-Umfeld gern als unstrukturiert bezeichnet werden und zur Vielfalt (variety) von Big Data beitragen.

Abbildung 8: Überblick über die wichtigsten Phasen im Zusammenspiel von KI und ML, eigene Darstellung



1.4 Wichtige Lernstile des Maschinellen Lernens: Überwachtes, unüberwachtes und bestärkendes Lernen

Bei maschinellen Lernverfahren unterscheidet man Lernstile, die für jeweils andere Zwecke geeignet sind. Je nachdem, welche Zusatzinformation zur Verfügung steht, können andere Aufgaben gelernt werden. Beim überwachten Lernen müssen die richtigen Antworten zu den Beispielen als sogenannte Labels mitgeliefert werden. Die Angabe von Labels bedeutet meist mehr Arbeit für die Datenvorverarbeitung, ist aber notwendig, wenn Objekte klassifiziert und Werte geschätzt oder vorhergesagt werden sollen. Beim unüberwachten Lernen hingegen reichen die rohen Beispieldaten aus, um grundlegende Muster in den Daten zu entdecken. Beim bestärkenden Lernen nutzen Maschinen Feedback aus ihrer Interaktion mit der Umwelt, um ihre zukünftigen Aktionen zu verbessern und Fehler zu verringern. Diese Art des Lernens kommt häufig in der Robotik zum Einsatz, beispielsweise zum Erlernen der besten Greifbewegungen für Objekte.

1.4.1 Überwachtes Lernen

Beim überwachten Lernen (Englisch: supervised learning) liegt zu jedem Trainingsbeispiel gleich die richtige Antwort vor. Wenn etwa handgeschriebene Ziffern erkannt werden sollen, braucht man zu den Zeichen die richtige Ziffer, so dass die Richtigkeit der maschinellen Zuordnung sofort an den Lernalgorithmus zurückgespiegelt werden kann[20]. Alle Trainingsdaten sind hierzu mit der richtigen Ziffer als Label versehen.

[20] Jeder Datenpunkt in der Datenmenge des ML-Trainings-Sets besteht aus Eingabe- und Ausgabewerten (Input und Output), oder mathematisch ausgedrückt: f(X)=y, wobei X eine Matrix (Tabelle) mit Eingabewerten (z.B. eine Repräsentation von Bilddaten von Ziffern) und y ein Vektor mit Ausgabewerten (z.B. die Zuordnung zu einzelnen Zahlen, z.B. einer »Sieben«) ist.

Überwachte Lernverfahren werden oft verwendet, um Beispiele in bestimmte Kategorien oder Klassen einzuordnen: Zeichen als Ziffern interpretieren, Spam-Emails entfernen, in Bildern Gegenstände und Personen erkennen. Solche Aufgaben bezeichnet man als Klassifikationsaufgaben.

Eine andere wichtige Aufgabe, die man aus Beispielen mit bekannten Antworten lernen kann, sind Schätzungen oder Prognosen, wie zum Beispiel Stau-, Kosten-, Nachfrage- und Angebotsprognosen. Auch Stellgrößen in physikalischen Modellen für Anlagen und Maschinen können anhand von Beispielen gelernt werden. Die Algorithmen lernen dazu die Parameter einer numerischen Funktion, die die bekannten Beispiele möglichst gut treffen soll. Diese Lernaufgaben werden als Funktionsapproximations- oder Regressionsaufgaben bezeichnet.

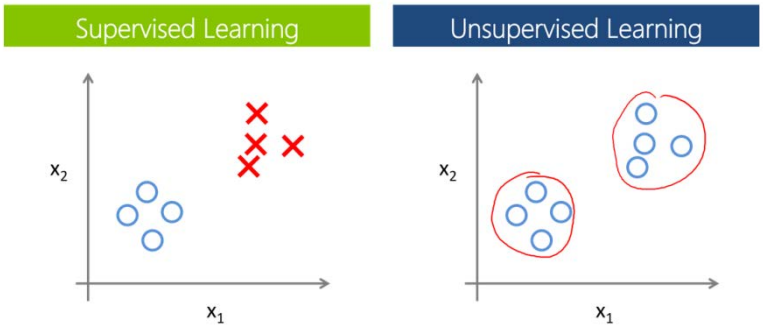
Ist die Qualität der zum Lernen verwendeten Beispieldaten schlecht, weil Label falsch sind oder wichtige Beispiele fehlen, lernt die Maschine folglich auch etwas Falsches. Wenn die Ziffer »4« selten vorkommt oder gelegentlich als »Sieben« markiert ist, wird die Maschine fälschlicherweise auch handschriftliche »Vieren« als »Sieben« interpretieren. Obwohl mit überwachtem Lernen der Fortschritt im ML-Trainingsprozess gut nachvollziehbar ist, ist das Auszeichnen der Trainingsdaten mit Labels oft mit einem erheblichen manuellen Aufwand verbunden. Die Plattform »Clickworker« mit 800 000 menschlichen Mitgliedern bietet unter anderem das Sammeln und Labeln von ML-Trainingsdaten an[21].

1.4.2 Unüberwachtes Lernen

Beim unüberwachten Lernen (Englisch: unsupervised learning) gibt es hingegen keine Labels zu den Trainingsdaten. Der Grund hierfür ist meistens, dass es sich um sehr große, unstrukturierte Datenmengen handelt, von denen man oftmals im Vorfeld noch gar nicht weiß, wie sie gut beschrieben oder nach welchen Kriterien sie eigentlich aufgeteilt werden können. Man kann aber versuchen, Strukturen und Unterschiede in den Daten zu erkennen, um etwa Gruppen (Englisch: Cluster) ähnlicher Beispiele zu finden. Eine typische Anwendung ist die Segmentierung von Kundendaten, um Zielgruppen zu identifizieren, die man auf ähnliche Weise ansprechen möchte. Clustering wird oft auch zur Datenexploration eingesetzt. Meist folgen darauf weitere eingehende Analysen des Datenbestands. Kennt man erst einmal die wichtigsten Cluster, kann man anschließend lernen, die Beispiele anhand ihrer Eigenschaften genau diesen Clustern zuzuordnen. Das ist dann eine Klassifikationsaufgabe.

[21] Clickworker 2017

Abbildung 9: Intentionen bei überwachtem und unüberwachtem ML[22]



Clustering vs. Klassifikation:
Beim Clustering werden Gruppen von ähnlichen Daten gefunden. Dabei steht noch gar nicht fest, welche Merkmale genau diese Ähnlichkeiten und Unterschiede ausmachen. In einer Menge von Emails können sich zum Beispiel zwei Cluster herausbilden, die ein Experte anschließend als »Spam« und »Wichtig« erkennt.

Bei einer Klassifikation steht dagegen schon im Vorfeld fest, in welche Gruppen ein Objekt eingeordnet werden kann. Hier geht es darum, die Merkmale herauszufinden, die für die Zuordnung am signifikantesten sind. Im Fall der Emails unterscheiden sich Spam und wichtige Emails zum Beispiel in den Absendern und den verwendeten Wörtern.

Die Daten, aus denen die Beispiele bestehen, können redundant sein, zum Beispiel, weil es Abhängigkeiten (Korrelationen) zwischen den Merkmalen gibt. Viele Lernalgorithmen funktionieren aber besser, wenn Merkmale möglichst unabhängig sind. Es gibt verschiedene Lernalgorithmen, um Beispieldaten in eine kompaktere Form zu überführen. Eine Möglichkeit ist die aus der Statistik stammende Hauptkomponentenanalyse, bei der Daten mit vielen, vermutlich korrelierten Merkmalen in eine Darstellung mit wenigen, (linear) unkorrelierten Merkmalen transformiert werden. Weil jedes Merkmal mathematisch als eine Dimension aufgefasst werden kann, bezeichnet man diese Aufgabenstellung des unüberwachten Lernens als Dimensionsreduktion.

[22] Ng 2017

1.4.3 Semi-überwachtes Lernen

Einen Kompromiss zwischen überwachtem und unüberwachtem Lernen bildet das semi-überwachte Lernen (Englisch: semi-supervised learning). Die Lernaufgaben sind hier prinzipiell die gleichen wie beim überwachten Lernen, jedoch sind nicht alle Trainingsdaten, sondern nur ein paar mit dem Ergebnis in Form des Labels versehen. Ein Hauptgrund hierfür ist der hohe Aufwand, der mit dem Ausstatten mit Labels verbunden ist und der bei sehr großen Datenmengen oftmals nicht praktikabel ist. Das halbüberwachte Lernen stellt Lernalgorithmen zur Verfügung, die auch Beispieldaten ohne Label für das Training verwendet werden können. Ein einfaches Beispiel ergibt sich aus der Kombination von Clustering und Klassifikation, indem man erst einen Clustering-Algorithmus anwendet und anschließend die wenigen vorhandenen Beispiele mit Label nutzt, um den Clustern, und damit allen Beispielen des Clusters, eine Klasse zuzuweisen.

1.4.4 Bestärkendes Lernen und sequentielles Entscheiden

Ein weiterer Lernstil von wachsendem Interesse ist das bestärkende Lernen (Englisch: reinforcement learning) für Maschinen, die mit ihrer Umgebung interagieren. Dabei nutzen sie Feedback, das sie auf ihre Aktionen von der Umwelt erhalten, um die Erfolgsaussichten der einzelnen Aktionen in den verschiedenen Situationen besser einschätzen zu lernen. Das Feedback erhält die Maschine in Form eines mathematischen Äquivalents zu »Belohnung« und »Tadel«, wenn sie ihr Ziel erreicht oder verfehlt hat. Sie soll lernen, welche Aktionen sie jeweils auswählen soll, um das Feedback, sprich die Nutzenfunktion, zu maximieren. Da die Maschine die Aktionen schrittweise auswählt und durchführt, bezeichnet man die Lernaufgabe auch als »sequentielles Entscheiden« (Englisch: sequential decison making).

Wenn die Maschine verspätet Feedback für vorherige Aktionen erhält, muss sie es auf ihre vergangenen Aktionen zurückrechnen können. Diese Art Lernen entspricht dem experimentellen Schachspieler, der eine neue Taktik testet und erst beim Spielausgang den Erfolg der Taktik beurteilen kann.

Bestärkendes Lernen wurde von der Google-Tochter »Deep Mind« angewendet, um eine Maschine dazu zu bringen, eigenständig sieben Atari 2600-Spiele bis auf »Meisterniveau« zu lernen[23]. Ähnlich wie für Menschen diente der Maschine der Punkteerwerb beim Spiel als Motivator für die Entwicklung immer besserer Strategien. Auch in AlphaGo, das 2016 einen der weltbesten Go-Meister besiegte, hat DeepMind Reinforcement-Lerntechniken eingesetzt, unter anderem, indem die Maschine gegen sich selbst gespielt hat. Bestärkendes Lernen ist außer für Spiele auch sehr wichtig für Bots, die in digitalen Umgebungen agieren und für Roboter, die sich in einer physischen Umgebung bewegen, Ereignisse erkennen und Dinge manipulieren.

[23] Mnih et al. 2013

1.5 Modelltypen und Algorithmen des Maschinellen Lernens

Es gibt eine Vielzahl Modelltypen und Algorithmen des Maschinellen Lernens, die jeweils für unterschiedliche Aufgaben besonders gut geeignet sind und heute immer noch breiten Einsatz finden. Dieser Abschnitt bietet einen Überblick über Regressionsmethoden, Entscheidungsbäume, Clustering, Kernmethoden, Bayessche Modelle und unterschiedliche Konfigurationen der Künstlichen Neuronalen Netze, mit Informationen darüber, wie sie jeweils funktionieren, was sie leisten und für welche Aufgaben sie sinnvoll eingesetzt werden können. Insbesondere die tiefen Neuronalen Netze haben besondere Fähigkeiten in der Datenrepräsentation und ermöglichen ganz neue Lernaufgaben.

Die Algorithmen des Maschinellen Lernens kann man nach verschiedenen Kriterien einteilen: Lernaufgabe und Lernstil, was sich danach richtet, ob die Beispiele Labels haben und ob es gelegentliches Feedback gibt. Weitere Aspekte sind die Art des Modells und die damit zusammenhängende Denkschule. Schließlich gibt es gemeinsame Lösungsprinzipien. Im Folgenden werden einige der häufig verwendeten ML-Modelle mit ausgewählten Lernverfahren vorgestellt.

1.5.1 Regressionsmodelle

Lineare Regression[24]

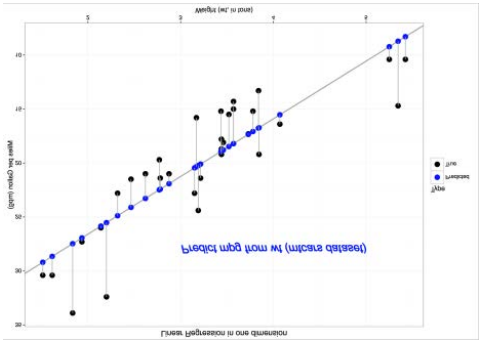


Abbildung 10: Lineare Regression[25]

Alle Merkmale müssen numerisch sein. Das Modell ist eine lineare Funktion. Ihre Parameter (Abstand vom Nullpunkt und Steigung) werden gelernt, indem die Abweichungen zu den bekannten Antworten der Beispiele möglichst minimiert werden. Als Antwort auf neue Beispiele wird der Funktionswert zurückgegeben.

Besonderheiten:
Lernaufgabe: Regression, Lernstil: überwacht, analogistischer Ansatz

[24] INWT Statistics 2017
[25] Kühn 2017

Logistische Regression[26]

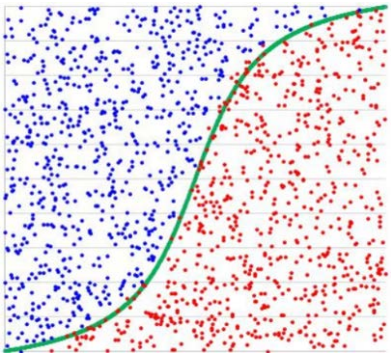


Abbildung 11: Logistische Regression[27]

Die abhängigen Merkmale müssen alle numerisch sein. Das Modell ist eine Trennungslinie, die zwei Klassen voneinander trennt. Hierzu werden die Parameter einer linearen Transformation gelernt, so dass eine anschließende logistische Funktion jedem Eingabewert eine von den zwei Klassen zuordnet. Dies ergibt eine Entscheidungsgerade, mit der man ungesehene Daten trennen kann. Über geeignete Transformationen können auch nichtlineare Entscheidungslinien gelernt werden.

Besonderheiten:
Lernaufgabe: Klassifikation, Lernstil: überwacht, analogistischer Ansatz
Varianten: Die Anzahl der Klassen kann erhöht werden.

1.5.2 Entscheidungsbäume

Lernverfahren Iterative Dichotomiser 3 (ID3)[28]

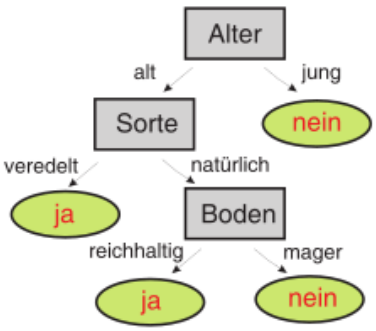


Abbildung 12: Iterativer Dichotomiser 3 (ID3)[29]

Das Modell ist ein sogenannter Entscheidungsbaum, der in einer Baumstruktur Entscheidungskriterien und ihren Ausgang darstellt. An jedem Verzweigungsknoten wird ein Merkmalswert abgefragt, in den Endknoten (Blatt) steht eine Klasse. Die Entscheidungskriterien an den Verzweigungsknoten werden so gelernt, dass die neue Aufteilung maximalen Informationsgewinn (kleinsten Entropiewert) hat. Ein neues Beispiel wird klassifiziert, indem man den Entscheidungskriterien von der Wurzel bis zu einem Blatt folgt.

Besonderheiten:
Lernaufgabe: Klassifikation, Lernstil: überwacht, symbolischer oder analogistischer Ansatz.

[26] INWT Statistics 2017
[27] Snider 2017
[28] Quinlan 1986
[28] Breiman et al. 1983
[29] Snider 2017

Klassifikations- und Regressionsbäume (CART)[30]

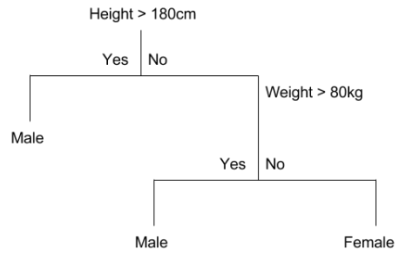


Abbildung 13: Klassifikationsbaum[31]

Das Modell ist ein Entscheidungsbaum, der jedem Endknoten einen numerischen Wert zuweist. Dieser soll bestmöglich für genau die Beispiele gewählt werden, die gemäß den Entscheidungen im Baum in diesem Blatt landen. Zum Beispiel kann der Durchschnitt der Labels von den Trainingsbeispielen dieses Blattes gewählt werden. Es liegt am Modellentwickler, den Baum bei einer geeigneten Tiefe zu kappen. Neuen Beispielen werden die Werte ihres zugehörigen Blattes zugewiesen.

Besonderheiten: Lernaufgabe: Regression, Lernstil: überwacht
Varianten: Es gibt verschiedene Strategien, die Entscheidungskriterien an den Knoten zu wählen.

Random Forests[32]

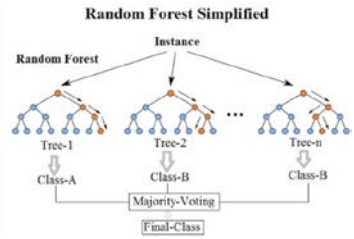


Abbildung 14: Random Forest[33]

Das Modell besteht bei Random Forests aus einer Gruppe (Wald, Forest) von Entscheidungsbäumen, die parallel laufen. Jeder Baum kommt am Ende zu einem Ergebnis (z.B. einer Klassifikation). Die Klasse, die von den meisten Bäumen »gewählt« wurde, ergibt die Antwort. Damit die einzelnen Bäume nicht genau dasselbe lernen, werden sie mit unterschiedlichen Teilmengen der Trainingsbeispiele trainiert.

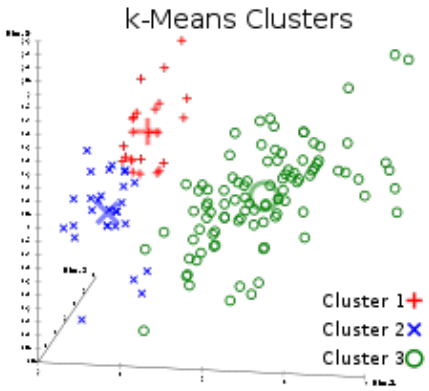
Besonderheiten: Lernaufgabe: Regression oder Klassifikation, Lernstil: überwacht
Varianten:
Random Forest gehört zu einer allgemeinen Klasse von Ensemble-Methoden. Sie nutzen mehrere Modelle von einem oder mehreren Lernalgorithmen, um bessere Ergebnisse zu erhalten. Die einzelnen Lernalgorithmen werden meist als »schwach« bezeichnet, da fehlerhafte Klassifizierungen für die einzelnen Entscheidungsbäume tolerierbar sind und im Gesamtmodell (zum Beispiel durch eine Mehrheitsentscheidung) ausgeglichen werden können. Beim »Boosting« (Deutsch: Verstärken) werden die schwachen Lernalgorithmen nacheinander trainiert. Dabei werden die vom Vorgänger fehlklassifizierten

[30] Rao 2013
[31] Brownlee 2017
[32] Wikipedia 2017c
[33] Jagannath 2017

Beispiele stärker gewichtet, um dessen Unzulänglichkeiten auszugleichen.

1.5.3 Cluster

k-means Clustering[34]



Beim Clustering soll die Beispielmenge so in Gruppen (Cluster) aufgeteilt werden, dass die Beispiele in einem Cluster möglichst ähnlich und Beispiele aus verschiedenen Clustern möglichst unähnlich sind. Die Ähnlichkeit wird durch eine vorzugebende Distanzfunktion auf den Beispieldaten ausgedrückt.

Abbildung 15: k-means Clustering[35]

Beim k-means Clustering wird die Anzahl k der Cluster vorgegeben. Es wird mit beliebigen k Punkten als Clusterzentren gestartet und alle Beispiele ihrem jeweils ähnlichsten Clusterzentrum zugeordnet. Nun werden wiederholt die Mittelpunkte der aktuellen Cluster berechnet, als neue Clusterzenten gewählt, und anschließend alle Beispiele neu zugeordnet. Im Einsatz wird ein neues Beispiel einfach dem Cluster zugeordnet, dessen Zentrum am nächsten ist.

Besonderheiten:
Lernaufgabe: Clustering, Lernstil: unüberwacht, analogistischer Ansatz
Varianten:
Es gibt eine Vielzahl an Clustering-Methoden, die auf unterschiedliche Weise die Gruppen bestimmen. Bei einigen muss die Anzahl der Cluster als Parameter vorgegeben werden, andere Algorithmen bestimmen ihn selbst. Ein bekanntes Beispiel ist DBSCAN, bei dem Cluster dichte-abhängig gefunden werden. Hier kann es sein, dass Beispiele als Ausreißer markiert werden, die keinem Cluster angehören.

Hierarchische Verfahren bestimmen ganze Hierarchien von Clustern, etwa, indem sie zunächst alle Beispiele in ein einziges Cluster packen und die Cluster nach und nach bezüglich eines vorgegebenen Kriteriums teilen.

[34] Wikipedia 2017d
[35] Wikipedia 2017d

1.5.4 Kernmethoden

Stützvektormaschine (SVM)[36]

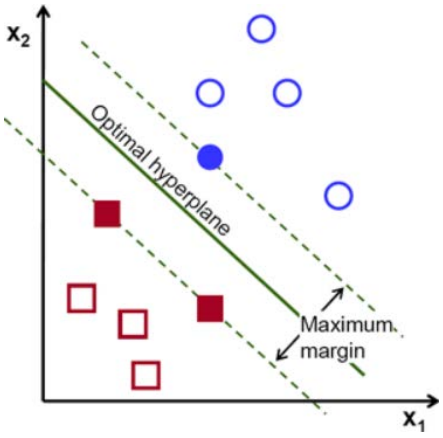


Abbildung 16: Stützvektormaschine (SVM)[37]

Eine Stützvektormaschine für die Klassifikation lernt eine Entscheidungsebene in dem Raum der Eingabedaten, und zwar genau die, die den maximalen Abstand zu den am nächsten liegenden Datenpunkten hat. Per »Kernel-Trick« lassen sich auch nichtlineare Entscheidungsebenen recheneffizient lernen, indem die Ebene in einem impliziten höherdimensionalen Raum bestimmt wird.

Besonderheiten:
Lernaufgabe: Klassifikation, Lernstil: überwacht, analogistischer Methode

Varianten:
Es existieren Alternativen für die Unterscheidung in mehr als zwei Klassen (Multiclass SVM), für die Regression (SVR, Support Vector Regression) und für halbüberwachtes Lernen (Transductive SVM).

Kernel Principal Component nalysis (PCA)[38]

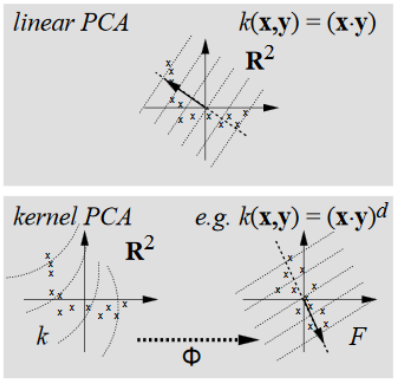


Abbildung 17: Kernel Principal Component Analsis (PCA)[39]

Bei der klassischen PCA wird die Darstellung der Daten mit vielen, vermutlich korrelierten Variablen in eine Darstellung mit wenigen, (linear) unkorrelierten Variablen reduziert. Die zugehörige Transformation kann neue Beispiele strukturell einordnen. Per »Kernel-Trick« wird auch dieses Verfahren nichtlinear.

Besonderheiten:
Lernaufgabe: Dimensionsreduktion, Strukturerkennung; Lernstil: unüberwacht, statistische Methode

[36] Cortes, C./Vapnik, V. 1995
[37] Open CV 2017
[38] Schölkopf/Smola/Müller 1998

1.5.5 Künstliche Neuronale Netze

Feed-forward Network (FF oder FFNN)[40]

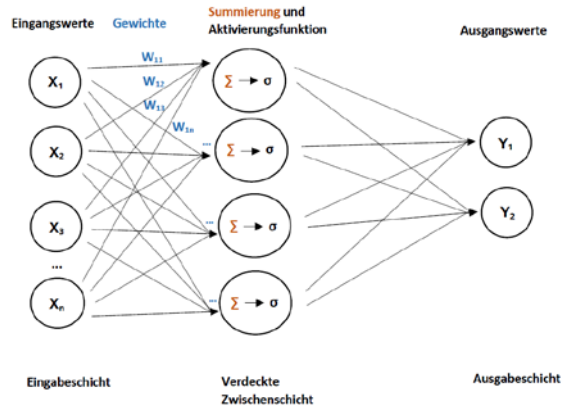


Abbildung 18: Feed-forward Network, eigene Darstellung

Numerische Eingaben werden als Signale über mehrere Schichten in Ausgaben umgewandelt. Der Aufbau des Modells ist von der Funktionsweise des menschlichen Gehirns frei abstrahiert und besteht aus einer Vernetzung von »künstlichen Neuronen« mit Aktivierungsfunktionen. Gelernt wird die Stärke der Verbindungen zwischen den Knoten benachbarter Schichten, indem Fehler zwischen berechneter und richtiger Ausgabe zurückgerechnet werden.

Neue Eingaben werden durch das Netz propagiert, indem sie mit den Gewichten multipliziert werden und vor jedem Knoten aufsummiert an die jeweilige Aktivierungsfunktion übergeben werden. Die Aktivierungsfunktionen sind nichtlinear, so dass die insgesamt gelernte Funktion komplizierte nichtlineare Funktionen approximieren kann

Besonderheiten:
Lernaufgabe: Klassifikation oder Regression, Lernstil: überwacht, konnektionistischer Ansatz

Varianten:
Man unterscheidet KNN mit höchsten einer Zwischenschicht von tiefen KNN mit mehreren Zwischenschichten. Hierzu mehr in Kapitel 1.6.

[39] Schölkopf/Smola/Müller 1998

[40] Goodfellow/Bengio/Courville 2017

1.5.6 Bayessche Modelle

Lernen eines Bayesschen Netzes (BN)[41]

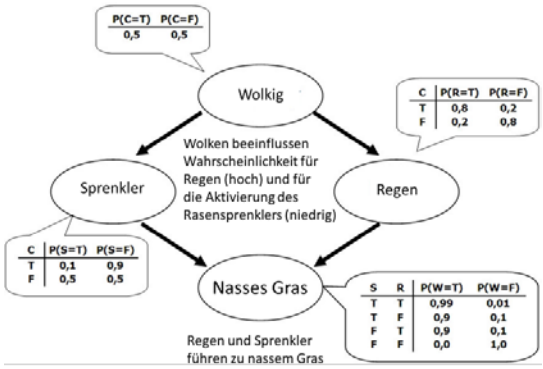


Abbildung 19: Bayessches Netz (BN)[42]

Ein Bayessches Netz besteht aus Knoten für Zufallsvariablen und Pfeilen, die Abhängigkeiten zwischen den Variablen darstellen. Zu jedem Knoten gibt es Tabellen mit bedingten Wahrscheinlichkeiten. Damit lässt sich die Wahrscheinlichkeit von unbeobachteten bedingten Variablen aus Wahrscheinlichkeiten von beobachteten Größen bestimmen. Die Wahrscheinlichkeitstabellen können von Parametern abhängen, die aus Daten gelernt werden.

Besonderheiten:

Lernaufgabe: Klassifikation, Lernstil: überwacht, Bayesscher Ansatz

Varianten:

In anderen Ansätzen werden nicht nur die Wahrscheinlichkeiten an den Knoten gelernt, sondern auch die zu den vorliegenden Beispielen passende Graphstruktur. Lernen ist auch möglich, wenn nicht zu jedem Knoten Messungen vorliegen.

1.5.7 Sequentielle Entscheidungsmodelle

Q-Lernen[43]

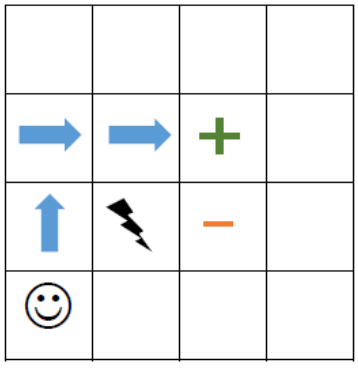


Abbildung 20: Q-Lernen nach[44]

Das Modell besteht aus einer Entscheidungsfunktion, die einem Agenten in einem Zustand eine Aktion zuweist. So bewegt sich der Agent von Zustand zu Zustand und trifft sequentiell die jeweils aussichtsreichste Entscheidung. Beim Q-Lernen wird dafür eine Bewertungsfunktion gelernt, die jedem Paar von Zustand und Aktion einen Zahlenwert zuweist. Dieser Wert wird gelernt, indem der Agent viele verschiedene Entscheidungssequenzen durchgeht. Im Einsatz wird dann in jedem Zustand die Aktion mit dem höchsten zugewiesenen Wert gewählt.

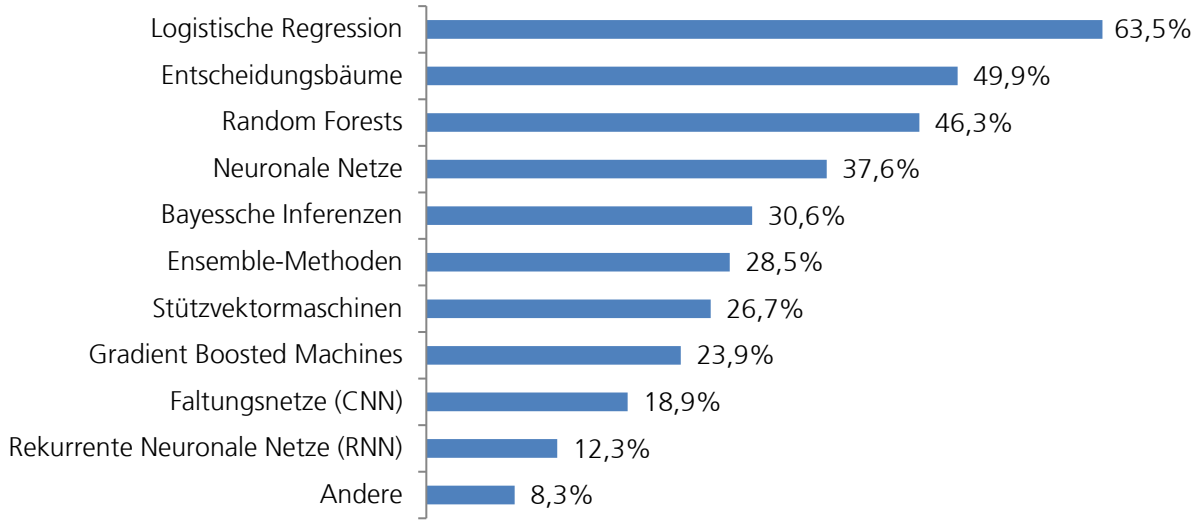
[41] Neapolitan 2003
[42] Goodman/Tenenbaum 2016
[43] Watkins/Dayan 1992
[44] Quantitative Journey 2015

Besonderheiten:
Lernaufgabe: Sequentielles Entscheiden, Lernstil: bestärkendes Lernen

Varianten:
Eine Stellschraube im Algorithmus wägt ab, inwieweit kurzfristige kleine Belohnungen über langfristige größere Belohnungen bevorzugt werden sollen.
Die Bewertungsfunktion kann gemäß konnektionistischem Ansatz mit einem Neuronalen Netz gelernt werden.

Kaggle, eine Plattform für ML-Wettbewerbe, hat 2017 in einer Umfrage nach den verwendeten Methoden gefragt. Das Ergebnis auf Basis von 7 301 Antworten zeigt die nächste Abbildung. Ensemble-Methoden und »Gradient boosted machines« kombinieren mehrere Modelle, meist Entscheidungsbäume, für Klassifikations- und Regressionsaufgaben. CNN und RNN sind tiefe Neuronale Netze.

Abbildung 21: Verwendete Methoden der von Kaggle befragten Data Scientists und ML-Fachleute [45]



1.6 Die Renaissance Künstlicher Neuronaler Netze

Insbesondere durch die Fortschritte in der Computertechnologie und das Anwachsen von Big-Data erlangten die KNN seit der Jahrtausendwende wieder zusehends Interesse in der Forschung. Anwendungen, die zuvor nur in der Theorie existierten und an geringer Rechenkapazität und spärlicher Datenlage scheiterten, verzeichneten sukzessive Erfolge, insbesondere in der Bilderkennung und Verarbeitung natürlicher Sprache (Englisch: Natu-

[45] Kaggle 2017

ral Language Processing oder NLP). In einigen Fällen können Maschinen inzwischen Gesichter und Objekte mit einer geringeren Fehlerquote identifizieren als es Menschen oder sogar Experten bewerkstelligen[46].

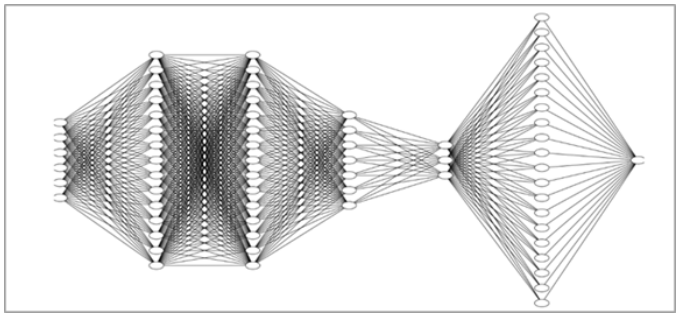
Während die ersten Künstlichen Neuronalen Netze aus einer Eingabe-, Ausgabe- und vielleicht einer verdeckten Zwischenschicht bestanden, findet man heute schon »tiefe« KNN mit Hunderten von Schichten und Milliarden von zu lernenden Gewichten zwischen den Knoten. Dabei reicht in der Theorie auch eine einzige verdeckte Zwischenschicht aus, um jede beliebige Funktion von Ein- zu Ausgabe zu lernen[47]. Allerdings würde sie dafür auch beliebig viele Knoten benötigen und es müssten entsprechend viele Gewichte gelernt werden. Experimente haben gezeigt, dass mehrere Zwischenschichten effizienter zu trainieren sind.

1.6.1 Funktionsweise von tiefen KNN

In einem tiefen KNN können die einzelnen Schichten unterschiedlich viele Knoten haben. Auch die Wahl der Verbindungen zwischen den Schichten ist ein wichtiger Bestandteil in der Gestaltung des Netzes und erfordert viel Erfahrung und einige Experimente.

Die nächste Abbildung zeigt schematisch ein tiefes KNN aus einer konkreten Anwendung: Es sollte die Stromnachfrage in der nächsten Stunde vorhergesagt werden. Die Eingabe besteht aus einem Vektor mit 112 Werten für Datum, Zeit, Wetter und aktuellem Verbrauch. Es folgen vier voll vernetzte Schichten, die zunächst 256 Knoten auf 16 Knoten verdichten. Die vorletzte Ebene hat eine andere Aktivierungsfunktion und schätzt die Wahrscheinlichkeiten, dass der Stromverbrauch in verschiedene Intervalle von 20 Megawatt fallen wird. Der letzte Knoten gibt das wahrscheinlichste Intervall aus.

Abbildung 22: Tiefes KNN zur Vorhersage der Stromnachfrage[48]



[46] He et al. 2015
[47] Hornik 1991
[48] Quantup 2016

Bilder wie in Abbildung 22 sind nur eine Veranschaulichung von programmierten Datenstrukturen und Funktionen. Abbildung 23 zeigt ein Stück Code, um ein Netz mit drei Schichten für eine Klassifikationsaufgabe zu erstellen.

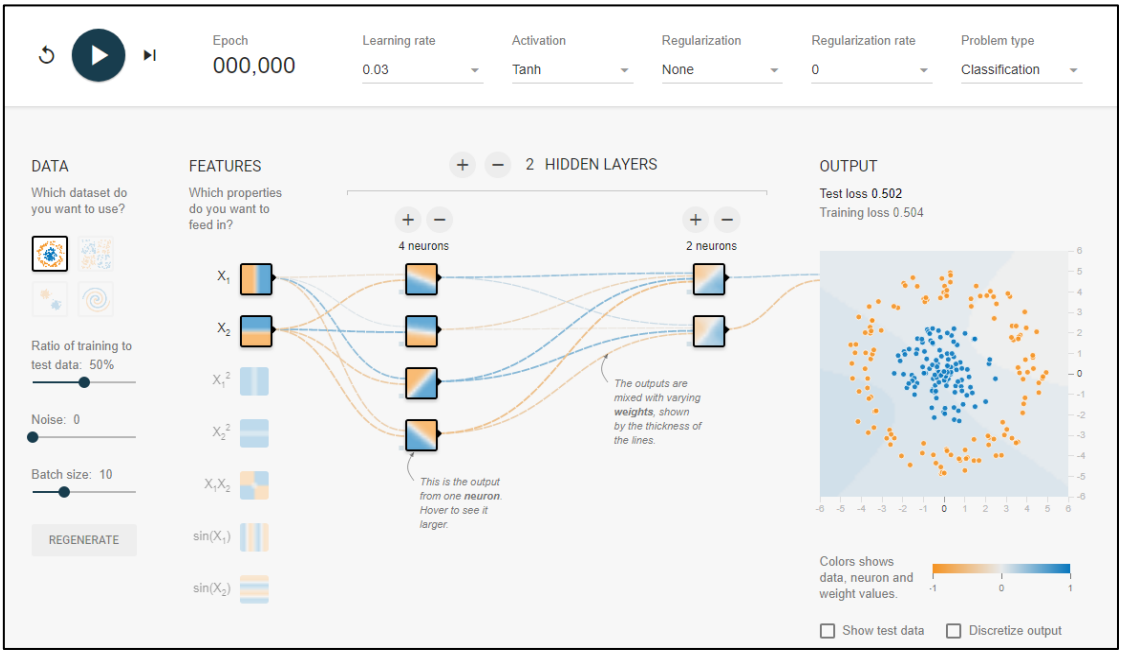
Abbildung 23: Aufbau eines Künstlichen Neuronalen Netzes in TensorFlow[49]

```
## Aufbau eines neuronalen Netzes mit einer versteckten Schicht
## zur Klassifizierung in 4 verschiedene Klassen
x = tf.placeholder(tf.float32, [None, 64]) # Datenbeispiele
y = tf.placeholder(tf.float32, [None, 4]) # Zielvorgaben
W1 = tf.Variable(tf.zeros([64, 128])) # Matrix für Gewichtungen
b1 = tf.Variable(tf.zeros([128]))
hidden = tf.nn.relu(tf.matmul(x, W1) + b1)  # Berechnung der versteckten Schicht
W2 = tf.Variable(tf.zeros([128, 4]))
b2 = tf.Variable(tf.zeros([4]))
predict = tf.matmul(hidden, W2) + b2 # Berechung der Vorhersage

## Kostenfunktion und Minimierungsverfahren
cross_entropy = tf.reduce_mean(
    tf.nn.softmax_cross_entropy_with_logits(labels=y, logits=predict))
train_step = tf.train.GradientDescentOptimizer(0.5).minimize(cross_entropy)
```

Wie sich die Anzahl der Schichten und Knoten in einem übersichtlichen Netz auswirkt, kann man auf einer interaktiven Seite ausprobieren.

Abbildung 24: Seite zum Testen von Parametern eines KNN [50]



[49] Tensorflow 2017a
[50] Tensorflow 2017b

Prinzipiell geht es beim »Trainieren« eines KNN um die Optimierung der Verbindungsgewichte. Um die Differenz zwischen der korrekten und der jeweils aktuellen Ausgabe möglichst klein zu halten, wird oft die Methode der sogenannten Fehlerrückführung (Englisch: Back-Propagation) eingesetzt. Ausgehend von der Ausgabeschicht wird der Fehler an den Verbindungen entlang durch das gesamte KNN unter Verwendung mathematischer Berechnungen (Kettenregel der Differentialrechnung) »rückgeführt«. So können die Gewichtungen der Verbindungen schrittweise gezielt verändert werden, damit in den nächsten Versuchen der Fehlerwert immer geringer ausfällt.

1.6.2 Lernen von Datenrepräsentationen in tiefen KNN

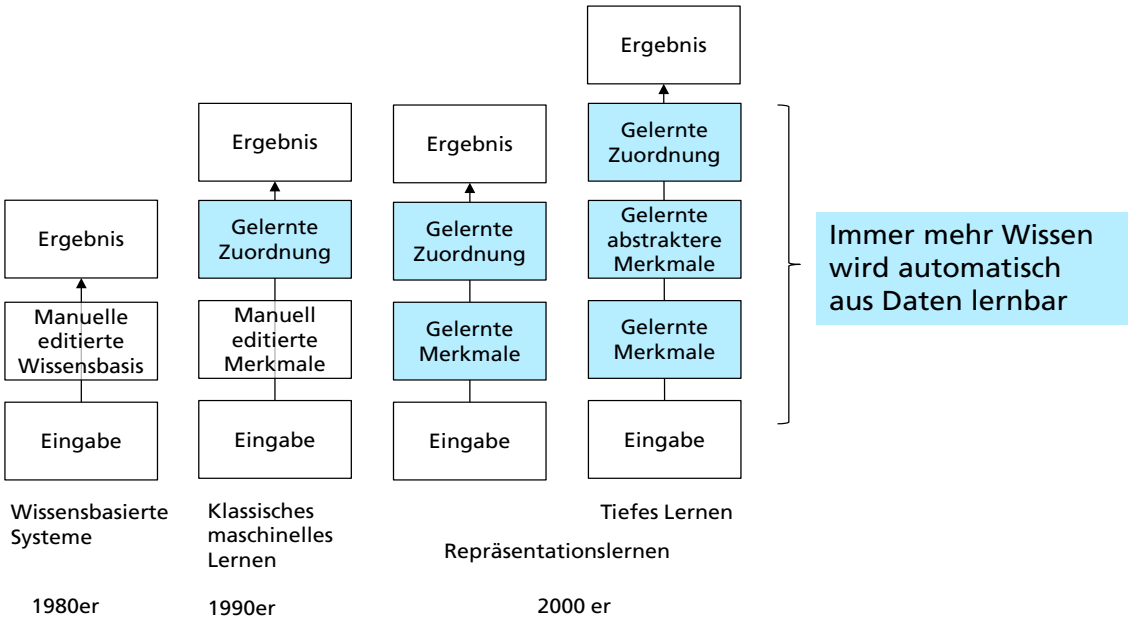
KNN können jegliche Art von Rohdaten verarbeiten, die als Zahlenvektoren kodiert werden. Bei einem Schwarz-Weiß-Bild kann jedem Pixel seine Graustufe zugewiesen werden, bei einem Farbbild braucht man für jedes Pixel drei Zahlen für die Intensität von Rot, Grün und Blau. Bei Audiosignalen kann man Fourier-Koeffizienten (mathematische Zerlegung in Sinus- und Kosinusfunktionen) für die Kodierung benutzen.

Wörter kann man per Wörterbuch auflisten und dann jedem Wort einen Vektor mit genau einer »1« und sonst lauter Nullen zuweisen. (1,0, ..., 0) würde also das erste Wort repräsentieren und (0, ..., 0, 1) das letzte. Solche Darstellungen bezeichnet man auf Englisch als »one-hot encoding«. Tiefe KNN können aber kompaktere Darstellungen lernen, wobei die Vektoren viel kleiner werden und mehrere von Null verschiedene Stellen haben. Bei ca. 75 000 Wörtern der deutschen Standardsprache kommt man damit von 75 000 Dimensionen auf wenige Hundert. Ein geschickter Lernalgorithmus kann erreichen, dass Wörter mit ähnlicher Bedeutung ähnlich dargestellt werden. Solche Darstellungen werden als Worteinbettungen (Englisch: word embedding) bezeichnet.

Tiefe KNN sind deshalb so erfolgreich, weil sie aus Rohdaten selbstständig Darstellungen lernen können, die die eigentliche Aufgabe erleichtern, eben weil ähnliche Darstellungen ähnliche Bedeutung haben. Sie finden automatisch Strukturen in den gegebenen Beispielen, die passende Merkmale für die eigentliche Lernaufgabe liefern. Damit führen die tiefen Netze implizit eine Dimensionsreduktion durch, wobei etliche Datenvorverarbeitungsschritte, zum Beispiel zur computergraphischen Erkennung von Kanten und Flächen oder zur linguistischen Erkennung von Lauten und Wörtern entfallen. Da das Modell im Ganzen, von der Eingabe in Form von Rohdaten bis hin zur Ausgabe, trainiert wird, spricht man auch von Ende-zu-Ende-Lernen (Englisch: end-to-end machine learning).

Der Fortschritt gegenüber der Künstlichen Intelligenz in den 1980er Jahren wird in Abbildung 25 deutlich. Bei einem Expertensystem musste die gesamte Wissensbasis manuell konstruiert werden. Beim Maschinellen Lernen der 1990er waren es nur noch die Merkmale, die mit großer Sorgfalt ausgewählt werden mussten. Nun hat man Neuronale Netze, die automatisch in ihren verdeckten Schichten immer abstraktere Repräsentationen lernen und die Arbeit der Merkmalswahl selber miterledigen.

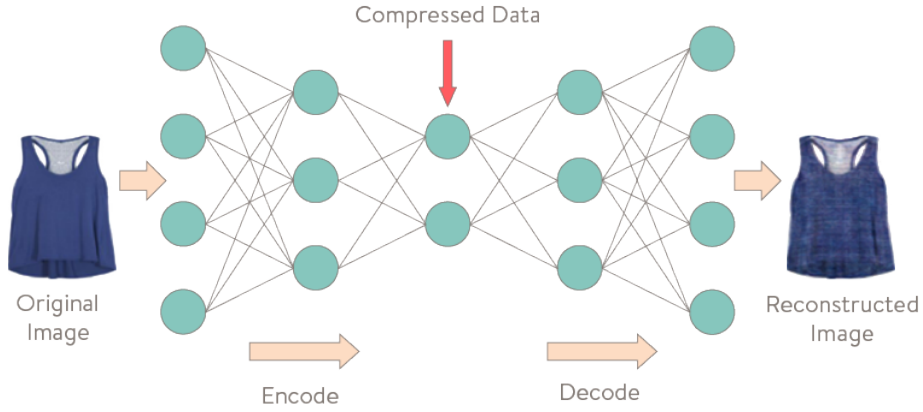
Abbildung 25: Fortschritte durch Maschinelles Lernen, adaptiert[51]



Das Lernen von geeigneten Repräsentationen findet automatisch als Teil des Netzes statt. Es kann aber auch ausgelagert und vorgelernt werden, damit das Netz für die Lösung der Lernaufgabe weniger Ressourcen benötigt. Eine Möglichkeit, kompakte Darstellungen zu lernen, besteht darin, ein tiefes Netz zu trainieren, das die Eingabe reproduziert. Wenn man die inneren Schichten erst verkleinert und dann wieder vergrößert, erhält man in der Mitte eine kompakte Repräsentation. KNN, die so funktionieren, nennt man Autoencoder.

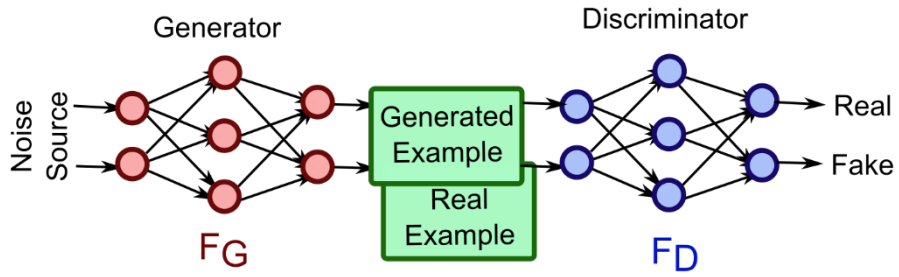
[51] Goodfellow/Bengio/Courville 2017

Abbildung 26: Autoencoder [52]



Eine andere Möglichkeit sind generative gegnerische Netze (Englisch: Generative Adversarial Networks). Sie bestehen aus zwei KNN, einem sogenannten Generator und einem Diskriminator, die parallel lernen. Der Generator versucht, realistische Beispiele zu erzeugen. Der Diskriminator sieht sowohl die generierten Beispiele als auch reale Daten und lernt, diese zu unterscheiden. Da der Generator Zugriff auf die Entscheidungsfindung des Diskriminators hat, werden die erzeugten Beispiele immer realitätsnaher.

Abbildung 27: Generative Adversarial Networks [53]



In Abbildung 27 sehen wir von einem Generator erzeugte Bilder. Er wurde als Bestandteil eines generativen gegnerischen Netzes auf Bildern von Schlafzimmern trainiert. Die neu erzeugten, realistisch aussehenden Bilder können zum Beispiel in einem nächsten Schritt als Trainingsbeispiele für eine Klassifikationsaufgabe genutzt werden.

[52] Torres 2015
[53] Guttenberg 2017

Abbildung 28: Künstlich generierte Bilder [54]

1.6.3 Neue Aufgaben für tiefe KNN

Weil tiefe KNN so gut Repräsentationen lernen, kann man sie nicht nur für die bekannten, eher analytischen Aufgabenstellungen nutzen, sondern auch für neue, eher konstruktive oder generative Lernaufgaben.

So gibt es in der Sprach- und Textverarbeitung viele Aufgaben, wo Folgen auf Folgen abgebildet werden müssen: die Transkription gesprochener Sprache in Text, die Übersetzung von Text in eine andere Sprache, die Aussprache von Text mit richtiger Betonung, die Produktion von Text in einer Handschrift, das Beantworten einer Frage, die Fortsetzung eines Dialogs, das automatische Beantworten einer E-Mail etc.

In der Bild- und Videoverarbeitung gibt es zudem viele Ergänzungsaufgaben: Bilder können rechnerisch vergrößert werden, Schwarz-Weiß-Bilder können eingefärbt werden, Objekte in künstlichen Welten können texturiert, und ein Video automatisch synchronisiert oder ein paar Sekunden in die Zukunft fortgesetzt werden.

Eine weitere Lernaufgabe ist das Generieren von neuen Beispielen. Hierunter fallen viele kreative Aufgaben wie das Schreiben von Gedichten, das Malen von Bildern und Komponieren von Musikstücken, die Animation von Figuren in digitalen Spielen und die Zusammenfassung von Meldungen zu einem Bericht. Mit realistisch generierten Beispielen kann man die Trainingsmenge für andere Lernaufgaben vergrößern.

Tiefe KNN eignen sich auch zum bestärkenden Lernen in der Welt der Spiele und Roboter. Die Google-Tochter DeepMind hat Maschinen das Atari- und Go-Spielen mit ihren paten-

[54] Radford/Metz/Chintala 2016

tierten »Deep Q-Networks (DQN)« beigebracht[55]. Diese Netze lernen eine Funktion, die den erwarteten Nutzen einer Aktion in einem Zustand schätzt. Damit kann jederzeit die perspektivisch beste Aktion gewählt werden.

1.6.4 Typen von tiefen Neuronalen Netzen

Man unterscheidet KNN nach der Anzahl und Breite der Schichten und den Verbindungen dazwischen. Bei tiefen Neuronalen Netzen sind die Möglichkeiten theoretisch unerschöpflich, und es kommen immer wieder Netze mit neuen Strukturen hinzu.[56] Der Aufbau der Netze und Teilnetze richtet sich nach Lernstil und Lernaufgabe, aber auch besonders nach Art und Bedeutung der Eingabe. Eine wichtige Rolle spielt hier die Frage, ob die Eingabe eine feste Größe hat, eine sukzessiv abzuarbeitende Folge ist (z.B. mehrere Wörter, die aufeinander folgen und einen Sinn ergeben sollen) oder eine noch komplexere dynamische Struktur hat, so dass sich das Netz Kontexte und vorherige Werte merken muss.

Deep feedforward network (DFF)[57]

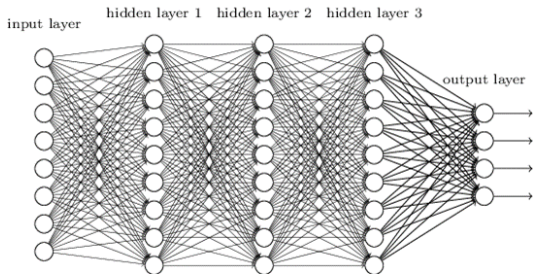


Abbildung 29: Deep feedforward network (DFF)[58]

Wie bei seinen Vorgängern, den klassischen KNN, werden bei einem tiefen Neuronalen Netz die Signale über Schichten von künstlichen Neuronen verarbeitet. Die Stärke der Signalübertragungen wird überwacht angelernt. Mehr Rechenkapazität und der Algorithmus der Fehlerrückführung erlauben heute Neuronale Netze mit bis zu Hunderten von Schichten. Dementsprechend steht das »tief« (Englisch: deep), im Gegensatz zum klassischen Neuronalen Netz, für eine hohe Anzahl an Schichten.

Einsatz:
Klassifikation oder Regression, wenn eine große Anzahl an Beispieldaten zu Verfügung steht.

[55] Mnih/Kavukcuoglu 2013
[56] Van Veen 2016
[57] Ivakhnenko 1965
[58] Hky 2017
[58] Geitgey 2016

Convolutional neural network (CNN)[59]

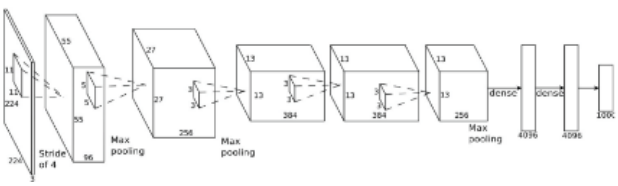


Abbildung 30: Convolutional neural network (CNN)[60]

Ein CNN ist ein tiefes Neuronales Netz mit mindestens einer Schicht, die eine mathematische Faltungsoperation durchführt. Wenn die Eingabedaten eine gewisse interne Struktur aufweisen (wie die zweidimensionale Struktur eines Bildes), kann ein Filter kleiner Größe über diese Struktur geschoben werden.

So können für die Aufgabe wichtige lokale Merkmale erkannt werden, egal, an welcher Position sie in den Eingaben vorkommen. Zum Beispiel können Ziffern an jeder Stelle auf dem Briefumschlag stehen. In der Gesichtserkennung können Augen erkannt werden, egal wo sich der Kopf innerhalb des Bildes befindet.

Einsatz: Daten mit einer für die Lernaufgabe wichtigen lokalen Struktur.

Recurrent network (RNN)[61]

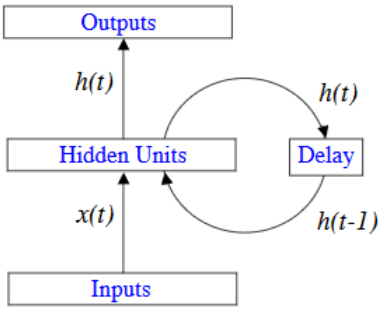


Abbildung 31: Recurrent network (RNN)[62]

Wenn eine Lernaufgabe auf Daten mit sequentieller Struktur vorliegt, verarbeiten RNNs die Eingaben nicht parallel und unabhängig voneinander, sondern nacheinander unter der Berücksichtigung der vorher gesehenen Eingaben. In dem Sinne besitzen RNNs so etwas wie ein Gedächtnis.

Einsatz:
Datenbeispiele haben eine sequentielle Struktur variabler, aber endlicher Länge (z.B. Wörter innerhalb eines Satzes)

[59] LeCun et al. 1998
[60] Smirnov/Timoshenko/Andrianov 2014
[61] Elman 1990
[62] Bullinaria 2015

Long short-term memory (LSTM)[63]

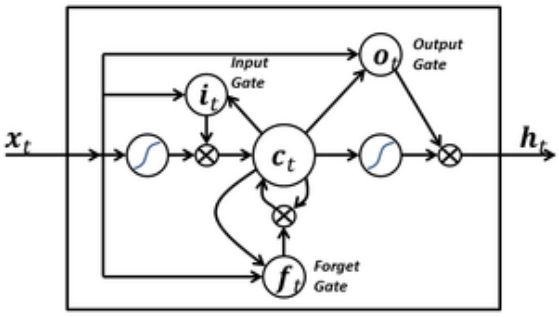


Abbildung 32: Long short-term memory (LSTM) [64]

LSTMs sind eine spezielle Art von RNN mit einem speziellen Modell für den Gedächtnisteil, das sowohl lokale als auch längerfristige Abhängigkeiten berücksichtigen kann.

Einsatz:
Datenbeispiele haben eine sequentielle Struktur variabler, aber endlicher Länge und es kommt vor allem auch auf weiter entfernte Abhängigkeiten in der sequentiellen Struktur an.

Autoencoder[65]

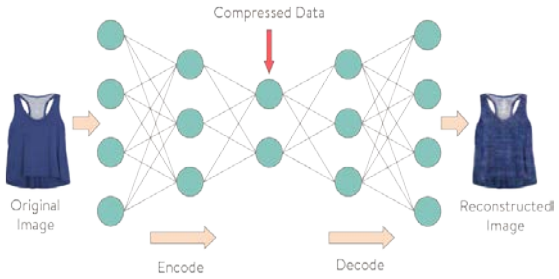


Abbildung 33: Autoencoder[66]

Ein unüberwachtes Lernverfahren zum Lernen einer komprimierten Repräsentation bestehend aus einem Neuronalen Netz mit mindestens drei Schichten. Die Schichten von Ein- und Ausgabe sind gleich groß, eine verdeckte, mittlere Schicht ist kleiner. Die Funktionen zwischen den Schichten werden so gelernt, dass die Ausgabeschicht die Eingabe bestmöglich reproduziert. In der kleinsten verdeckten Schicht müssen so die wichtigsten Merkmale auf reduzierte Weise kodiert werden.

Einsatz:
Zur Dimensionsreduktion, oft als Vorverarbeitung der Beispieldaten um deren Größe für die Hauptaufgabe zu reduzieren.

[63] Hochreiter/Schmidhuber 1997
[64] Wikipedia 2017e
[65] Boulard/Kamp 1988
[66] Torres 2015

Deep belief network (DBN) [67]

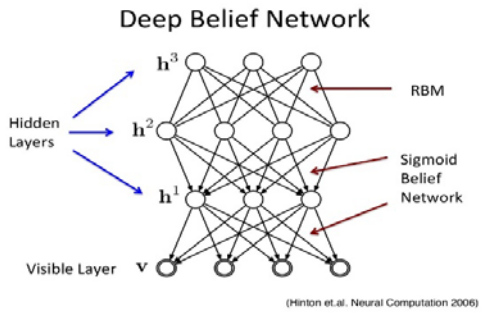


Abbildung 34: Deep belief network (DBN) [68]

Eine Komposition mehrerer unüberwachter Lernverfahren wie zum Beispiel Autoencoder. Hier werden Schicht für Schicht Repräsentationen der Daten gelernt, welche dann zur Generierung oder Vervollständigung von Beispielen dienen können.

Einsatz:
Lernen von Repräsentationen

Generative adversarial network (GAN)[69]

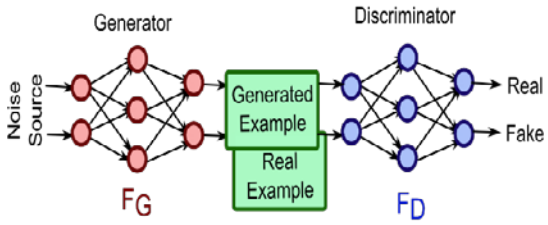


Abbildung 35: Generative adversarial network (GAN) [70]

Bei einem generativen gegnerischen Netz werden parallel zwei Neuronale Netze trainiert: ein Generator und ein Diskriminator. Der Generator versucht, neue Beispiele zu erzeugen. Der Diskriminator sieht sowohl die generierten Beispiele als auch reale Daten und lernt diese zu unterscheiden. Da der Generator Kenntnis über die Entscheidungsfindung des Diskriminators besitzt, werden die erzeugten Beispiele immer realitätsnaher.

Einsatz: Generierung von Beispielen.

[67] Bengio et al. 2007
[68] Grosse 2017
[69] Goodfellow et al. 2014
[70] Guttenberg 2017

1.7 Herausforderungen an die Güte und Qualität beim ML

Beim Maschinellen Lernen als datengestützter Technologie gibt es ganz andere Herausforderungen als bei der klassischen Programmierung. Generell gilt, je mehr Trainingsdaten ein Lernalgorithmus erhält, umso eher kann er sein Modell verbessern und die Fehlerquote verringern. Dabei besteht die Kunst darin, das Modell allgemein genug zu halten, damit es auch auf neuen Daten, die nicht in der Trainingsphase vorkamen, gut funktioniert. Außerdem sollen die Modelle robust sein, also auf ähnliche Eingaben auch ähnlich reagieren.

Die Qualität eines Modells hängt auch von der Qualität der Trainingsdaten ab. Werden dem Algorithmus zu viele falsche Beispiele gezeigt, kann er nicht die korrekten Antworten lernen. Wenn die Beispiele nicht repräsentativ sind, sind die Ausgaben bei neuartigen Eingaben auch mit größerer Unsicherheit behaftet. Manche Modelle können aber zusammen mit ihrer Ausgabe auch eine Einschätzung abliefern, wie fundiert die Ausgabe ist.

Dieser Abschnitt widmet sich wichtigen Herausforderungen des Maschinellen Lernens hinsichtlich Güte und Qualität. Weitere wichtige Herausforderungen, die die Forschung stark beeinflussen, werden in Kapitel 2 eingehend behandelt.

1.7.1 Qualität der Daten

»Garbage in – garbage out.« Selten trifft dieser Spruch so gut zu wie auf das Maschinelle Lernen. Die Qualität eines Modells hängt auch von der Qualität der Trainingsdaten ab. Werden dem Algorithmus zu viele falsche Beispiele gezeigt, kann er nicht die korrekten Antworten lernen.

Die Qualität der Daten versucht man vorab mit statistischen Methoden und Visualisierungen zu prüfen. Fehlen Werte, liegen die Werte im zulässigen Bereich, gibt es Ausreißer, verteilen sich die Daten erwartungsgemäß? Da man aber nicht alle Möglichkeiten antizipieren kann, können Qualitätsprobleme auf die Performanz durchschlagen.

Eine wenig repräsentative Datenlage (Englisch: Bias) führt zu falsch gelernten Modellen. Man nehme an, jemand will ein KNN trainieren, um auf Bildern Flugzeuge zu erkennen und nutzt als Trainingsdaten ausschließlich Bilder von Flugzeugen in der Luft. In diesem Fall kann es auch passieren, dass das gelernte Modell den blauen Hintergrund als relevantes Merkmal mit heranzieht. Wird das Modell nun auf neue Bilder von Flugzeugen im Hangar angesetzt, kann es sein, dass es gar nichts mehr erkennt, weil der passende Hintergrund fehlt. Neue Methoden erlauben es beispielsweise sogar in KNN diejenigen Bereiche aufzuzeigen, die für an der Generierung des Ergebnisses maßgeblich beteiligt waren, um solche Fehler wie mit dem blauen Hintergrund aufzudecken und zu vermeiden. Wenn die Beispiele nicht repräsentativ sind, sind die Ausgaben bei neuartigen Eingaben auch mit größerer Unsicherheit behaftet. Manche Modelle können aber zusammen mit ihrer Ausgabe auch eine Einschätzung abliefern, wie fundiert die Ausgabe ist.

Nicht repräsentative Beispieldaten führen schnell zu Verzerrungen und Fehlern in den Antworten. Selbst bei umfangreichen Tests können sie unbemerkt bleiben und im Einsatz

– oder noch schlimmer – überhaupt nicht erkannt werden. Darum ist es richtig, dass die neue europäische Datengrundverordnung Bürgern das Recht einräumt, sich die Algorithmen erklären zu lassen, die zu wichtigen, sie betreffenden Entscheidungen geführt haben.

Somit hängt die Performanz eines Modelles auch von den Menschen ab, die das Trainingsmaterial, die Labels und das Feedback beim ML bereitstellen. Was passieren kann, wenn schlechte oder einseitige Daten zur Verfügung stehen, zeigt sich in den medienwirksamen Beispielen. Ein Objekterkennungssystem von Google hat fälschlicherweise dunkelhäutige Personen als »Gorilla« gekennzeichnet. Das kann an der Auswahl der Trainingsdaten gelegen haben: z.B. überwiegend helle Menschen und dunkle Affen. Selbstständiges Weiterlernen im Einsatz ist auch nicht unproblematisch. So lernte ein Twitter Bot von Microsoft rassistische Ausdrucksweisen von seinen Gesprächspartnern im Internet und musste deshalb deaktiviert werden.

Im klassischen Data Mining schätzt man, dass 50 bis 70 Prozent des Arbeitsaufwands in die Datenvorverarbeitung fließt. Das beinhaltet nicht nur die Säuberung der Daten, sondern je nach Verfahren müssen abhängige Merkmale entfernt werden, Werte abstrahiert, transformiert und normiert werden. Zum Data Mining zählt auch die Exploration der Daten und die Suche nach auffälligen Mustern. Bei der Bild-, Sprach- und Textverarbeitung gab es vor dem tiefen Lernen ganze Vorverarbeitungsketten, um etwa Kontraste, Kanten und Umrisse zu erkennen, Laute und Worte zu separieren oder Texte syntaktisch zu analysieren.

1.7.2 Overfit, Underfit und Generalisierbarkeit

Die Generalisierung auf neue Beispiele kann in zwei Richtungen fehlschlagen: Das Modell kann an die gegebenen Beispieldaten überangepasst (overfit) sein oder es kann zu wenig angepasst sein (underfit). Ein ML-Modell ist überangepasst, wenn die Resultate auf den Trainingsdaten gut und auf den Testdaten schlecht sind. Es ist dann zu stark auf die gesehenen Beispiele abgestimmt und bezieht auch irrelevante Unterschiede oder gar statistisches Rauschen mit in die Entscheidung ein.

Ein überangepasstes Modell, das Flugzeuge identifizieren soll, könnte beispielsweise die einzelnen Logos der Betreibergesellschaften mit einbeziehen, wenngleich diese für die Geometrie von Flugzeugen unwichtig sind. Kommen neue Objekte der jeweiligen Klassen – insbesondere Flugzeuge ohne Logo – hinzu, würden diese Objekte auf einmal nicht mehr korrekt identifiziert, da das Flugzeug kein oder ein unbekanntes Logo besitzt.

Bei Anzeichen von Überanpassung kann man Modelle mit weniger Parametern betrachten oder das Lernverfahren so verändern, dass einfachere Modelle gegenüber komplexeren bevorzugt werden. Das nennt man Regularisierung. Häufig hilft gegen Überanpassung das Einbeziehen von mehr Beispieldaten – beispielsweise mehr Flugzeuge ohne Logo – sofern das möglich ist.

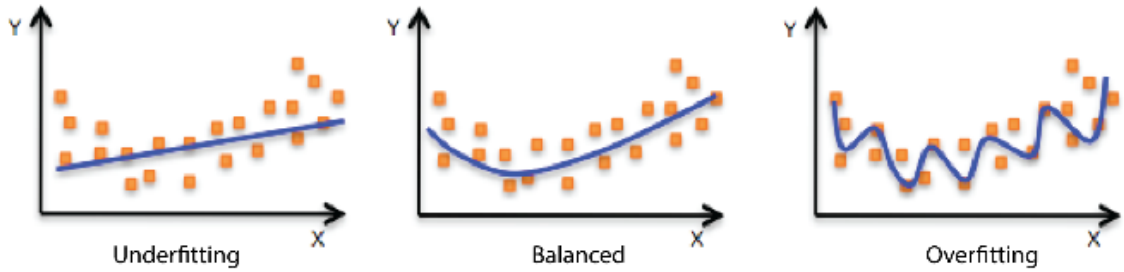
Mit ihrer Vielzahl an festzulegenden Gewichten neigen KNN zwar zu einer Überanpassung mit schlechter Generalisierung auf neue Beispiele, aber interessanterweise legen experimentelle Resultate nahe, dass sie aus noch unbekannten Gründen zu weniger Überanpassung neigen als es ihre Komplexität erwarten lässt.[71]

Bei einer Unteranpassung beantwortet ein Modell schon die Trainingsbeispiele nicht gut genug. Es hat zu wenig Ausdruckskraft, um relevante Unterschiede zu berücksichtigen. Wenn drei Fragen nötig sind, um eine vernünftige Entscheidung zu treffen, kommt ein Entscheidungsbaum der Tiefe zwei einfach nicht weit genug. Wird beispielsweise nur die Größe und Anzahl der Beine herangezogen, könnte das evtl. nicht ausreichen, um Stiere von Pferden zu unterscheiden. Die Hinzunahme eines weiteren Merkmals, z.B. das Vorhandensein von Hörnern, würde das Modell verbessern.

Bei tiefen Modellen ist die Gefahr eines Underfit weniger gegeben. Im Zweifel kann man das Modell einfach vergrößern.

Wenn die Performanz eines Modells auch nach Einführung von mehr Parametern niedrig ist, kann das an der Auswahl des Lernalgorithmus, der Modellklasse oder an der Wahl der Merkmale liegen. Außerdem kann die Qualität oder Quantität der Daten zu schlecht sein, um ein gutes Modell zu lernen.

Abbildung 36: Unteranpassung (links) und Überanpassung (rechts)[72]



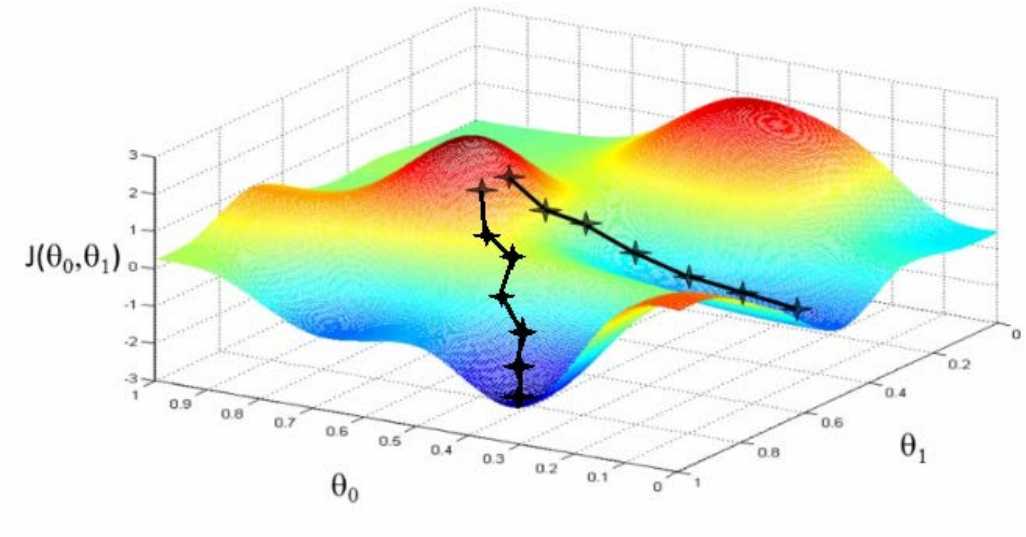
1.7.3 Performanz und Kostenfunktion

Um beurteilen zu können, wie gut ein Modell ist, benötigt man ein Maß. Solche »Performanzmaße« sind abhängig vom Lernstil und müssen an die konkrete Fragestellung angepasst werden. Beim überwachten Lernen etwa, wo man ja die richtigen Antworten für die Beispiele kennt, vergleichen die Performanzmaße den Anteil der richtigen und falschen Treffer miteinander.

[71] Zhang et al. 2016
[72] Amazon 2017b

Die meisten Lernverfahren nutzen Such- oder Optimierungsverfahren, die eine Kostenfunktion minimieren. Eine solche Kostenfunktion quantifiziert die Performanz auf den Beispieldaten und berücksichtigt natürlich das vereinbarte Performanzmaß. Man kann sich die Minimierung einer Kostenfunktion vorstellen wie die Suche nach der tiefsten Stelle in einer hügeligen Landschaft bei schlechter Sicht. Hierzu kann man von der jeweiligen Position die Gefälle in der unmittelbaren Umgebung messen und immer den Weg mit dem größten Gefälle einschlagen. Wählt man jedoch die Schritte beim Gehen zu groß, kann es sein, dass man ein Gebiet mit minimalen Werten der Kostenfunktion überspringt. Wählt man sie zu klein, kommt man nur sehr langsam voran. Es kann passieren, dass man glaubt, sich bereits auf dem tiefsten Punkt zu befinden und dabei einen noch tieferen Punkt in der ganzen Landschaft übersieht. In dieser Situation wäre man in einem lokalen Minimum gefangen. Die Wahl der Schrittlänge ist also bedeutend und verschiedene Methoden passen sie während der Suche nach dem Minimum adaptiv an.

Abbildung 37: Minimierung der Kostenfunktion[73]



Häufig ist die Kostenfunktion eine komplizierte Funktion, da sie von vielen Parametern des Modells abhängt. Es gibt unter Umständen viele lokale Minima, die innerhalb ihrer Umgebung die bestmögliche Parameterwahl für das Modell versprechen, global gesehen jedoch relativ schlecht abschneiden. So bleiben manche Verfahren in der Lernphase in lokalen Minima stecken und nähern sich dem globalen Minimum nicht weiter an. Man kann solche Situationen erkennen und teilweise umgehen, indem man unterschiedliche Startwerte für die Parameter bei gleichbleibenden Trainingsdaten verwendet. Die mathematische

[73] Amazon 2017a

Optimierung beschäftigt sich mit dieser Problematik, um lokale Minima in vertretbarer Zeit zu finden. Dort, wo es an passenden Methoden für das Maschinelle Lernen fehlt, ergeben sich neue Forschungsfragen für die Optimierung.[74]

Im Gegensatz zur Optimierung besteht im Maschinellen Lernen das Ziel aber nicht nur in der Minimierung der Kostenfunktion, sondern auch in einer guten Generalisierung auf ungesehene Beispiele. Genau diese gut generalisierende Performanz auf Ungesehenem macht das implizit gelernte »Wissen« beim Maschinellen Lernen aus. Sonst könnte die Maschine die endlichen Beispieldaten ja auch einfach »auswendig lernen«, d.h. abspeichern und abrufen. Also berücksichtigt die Kostenfunktion auch andere Kriterien, die zum Beispiel verhindern sollen, dass ein Modell zu komplex wird.

Um zu kontrollieren, wie gut ein Modell auf ungesehenen Daten funktioniert, teilt man die vorliegenden Beispieldaten auf. Der eine Anteil dient dem Training des Modells, der andere Anteil dient dem Testen des gelernten Modelles auf bis dahin ungesehenen Testdaten.

Außer den Modellparametern, die ein Lernalgorithmus zu optimieren versucht, kann es weitere Parameter geben, die der Data Scientist oder ML-Ingenieur vorgeben muss. Solche sogenannten »Hyperparameter« sind experimentell verstellbare Freiheitsgrade, wie die Tiefe eines Entscheidungsbaums oder die Anzahl und Breite der Schichten oder die Aktvierungsfunktionen der Knoten in einem KNN. Die Festlegung dieser Hyperparameter hat viel mit der Erfahrung und geschickten Experimenten des Data Scientist zu tun.

1.7.4 Robustheit

Wenig robuste Modelle führen bei kleinen Änderungen in der Eingabe zu großen Änderungen in der Antwort. Die Problematik der Robustheit hat für das tiefe Lernen eine gewisse Bekanntheit erlangt, seitdem demonstriert wurde, dass beispielsweise Straßenschilder durch gezielt aufgebrachte Aufkleber und Graffiti oder auf andere Weise manipuliert werden können, so dass selbstfahrende Autos sie falsch interpretieren. Dies kann dazu führen, dass zum Beispiel ein Stoppschild als Geschwindigkeitsbegrenzung identifiziert wird und durch die darauffolgende falsche Aktion Unfälle passieren können.[75] Ähnlich könnten auch Störgeräusche in sprachgesteuerte Software eingespielt werden, die von einem Menschen nicht als Befehl für die zuhörende Software erkannt werden, aber das Herunterladen von Malware oder eine andere schädliche Aktion des Programmes auslösen. Diese Möglichkeiten einer Beeinflussung sind ein großer Nachteil der Methoden des tiefen Lernens.

[74] Bennet/Parrado-Hernandez 2006
[75] Evtimov et al. 2017

Wenn ein Mensch für die Ergebnisse des Maschinellen Lernens die Verantwortung übernehmen muss oder ein Unternehmen für Schäden haften muss, ist es wichtig, dass sie dem Modell vertrauen. Robustheit ist auch ein wichtiges Anliegen bei der Anwendung maschineller Intelligenz.

Daher gibt es einen aktiven Forschungszweig im Maschinellen Lernen, der versucht, Verfahren für robustere Modelle zu finden. Erste Ansätze für tiefe Netze bestehen aus dem Training mit absichtlich gestörten Daten oder dem zufälligen Ausschalten einzelner künstlicher Neuronen. Die bereits in 1.6.2 erwähnten Adversarial Networks finden in diesem Kontext auch Anwendung.

KAPITEL 2

Anforderungen an die ML-Forschung und aktuelle Forschungsthemen

Dr. Miriam Leis | Fraunhofer ZV
Dr. Henning Petzka | Fraunhofer IAIS
Dr. Stefan Rüping | Fraunhofer IAIS
Dr. Angelika Voss | Fraunhofer IAIS

Inhaltsverzeichnis Kapitel 2

2 Anforderungen an die ML-Forschung und aktuelle Forschungsthemen 55

2.0 Einführung 55

2.1 Datenlage 57
2.1.1 Lernen mit sehr großen Datenmengen 58
2.1.1.1 Verteilte Algorithmen 59
2.1.1.2 Lernen auf Datenströmen 59
2.1.1.3 Lernen in Quantencomputern 60
2.1.2 Lernen mit wenig Daten 61
2.1.2.1 Automatisches Lernen von Labels 62
2.1.2.2 Lernen aus Simulationen 62
2.1.2.3 Lernen mit einem oder keinem Beispiel 63

2.2 Fähigkeiten 64
2.2.1 Anpassungsfähigkeit und Flexibilität 65
2.2.1.1 Multi-Task-Lernen 66
2.2.1.2 Transfer-Lernen 67
2.2.1.3 Lebenslanges Lernen 68
2.2.1.4 Interaktives Lernen in der Umwelt 69
2.2.2 Kollaboration 70
2.2.2.1 Interaktives Lernen vom Menschen (human in the loop) 70
2.2.2.2 Auto-ML 71
2.2.3 Lernen mit zusätzlichem Wissen 71
2.2.3.1 Lernen mit physikalischen Modellen 72
2.2.3.2 Lernen mit symbolischem Wissen 72
2.2.3.3 Lernen mit Wissensgraphen 73
2.2.3.4 Regellernen 75
2.2.3.5 Lernen mit Aufmerksamkeit und Gedächtnis 77
2.2.3.6 Selbstlebende Maschinen und die Verknüpfung von Sinnen 79

2.3 Akzeptanz, Sicherheit und Verlässlichkeit 80
2.3.1 Akzeptanz 81
2.3.1.1 Erklärbare Modelle 82
2.3.1.2 Fairness 83
2.3.2 Sicherheit und Verlässlichkeit 84
2.3.2.1 Safety by design für ML 84
2.3.2.2 Robuste Lernverfahren 85
2.3.2.3 Lernen unter Beschränkungen 85
2.3.2.4 Lernen von Kompetenzgrenzen 85

Tabellarische Zusammenfassung der offenen Forschungsfragen 87
Anhang A: Glossar: ML-Fachbegriffe 89
Index: Kapitel 1 und 2 93

2 Anforderungen an die ML-Forschung und aktuelle Forschungsthemen

2.0 Einführung

Im Folgenden werden Themen und Ansätze aus der Forschung zum Maschinellen Lernen zusammengestellt, die in den letzten Jahren auf den Workshops und Tutorials der beiden wichtigsten ML-Tagungen[76,77] in den Vordergrund getreten sind. Wichtig für die Forschungsförderung sind die Ziele, die damit verfolgt werden sollen. In Bezug auf das Maschinelle Lernen betreffen sie die Datenlage und Datennutzung, die Fähigkeiten oder Funktionalität der zukünftigen intelligenten Systeme, und ihre Akzeptanz durch Kunden und Nutzer.

Wir haben die Experten sowohl in unseren Interviews als auch auf dem Validierungsworkshop nach der Relevanz dieser Themen befragt, wobei sich eine recht hohe Übereinstimmung ergab. Die nächste Tabelle zeigt die resultierenden Rangfolgen. Die Felder mit der höchsten Relevanz (1) sind intensiv grün eingefärbt, die weniger relevanten heller. Die mit einem Stern (*) gekennzeichneten Themen wurden vom BMBF bereits zur Förderung ausgeschrieben. Dennoch empfahlen die Experten auf dem Validierungsworkshop, diese Themen auch weiterhin zu fördern. Ebenfalls als relevant erachtet wurden die Themen Sicherheit und Verlässlichkeit von ML-basierten Systemen, die jedoch eine Überschneidung mit den unter Nachvollziehbarkeit genannten Themen aufweisen.

Tabelle 2: Zukünftige Forschungsthemen und die Bewertung der Fachleute

Forschungsziele	**Forschungsansätze**	**Relevanz (1 = höchste Relevanz)**
Verbesserung der Akzeptanz		
• Nachvollziehbarkeit*	• Erklärbare KI • Erkennung von Diskrimination • Adversarial Training • Robustes Lernen	1
Ausbau der Fähigkeiten		
• Lernen mit zusätzlichem Wissen*	• Grey-Box-Modelle • Lernen mit symbolischem Wissen	2

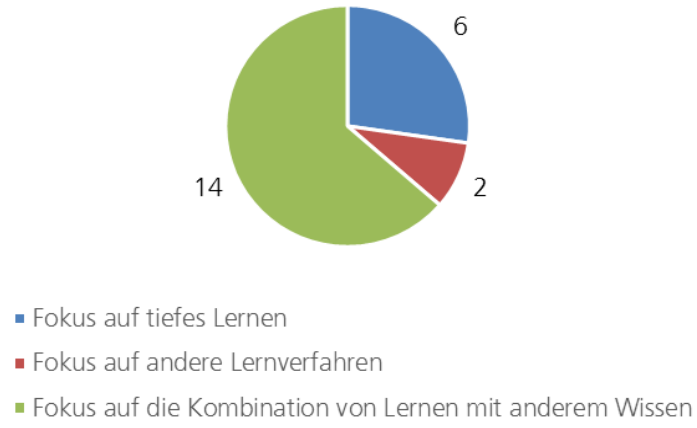
[76] NIPS 2018
[77] ICML 2018

Forschungsziele	Forschungsansätze	Relevanz (1 = höchste Relevanz)
• Kollaboration	• Interaktives Lernen vom Menschen • Metalernen (Auto-ML)	4
• Anpassungsfähigkeit und Flexibilität	• Multitask-Lernen • Transfer-Lernen • Lebenslanges Lernen • Multimodales Lernen	5
Datennutzung		
• Lernen mit wenigen Daten*	• Lernen aus Simulationen • One-Shot- und Zero-Shot-Lernen • Unüberwachtes Lernen von Labels	3
• Lernen mit sehr großen Datenmengen	• Verteilte Algorithmen • Lernen aus Datenströmen • Lernen in Quantencomputern	6

Am wichtigsten ist den Experten zufolge die Forschung zur besseren Nachvollziehbarkeit von ML-Anwendungen. Des Weiteren werden das Lernen mit zusätzlichem Wissen durch Einbeziehen von physikalischen, semantischen oder strukturellen Beschreibungen, sowie das Lernen mit wenigen Daten als wichtig angesehen. Dem Forschungsthema der Mensch-Maschine-Kollaboration wird eine steigende Bedeutung beigemessen, während der Aspekt der Anpassungsfähigkeit und Flexibilität als weniger wichtig erachtet wurde

Die gute Skalierbarkeit mit steigenden Datenmengen einerseits und die schlechte Nachvollziehbarkeit andererseits sind Gründe, weshalb die konsultierten Fachleute das tiefe Lernen für notwendig, aber nicht als ausreichend für erfolgreiche ML-Anwendungen halten. Die Wahl der Methode sollte sich immer nach den Anforderungen der Aufgabe richten. In Deutschland gibt es auch zukünftig viele Einsatzbereiche für klassische Lernverfahren, die weniger Daten benötigen, wie die hier stark vertretenen Stützvektormaschinen und Kernmethoden. Noch größeres Potenzial sehen die Fachleute aber in der Verbindung von maschinellen Lernverfahren mit anderen Wissensformen. Die nächste Graphik zeigt die Einschätzung der Experten auf dem Validierungsworkshop.

Abbildung 38: Anzahl der Expertenmeinungen zur ML-Technologie, auf die fokussiert werden sollte



Die Forschungsansätze aus Tabelle 2 werden im Folgenden vorgestellt. Manche können auch einen Beitrag zur Lösung mehrerer Herausforderungen leisten. Interaktives Lernen vom Menschen kann sowohl eingesetzt werden, um unzureichende Daten zu kompensieren, als auch zur Anpassung an eine veränderliche Umwelt. Multi-Task- und Transfer-Lernen beruhen auf der Wiederverwendung von Modellteilen und eignen sich deshalb sowohl zum Lernen mit wenigen Daten als auch zur Anpassung an eine veränderliche Umwelt. Ansätze zur Einbeziehung von zusätzlichem Wissen können dabei helfen, mit wenigen Daten zu lernen und auch die Nachvollziehbarkeit zu erhöhen. Lernen mit zusätzlichem Wissen spielt also eine zentrale Rolle, die durch die hohe Bewertung der Experten für diese Anforderung bestätigt wird.

2.1 Datenlage

Als datengetriebene Technologie stößt ML an Grenzen, wenn Daten in noch größeren Mengen noch schneller anfallen und zeitgleich daraus gelernt werden soll, oder wenn aus sehr kleinen Datenmengen verlässlich gelernt werden soll.

Insbesondere bei sensorischen Messungen in der vorausschauenden Maschinenüberwachung, bei Finanztransaktionen und anderen Prozessen fallen kontinuierlich extrem große Datenmengen an, die praktisch in Echtzeit ausgewertet werden müssen, ohne dass die Daten überhaupt gespeichert werden können. Hier bietet sich das sogenannte Online-Lernen an, das ohne die konventionelle Einteilung in Trainingsdaten für die Modellentwicklung und Testdaten für die Modellbewertung auskommt.

Ein Forschungsfeld, das wohl erst mittelfristig praktische Relevanz gewinnen wird, ist das Lernen in Quantencomputern, was die Möglichkeit für eine extrem hohe Parallelisierung

von Berechnungen verspricht. Dafür werden allerdings ganz neue ML-Algorithmen notwendig.

Gerade für Anwendungen, die in Deutschland von Interesse sind, liegt das Problem oftmals darin, dass nicht genügend, beziehungsweise nicht genügend brauchbare Daten verfügbar sind. In der Robotik, der industriellen Produktion, in medizinischen und sicherheitsrelevanten Anwendungen treten manche Ereignisse nur selten auf, weshalb sie in den Daten unterrepräsentiert sind. Auch kann es aus Gründen des Datenschutzes oder Urheberrechts zu Restriktionen bei der Datennutzung kommen.

In manchen Fällen existieren zwar Daten, aber keine Labels. Unüberwachte Lernverfahren könnten hier helfen und fehlende Labels automatisch erzeugen. Für autonome Agenten, Fahrzeuge und Roboter, die beim bestärkenden Lernen Feedback zu ihren Aktionen benötigen, steigt das Interesse am Lernen in Simulationen, da hier genug Trainingsdaten aller Art generiert werden können.

Performantes Lernen mit wenigen Trainingsdaten ist ein noch offenes Forschungsfeld, dem eine große Bedeutung zugesprochen wird. Idealerweise möchte man Maschinen dazu befähigen, anhand weniger Beispiele oder durch die Kombination bekannter Beispiele zu lernen.

2.1.1 Lernen mit sehr großen Datenmengen

Die Verfahren des Maschinellen Lernens sind zwar sehr leistungsfähig, aber sie benötigen zum Teil sehr viele Trainingsdaten. Das gilt insbesondere für die Verfahren des tiefen Lernens. Wenn nicht ausreichend Daten vorliegen, kann das Lernen der vielen Gewichte in einem tiefen Netz problematisch werden.

Um gut auf ungesehene Daten verallgemeinern zu können, müssen sie in der Nähe von gelernten Beispielen liegen. Wenn man nun kein Vorwissen über »leere« Gebiete im Eingaberaum hat, muss man versuchen, ihn ausreichend mit Beispielen abzudecken. Mit jedem neuen Merkmal vergrößert sich der Eingaberaum aber um eine Dimension. Der sogenannte »Fluch der Dimensionalität« im Maschinellen Lernen bezieht sich darauf, dass mit jedem neuen Merkmal die Anzahl der Beispiele exponentiell steigt, mit denen man den Eingaberaum gleichbleibend abdecken kann. Reichen 100 Beispiele für ein Zahlenmerkmal, so sind es schon 10.000 bei zwei solchen Merkmalen.[78]

Selbst wenn man Methoden zur Dimensionsreduktion vorschaltet, kann so die Menge an gegebenen Beispielen immer noch zu klein sein. Für viele Arten von Modellen haben die Beispieldaten, die vor der Big-Data-Welle verfügbar waren, einfach nicht ausgereicht. Aber selbst in Zeiten von Big Data kann es zu wenig Daten mit Labels geben und es kann

[78] Wikipedia 2017f

nötig sein, Daten aus verschiedenen Quellen zu kombinieren und in eine gemeinsame Form zu überführen. Ein Mehr an Daten bedeutet aber meist auch ein Mehr an Vorarbeit.

2.1.1.1 Verteilte Algorithmen

Wenn Datenmengen so groß sind, dass ihre Verarbeitung zeitlich aufwändig wird, hilft es, wenn Berechnungen auf mehreren Maschinen verteilt stattfinden können, ohne lange auf Resultate von anderen Rechnern warten zu müssen. Nur gelegentlich sollten die Ergebnisse wieder zusammengeführt werden. Eine lediglich sporadische Kommunikation verteilter Geräte kann nicht nur zeitlich begründet sein, sondern auch durch Ressourcenknappheit, wenn die Kommunikation aus entlegenen Gebieten etwa über Funk auf begrenzte Energiereserven zugreift.

Auch kann es vorkommen, dass aus rechtlichen oder Datenschutzgründen die Trainingsdaten auf den einzelnen Geräten bleiben sollen, die jeweiligen Daten aber für die Verbesserung eines Gesamtmodells genutzt werden sollen. Hier können verteilte Ansätze gute Lösungen bieten. Anonymisierte Datensätze können Label teils nur in Proportionen auf Untergruppen angeben, was wiederum Methoden abseits der Klassischen benötigt.[79]

Beim *Federated Learning* nutzen die einzelnen lokalen Systeme die ihnen jeweils vorliegenden (Roh)daten, um das bestehende Modell vorerst lokal zu verbessern. Nur die akkumulierten Veränderungen werden an einen Server gesendet, der die Änderungen aller Geräte synchronisiert und das Endresultat wieder an alle Systeme zurücksendet. So kann ohne Austausch der lokalen Rohdaten ein globales Modell verbessert werden.[80]

Aber nicht alle bekannten Algorithmen sind gut parallelisierbar, was ihre Anwendungsmöglichkeiten auf großen Datenmengen und auf verteilten Geräten stark einschränkt. Die Entwicklung neuer Ansätze in der algorithmischen Parallelität kann helfen, die statistischen Vorteile von »Big Data« weiter auszunutzen.[81]

2.1.1.2 Lernen auf Datenströmen

Hinter dem Begriff des *Online-Lernens*[82] stehen Algorithmen, die versuchen, in Realzeit aus Datenströmen zu lernen. Neue Daten fallen kontinuierlich an, und die Datenströme können viel zu groß sein, als dass man sie für eine spätere oder wiederholte Behandlung zwischenspeichern könnte. Häufig ist eine beinahe unmittelbare Reaktion des Systems auf die Eingabe gewünscht. In diesem Sinne gibt es beim Online-Lernen nicht die klassische Einteilung in Trainings- und Testdaten, sondern das Modell sollte für die eintreffenden Daten in kürzester Zeit eine Antwort möglichst hoher Qualität generieren.

[79] Stolpe/Morik 2011

[80] McMahan/Ramage 2017

[81] Xing et al. 2016

[82] Ng 2017

Einsatzbereiche für Online-Lernen sind bspw. die Verarbeitung von Sensordaten von Maschinen in der Produktion oder Assistenzsysteme für Finanztransaktionen. Neue Daten werden fortlaufend generiert, und das Modell soll sie nutzen, um Fehler frühzeitig immer besser vorherzusagen. Ein System, das für seine Vorhersagen mit der Zeit immer länger braucht oder erst nach Jahren im Einsatz gute Vorhersagen macht, ist allerdings nicht brauchbar.

Bei der Entwicklung gibt es einige wichtige Beschränkungen. So sollte die Verarbeitung der neuen Daten nicht länger dauern, wenn mit der Zeit immer mehr Daten in das Modell eingebracht werden. Bei begrenzt verfügbarem Speicher darf auch die Größe des gelernten Modells mit der Zeit nicht ansteigen. Im Idealfall soll das System schnell lernen, und nicht erst ab einer großen Menge von gesehenen Daten gute Vorhersagen machen oder gar erst, wenn die gesehenen Daten die Gesamtheit der möglichen Daten statistisch approximieren.

Somit spielt beim Online-Lernen eine geeignete Repräsentation der Daten eine wichtige Rolle. Sie hilft bei der Dimensionsreduktion und erlaubt, für die Aufgabe wichtige Varianzen in den Daten von unwichtigen zu unterscheiden oder möglichst unabhängige Merkmale zu finden.

Wenn sich die Datengrundlage mit der Zeit fundamental ändert, können Merkmale an Bedeutung verlieren, während andere Merkmale wichtiger werden. Die damit verbundene Schwierigkeit liegt in der algorithmischen Entscheidung, welche der gelernten Informationen vergessen werden sollten.[83] Auch hier ist man auf der Suche nach geeigneten Repräsentationen, die helfen, diese Entscheidungen zu treffen.

2.1.1.3 Lernen in Quantencomputern

Mit der Aussicht, in nicht allzu ferner Zukunft Quantencomputer einsetzen zu können, gibt es bereits seit über 20 Jahren einen aktiven Forschungszweig, der Methoden des Maschinellen Lernens unter der Verwendung von Quanteneffekten entwickelt. Quantencomputer ermöglichen eine hochparallelisierte Verarbeitung von Daten durch die gezielte Manipulation von Quantensystemen. Da die Quantenmechanik die Grenzen des Möglichen in der klassischen Physik verschiebt, verspricht das Quantenrechnen durch fundamentale Veränderungen von Algorithmen erhebliche Verbesserungen in der Performanz. Das gilt vor allem für Optimierungsprobleme, welche die Grundlage vieler maschineller Lernverfahren sind. Eine große, aber noch unvollständige Anzahl an stark beschleunigten klassischen Lernverfahren steht schon jetzt bereit für die Anwendung auf zukünftigen Quantencomputern. Die aktuelle Forschung beschäftigt sich vor allem mit einer Vervollständigung dieser Liste und der systematischen Umschreibung von Algorithmen.

[83] Gama 2012

Wann genügend Rechenpower in effizienter Weise auf Quantencomputern erreicht wird, ist im Moment noch sehr unklar. Bisher sind die hochentwickelten klassischen Methoden den bisher experimentellen Quantencomputern noch überlegen. Aber große Investitionen in die Forschung für die Verwirklichung dieser Technologie lassen erahnen, dass *Quanten-Lernen* in der Zukunft eine bedeutende Rolle im Bereich des ML spielen könnte.

2.1.2 Lernen mit wenig Daten

Wie mehrfach erwähnt, benötigen die meisten Verfahren des Maschinellen Lernens große Datenmengen. Um bei einer Klassifikation ein Objekt zu erkennen, muss dieses Objekt sehr viele Male in leichter Variation in den Beispieldaten vorkommen. Manchmal sind dafür wirklich nicht genügend Daten vorhanden. In der Robotik, in der personalisierten Medizin und in sicherheitsrelevanten Anwendungen treten manche Ereignisse nur selten auf, weshalb sie in den Daten unterrepräsentiert sind. Auch kann es aus Gründen des Datenschutzes oder Urheberrechts zu Restriktionen bei der Datennutzung kommen.

Im deutschen Kontext haben die Fachleute das Maschinelle Lernen mit wenigen Daten als besonders wichtig eingeschätzt, da hier in manchen Bereichen zu wenig brauchbare Daten vorliegen. Gleichzeitig wurde auf das sehr umfangreiche Hintergrund- und Expertenwissen vor allem in der deutschen Industrie hingewiesen, welches für das Lernen nutzbar gemacht werden sollte. Deshalb wird daran gearbeitet, bereits vorhandenes Expertenwissen und Naturgesetze in ML-Verfahren zu integrieren, um im Vorhinein sinnvolle Einschränkungen festzusetzen und das Trainieren der Maschine effizienter und effektiver zu gestalten.

Mit einem unüberwachten Verfahren kann eine Repräsentation vorgelernt werden, so dass überwacht nur noch ein kleineres Netz mit weniger Parametern trainiert werden muss. Oder man kann Schichten wiederverwenden und so die auf ähnlichen Eingabedaten bereits gelernten Repräsentationen nutzen. Aber auch wenn in beiden Fällen wenige Daten mit Label für die Lernaufgabe benötigt werden, muss man zumindest sehr viele Beispieldaten ohne Zielvorgabe haben.

Wenn es nicht möglich ist, mehr Daten der gewünschten Form zu beschaffen, kann man versuchen, andere Daten aus verwandten Aufgaben zu nutzen. Selbst wenn viele eigene Daten vorliegen, kann man den eigenen Trainingsaufwand verringern und die Resultate verbessern, wenn man auf bestehenden Modellen aufsetzt, in die bereits viele Daten eingeflossen sind. Ein System kann auch schrittweise an komplexere Aufgaben herangeführt werden, zum Beispiel von der Erkennung einzelner Objekte in Bildern zur Gesamtinterpretation einer Szene oder von der Vorhersage des nächsten Wortes zum Verstehen der Äußerung insgesamt.

Idealerweise möchte man Maschinen dazu befähigen, anhand weniger Beispiele oder durch die Kombination bekannter Beispiele zu lernen. Diese Fähigkeit in ML-Anwendungen zu implementieren wird als »One-shot« oder gar »Zero-shot«-Lernen bezeichnet.

2.1.2.1 Automatisches Lernen von Labels

Beim überwachten Lernen sind Trainingsdaten mit Label notwendig, was in der Regel mit einem erheblichen manuellen Aufwand im Vorfeld verbunden ist. Daher werden Ansätze benötigt, die mit weniger Daten auskommen. In manchen Fällen existieren zwar Daten, aber keine Labels. Unüberwachte Lernverfahren könnten hier zum Lernen von Labels und geeigneten Repräsentationen verwendet werden.

Wenn die Label nicht zuverlässig gelernt werden können, wird auf der anderen Seite auch geforscht, wie mit verrauschten Daten effizient gelernt werden kann. Mit solchen Methoden zur Hand kann man zum Beispiel einfache, grobe Regeln aufstellen, die in vielen Fällen das richtige Label vorhersagen. Dabei müssen die Regeln nicht für alle Beispiele gelten oder konsistent sein, sondern es genügt, wenn ihre Kombination zu einem ausreichend verlässlichen Label führt. Das ist übrigens eine erste Art, Hintergrundwissen einzubringen.

2.1.2.2 Lernen aus Simulationen

Für autonome Agenten, Fahrzeuge und Roboter, die beim bestärkenden Lernen Feedback zu ihren Aktionen benötigen, steigt das Interesse am Lernen in Simulationen, da hier genug Trainingsdaten aller Art generiert werden können. In einer Simulationsumgebung sammelt das lernende System in kontrollierter Umgebung Trainingsdaten. Ein System für selbstfahrende Autos kann in einer Simulation viele Verkehrssituationen durchspielen und aus seinen Fehlern lernen.

Beim bestärkenden Lernen hilft es, wenn eine Umgebung mit positivem und negativem Feedback simuliert wird, in der das lernende Modell über viele Versuche einen signifikanten Anteil der möglichen Schritte durchprobieren kann, um das positive Feedback zu maximieren. Bei zwei handelnden Agenten kann das Modell aus vielen Partien gegen sich selbst lernen, wie AlphaGo, das sich das Go-Spiel durch Spielen gegen sich selbst beigebracht hat.

Simulationen können aber nie die Komplexität der Realität wiederspiegeln. In der Anwendung treten dann neue Probleme auf. Lösungsansätze, die die Lücke zwischen Simulation und Anwendung zu schließen versuchen, versprechen effektiveres Lernen im Simulationslabor.[84]

Das Ziel des Lernverfahrens ist entweder das Lernen einer Bewertung von möglichen Zuständen, anhand derer eine gute Strategie abgeleitet wird, oder das direkte Lernen der Handlungsstrategie.[85] Welche Algorithmen einen performanten Übergang von der Simulation zum Einsatz schaffen, ist Bestandteil der aktuellen Forschung.

[84] Bousmalis et al. 2017

[85] Arulkumaran et al. 2017

Die modernen Methoden des tiefen Lernens mit ihrem automatisierten Lernen von Repräsentationen sind besonders gut darin, Fehler in der Simulation auszunutzen, um ihre Aufgabe zu bewältigen. Wenn beispielsweise in einer zu einfach aufgebauten Simulation für das automatisierte Fahren in jeder Linkskurve ein bestimmtes Haus an der Straßenseite steht, wird der Algorithmus dieses Haus als Merkmal zur Voraussage von Lenkbewegungen ausnutzen. Es ist eine große Herausforderung, so etwas zu erkennen oder zu vermeiden.

2.1.2.3 Lernen mit einem oder keinem Beispiel

Menschen können aufgrund ihrer Erfahrung schnell neue Begriffe und Konzepte lernen. Ein einziges oder wenige Beispiele reichen, um etwa zu lernen, wie ein Olinguito aussieht, um es anschließend auf anderen, ungesehenen Bildern zu erkennen. Diese Lernfähigkeit des Menschen würde man gerne auf maschinelle Lernverfahren übertragen.

Wie kann also ein Modell erweitert werden, wenn eine neue Klassifikationsklasse hinzugefügt werden soll? Braucht man auch für die neue Klasse wieder viele Beispiele und muss das System das neue Modell ganz neu lernen? Wenn ein einziges Beispiel (oder sehr wenige) der neuen Klasse reichen soll, spricht man von One-shot Learning[86].

Erste Ansätze lernen einfachere Strukturen, aus denen komplexere Objekte zusammengesetzt werden. Bei der Einführung einer neuen Klasse muss anschließend nur die Zusammensetzung der bekannten Strukturen trainiert werden. Dafür reichen ein oder wenige Beispiele.[87]

Menschen können auch Aufgaben bewältigen, die sie in der Form zuvor noch nicht gesehen haben, wenn sie sich aus bekannten Teilen zusammensetzen. Ein Mensch kann sich bspw. ein Olinguito vorstellen, wenn er erfährt, dass das Tier wie eine Mischung aus Teddybärkopf, Wildkatzenkörper mit kurzen Beinen und langem Schwanz aussieht. Die Ansätze, diese Fähigkeit in Modellen des Maschinellen Lernens zu ermöglichen, nennt man Zero-shot Learning[88]. Ein Modell wird hier nur auf verwandten Teilaufgaben trainiert.

[86] Rezende et al. 2016
[87] Lake/Salakhutdinov/Tenenbaum 2015
[88] Xian et al. 2017

Abbildung 39: Olinguito[89]

Die Schwierigkeit beim One-shot Learning liegt in der Repräsentation von Konzepten. Wenn ein Mensch gelernt hat, Kartoffeln zu schälen, dann kann er auch ohne weiteres eine Gurke schälen. Hierbei kommt ihm natürlich seine Abstraktionsfähigkeit zugute: er weiß, dass Schälen generell das Abtragen einer anders gearteten oberen Schicht bedeutet, und welche Werkzeuge hierfür sinnvoll eingesetzt werden können. Mit Modellen aus maschinellen Lernverfahren sind solche Aufgaben der Generalisierung nicht einfach lösbar. Hier experimentieren aktuelle Ansätze zu tiefen Neuronalen Netze mit Aufmerksamkeits- und Gedächtnismechanismen, wobei die Aufmerksamkeit in Form von gewichteten Verbindungen zu intern abgespeicherten Repräsentationen den Zugriff auf die Gedächtnisinhalte steuert.

2.2 Fähigkeiten

Es besteht das Interesse, ML-basierte und "kognitive" Systeme weiterhin zu verbessern und noch effektiver, effizienter und "intelligenter" zu gestalten. Heutige ML-basierte Systeme besitzen noch nicht die Fähigkeit, wirklich zuverlässig in Situationen zurechtzukommen, die außerhalb ihrer recht begrenzten Trainingskonditionen liegen. In der Realität ist jedoch kaum ein Einsatzgebiet statisch: Die Umwelt ändert sich kontinuierlich, es kommen neue Daten hinzu und ihre statistische Verteilung ändert sich. Damit verändern oder erweitern sich auch die Aufgaben des Systems, weshalb sich das gelernte Modell flexibel auf die neue Situation einstellen sollte. Systeme, die nichts dazulernen oder alte Fehler immer wiederholen, sind wenig nutzbringend.

In diesem Kontext existiert ein weiterer Block an Forschungsfragen und Herausforderungen, der sich damit befasst, die Fähigkeiten zukünftiger "kognitiver Systeme" so auszuweiten, dass sie eine bessere Anpassungsfähigkeit und Flexibilität in neuen Situationen aufweisen, ohne erst lange und aufwendig "umtrainiert" zu werden, was auch ihren effektiven Einsatz in der realen Welt erhöhen würde. Idealerweise könnte ein zukünftiges ML-basiertes System beispielsweise bereits existierendes Wissen zur Lösung einer neuen Aufgabe nutzen können, oder anhand weniger Beispiele bzw. durch Vormachen schnell eine neue Fähigkeit erwerben. Heutige KNN tendieren jedoch noch dazu, bereits Gelern-

[89] N-tv 2013

tes wieder zu vergessen. Diesem sogenannten katastrophalen Vergessen versuchen neue Forschungsansätze entgegen zu wirken.

In vielen Anwendungsbereichen nutzen Fachleute eigene Formalisierungen und mathematische Gleichungen, um typische Eigenschaften oder das Verhalten der untersuchten Systeme möglichst genau zu beschreiben. Solche Verfahren sind jedoch sehr aufwendig in ihrer Erstellung und können meist auch nicht vollständig beschrieben werden. Hier können datengetriebene Verfahren die Lücken im Modell schließen. Aber auch das Einbringen von Expertenwissen kann rein datengetriebene Modelle verständlicher machen und ihre Akzeptanz erhöhen.

Ein weiteres großes Ziel für zukünftige intelligente Anwendungen ist eine möglichst flexible Kollaboration zwischen Mensch und Maschine. Hier soll die Maschine aus der Interaktion mit dem Menschen dazulernen und nicht immer wieder die gleichen Fehler machen. Bei unsicheren Antworten könnte sie vom Experten eine Bestätigung oder die Angabe eines Labels anfordern.

Eine andere Art der Zusammenarbeit liegt in der Unterstützung von Datenwissenschaftlern. Der Workflow von einer initialen Fragestellung bis zum gut trainierten Modell beinhaltet eine Reihe von Arbeitsschritten, die mehrfach durchlaufen werden müssen, bis die gewünschte Qualität erreicht ist. »Metalernen« oder »Auto-ML« entwickelt Algorithmen, die lernen, selber ML-Modelle zu trainieren, wodurch der Aufwand für Data Scientists reduziert wird.

2.2.1 Anpassungsfähigkeit und Flexibilität

Eng verbunden mit dem Lernen aus wenigen Daten ist das Ziel, trainierte Modelle leichter an neue Kontexte anzupassen. Ein intelligenter Roboter im Servicebereich sollte je nach Bedarf in unterschiedlichen Räumen assistieren, ohne dass er für jeden Wechsel aufwändig neu trainiert oder umprogrammiert werden muss. ML-Systeme kommen heute jedoch noch nicht wirklich zuverlässig in Situationen zurecht, die außerhalb ihrer Trainingskonditionen liegen. Die oftmals erstaunlichen Leistungen, die sie im Labor demonstrieren, funktionieren unter Realbedingungen oft erheblich schlechter.

ML-Systeme werden flexibler, wenn vergangene Informationen im Modell abgespeichert und bei Bedarf zielgerichtet wieder abgerufen werden können. Beim Multi-Task-Lernen wird versucht, verschiedene Aufgaben gleichzeitig auf einer gemeinsamen Repräsentation zu lernen, die dadurch tragfähiger wird. Das Transfer-Lernen zielt darauf ab, in einem Neuronalen Netz Teilbereiche zu finden, die bereits für eine bestimmte Aufgabe trainiert worden sind, und sie für eine nächste, verwandte Aufgabe wieder zu verwenden. Noch weiter geht die Forderung nach Maschinen, die »lebenslang lernen«, also ihr Modell fortlaufend an Veränderungen anpassen. Bei starken Veränderungen tendieren Künstliche

Neuronale Netze noch dazu, bereits Gelerntes wieder zu vergessen. Diesem »katastrophalen Vergessen« versuchen erste Ansätze entgegen zu wirken. Zum Beispiel versuchen sie, das Neusetzen von Gewichten in »erinnerungswürdigen« Teilen des Netzes zu verlangsamen[90].

Menschen nutzen zur Situationseinschätzung meist mehrere Informationskanäle gleichzeitig: auditive, visuelle und haptische. Ähnliches soll auch im ML gelingen, wobei die Herausforderung vor allem in der semantischen Verknüpfung dieser unterschiedlichen Signale liegt. Ein verwandtes Forschungsfeld in der Robotik ist das sogenannte Embodiment, das sich mit der Frage befasst, inwieweit eine sensorische »Körpererfahrung« künstlichen Systemen beim Lernen helfen kann. Hierfür können aber auch die bereits erwähnten Simulationen eingesetzt werden.

2.2.1.1 Multi-Task-Lernen

Beim *Multi-Task-Lernen* lernt man verwandte Aufgaben von vornherein gleichzeitig aus einer Gesamtmenge von Trainingsdaten[91]. Zum Beispiel kann man verschiedene Wirtschaftsindikatoren, Symptome für verschiedene Krankheiten oder aktive chemische Komponenten für neue Medikamente gleichzeitig lernen, anstatt dass Vorhersagen für jeden Indikator, jedes Symptom oder jede chemische Komponente einzeln gelernt werden. Das funktioniert gut, wenn für jede Aufgabe etwas andere Aspekte der Repräsentation wichtig sind, so dass wechselseitig eine Überanpassung vermieden wird. Gleichzeitig helfen Gemeinsamkeiten in der Repräsentation, so dass automatisch Abhängigkeiten der Aufgaben in der gelernten Repräsentation entdeckt und genutzt werden. So verbessert man auch die Bildinterpretation für autonom fahrende Autos, wenn man sie gleichzeitig trainiert, verschiedene Eigenschaften einer Straße zu erkennen, oder man verbessert die Gesichtserkennung, indem man gleichzeitig das Erkennen einzelner Gesichtspartien trainiert.
Eine Herausforderung sind mögliche negative Effekte, wenn unpassende Lernaufgaben einbezogen werden, die dem Lernen der anderen Aufgaben eher schaden als nutzen. Negative Auswirkungen zu erkennen und zu minimieren ist Gegenstand aktueller Forschung.

Die meisten Ansätze des Multi-Task-Lernens beziehen sich auf überwachte Verfahren. Sowohl unüberwachte Verfahren als auch das bestärkendes Lernen könnten vom parallelen Lernen verschiedener Einzelaufgaben profitieren. Hierzu gibt es bislang noch relativ wenige Ansätze.

[90] Kirkpatrick et al. 2017

[91] Zhang/Yang 2017

2.2.1.2 Transfer-Lernen

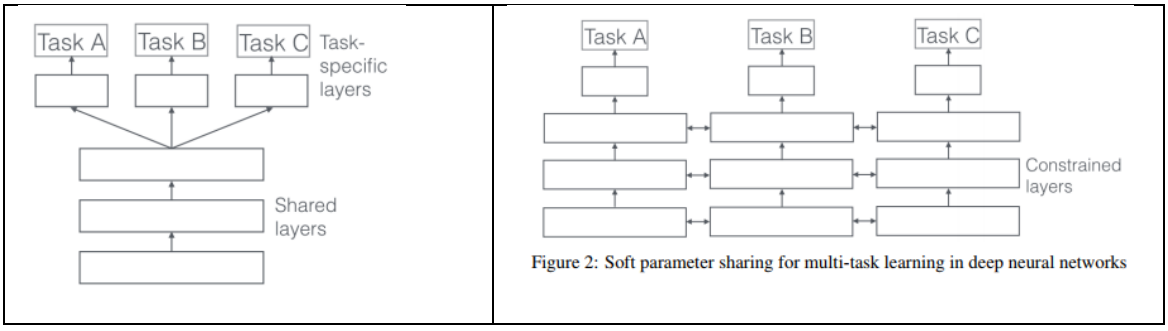
Beim *Transfer-Lernen* geht es um die Frage, wie man gelerntes Wissen aus einer Aufgabe nutzen kann, wenn sich die Aufgabenstellung oder das Anwendungsgebiet oder die Datengrundlage leicht verändern[92] [93]. Im Gegensatz zum Multi-Task-Lernen sind die Aufgaben nicht gleichberechtigt, sondern in eine Hilfs- und eine Zielaufgabe unterteilt.

Eine Herangehensweise, speziell bei Neuronalen Netzen, besteht in der *Wiederverwendung von Teilen früherer Netze,* die bereits erfolgreich eine ähnliche Aufgabe gelöst haben, samt ihren Gewichten und anderen Parametern. So können implizit große Mengen an bereits verarbeiteten Beispieldaten genutzt werden. Diese Wiederverwendung von Netzen geschieht bislang noch nicht systematisch.

Konkret könnte man sich ein System vorstellen, das bereits gelernt hat, Tauben zu erkennen, um anschließend auch Krähen zu identifizieren. Wahrscheinlich wird das nicht direkt funktionieren, da das gelernte Modell nicht von den gelernten taubenspezifischen Eigenschaften generell auf Vögel zu abstrahieren vermag, um sich dann dem Erkennen von Krähen anzupassen. Jedoch ist es wohl möglich, Teile wiederzuverwenden, um schon mal ohne weiteres Training vogelförmige Umrisse, Flügel und Schnäbel zu erkennen. Wie das generalisiert funktionieren kann, ist noch eine offene, aber aktuelle Forschungsfrage.

Die nächste Abbildung veranschaulicht zwei Ansätze zur Wiederverwendung von Netzen. Links das Prinzip eines Netzes, bei dem eine gemeinsame Repräsentation für verschiedene Aufgaben wiederverwendet wird, rechts ein Netz mit verschiedenen Repräsentationen, wo zwischengeschaltete Knoten die Relevanz verschiedener Repräsentationen für unterschiedliche Tasks lernen sollen.

Abbildung 40: Ansätze zur Wiederverwendung von tiefen Netzen[94]



[92] Pan/Yang 2010
[93] Ruder 2017b
[94] Ruder 2017

Das Hauptkriterium zur Wiederverwendung von Gewichten ist die Performanz auf verwandten Aufgaben. Bei einem Neuronalen Netz, das eine neue Aufgabe auf Bildern lösen soll, benutzt man etwa die ersten Schichten eines anderen Netzes zur Bildverarbeitung, das mit vielen Daten trainiert wurde, und baut darauf weitere Schichten auf. Weitergehend kann man ein riesiges Netz bauen, das unterschiedliche Aufgaben über unterschiedliche Pfade löst[95]. Bei einer neuen Lernaufgabe wird dieses riesige Netz nach Teilnetzen durchsucht, die wiederbenutzt werden können. Dann müssen nur noch wenige Gewichte angepasst werden. Um passende Teilnetze zu finden, muss man versuchen, die Ähnlichkeit von Aufgaben und die Beziehungen zwischen ihnen zu bestimmen. Dazu gibt es verschiedene Ansätze, darunter auch die Nutzung von Bayesschen Netzen und das Lernen von Aufgabenhierarchien. Die Neuroforschung an Fruchtfliegen hat sogar neue Erkenntnisse geliefert, wie auf effiziente Weise Ähnlichkeiten zwischen Strukturen, Objekten, Situationen etc. identifiziert werden können, die evtl. auch zukünftige ML-Algorithmen inspirieren könnten.[96,97] Tatsächlich muss hierzu aber weiter geforscht werden.

Viele existierende Modelle des Maschinellen Lernens sind nicht modular aus Teilkomponenten zusammengesetzt, die jeweils eine bestimmte Aufgabe lösen. Um eine Wiederverwendung von bereits gelernten Modellen zu ermöglichen, wäre solch eine Modularität jedoch sehr hilfreich. Andersherum kann man untersuchen, wie man es mit maschinellen Lernmethoden erreichen kann, dass aus einer Palette gegebener Module mit klar definierter Aufgabe automatisch ausgewählt und geeignet zu einem großen Ganzen verbunden werden können.

Beim Transfer-Lernen wird eine Ähnlichkeit der involvierten Lernaufgaben Lerndaten vorausgesetzt. So müssen die jeweiligen Beispiele auf ähnliche Weise erhoben werden, und sie sollten ähnlichen Grundannahmen folgen. Andernfalls kann es passieren, dass die Zielaufgabe mit Transfer-Lernen schlechter gelöst wird als ohne. Es ist eine große Herausforderung, solch einen negativen Transfer vorherzusehen und zu verhindern. Hier wird an Metriken geforscht, um die Ähnlichkeiten passend zu charakterisieren.

2.2.1.3 Lebenslanges Lernen

Beim *lebenslangen Lernen* lernt das System im Einsatz stets weiter und passt sein Modell fortlaufend an[98]. Wenn sich die Umgebung langsam aber stetig ändert, soll sich das trainierte Modell auf die Veränderungen einstellen können. So wird ein vortrainiertes Modell beispielweise am Kunden oder Patienten personalisiert werden.

[95] Fernando et al. 2017
[96] Salk Institute 2017
[97] Dasgupta 2017
[98] Silver/Yang/Li 2013

Für das lebenslange Lernen existieren schon gut funktionierende Ansätze, wenn die Veränderungen aus kleineren Anpassungen bestehen. Problematisch wird es, wenn sich die Aufgabenstellung zu stark verändert, Ein großes Problem, das als »catastrophic forgetting« bezeichnet wird, besteht darin, dass Modelle, die sich einer neuen Aufgabe anpassen, in der Regel ihre alten Fähigkeiten vergessen[99]. Obwohl es hier bereits erste Lösungsansätze gibt, handelt es sich noch um eine offene Forschungsfrage, wie man dieses Problem beheben könnte.

2.2.1.4 Interaktives Lernen in der Umwelt

Beim Einsatz in der realen Welt muss ein Modell mit ungesehenen Situationen umgehen können. Die beeindruckenden Resultate des bestärkenden Lernens, bei dem Handlungen aus Feedbacks der Umwelt gelernt werden, spielen sich bisher nur in begrenzten Szenarien ab. Die reale Welt ist in vielen Fällen zu komplex, als dass sich die möglichen Zustände und Aktionsmöglichkeiten durch das Sammeln von Daten abdecken lassen. Beim Go-Spiel beispielsweise war die Anzahl der Spielmöglichkeiten zu groß für die klassischen Methoden. Dennoch ist es in solch einer Spielumgebung relativ einfach, viele Trainingsbeispiele zu erzeugen. Noch viel Forschungsbedarf besteht in der Übertragung dieser Lösungen auf komplexere Umgebungen und Aufgaben. Damit verbunden ist die Problematik, allgemeine Evaluationsumgebungen zu schaffen, die die Komplexität der realen Welt widerspiegeln und in denen verschiedene Modelle verglichen werden können.

Beim bestärkenden Lernen erlauben die Methoden des tiefen Lernens den Übergang von einfachen Paaren aus Zustand und Aktion hin zu Handlungsentscheidungen auf Basis von komplexem visuellem Input (wie die Pixel bei Atari-Spielen). Dies erlaubt gute Repräsentationen der Eingabe, um auf leicht abweichende Situationen generalisieren zu können. Dennoch bleibt die Anzahl an Kombinationsmöglichkeiten von Handlungen ein großes Problem. Sie ist zu groß, als dass alle Möglichkeiten ausgetestet und verglichen werden könnten. Das führt zur »Exploration-Exploitation« Problematik. Dabei müssen Strategien entwickelt werden, wann die bisher erfolgreichsten Handlungen wiederverwendet, wann bisher scheinbar suboptimale Handlungen wiederversucht, und wann risikofreudig ganz neue Wege ausprobiert werden sollten.[100]

Einige Ansätze, die an anderer Stelle erwähnt werden, helfen auch beim interaktiven Lernen aus der Umwelt. So erlauben auch das bereits erwähnte Transfer-Lernen und modulares Lernen, bereits Gelerntes wiederzuverwenden. Hintergrundwissen und physikalische Modelle helfen, das Lernen zu beschleunigen und die Anzahl an Aktion-Feedback-Schritten zu verringern, indem »vorausgedacht« und die Konsequenzen einer Handlung simuliert werden.

[99] Kirkpatrick et al. 2017
[100] Audibert/Munos 2011

Des Weiteren werden Umgebungen mit mehr als einem handelnden Agenten betrachtet. Die Agenten handeln dann als Gegner oder auch in Kooperation miteinander. Dabei werden Kommunikationskanäle benutzt, über die die einzelnen Agenten lernen, sich für die Bewältigung der Aufgaben abzustimmen.

2.2.2 Kollaboration

Ein großes Ziel für zukünftige intelligente Anwendungen ist eine sichere, effiziente, intuitive und möglichst flexible Kollaboration zwischen Mensch und Maschine. Indem der Mensch interaktiv mit einbezogen wird, trifft das ML-System nicht ausschließlich statistische Entscheidungen. ML-Anwendungen, die sich auf ihren Benutzer einstellen, wurden von den interviewten Fachleuten hoch bewertet. Auf dem Validierungsworkshop wurde ihnen trotz einer mittleren Priorität ein hohes Potenzial zugesprochen.

2.2.2.1 Interaktives Lernen vom Menschen (human in the loop)

Der Workflow von einer Fragestellung bis zum gut trainierten Modell, den Datenwissenschaftler durchlaufen, ist oft lang und beinhaltet Iterationen auf vielen Stufen. Soll das System im Einsatz weiter lernen, muss wieder der Data Scientist eingeschaltet werden. Das versuchen Methoden des interaktiven Lernens zu vermeiden. Intelligente Systeme sollen direkt durch Interaktion mit dem Nutzer lernen. Einfachste Beispiele sind Empfehlungssysteme, die bspw. direkt aus den Eingaben der Endnutzer lernen.
Mensch und Maschine sollen flexibel zum gegenseitigen Vorteil zusammenarbeiten. Einerseits soll die Maschine aus der Interaktion mit dem Menschen dazulernen und nicht immer wieder die gleichen Fehler machen. Bei unsicheren Antworten könnte sie vom Experten eine Bestätigung oder die Angabe eines Labels anfordern. Andererseits soll die Maschine ihre Aktionen und Antworten für den Menschen nachvollziehbar gestalten.

Eine andere Art der Zusammenarbeit ist die Unterstützung für Datenwissenschaftler. Der Workflow von einer Fragestellung bis zum gut trainierten Modell beinhaltet Iterationen auf vielen Stufen. »Metalernen« oder »Auto-ML« entwickeln Algorithmen, die lernen, selber ML-Modelle zu trainieren und dadurch den Aufwand für den Data Scientist zu reduzieren.

Wenn ein ML-System lernt, eigene Schwachpunkte zu identifizieren und gezielt in sinnvoller Weise fehlende Beispiele vom Nutzer oder gar von vielen Experten in einer Crowd zu erhalten, spricht man von aktivem Lernen bzw. vom maschinellen Lernen mit Crowd Sourcing. Weitere Ansätze zum interaktiven ML sind Relevanz-Feedback, Debugging von Modellen und »socially guided machine learning«.

Personalisierte Anwendungen verspricht das Imitationslernen, bei dem das Lernverfahren eine Handlung seines Lehrers imitieren soll. Der Mensch macht eine Aktion vor und wenige Iterationen des Vormachens sollen reichen, damit ein Roboter die Schrittfolgen nachahmen kann. Eine Lösung dieser Aufgabe verspräche ganz neue kundenorientierte Anwendungen des Maschinellen Lernens.

2.2.2.2 Auto-ML

Um ML schneller in die Breite zu tragen, wäre es vorteilhaft, den Data Scientist soweit wie möglich zu entlasten. »Auto-ML« oder »Metalernen« versucht, den Entwicklungsaufwand durch Teilautomatisierung in der Modellentwicklung zu verringern.[101] Ein Beispiel sind ML-Systeme zum Trainieren von ML-Systemen.

Die geeignete Wahl von Hyperparametern hat einen enormen Effekt auf die Güte der gelernten Modelle. Welche Werte hier gute Resultate liefern, hängt von Modell, Aufgabenstellung und Daten ab und muss in jeder neuen Anwendung neu justiert werden. Ausgiebige Erfahrung hilft dem Data Scientist. Es bleibt dennoch ein zeitaufwändiges Unterfangen, das über Suchstrategien automatisch ablaufen könnte. In Ansätzen reichen dann oft Vorbestimmungen für Suchintervalle und eine Historie der bisherigen Resultate verschiedener Parameterwahlen, um die nächsten geeigneten Parameterkandidaten und Werte auszuwählen. Auto-ML geht dabei aber weit über die Wahl geeigneter Hyperparameter hinaus. Auch Teile der Modellierung können variiert werden und verschiedene Verfahren ausgewählt werden. Hier gibt es Überschneidungen mit dem Transfer-Lernen, wenn passende Teilkomponenten zu einem großen Modell zusammengesetzt werden (siehe Abschnitt 2.2.1.2).

Erfolgskritisch ist die Entwicklung guter Suchstrategien nach geeigneten Modellen. Neben Hill Climbing (dem umgekehrten Vorgang des iterativen Optimierungsverfahren aus Abschnitt 1.7.3) angewendet auf den Hyperparameterraum, oder genetischen Algorithmen gibt es beispielsweise *Racing*-Algorithmen[102], bei denen mehrere Modellwahlen parallel getestet werden und schlechtere frühzeitig herausgenommen werden, um die Rechenkapazitäten auf die besseren zu fokussieren. Es können verschiedene Kriterien bestimmt werden, um den Modellen jeweils einen Rang zuzuordnen und anschließend über die gemittelten Ränge die besseren von den schlechteren Modellen zu unterscheiden.

Alternativ gibt es den Ansatz, rechnerisch günstigere Lernverfahren zu benutzen, deren Ausgang die Resultate der gewünschten Lernverfahren approximieren. So können schnell Hyperparameter oder Modellstrukturen getestet werden, bevor die eigentliche Methode zum Einsatz kommt.

2.2.3 Lernen mit zusätzlichem Wissen

In vielen Fällen ist die Welt zu komplex für eine genaue physikalische Modellierung. In diesen Fällen sind statistische Methoden überlegen. Andererseits lassen sich rein statistische Lösungen schwer kontrollieren. Zudem sind sie abhängig von großen Mengen an Beispieldaten. Hier wäre also eine Symbiose ideal, um fehlende Daten durch Expertenwis-

[101] Hutter et al. 2017
[102] Maron/Moore 1997

sen auszugleichen, um Einschränkungen der möglichen Lösungen verlässlich vorzugeben und auch, um Lösungen verständlicher zu machen.

Ein anderes hybrides Prinzip nutzt symbolisches Wissen über Objekte und ihre Beziehungen, das in Wissensgraphen formalisiert wird. Je vollständiger ein Wissensgraph ist, umso komplexere Zusammenhänge lassen sich darin verfolgen und umso anspruchsvollere Fragen beantworten. Dabei steigt allerdings auch der Erstellungsaufwand drastisch. Erste gute Lernverfahren versuchen, Wissensgraphen automatisch zu vervollständigen, indem sie Relationen zwischen Knoten vorhersagen, Knoten identifizieren, die dasselbe Objekt beschreiben, und einem Knoten eine Klasse zuordnen.

In bestehenden Ansätzen laufen Lernvorgang und Schlussfolgerungen aus gelernten Repräsentationen getrennt voneinander ab. Idealerweise sollten diese beiden Schritte stärker verzahnt sein. Dafür werden Modelle benötigt, die einerseits das Einbauen von Hintergrundwissen erlauben und andererseits statistisch lernbar sind.

Die Einbeziehung von Wissensgraphen, logischen Regeln und White-Box-Elementen in das Training eines ML-Modells kann auch dazu beitragen, die ML-Anwendung insgesamt effizienter, zuverlässiger und sicherer zu gestalten, da von Anfang an Domänen-, Erfahrungs- oder Expertenwissen integriert wird. Wissen in die Methoden des ML zu integrieren ist auch ein Ansatz, um mit wenigen Daten besser zurecht zu kommen. Wo Wissen eingebracht wird, können Daten zur statistischen Beschreibung eingespart werden.

2.2.3.1 Lernen mit physikalischen Modellen

Zum Verständnis, zur Überwachung und Steuerung von physischen Systemen nutzen Ingenieure oft Modelle mit komplexen Differentialgleichungen, die die physikalischen Eigenschaften des Systems beschreiben. Solche Modelle bezeichnen sie als White-Box-Modelle, weil sie das Systemverhalten kausal beschreiben. Im Gegensatz dazu bezeichnen sie die rein statistischen datengetriebenen Modelle des Maschinellen Lernens, insbesondere das Deep Learning, als Black-Box-Modelle, da sie die interne Funktionsweise des Systems ignorieren. Wann immer Prozesse zu komplex sind, um sie physikalisch exakt zu beschreiben, ist das maschinelle Lernen jedoch eine gute Alternative, um die Modellierung aus statistischen Beobachtungen zu lernen.

Um beide Welten miteinander zu vereinbaren, behelfen sich die Ingenieure mit »Grey-Box-Modellen«. Hier werden auf Basis von Expertenwissen physikalische Modelle erstellt, deren noch unvollständige Modellparameter und Fehlerterme mit maschinellen Lernverfahren aus beobachteten Daten gelernt werden

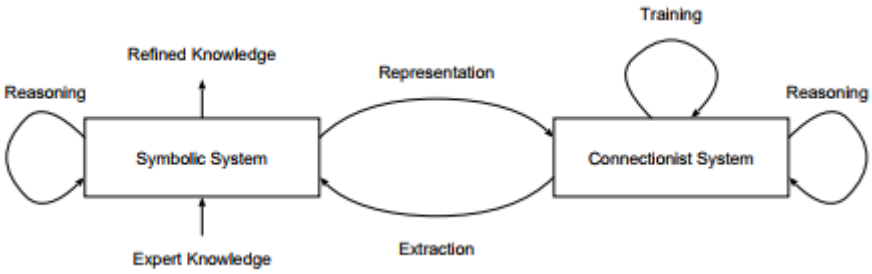
2.2.3.2 Lernen mit symbolischem Wissen

Maschinelle Lernverfahren haben sich als statistische Mittel etabliert, weil sie gut mit verrauschten Rohdaten umgehen können. Die tiefen Modelle in der Bild- und Sprachverarbeitung sind besonders erfolgreich, wenn genügend Beispiele für das Training vorhanden sind. Sie funktionieren aber nur für sehr begrenzte Aufgabenbereiche und sind schwer nachvollziehbar. Genau umgekehrt im Hinblick auf die Nachvollziehbarkeit verhält es sich

mit symbolischem Wissen, also Regeln und Fakten, die sich aus Naturgesetzen ergeben, die von Experten geliefert werden, oder die aus fachlichen Textquellen extrahiert werden können. Eine sprachbasierte Wissensrepräsentation hat zudem den Vorteil, dass sie immer neue Konzepte für komplexere Sachverhalte und neue Konzepte bilden kann und deshalb eine sehr vielseitige Ausdrucksfähigkeit besitzt. Wie jedoch bereits in Abschnitt 1.3.4 beschrieben, stoßen symbolische Modellierungen auch recht schnell an ihre Grenzen, da die zu bewältigenden Aufgaben extrem komplex werden können und laufend Erweiterungen und Sonderfälle hinzukommen. Hier hat ja gerade das Maschinelle Lernen seine Stärken.

Es liegt also nahe, Maschinelles Lernen mit symbolischem Wissen zu kombinieren. In so einem hybriden System könnten symbolisch ausgedrückte Konzepte und Regeln genutzt werden, um das Lernen zu steuern, zu unterstützen, einzugrenzen und zu kontrollieren.[103] Umgekehrt kann neu Erlerntes aus dem Modell extrahiert und in das vorhandene symbolische Wissen integriert werden. Ein solcher Transfer würde nicht nur das vorhandene Wissen ergänzen, sondern auch das trainierte Modell für die Nutzer nachvollziehbarer machen. Ein gutes Beispiel hierfür ist die Ergänzung von Wissensgraphen.

Abbildung 41: Integration von Lernen und Wissen Bildquelle[104]



2.2.3.3 Lernen mit Wissensgraphen

Graphen aus Knotenpunkten und Kantenverbindungen sind eine bewährte Art, strukturierte Informationen zu repräsentieren. Beispiele sind soziale Netzwerke, bei denen Kanten geschlossenen Freundschaften entsprechen, biologische Korrelationsnetzwerke, bei denen die Knoten Proteine oder Gene kodieren und Automatemodelle aus der Informatik, bei denen die Knoten Zuständen des Systems und die Kanten den Übergängen entsprechen. Die Knoten und Kanten können mit einem symbolischen oder numerischen Label versehen sein. Bei den Wissensgraphen in Abschnitt 1.3.4 waren sowohl Knoten als auch Kanten mit einer Bedeutung versehen, bei Netzwerken von Proteinen können Zahlen an den Kanten aus einem Korrelationskoeffizienten bestehen, der die gemeinsame Anreicherung der beteiligten Proteine beschreibt.

[103] Shavlik 1994
[104] Bader, Hitzler 2014

Große Graphen können unterteilt werden[105]. Beim Lernen kann jedes Beispiel aus einem kleinen Graphen bestehen und zum überwachten Lernen mit einem Label versehen sein, das man voraussagen will. Beispielsweise beschreibt jeder kleine Graph ein Molekül und die Aufgabe ist es, eine bestimmte biochemische Aktivität des Moleküls vorherzusagen. Das könnte die Fähigkeit des Moleküls sein, die Vermehrung des HI-Virus in menschlichen Zellen zu verlangsamen. Grundannahme ist hier, wie bei jeglicher passender Repräsentation, dass Moleküle, die durch strukturell ähnliche Graphen beschrieben werden, auch ähnliche Eigenschaften besitzen. Die Forschung beschäftigt sich damit, effizient berechenbare Ähnlichkeitsmaße oder Merkmalsdarstellungen für solche Graphen zu definieren. Einfache Wahlen wie die Anzahl der Knoten sind jedoch nicht sehr aussagekräftig, und es werden komplexere Merkmale benötigt[106] [107].

Mit Kernel-Methoden kann man über eine Ähnlichkeitsfunktion implizit die Merkmale bestimmen.[108] Es ist vielversprechend, Ähnlichkeit an strukturellen Überlappungen von Teilgraphen der Molekülrepräsentation festzumachen. Alle Moleküle jedoch untereinander nach beliebig großen strukturell gleichen Teilen zu durchsuchen ist rechnerisch viel zu aufwändig. Hier muss noch viel nach Algorithmen gesucht werden, die auf theoretischen Überlegungen basieren und so standardisiert implementiert werden können, dass sie über das Anwendungsbeispiel der chemischen Moleküle hinaus zum Lernen auf kleinen Graphen geeignet sind.

Wissensgraphen beinhalten im Allgemeinen zu den Objekten auch Eigenschaften mit numerischen Daten und Text, wie das Alter einer Person oder ihre Adresse. Es ist eine Herausforderung für Lernverfahren, mit diesen verschiedenartigen Datentypen umzugehen.[109] Oftmals liegen Daten zudem unvollständig vor, wenn zum Beispiel nicht für alle Personen eine Adresse eingetragen wurde. Da die klassischen Lernverfahren vorbestimmte Merkmale benötigen, werden auch hier neue Ansätze benötigt.

Eine weitere Herausforderung besteht darin, dass dieselbe Information unterschiedlich dargestellt sein kann. Zum Beispiel kann direkt formuliert werden, dass »Person A Großmutter von C« ist. Andererseits kann dieselbe Information auch über die zwei Relationen »A ist Mutter von B« und »B ist Elternteil von C« beschrieben werden. Wenn Beispiele aus verschiedenen Quellen zusammengefügt werden, müssen die Lernverfahren flexibel mit diesen Unterschieden in der Darstellung umgehen können.

[105] Von Luxburg 2007
[106] Deshpande et al. 2005
[107] Horvath et al. 2004
[108] Gärtner et al. 2002
[109] Wilcke et al. 2017

Momentan wird stark an Lernverfahren geforscht, die nicht nur neue Relationen vorhersagen können, sondern auch ihre Entscheidungsgrundlage verraten[110]. Das bedeutet das Lernen von Regeln, die sich aus mehreren Relationen zusammensetzen. Auch der Nutzen von Wissensgraphen in der Erkennung von Fake News ist ein aktueller Forschungsbereich[111].

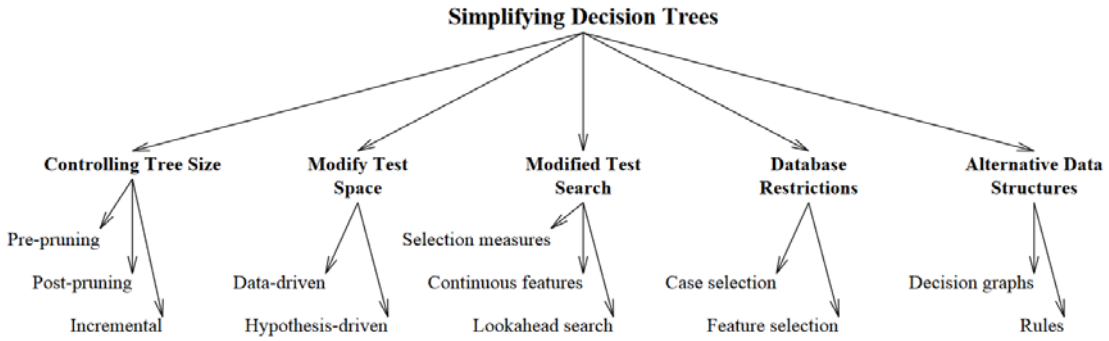
2.2.3.4 Regellernen

Regeln repräsentieren logische Wenn-Dann-Formeln. Sie können statistische Auffälligkeiten beschreiben, die herausgefiltert werden, oder sie klassifizieren Objekte. Meistens liegen die Daten dafür in Tabellen, die Objekte und ihre Eigenschaften und Relationen beschreiben. Es kann eine große Tabelle geben oder viele kleine Tabellen für jeweils eine bestimmte Relation zwischen den Objekten. Wenn beispielsweise die Relation »ist Elternteil von« zwischen zwei Personen vorliegt, und den Personen ein Geschlecht zugeordnet wird, möchte man die Regel lernen, dass »ist Elternteil von« gemeinsam mit »ist männlich« der Relation »ist Vater von« entspricht. Diese Regel soll aus den vorliegenden Beispielen extrahiert werden. Auch auf Wissensgraphen werden Regeln gelernt.

An Regeln kann man über Entscheidungsbäume gelangen[112] (siehe Abschnitt 1.5.2). Folgt man den Entscheidungen in den Knoten bis in die Blätter, ergeben sich Regeln. Da tiefe Entscheidungsbäume sehr detaillierte Beschreibungen liefern, die durch ihre Genauigkeit schlecht auf kleine, statistische Abweichungen generalisieren (siehe Überanpassung in Abschnitt 1.7.2), werden die Bäume für allgemeinere Regeln abgeschnitten[113]. Neben dem einfachen Stutzen der Bäume, gibt es aufwändigere Methoden, die schon bei der Generierung manche Beispiele strategisch auslassen, so dass die resultierenden Regeln nicht unbedingt die Gesamtheit richtig beschreiben. Abbildung 42 zeigt eine Unterteilung in verschiedene Strategien für die Kontrolle der Komplexität von Regeln aus Entscheidungsbäumen.

[110] Fan et al. 2017
[111] Shu 2017
[112] Quinlan 1986
[113] Fürnkranz 1999

Abbildung 42: Übersicht zu Methoden zum Verkleinern von Entscheidungsbäumen[114]



Alternativ können Regeln auch ohne den Einsatz von Entscheidungsbäumen gelernt werden. Beispielsweise startet man mit einer leeren Liste von Regeln und fügt iterativ neue Bedingungen hinzu. Dabei können verschiedenste Kriterien den Ausschlag geben, welche Bedingungen bevorzugt werden. Hat man auch negative Beispiele, werden diese anschließend ausgesondert und separat betrachtet. Andersherum kann mit sehr spezifischen Regeln gestartet werden, die in Folge generalisiert werden. Bei der *Subgruppensuche* generiert man Regeln für Teilpopulationen, deren Labels in besonderen Bereichen liegen.

Allgemein gilt, dass Regellernen relativ schlecht skaliert. Dabei spielt auch die Komplexität der Ausdrucksmöglichkeiten in den Regeln eine Rolle. Ist das Ziel eine einfache Regel oder eine Liste von Regeln? Lässt man numerische Werte zu? Darf ein Objekt über mehrere Werte derselben Eigenschaft beschrieben werden (zum Beispiel kann eine Person mehrere Kinder haben)? Sind Funktionen erlaubt? Je ausdrucksstärker die Regeln, desto aufwändiger werden die Berechnungen. Außerdem sind passende Datensätze sehr aufwändig zu beschaffen. Große Datensätze, die aus kollaborativen Crowd-Sourcing-Projekten entstehen, enthalten oft fehlerhafte oder inkonsistente Daten.

Die Stärke des Regellernens ist die Interpretierbarkeit der Lösungen. Die gelernten Regeln sind leicht zu verstehen und nachzuvollziehen. Daher bedarf es Weiterentwicklungen, die skalierbar sind und die Beschreibungssprache, die Suchstrategie und den Ansatz zur Vermeidung von Überanpassung passend wählen.

Um symbolisches Wissen mit statistischen Methoden zu vereinbaren, forscht man an Ansätzen zu abgeschwächten Regeln. Sogenannte schwache Regeln (Englisch: fuzzy logic) beschreiben symbolisches Wissen mit Unsicherheitswerten. So wird die harte Unterteilung von Regeln in fließende, kontinuierliche Werte abgeschwächt, die auf natürliche Weise in statistische Modelle integriert werden können[115].

[114] Breslow/Aha 2006
[115] Ishibuchi, Yamamoto 2005

2.2.3.5 Lernen mit Aufmerksamkeit und Gedächtnis

Klassische Lernverfahren bestimmen eine Funktion vom Eingaberaum in den Ausgaberaum anhand von gelernten Parametern. Ihnen fällt es aber schwer, algorithmische Strukturen zu kodieren. Das liegt nicht an der Schwierigkeit des Algorithmus an sich, da die Lösungen leicht iterativ zu programmieren wären.

Theoretisch wurde gezeigt[116], dass rekursive Neuronale Netze (RNN, siehe Abschnitt 1.6.4) Turing-vollständig sind. Das bedeutet, dass man mit ihnen theoretisch jede von einem Computer durchführbare Berechnung simulieren kann. In der Praxis wurden Limitationen festgestellt. So kann zum Beispiel ein RNN, dessen Struktur schon auf das Bearbeiten von Folgen spezialisiert ist, sehr erfolgreich trainiert werden, um Folgen von maximal 20 Zahlen zu wiederholen oder aufsteigend zu sortieren. Wenn im Anwendungsfall jedoch deutlich längere Sequenzen auftreten, dann versagt es komplett. Statt ein Netz mit Folgen in allen möglichen Längen zu trainieren, sollte es besser an kleinen Folgen die prinzipielle Vorgehensweise lernen.

Die klassischen Lernverfahren können die Suche nach relevanter Information und ihre Verwendung nicht getrennt voneinander behandeln. Sie organisieren Information nicht in einem Zwischenspeicher, um darauf Berechnungsschritte durchzuführen. Eine Art Zwischenspeicher würde auch das Abspeichern von Wissen in Form von komplexen, strukturierten Daten wie zum Beispiel Graphen ermöglichen[117].

Inspiriert von der Turing-Maschine als Abstraktion eines Computers wurde ein tiefes Netz mit der Bezeichnung »Neuronale Turing-Maschine«[118] entwickelt. Hier simuliert ein tiefes Netz einen Prozessor oder eine Steuereinheit mit Lese- und Schreibzugang zu einer Art Speicher. Die Neuronale Turing-Maschine kann bereits besser als vorherige Methoden einfache Algorithmen lernen und zeigt beeindruckende Ergebnisse, wenn aus einem Text einfache Folgerungen gezogen werden sollen.

[116] Siegelmann/Sontag 1995
[117] Graves et al. 2016
[118] Graves/Wayne/Danihelka 2014

Abbildung 43: Teil des bAbI Datensatzes[119]

```
1 Mary moved to the bathroom.
2 John went to the hallway.
3 Where is Mary?	bathroom	1
4 Daniel went back to the hallway.
5 Sandra moved to the garden.
6 Where is Daniel?	hallway	4
7 John moved to the office.
8 Sandra journeyed to the bathroom.
9 Where is Daniel?	hallway	4
10 Mary moved to the hallway.
11 Daniel travelled to the office.
12 Where is Daniel?	office	11
13 John went back to the garden.
14 John moved to the bedroom.
15 Where is Sandra?	bathroom	8
1 Sandra travelled to the office.
2 Sandra went to the bathroom.
3 Where is Sandra?	bathroom	2
```

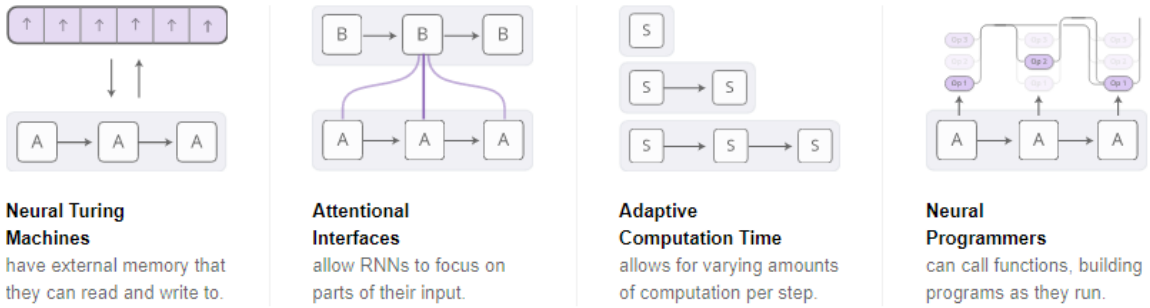
Schlussfolgerungen werden hier aus abgespeichertem Wissen abgeleitet, anstatt wie bisher bei Neuronalen Netzen nur aus günstigen Repräsentationen. Das erlaubt das automatische Lernen von iterativen Lösungsschritten.

Zielführende Berechnungsschritte zu lernen, bedeutet ähnlich wie beim bestärkenden Lernen, Aktionen auszuprobieren und ihre Resultate zu interpretieren. Diese Aktionen sind aber etwas sehr Diskretes, was sich nicht mit den statistischen Lernmethoden verträgt. Darum schwächt man die diskreten 0-oder-1-Werte zu einem schwachen Konfidenzwert im Intervall [0,1] ab und kann dann im Fall von Neuronalen Netzen den erfolgreichen Rückpropagierungsalgorithmus anwenden. Der Konfidenzwert zwischen Null und Eins bestimmt, wie die Aufmerksamkeit des Netzes auf die verschiedenen Aktionen verteilt ist, und ist höher, wenn eine Aktion für das Lösen der Aufgabe relevanter scheint. Solche Aufmerksamkeitswerte können auch benutzt werden, um selektiv Teile des internen Spei-

[119] Facebook Research 2016

chers auszulesen. Das hat den nützlichen Nebeneffekt, dass die Entscheidungsgrundlage des Modells in Teilen nachvollzogen werden kann (siehe Abschnitt 2.3.1.1).

Abbildung 44: Netzarchitekturen zum Lernen von Algorithmen[120]



Da zu jeder Zeit jeder Aktion ein Aufmerksamkeitswert zugewiesen wird, ist dieser Ansatz rechnerisch teuer und widerspricht der Idee eines effizienten Verfahrens, das durch seine Aufmerksamkeit Ressourcen spart[121]. Die Entwicklung eines ressourcen-günstigeren Aufmerksamkeitsansatzes wäre ein großer Sprung in diesem aktuellen Forschungsbereich, der Lösung algorithmischer Aufgaben mit Hilfe von Neuronalen Netzen.

2.2.3.6 Selbstlebende Maschinen und die Verknüpfung von Sinnen

Die steigende Popularität von Produkten mit Sprachassistenten wie Google Now, Siri, Cortana oder Alexa bestätigen den fortgeschrittenen Stand der Sprachverarbeitung. Allerdings werden Wörter als Audiosignale bzw. Buchstabenfolgen aufgefasst und nach statistischen Zusammenhängen verarbeitet, ohne dass ein tiefergehendes Verständnis erlangt wird.

Ein wichtiger Aspekt des menschlichen Verstehens und Handelns ist die Verbindung von Sprache, Sehleistung und Aktion, die man auch auf Maschinen übertragen möchte. Beim *multimodalen Lernen* werden Eingabesignale von unterschiedlichen Quellen (wie beispielsweise Audio- und Bildsignale) für die Bewältigung einer Aufgabe herangezogen. Darüber hinaus sollen die verschiedenen Eingabesignale in Zusammenhang gebracht und darauf aufbauend eine passende Aktion abgeleitet werden. Hier geht es um mehr als die gleichzeitige Verarbeitung verschiedener Signale, sondern um die semantische Verknüpfung der Signale in einer einheitlichen Repräsentation.

Dazu sind mindestens zwei Fähigkeiten unabdingbar. Erstens müssen Wörter mit Erfahrungen zusammengebracht werden, die das System selber gewinnen muss. Zweitens müssen nicht nur die einzelnen Wörter, sondern die gesamte Äußerung verstanden wer-

[120] Olah/Carter 2016
[121] Britz 2016

den. Beides zusammen würde zu einer internen maschinellen Repräsentation führen, die Sprache mit Erfahrungen verknüpft. OpenAI forscht in dieser Richtung und meldet erste Erfolge[122].

2.3 Akzeptanz, Sicherheit und Verlässlichkeit

Eine wünschenswerte und wichtige Eigenschaft ist die Nachvollziehbarkeit der Modelle im Allgemeinen und ihrer Ergebnisse im Einzelfall. Entscheidungsbäume lassen sich besonders gut von Nutzern interpretieren, tiefe Neuronale Netze hingegen schlecht. In Fällen, wo die Gesetzgebung verlangt, dass automatisierte Entscheidungen begründet werden und die Betroffenen beispielsweise nicht aufgrund ihres Geschlechts oder ihrer Religion benachteiligen dürfen, sind Nachvollziehbarkeit, Fairness und Sicherheit einer ML-Anwendung essenziell und erhöhen außerdem die Akzeptanz.

Die Nachvollziehbarkeit von ML-Anwendungen ist aus Sicht der Fachleute das wichtigste Forschungsziel. Für diese Thematik hat sich der Begriff »Explainable AI« (XAI oder erklärbare KI) etabliert. Es wird dabei zwischen Transparenz und Erklärbarkeit unterschieden. Transparenz bedeutet, dass das Verhalten der Anwendung vollständig nachvollziehbar ist. Praktisch ist diese Forderung jedoch nur schwer erfüllbar, da viele Modelle notwendigerweise sehr komplex sind. Erklärbarkeit hingegen bedeutet, dass für eine konkrete Einzelentscheidung der Anwendung die wesentlichen Einflussfaktoren aufgezeigt werden können, was im Sinne der EU-Datenschutzverordnung ausreichend ist. Technisch ist sie zudem deutlich einfacher zu erfüllen als Transparenz. Eine wissenschaftliche Herausforderung besteht darin, die Eingabedaten zu Konzepten zu abstrahieren, die für den Menschen sinnvoll sind. Andere prototypische Ansätze liefern repräsentative Einzelfälle zur Erklärung von Entscheidungen.

Neben der Nachvollziehbarkeit und Transparenz geht es auch darum, KI-Systeme hinreichend robust zu gestalten, so dass sie im Alltagseinsatz keine unakzeptablen Risiken und Schäden verursachen. Wobei das Problem der »KI-Sicherheit« nicht, wie oftmals in den Medien verbreitet, in der Intelligenz von KI-Systemen liegt, sondern eher am Gegenteil - in falsch gelernten Modellen, fehlerhaftem Design des Systems, dem Nicht-Vorhandensein von Kontextwissen in den Systemen oder schlechten Trainingsdaten.

Das Ergebnis einer ML-Anwendung kann als unethisch und diskriminierend erachtet werden, wenn beispielsweise Ethnie, Religion, Geschlecht oder Alter die Entscheidung unerlaubt oder falsch beeinflussen. Hier ist es notwendig, bereits in den Trainingsdaten für hinreichende Anonymisierung und Repräsentativität zu sorgen, und besonders sorgsam darauf zu achten, dass die Labels die richtigen Antworten darstellen. Ebenfalls kann es

[122] Lowe et al. 2017

sinnvoll sein, von Vornherein in das Verfahren Beschränkungen hinsichtlich der erlaubten Modelle zu integrieren. Eine technische Herausforderung besteht auch darin, den rechtlichen Begriff der Diskriminierung anwendungsabhängig in eine mathematische Definition zu überführen, die algorithmisch überprüft und umgesetzt werden kann.

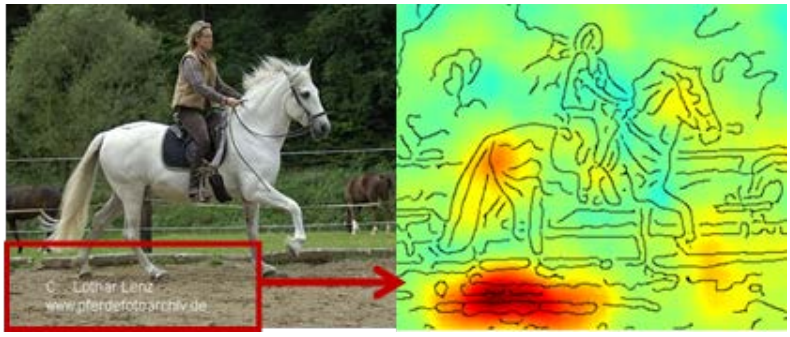
2.3.1 Akzeptanz

Wenn ein Mensch oder ein Unternehmen die Verantwortung für die Antworten eines Modells übernehmen muss, ist Transparenz eine wichtige Forderung. Die Entscheidungsfindung im Modell sollte in bestimmten Kontexten nachvollziehbar sein. Das geht, wenn symbolische Lernverfahren lesbare Entscheidungsbäume oder Modelle mit einer überschaubaren Anzahl möglichst unabhängiger sinnhafter Merkmale liefern.

Obwohl tiefes Lernen gerade in der Mustererkennung (Bilder, natürliche Sprache) sehr gute Resultate erzielt, ist für den Menschen nicht direkt nachvollziehbar, wieso ausgerechnet bestimmte Neuronen-Gewichtungen sehr gute Ergebnisse liefern. Tiefe Neuronale Netze sind also eine »Black Box«, die zwar gut funktioniert, aber nicht nachvollziehbar ist. Besonders für kritische Anwendungen wie im Medizinbereich, oder dort, wo der Gesetzgeber ein gewisses Maß an Erklärbarkeit der Entscheidung vorgibt (z.B. bei Kreditvergaben oder im Sicherheitsbereich), wird tiefes Lernen aufgrund seiner Intransparenz mit Skepsis betrachtet. Das gilt übrigens auch bei der Automatisierung von Maschinen und industriellen Prozessen, wo es kaum Fehlertoleranzen gibt. Hier müssen gelernte Modelle wahrscheinlich mit anderen Mechanismen kombiniert werden, die eingreifen, wenn ein Ergebnis zu unsicher ist oder die grundsätzlich prüfen, ob ein Ergebnis gewisse Mindeststandards erfüllt.

Für tiefe Neuronale Netze zur Objekterkennung gibt es inzwischen schon Verfahren, um die Beiträge der einzelnen Eingaben zur einer bestimmten Antwort über die Schichten hinweg zurückzuverfolgen und auf einer Hitzekarte (»heatmap«) darzustellen. So kann durch die Visualisierung in Abbildung 45 deutlich erkannt werden, dass das Neuronale Netz bei der Erkennung des Pferdebildes anscheinend ein Merkmal – nämlich das Copyright Logo – herangezogen hat, das für die eigentliche Identifikation des Objekts »Pferd« unbedeutend ist und sogar zu Fehlklassifikationen neuer Daten führen kann.

Abbildung 45: Heatmap für die Visualisierung der Parameter [123]



Für etliche Anwendungen werden in der heutigen Praxis deshalb auch weiterhin klassische Modelle wie Entscheidungsbäume präferiert, wenn auf wenige Prozent in der Genauigkeit in der Anwendung verzichtet werden kann.

2.3.1.1 Erklärbare Modelle

Um Schäden zu vermeiden und automatisierte Entscheidungen verantwortungsvoll einzusetzen, möchte man verstehen, was die Stärken und Schwächen eines Systems sind, nach welchen Grundprinzipien es sich verhält und wie es im Einzelfall zu seiner Antwort gekommen ist. Verständlichkeit wirkt dabei auf mehreren Ebenen: sie hilft dem Data Scientist, Modelle zu bereinigen und zu verbessern; sie hilft dem Anwender, die Relevanz der Ergebnisse zu beurteilen; und sie hilft Entscheidungsträgern und Regulierern, die Zertifizierbarkeit und Zulassungsfähigkeit von lernenden Systemen in kritischen Anwendungen zu bewerten. Um solchen Fragestellungen nachzugehen, hat 2016 auch die DARPA ein Forschungsprogramm für »Erklärbare KI« aufgesetzt und möchte damit insbesondere erklärbares Maschinelles Lernen fördern[124].

Der Wunsch nach Transparenz, also der Möglichkeit, das Verhalten des Systems vollständig nachvollziehen zu können, ist häufig in der Praxis nur schwierig erfüllbar, da viele Modelle sehr komplex sein müssen, um die gegebene Lernaufgabe hinreichend zu bewältigen. Eine Erklärbarkeit hingegen im Sinne der EU-Datenschutzverordnung bedeutet, für eine konkrete Einzelentscheidung des Systems die wesentlichen Einflussfaktoren aufzuzeigen. Dieses ist deutlich einfacher zu erfüllen[125].

Zur Erzeugung transparenter Systeme gibt es im Wesentlichen zwei Ansätze: zum einen kann für beliebige Modelle versucht werden, verständliche Approximationen zu generieren. Sie sollen das Verhalten des originalen Modells bis auf eine möglichst kleine Abwei-

[123] Binder et al. 2016
[124] Gunning 2016
[125] Doshi-Velez 2017

chung nachvollziehen, dabei aber eine für den Menschen verständliche Darstellung besitzen. Zum anderen existieren viele Ansätze, direkt verständliche Modelle aus Daten zu erzeugen, wie etwa Entscheidungsbäume. Diese Arten von Modellen sind nicht immer optimal für die Lernaufgabe geeignet. Falls die Transparenz aber gegenüber der Ergebnisqualität übergeordnete Bedeutung hat, sind solche Verfahren vorzuziehen.

Auch bei den interaktiven Ansätzen des »Human in the loop« (Abschnitt 2.2.2.1) spielt die Transparenz eine wichtige Rolle, um das Modell soweit zu verstehen, dass der Mensch effektiver eingreifen oder delegieren kann. Unklar ist allerdings oft, wo die Ursachen eines Fehlers liegen: im Modell, beim Benutzer oder dem Interaktionssystem. Wie überprüft man beim interaktiven Lernen die Sicherheit und Qualität des geänderten Modells? Kann die Interaktion offener gestaltet werden, so dass der Nutzer steuern kann, welche Fälle er und welche das System übernimmt?

Das Generieren von Erklärungen ist zurzeit ein aktives Forschungsfeld des Maschinellen Lernens. Bekannte Ansätze sind der SHAP[126]- und der LIME[127]-Algorithmus, die für den zu erklärenden Einzelfall und ähnliche Datenpunkte mittels einfacherer Verfahren ein lokales Erklärungsmodell bilden. Dabei ist es notwendig, dass Merkmale zur Verfügung stehen, die eine Interpretation durch den Menschen zulassen.

2.3.1.2 Fairness

Das Ergebnis eines maschinell trainierten Systems kann aus mehreren Gründen unethisch und diskriminierend sein, weil Entwickler oder Benutzer ihm dieses Verhalten bewusst oder unbewusst beigebracht haben oder weil das System in den Daten vorhandene Vorurteile gelernt und übernommen hat. So resultierte das Lernen auf großen Textmengen aus dem Internet in einem Modell, in welchem »Mann« mehr mit Programmierung und »Frau« mehr mit Hausarbeit assoziiert wurden, da die jeweiligen Begriffe in den Texten häufiger in einem Zusammenhang standen[128]. Solche vorhandenen Problematiken treten erst durch das ML-Verfahren zum Vorschein.

Sollen die Ausgaben der Verfahren nicht in einer Nachbearbeitung manuell verändert werden, indem etwa kritische Muster entfernt oder Prognosen für benachteiligte Gruppen manuell verbessert werden, so können Daten vor der Analyse bearbeitet werden, etwa indem bei der Gesichtserkennung darauf geachtet wird, dass Geschlechter, Hautfarben und Alter repräsentativ in den Daten vertreten sind.

Eine weitere Möglichkeit ist die Vorgabe von ethischen Beschränkungen bezüglich der erlaubten Lösungen. Hier besteht die technische Herausforderung darin, den rechtlichen Begriff der Diskriminierung in eine mathematische Definition zu überführen, die algorithmisch überprüft und erkannt werden kann. Soll beispielsweise bei Versicherungstarifen

[126] Lundberg/Su-In 2017
[127] Ribeiro/Singh/Guestrin
[128] Bolukbasi et al. 2016

nicht aufgrund des Geschlechts diskriminiert werden, so kann als Randbedingung vorgegeben werden, dass im Mittel für Männer und Frauen die gleiche Menge von Positiventscheidungen gebildet wird. Ein reines Löschen des Merkmals Geschlecht reicht hingegen nicht aus, da in den meisten Fällen das Geschlecht auch durch korrelierte Merkmale wie etwa Einkommen oder Berufswahl mit hinreichender Genauigkeit erschlossen werden kann.

Ist in anderen Fällen eine Korrelation eines kritischen Merkmals mit der Prognose gewünscht, kann zumindest vorgegeben werden, dass die Fehler auf den verschiedenen Werten des kritischen Merkmales vergleichbar sein sollen. Wenn sich zum Beispiel Bewerbungen auf einen Job in der Anzahl je Geschlecht massiv unterscheiden, dann möchte man vielleicht nicht, dass ein Algorithmus zur Anstellung gleich viele Männer wie Frauen vorschlägt, sondern dass die Prognose der passenden Kandidaten für Mann und Frau vergleichbar gut ist. Beim Begriff »counterfactual fairness« wird mit Hilfe von Bayesschen Modellen per Einzelfall verglichen, wie sich die Entscheidung in einer hypothetischen Welt geändert hätte, in der die Person beispielsweise eines anderen Geschlechts, Nationalität oder Alters wäre.[129]

Allgemein gilt, dass die jeweilige mathematische Definition von Fairness stark anwendungsbezogen ist und nicht ohne eine Entscheidung eines verantwortlichen Menschen getroffen werden kann.[130]

2.3.2 Sicherheit und Verlässlichkeit

2.3.2.1 Safety by design für ML

Sicherheit bedeutet, dass ein System keine gravierenden Fehler macht, die zu finanziellen oder physischen Schäden führen. Unter diesem Aspekt können drei Arten von Maßnahmen unterschieden werden, für die geeignete Herangehensweisen beim Einsatz lernender Systeme gesucht werden müssen[131]. Man kann inhärent sichere Systeme bauen, indem man Schäden durch die Konstruktion vermeidet (*safety by design*). Im Hardware-Design kennt man Verfahren, die effizient prüfen, ob die einzelnen Teile des Entwurfs genutzt werden (CDV, Constraint-based verficiation). Übertragen auf das Maschinelle Lernen könnte man prüfen, ob alle Teile der Repräsentation genutzt werden. Wenn ungenutzte Teile im Einsatz plötzlich genutzt werden, könnten die Ergebnisse überraschen. Eine weitere, aus dem Software-Design bekannte Methode ist die Modularisierung nach dem Prinzip »Teile und Herrsche«, womit die Teilung der Gesamtaufgabe in kleinere Einzelteile gemeint ist, die einfacher lösbar sind. Tiefe Netze werden zunehmend modular zusammengesetzt (siehe Abschnitt 2.2.1.2). Hier würde man viel gewinnen, wenn man die Schnittstellen der Module standardisiert testen könnte.

[129] Kusner et al. 2017
[130] Barocas/Hardt 2017
[131] Varshney 2016

2.3.2.2 Robuste Lernverfahren

Eine zweite Art von Maßnahmen für die Sicherheit in ML-Systemen ist der *Einbau von Sicherheitspannen*. Hier gibt es im System Toleranzen zwischen den Eingaben, für die es funktionieren soll, und weitergehenden Eingaben. Im Maschinellen Lernen fallen in diese Gruppe Tests auf Robustheit, um zu prüfen, dass kleine Änderungen in der Eingabe nur kleine Änderungen im Ergebnis bewirken. Schlagzeilen machten tiefe Modelle, die durch Veränderung weniger Pixel zu Fehlklassifikationen kamen, etwa bei der Erkennung von Verkehrszeichen[132]. Ein Ansatz im *gegnerischen Training* ist es, Modelle absichtlich mit Daten mit Störsignalen zu trainieren, so dass sie im Einsatz robuster generalisieren.

2.3.2.3 Lernen unter Beschränkungen

Als dritte Art von Maßnahmen für Sicherheitsgarantien kann man versuchen, harte Beschränkungen vorzugeben. Entweder soll beim Lernen direkt sichergestellt werden, dass das System diese Grenzen nicht überschreitet, oder dass beim Einsatz ein Fallback-System einspringt, das rechtzeitig die sicherheitskritische Aktion verhindert. Als Fallback-System könnte sich ein regelbasiertes System eignen, dass Expertenwissen anwendet, um das Verhalten in die richtige Bahn zu lenken. Feste Beschränkungen kann man auch nutzen, um Lernergebnisse zu plausibilisieren (siehe Abschnitt 0).

2.3.2.4 Lernen von Kompetenzgrenzen

Entscheidend für den Einsatz eines Fallback-System ist, dass ein trainiertes System überhaupt erkennt, dass seine Antwort unsicher sein könnte. Das entspricht einem *Lernen von Kompetenzgrenzen,* die als Kriterium dienen sollen, wann das Fallback-System aktiviert werden muss. Dazu könnte das System Anomalien erkennen. Auch wenn es sich unsicher ist oder wenn es alternative, ähnlich sichere Antworten gibt, könnte es sich melden. Dabei genügt es jedoch nicht, per gelernter Funktion jedem Beispiel einen Konfidenzwert zuzuweisen. Da die Gesamtheit der Beispiele und Situationen im Einsatz nicht überschaubar ist, muss auch die Unsicherheit in der Bestimmung der angegebenen Konfidenzwerte berücksichtigt werden.

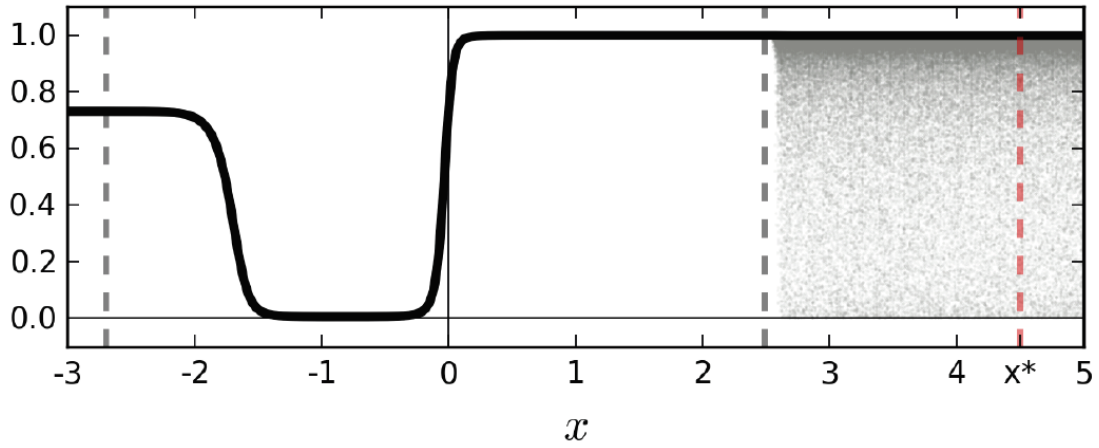
Von zentraler Bedeutung ist dabei, verschiedene Arten von Unsicherheit auseinander zu halten. Unsicherheit kann auf Zufälligkeit eines Ereignisses im Sinne der Wahrscheinlichkeitstheorie beruhen, sie kann aber auch aus fehlendem Wissen resultieren. Gelegentlich kommen noch Unsicherheiten durch das Fehlen strikter Konzepte (alt, jung, …) hinzu, oder Unsicherheiten, die durch Mehrdeutigkeiten gegeben sind.

Wenn man nicht weiß, wie repräsentativ die Beispieldaten sind, dann bleibt unklar, wie das System in Fällen reagiert, für die es keine Beispiele gesehen hat. Daher muss berücksichtigt werden, auf wie vielen vergleichbaren Beispielen trainiert wurde. Die nächste Ab-

[132] Evtimov et al. 2017

bildung zeigt eine binäre Klassifikationsfunktion mit Ausgabe eines Konfidenzwertes. Die Beispieldaten liegen zwischen den gestrichelten schwarzen Linien und auf x* wird mit einem Konfidenzwert von 1 generalisiert, ohne Rücksichtnahme auf Unsicherheiten weit weg von den Trainingsdaten.

Abbildung 46: Binäre Klassifikationsfunktion mit Ausgabe eines Konfidenzwertes[133]



[133] Gal/Ghahramani 2016

Tabellarische Zusammenfassung der offenen Forschungsfragen

Tabelle 3: Anforderungen aus der Praxis und Forschungsthemen, die diese Anforderungen aufnehmen

Forschungsthemen	Offene Fragestellungen
Lernen mit sehr großen Datenmengen	
Verteilte Algorithmen	Konvergenzgarantien aufgespaltener Algorithmen bei minimiertem Datenaustausch, Umgang mit knappen Ressourcen
Lernen aus Datenströmen	Geeignete Repräsentationen für den sofortigen Erkenntnisgewinn, Verbindung von inkrementellem Lernen und der Möglichkeit veraltete Daten vergessen zu können
Lernen in Quanten-Computern	Erkenntnisse über die Möglichkeiten durch quantentheoretische Beschreibungen
Lernen mit wenig Daten	
Automatisches Lernen von Labels	Zuverlässige Zuweisung von Zielvorgaben unter Unsicherheit, Lernen mit verrauschten Daten
Lernen aus Simulationen	Diskrepanz zwischen Realität und Simulation, Verständnis der Eigenschaften der Generalisierung, passende Simulationsumgebungen entwerfen
Lernen mit einem oder keinem expliziten Beispiel	Unterteilung und Wiederverwendung von Teilaufgaben, Repräsentation von Konzepten
Transfer- Lernen	Charakterisierung von Ähnlichkeiten, Verhinderung von negativem Transfer
Multi-Task-Lernen	Geschickte Kombination von Aufgaben, Anwendung auf unüberwachte Verfahren und bestärkendes Lernen, Verhinderung von Negativeffekten
Lernen mit zusätzlichem Wissen	Entwicklung neuer Lernverfahren, vereinigende Repräsentationen, Verbindung von Lernvorgang und Schlussfolgerung
Interaktives Lernen vom Menschen (human in the loop)	Entwicklung neuer Lernverfahren, Erkennung der eigenen Unsicherheiten, Bestimmung zielführender Interaktionsanforderungen, Imitationslernen
Anpassungsfähigkeit und Flexibilität	
Multi-Task-Lernen	s.o.
Transfer-Lernen	s.o.
Lebenslanges Lernen	Verhinderung des Vergessens von gelernten Fähigkeiten (catastrophic forgetting)
Interaktives Lernen in der Umwelt	Exploration/Exploitation, Bestimmung von guten Bewertungsfunktionen von Zustand-Aktion-Paaren (Q-value functions), kompakte Darstellung der Umgebung
Kollaboration	
Interaktives Lernen vom Menschen (human in the loop)	s.o.

Forschungsthemen	Offene Fragestellungen
Metalernen (Auto-ML)	Theoretische Sicherheiten für die Güte von Parameterwahlen, Heuristiken zur Auswahl geeigneter Architekturen
Lernen mit zusätzlichem Wissen	
Lernen mit physikalischen Modellen	Forschung nach geeigneten Teilmodellierungen
Lernen aus Simulationen	s.o.
Lernen mit symbolischem Wissen	Entwicklung neuer Lernverfahren, Umwandlung von gelernten Resultaten in symbolisches Wissen
Lernen mit Wissensgraphen	Entwicklung neuer Lernverfahren, Umgang mit heterogenen Datenstrukturen, unterschiedlichen Modellierungen und fehlenden Daten
Lernen mit Aufmerksamkeit und Gedächtnis	Verwaltung und Lenkung der computerinternen Ressourcen, Entwicklung neuer Ansätze
Lernen mit Graphen	Bestimmung passender Ähnlichkeitsfunktionen von Graphen, Kontrolle der Komplexität, Entwicklung neuer Ansätze
Regellernen	Skalierbarkeit bezüglich der Komplexität der Regeln und großen Regelmengen (Konsistenz)
Selbsterlebende Maschinen und Verknüpfung von Sinnen	Verbindung Signal mit Erfahrung, semantisches Verstehen
Nachvollziehbarkeit, Sicherheit und Akzeptanz	
Erklärbare KI-Modelle	Statistische Ergebnisfindung nachvollziehbar gestalten
Erkennung von Diskriminierung und Fairness	Modellierung von Fairness, Erkennung und gezielte Verhinderung von Diskrimination
Lernen mit zusätzlichem Wissen	s.o.
Interaktives Lernen vom Menschen (human in the loop)	s.o.
Safety by design für ML	Konstruktion neuer Lernverfahren mit Handlungsgrenzen
Robuste ML-Verfahren	Einschränken von Manipulationsmöglichkeiten, theoretisches Verständnis der Generalisierung auf ungesehene Beispieldaten
Lernen von Kompetenzgrenzen	Modellierung von Unsicherheiten, Verbesserung Bayesscher Modelle
Lernen unter Beschränkungen	Konstruktion neuer Lernverfahren mit Handlungsgrenzen

Anhang A: Glossar: ML-Fachbegriffe

Adversarial Learning oder gegnerisches Lernen	Beim gegnerischen Lernen wird versucht, ein Modell durch Lernen mit sogenannten gegnerischen Beispielen (Englisch: Adversarial Examples) robuster gegenüber Angriffen zu machen. »Adversarial Examples« sind absichtlich gestört, um gezielt falsche Ergebnisse herbeizuführen.
Auto-ML, automatisiertes maschinelles Lernen	Auto-ML steht für »automatisiertes maschinelles Lernen« und bezeichnet Verfahren, die den Datenwissenschaftler unterstützen, indem sie automatisch den Datenanalyseprozess aufsetzen, inklusive der Schritte zum Maschinellen Lernen.
Algorithmus, Lernalgorithmus	In der Informatik ist ein Algorithmus eine genaue Berechnungsvorschrift zur Lösung einer Aufgabe. Ein Lernalgorithmus ist ein Algorithmus, der Beispieldaten (Lerndaten, oder Trainingsdaten) erhält und ein Modell für die gesehenen Daten berechnet, das auf neue Beispieldaten verallgemeinert.
Bestärkendes Lernen oder Reinforcement Learning	Beim bestärkenden Lernen erhält der Lernalgorithmus gelegentliches Feedback für Interaktionen mit der Umwelt und lernt, die Erfolgsaussichten der einzelnen Aktionen in den verschiedenen Situationen besser einzuschätzen.
Bild- und Videoanalyse	Bei der Bild- und Videoanalyse werden visuelle Daten von optischen Sensoren und Kamerasystemen verarbeitet, um Objekte, Szenen und Aktivitäten in der Umgebung wahrzunehmen und identifizieren.
Black-Box-, White-Box-, Grey-Box-Modelle	Black-Box-Modelle des Maschinellen Lernens sind Modelle rein statistischer Art. White-Box-Modelle dagegen bezeichnen analytische und physikalische Beschreibungen, deren Modellierung meist sehr aufwändig ist. Bei Grey-Box-Modellen kombiniert man beide Ansätze, um die jeweiligen Vorteile zu vereinen.
Bot	Unter einem Bot versteht man ein Computerprogramm, das weitgehend automatisch wiederkehrende Aufgaben abarbeitet. Beispiele, die vom maschinellen Lernen profitieren könnten, sind Chatbots, Social Bots und Gamebots.
Tiefes Lernen oder Deep Learning (DL)	Tiefes Lernen bedeutet das Lernen in Künstlichen Neuronalen Netzen mit mehreren bis sehr vielen inneren Schichten. Tiefes Lernen ist verantwortlich für die Erfolge in der Sprach- und Text-, Bild- und Videoverarbeitung.

Echtzeit	Echtzeit bedeutet die ständige Betriebsbereitschaft eines Systems, bei der alle Reaktionen und Rechenschritte in einer bestimmten kurzen Zeitspanne ablaufen.
Ende-zu-Ende Lernen	Beim Ende-zu-Ende-Lernen werden alle nötigen Zwischenschritte von Eingabe zu Ausgabe innerhalb eines Modelles integriert.
Erklärbare KI	Black-Box-Modelle, wie insbesondere die tiefen Künstlichen Neuronalen Netze, sind für Menschen nicht nachvollziehbar. Die »Erklärbare KI« sucht nach Möglichkeiten, die versteckte Logik oder die einzelnen Ausgaben besser nachvollziehbar oder erklärbar zu machen.
KI-gestützte Sensordatenfusion	Bei der KI-gestützten Sensordatenfusion werden verschiedene Sensortypen zusammengeführt und statistisch gegenseitige Abhängigkeiten zwischen unterschiedlichen Datenquellen gelernt.
Klassische Lernverfahren, traditionelle Lernverfahren	Zu den klassischen oder traditionellen Lernverfahren gehören symbolische Verfahren und ältere statistische Verfahren. Nicht dazu zählen Verfahren für tiefe neuronale Netze.
Kognitive Maschinen oder kognitive Systeme	Kognitive Maschinen oder Systeme sind alternative Begriffe für künstliche intelligente Systeme. Sie zeichnen sich aus durch Fähigkeiten des Lernens und Schlussfolgerns sowie der Sprachverarbeitung, Bildverarbeitung und Interaktion mit dem Nutzer.
Künstliche Intelligenz (KI)	Künstliche Intelligenz ist ein Teilgebiet der Informatik mit dem Ziel, Maschinen zu befähigen, Aufgaben »intelligent« auszuführen. Dabei ist weder festgelegt, was »intelligent« bedeutet, noch welche Technologien zum Einsatz kommen.
Label	Ein Label, auch Zielmerkmal, markiert die korrekte Antwort zu einem Beispiel. Überwachte Lernverfahren lernen durch Verallgemeinerung Modelle, die auch neuen Beispielen eine Antwort zuordnen. Oft müssen Labels von Menschen vergeben werden.
Künstliche Neuronale Netze (KNN)	Künstliche Neuronale Netze sind Modelle des Maschinellen Lernens, die durch Aspekte des menschlichen Gehirns motiviert wurden. Sie bestehen aus vielen in Datenstrukturen realisierten Schichten von Knoten, die als künstliche Neuronen bezeichnet werden. Der Lernalgorithmus verändert die Gewichte, das sind Zahlenwerte an den Verbindungen zwischen den Knoten, solange, bis die Ergebnisse für die Aufgabe gut genug sind.
Maschinelles Lernen (ML)	Maschinelles Lernen bezweckt die Generierung von »Wissen« aus »Erfahrung«, indem Lernalgorithmen aus Beispielen ein komplexes Modell entwickeln. Das Modell kann anschließend auf neue, unbe-

	kannte Daten derselben Art angewendet werden. Damit kommt das Maschinelle Lernen ohne manuelle Wissenseingabe oder explizite Programmierung eines Lösungswegs aus.
Maschinelles Lernverfahren, Lernmethode	Die Begriffe »maschinelles Lernverfahren« und »Lernmethode« werden hier synonym verwendet. Sie beschreiben auf einer abstrakteren Ebene eine Vorgehensweise, die unter Umständen durch mehrere konkrete Lernalgorithmen realisiert werden kann.
Multimodales Lernen	Beim multimodalen Lernen werden Eingabesignale von unterschiedlichen Quellen (wie Audio- und Bildsignale) herangezogen und in Zusammenhang gebracht, um darauf aufbauend eine passenden Aktion zur Bewältigung der Aufgabe abzuleiten.
Lernstile und Lernaufgaben	Lernstile unterscheiden sich in der Zusatzinformation und den dadurch möglichen Lernaufgaben. Die wichtigsten Lernstile sind überwachtes Lernen, unüberwachtes Lernen und bestärkendes Lernen. Je nachdem, welche Zusatzinformation zur Verfügung steht, können andere Lernaufgaben gelernt werden.
Modell	Ein Modell ist eine Abstraktion der Wirklichkeit. Im Maschinellen Lernen erzeugt der Lernalgorithmus ein Modell, das Beispieldaten generalisiert, so dass es anschließend auch auf neue Daten angewendet werden kann.
Multitask-Lernen	Beim Multitask-Lernen wird versucht, verschiedene Aufgaben gleichzeitig auf einer gemeinsamen internen Repräsentation zu lernen, die dadurch tragfähiger wird.
One-Shot-, Zero-Shot-Lernen	Idealerweise soll anhand weniger Beispiele einer Klasse oder ausschließlich durch die Kombination bekannter Beispiele anderer Klassen automatisch gelernt werden. Diese Fähigkeit in ML-Anwendungen zu implementieren, wird als »One-Shot« oder gar »Zero-Shot«-Lernen bezeichnet.
Online Lernen, Lernen auf Datenströmen	Das Online-Lernen kommt ohne die konventionelle Einteilung in Trainingsdaten für die Modellentwicklung und Testdaten für die Modellbewertung aus. Online Lernen funktioniert also in Echtzeit und auf Datenströmen, die nicht abgespeichert werden.
Quantencomputer	Quantencomputer basieren ihre elementaren Rechenschritte auf quantenmechanischen Zuständen – sogenannte Qubits – anstelle der binären Zustände (Bits) in digitalen Computern und verarbeiten sie gemäß quantenmechanischer Prinzipien. Hierdurch wird für manche Anwendungen ein enormer Geschwindigkeitsvorteil erwartet.

Repräsentationen, Lernen von Repräsentationen	Repräsentationen sind Darstellungen der Daten, die in subsymbolischen Modellen aus vielen Zahlenwerten bestehen. Beispielsweise kann das Bild eines Gesichtes durch alle Pixelwerte repräsentiert werden, oder durch die Angabe von Ort und Größe von Auge, Nase, Mund, usw. Beim Repräsentationslernen werden durch Transformation der Eingabe kompaktere Repräsentationen gelernt, die die eigentliche Lernaufgabe erleichtern.
Subsymbolisch	In subsymbolischen Modellen, wie Künstlichen Neuronalen Netzen, sind die Merkmale der Beispiele und die erlernten Zusammenhänge in vielen Zahlen versteckt, die keinen Einblick in die erlernten Lösungswege erlauben.
Symbolisch	In symbolischen Modellen sind die Merkmale der Beispiele und die erlernten Zusammenhänge explizit und nachvollziehbar repräsentiert. Beispiele sind logische Regeln, Entscheidungsbäume und Wissensgraphen.
Text- und Sprachverarbeitung oder Natural Language Processing (NLP)	Text- und Sprachverarbeitung umfassen Technologien zur Interpretation und Erzeugung von natürlicher Sprache in Wort und Schrift. Dazu gehören die Vertextung gesprochener Sprache, Stimmungserkennung, Informationsextraktion aus Texten, maschinelle Übersetzung und das Führen von Gesprächen
Transfer-Lernen	Das Transfer-Lernen zielt darauf ab, in einem Künstlichen Neuronalen Netz Teilbereiche zu finden, die bereits für eine bestimmte Aufgabe trainiert worden sind, um sie für eine andere, verwandte Aufgabe wieder zu verwenden.
Überwachtes Lernen oder supervised learning	Beim überwachten Lernen müssen die richtigen Antworten als »Labels« mitgeliefert werden. Damit lassen sich Klassifikations- und Regressionsaufgaben lernen, bei denen Beispiele ihren jeweiligen Labels zugeordnet werden.
Unüberwachtes Lernen oder unsupervised learning	Beim unüberwachten Lernen gibt es nur die rohen Beispieldaten, ohne Labels oder Feedback. Damit können vereinfachende Beschreibungen der gesamten Beispielmenge gefunden werden. So können Daten in verschiedene Gruppen oder Cluster unterteilt oder die Dimensionen, also die Anzahl der Merkmale reduziert werden.
Wissensgraph	Wissensgraphen bestehen aus Knoten und Verbindungen, die symbolisches Wissen über Objekte und ihre Beziehungen repräsentieren. Die Knoten haben einen Typ, beispielsweise Tier, Möbel, Stadt, der die zulässigen Objekte und ihre Beziehungen einschränkt.

Index: Kapitel 1 und 2

aktives Lernen 70

Aktivierungsfunktion 19

analogistische Verfahren 18

Arten von Unsicherheit 85

Aufmerksamkeit des Netzes 78

Autoencoder 40, 45

Auto-ML 71

Bayessche Modelle 35

Bayessches Netz 17

bestärkendes Lernen 28

Bias 47

Big Data 24

Big Data Analytics 24

Bilderkennung 21, 36, 81

Bildverarbeitung 42, 48

Black Box 81

catastrophic forgetting 69

Cluster 26, 32

convolutional neural network 44

Data Mining 48

Data Scientist 51, 71

Daten 51

Datenqualität 47

deep belief network 46

deep feedforward network 43

deep Q-Network 43

Dimensionsreduktion 39, 58

Einbau von Sicherheitspannen 85

Entscheidungsbaum 22, 30

Erklärbare KI 82

Erklärbarkeit 82

Expertensystem 40

Exploration-Exploitation-Problematik 69

Fairness 84

Fallback-System 85

federated learning 59

Feedback 28

Fehlerrückführung 21, 39, 43

Fluch der Dimensionalität 58

gegnerisches Training 85

generative gegnerische Netze 41, 46

Grey-Box-Modelle 72

Hitzekarte 81

hybrides System 73

induktive Statistik 16

Kernmethoden 33

KI-Winter 21, 22

Klassifikation 26

kognitive Systeme 11

konnektionistische Ansätze 21

Kostenfunktion 50

Künstliche Intelligenz 9

Künstliche Neuronale Netze 34

Label 18, 25, 28, 58

lebenslanges Lernen 68

Lernalgorithmus 12, 29

Lernaufgabe 42

Lernen im Simulationslabor 62

Lernen mit Crowd Sourcing 70

Lernen von Kompetenzgrenzen 85

Lernen von Regeln 75

Long short-term memory 45

Maschinelles Lernen 9
Metalernen 71
Modell 13, 16, 39
modulares Lernen 68, 69
multimodales Lernen 79, 91
Multi-Task-Lernen 66
Netze mit Speicher 79
neuronale Turing-Maschine 77
One-shot Learning 63
Online-Lernen 59
Ontologie 21
OpenAI 80
Optimierung 39, 51
Parameter 26, 29, 30, 32, 35, 38, 50, 51, 71
Performanzmaß 49
Q-Lernen 35
Quanten-Lernen 61
Racing-Algorithmen 71
Recurrent network 44
Regellernen 76
Regression 18, 26, 29
Regularisierung 48
Repräsentation 40, 42, 61
Robustheit 51, 85
Rohdaten 39
safety by design 84
semi-überwachtes Lernen 28
sequentielles Entscheiden 28, 35
Sicherheit 84
Sprachverarbeitung 21, 36, 42, 48
stochastische Verfahren 17
Störsignale 85
Stützvektormaschine 33
Subgruppensuche 76
subsymbolische Modelle 21
symbolische Methoden 22
tiefe Neuronale Netze 21, 37, 81
tiefes Lernen 10, 58, 81
Trainingsdaten 13, 14, 24, 58
Transfer-Lernen 67, 69
Transparenz 81
Turing-vollständig 77
Überanpassung 48
überwachtes Lernen 25
Universalintelligenz 12
Unsicherheit 85
Unteranpassung 49
unüberwachtes Lernen 26
Vektoren 37
verteilte Algorithmen 59
Wiederverwendung von Teilnetzen 67
Wissensgraph 23
Worteinbettung 39
Zero-shot Learning 63

KAPITEL 3

Kompetenzlandkarte Maschinelles Lernen: Publikationen, Forschungsförderung, Patente und Akteure

Manuel Molina Vogelsang | Fraunhofer IMW
Dmitry Neustroev | Fraunhofer IMW
Inga Döbel | Fraunhofer IMW

Inhaltsverzeichnis Kapitel 3

3 Kompetenzlandkarte Maschinelles Lernen: Publikationen, Forschungsförderung, Patente und Akteure 97

3.1 Methodik und Forschungsdesign 97
3.1.1 Datengrundlage - Publikationen 99
3.1.2 Datengrundlage – EU-Förderung 100
3.1.3 Datengrundlage - Patentanalysen 101
3.1.4 Generelle methodische Ansätze 102

3.2 Bibliometische Analyse der ML-Publikationen 103
3.2.1 Publikationen von Forschungseinrichtungen und Unternehmen 105
3.2.2 Publikationsanalyse nach Anwendungsfeldern 108
3.2.2.1 Definitionen der Anwendungsfelder 108
3.2.2.2 Publikationsstatistiken zu Anwendungsfeldern 110
3.2.2.3 Forschungseinrichtungen nach Anwendungsbereichen 113

3.3 Maschinelles Lernen in Deutschland – Publikationen 113
3.3.1 Forschungseinrichtungen und Unternehmen 114
3.3.2 Regionen und deren Akteure 119

3.4 Analyse der ML-Projektförderung auf europäischer Ebene 121
3.4.1 ML-Projektförderung in den EU-Rahmenforschungsprogrammen 121
3.4.2 Projektkoordinatoren im Ländervergleich 121
3.4.3 Deutsche Projektkoordinatoren 123
3.4.4 Projektkoordination nach Anwendungsbereichen 124

3.5 ML-Patentanalyse 125
3.5.1 Generelle Patententwicklung 125
3.5.2 ML-Patente im internationalen Vergleich 125
3.5.3 Patententwicklung in Anwendungsbereichen 127
3.5.4 Regionen und deren Akteure - Patentanmeldungen 129

3.6 Maschinelles Lernen: Produkte, Märkte und Wirtschaftsakteure 130
3.6.1 Generelle Entwicklungen 130
3.6.2 Anwendungsbranchen und Produkte 134
3.6.2.1 Gesundheitswesen 135
3.6.2.2 Autonomes Fahren 137
3.6.2.3 Sensordatenanalyse in der Industrie und Automatisierung 140
3.6.2.4 Einzelhandel und Marketing 142

Anhang B: Schlagworte für die Suchanfragen 150

3 Kompetenzlandkarte Maschinelles Lernen: Publikationen, Forschungsförderung, Patente und Akteure

Im Zuge der fortschreitenden Digitalisierung in nahezu allen Lebensbereichen nehmen Maschinelles Lernen und Künstliche Intelligenz eine bedeutende strategische Rolle im wirtschaftlichen, zivilen und militärischen Bereich ein.[134] [135] Somit rücken sie auch immer stärker in den Fokus von Politik, Wissenschaft und Wirtschaft, um im globalen Wettbewerb hinsichtlich Forschungs- und Entwicklungserfolgen, Talenten, Daten sowie kommerziellen Anwendungen mithalten zu können. Auch in den USA äußern sich inzwischen besorgte Stimmen, in KI- und ML-basierter Forschung und -Technologie insbesondere von China, aber auch von Russland überholt zu werden.[136]

Ziel der Analyse im Kapitel 3 war es, einen Überblick über die Forschungsaktivitäten im Bereich ML zu erarbeiten, mit einem besonderen Schwerpunkt auf dem Standort Deutschland. Als Methode hierfür dienten statistische Analysen über wissenschaftliche Publikationen (Artikel in Fachzeitschriften und Konferenzbeiträge), Patentanmeldungen sowie öffentlich geförderte Forschungsprojekte. All diese Indikatoren bestätigen ein starkes Wachstum der ML-Thematik in den letzten 10 Jahren. Ebenfalls wurden wissenschaftliche und wirtschaftliche Aktivitäten von Akteuren in vier bedeutenden Anwendungsbereichen betrachtet, die ML auf bestimme Datentypen anwenden: Bild- und Videoanalyse, Text- und Sprachverarbeitung, Verarbeitung von Audiodaten sowie Datenverarbeitung aus mehreren Quellen.

3.1 Methodik und Forschungsdesign

Die Ergebnisse der vorliegenden Studie stützen sich auf einen Methodenmix aus quantitativen und qualitativen Ansätzen der empirischen Innovationsforschung. Im Folgenden wird ein kurzer Überblick der verwandten Methoden und Vorgehensweise skizziert.

Die Definition und Abgrenzung der relevanten Schlagworte für die Datenbankabfragen und die quantitative Analyse der Akteurs- und Forschungslandschaft stellt eine methodische Herausforderung dar. Um dieser zu begegnen, wurde das Technologiefeld in einem iterativen Prozess mit ML-Fachleuten aus dem Projektteam in Form von Suchbegriffen zunächst auf der Systemebene in Form von ML-Methoden, Lernstilen und -aufgaben sowie Modellen und Lernverfahren abgegrenzt. Neben dem Suchbegriff »Machine Learn-

[134] The Hague Centre for Strategic Studies 2017
[135] Button 2017
[136] Simonite 2017

ing« wurden insgesamt 15 englische Suchbegriffe definiert, die das internationale Forschungsfeld ML möglichst treffend und gut abdecken. Um die Trefferquote und damit die Aussagekraft der Ergebnisse weiter zu erhöhen, wurden insgesamt 85 Schreibvariationen für die Suchbegriffe definiert[137]. Die Ergebnisse werden im Folgenden unter ML-Technologie zusammengefasst und bilden jeweils die Grundgesamtheit für die anschließenden Analysen.

Abbildung 47: Analyserahmen ML-Akteure und Forschungslandschaft

Quelle: Eigene Darstellung

In einem zweiten Schritt wurden die vier Anwendungsbereiche »Bild- und Videoverarbeitung«, »Sprachverarbeitung«, »Audioverarbeitung« und »Signalverarbeitung« für ML-Technologien abgegrenzt. Zusätzlich wurden Suchergebnisse für zwei oder mehr Anwendungsbereiche unter »heterogene Daten« zusammengefasst. Analog zur Definition des Technologiefeldes wurden in einem iterativen Prozess mit ML-Fachleuten aus dem Projektteam insgesamt 41 Schreibvariationen für die Suchbegriffe definiert. Die Suchergebnisse werden im Folgenden unter ML-Anwendungsbereiche zusammengefasst.

In einem letzten Schritt wurden die ML-Technologien nach Produktkategorien differenziert. Es werden Anwendungsbereiche betrachtet, die ML auf bestimme Datentypen anwenden und ggf. mit Technologien kombinieren. Die Ergebnisse für relevante Branchen der deutschen Wirtschaft werden im Kapitel »Wirtschaftsakteure und ML-Produkte« exemplarisch dargestellt. Die Auswahl umfasst intelligente vernetzte Objekte und Umgebungen (z.B. Autonome Transportmittel oder Roboter), sowie Softwareprodukte und Dienstleistuntgen (z.B. Chatbots und Service-Assistenten, Entscheidungsunterstützungssysteme, vorausschauende Wartung, Sentimentanalyse).

[137] Für eine vollständige Liste der Suchbegriffe siehe Anhang.

3.1.1 Datengrundlage - Publikationen

Ein in der Forschung etablierter Indikator für die Betrachtung wissenschaftlicher Aktivitäten sind wissenschaftliche Publikationen (Artikel in Fachzeitschriften bzw. Fachjournals und Konferenzbeiträge), die mit Hilfe von bibliometrischen Statistiken in Publikationsdatenbanken erfasst und recherchiert werden können. Die Elsevier-Scopus-Datenbank bildet als etablierte Publikationsdatenbank[138] mit über 67 Millionen Aufsätzen in wissenschaftlichen Fachzeitschriften und rund 8 Millionen Konferenzbeiträgen die Grundlage für die bibliometrischen Auswertungen der vorliegenden Untersuchung.

Die Definition und Abgrenzung der relevanten Suchbegriffe für die Datenbankabfrage wurde in Zusammenarbeit mit ML-Fachleuten erstellt und erfolgte einerseits auf der technologischen Systemebene in Form von ML-Methoden, Lernstilen und –aufgaben, Modellen und Lernverfahren sowie der Auswahl großer Anwendungsfelder andererseits. Mit Hilfe der definierten Suchbegriffe wurde ein Suchstring für die Datenbankabfrage entwickelt und die Datenbankfelder zu Titel, Zusammenfassung und Schlüsselwörtern semantisch durchsucht. Die Suche wurde auf Publikationen in englischer Sprache im Zeitraum zwischen 2006 und 2016 eingegrenzt. Weiterhin wurde die Suche auf Aufsätze in wissenschaftlichen Fachzeitschriften und Konferenzbeiträge in den relevanten Disziplinen (Physik und Ingenieurwissenschaften, Lebenswissenschaften, Wirtschafts- und Verhaltenswissenschaften) begrenzt. Da das Analyseziel die Identifikation von relevanten Akteuren war, die in der Forschung und Entwicklung von ML-Methoden und Technologien tätig sind, sollten mit der Disziplinenfokussierung beispielsweise Beiträge zur philosophischen oder ethischen Auseinandersetzung mit ML gefiltert werden.

Es wurde hier keine Vorauswahl der Fachzeitschriften oder Konferenzbeiträgen vorgenommen. Um alle Aufsätze in wissenschaftlichen Zeitschriften und Konferenzen auf Qualität zu bewerten und letztlich für die Auswertung auszuwählen, ist ein eigenes Forschungsprojekt inkl. einer umfangreichen Expertenkonsultierung notwendig gewesen.

Im nächsten Schritt erfolgte eine Bereinigung des ermittelten Datensatzes in folgenden Teilschritten: alle unvollständigen Suchergebnisse, d.h. Datenbankeinträge ohne Titel, Zusammenfassung, Informationen über Akteure, Herkunftsangaben oder Jahresangaben, aus dem Datensatz gelöscht, da hierüber keine aussagekräftigen Analysen möglich gewesen wären. Die anschließende Aufarbeitung des Datensatzes hatte den Sinn, der Harmonisierung von Informationen über Akteure (bspw. das Zusammenführen identischer Einrichtungen, die mit unterschiedlicher Schreibweise in den Einträgen gelistet waren). Abschließend wurden weitere Angaben, wie zum Beispiel Zugehörigkeit zu außeruniversitären Forschungsorganisationen, ergänzt. Die Grundgesamtheit der ML-Technologien bilden in

[138] Darüber hinaus gibt es noch weitere Datenbanken (z.B. Web of Scince, Google Scholar, arXiv.org), die sich hinsichtlich Abdeckung und Qualität von der gewählten Datenbank unterscheiden.

dieser Untersuchung 315 817 in Scopus erfasste Publikationen zwischen 2006 und 2016, davon 171 239 Aufsätze in wissenschaftlichen Zeitschriften und 144 578 Konferenzbeiträge.

Generell ist hier anzumerken, dass für unsere Analysen wissenschaftliche Publikationen als Datengrundlage herangezogen wurden. Aus diesem Grund besteht die Annahme, dass insbesondere Unternehmen und stärker anwendungsorientierte Forschungseinrichtungen hier strukturbedingt unterrepräsentiert sein könnten, da sich ihre ML-bezogenen Aktivitäten nicht primär in wissenschaftlichen Publikationen äußern.

3.1.2 Datengrundlage – EU-Förderung

Die Cordis Projektdatenbank der Europäischen Kommission bildet die Datengrundlage für die statistischen Auswertungen der europäischen Projektförderung. Sie erfasst zentrale Projektinformationen der europäischen Forschungsrahmenprogramme wie z.B. Förderkennzeichen, Projekttitel, Zusammenfassung des Vorhabens, Start- und Enddatum, Gesamtkosten, maximale EU Förderung, Projektkoordinatoren und –partner sowie deren Nationalitäten. Für die Förderperiode 2007 bis 2014 wurden im Rahmen des 7. Forschungsrahmenprogramm (FP7) mehr als 25 000 Projekte mit einer Summe von rund 43,9 Mrd. Euro gefördert[139]. Aktuell läuft das 8. Forschungsrahmenprogramm (Horizon 2020), in dem bereits 11 048 Projekte mit einer Summe von über 20,3 Mrd. Euro gefördert werden[140].

Analog zum oben skizzierten Ansatz wurden die Datenbankfelder Titel und Zusammenfassung der Projektdatenbank semantisch durchsucht. Die Suche beschränkte sich auf den Zeitraum 2007 bis 2017. Im Sinne einer Akteursanalyse mit dem Ziel, Forschungs- oder Industrieakteure zu identifizieren, wurden die zentralen Akteure öffentlich geförderter Projekte ermittelt. Aufgrund der Beschaffenheit der verfügbaren Daten wurden ausschließlich die Projektkoordinatoren berücksichtigt. Methodisch ist dabei zu beachten, dass die gesamte Fördersumme eines jeden Projektes auf das jeweilige Jahr des Projektbeginns aggregiert wurde. Bspw. fing die Projektlaufzeit für das von der Fraunhofer-Gesellschaft geführte Projekt »A Novel Decision Support System for Intelligent Maintenance« (iMAIN) am 01.09.2012 an und lief bis zum 31.08.2015, wo in diesem Fall die Fördersumme auf das Jahr 2012 angerechnet wurde.

Der Rohdatensatz wurde analog zum Vorgehen bei den wissenschaftlichen Publikationen aufgearbeitet. Zunächst wurden alle unvollständigen Suchergebnisse, d.h. Datenbankeinträge ohne Titel, Zusammenfassung, Koordinator, Herkunft des Koordinators, EU-Förderung sowie Start- und Enddatum aus dem Datensatz gelöscht. Abschließend wurde der Rohdatensatz aufgearbeitet, so wurden etwa Informationen über Akteure harmoni-

[139] https://ec.europa.eu/research/evaluations/pdf/fp7_final_evaluation_expert_group_report.pdf
[140] Stand Oktober 2017.

siert und weitere Informationen wie Zugehörigkeit zu Forschungsorganisationen ergänzt. Die Grundgesamtheit bilden 595 Forschungsprojekte und 1,49 Mrd. Euro EU-Förderung.

3.1.3 Datengrundlage - Patentanalysen

Patente sind ein wertvoller Indikator, um die technologische Leistungsfähigkeit von Unternehmen, Regionen oder Ländern nachzuvollziehen und Technologietrends abzuleiten. Besonders vorteilhaft bei der Arbeit mit Patenten ist, dass sehr genau rekonstruiert werden kann, wo der Ursprung einer Erfindung liegt und wie sich diese über die Zeit entwickelt hat. Allerding muss an dieser Stelle auf die eingeschränkte Patentierbarkeit von Software bzw. computerimplementierten Erfindungen hingewiesen werden. Untersuchungen zeigen, dass die Patentierbarkeit von Software und damit auch die Patentierbarkeit von ML-Technologien sich in den jeweiligen Patentsystemen stark unterscheiden.[141]

Die Espacenet Patentdatenbank des Europäischen Patentamts (EPO), mit über 100 Mio. Patentschriften die umfassendste Informationsquelle weltweit, bildet die Grundlage für die statistische Analyse von Patentfamilien in dieser Studie. Patentfamilien setzen sich aus einer oder mehr nationalen oder internationalen Patentanmeldungen zusammen und schützen die gleiche technologische Erfindung in unterschiedlichen Patentsystemen[142]. Analog zum oben skizzierten Vorgehen wurden die Datenbankfelder Titel und Zusammenfassung semantisch durchsucht. Die Suche beschränkte sich auf Patentfamilien zwischen 2006 und 2015. Es ist darauf hinzuweisen, dass die Offenlegung von Patentanmeldungen in der Regel erst 18 Monate nach dem Prioritätsdatum (das früheste Stichdatum, zu dem eine Technologie als neu und erfinderisch zur Patentierung eingereicht wird) erfolgt. Daher ist davon auszugehen, dass lediglich für die Jahre 2006-2015 vollständige Informationen vorliegen.

Alle unvollständigen Suchergebnisse wurden bei der Analyse aus dem Rohdatensatz gelöscht, d.h. Datenbankeinträge ohne Titel, Zusammenfassung, Informationen über Akteure, Herkunftsangaben oder Jahresangaben. Anschließend wurde der Datensatz aufgearbeitet, so wurden etwa Informationen über Akteure harmonisiert und weitere Informationen wie Zugehörigkeit zu außeruniversitären Forschungsorganisationen ergänzt. Das Prioritätsdatum der ersten Patentanmeldung einer Patentfamilie wurde als Referenzdatum für die Patentfamilie festgelegt. Die Auswertung erfolgt auf Ebene der Patentanmelder mit dem Ziel, die Innovationsleistung der Akteure nachzuvollziehen, unabhängig davon, wo die Forschung ihren Ursprung hat. Die Grundgesamtheit der ML-Technologien bilden 7 252 Patentfamilien, die sich aus 45 387 nationalen und internationalen Patentanmeldungen zusammensetzten.

[141] Neuhäusler, P. Frietsch, R. Rothengatter, O. (2015): Patentierung computerimplementierter Erfindungen – Aktuelle Rechtslage und ökonomische Implikationen, Fraunhofer ISI Discussion Papers No. 46

[142] https://www.epo.org/searching-for-patents/helpful-resources/first-time-here/patent-families.html

3.1.4 Generelle methodische Ansätze

Insgesamt war für die Erstellung der Rohdatensätze wichtig, dass die in der Datenbank hinterlegten Informationen wie Titel und Zusammenfassung mindestens einen oder mehrere der definierten Suchbegriffe umfassen. Eine inhaltliche Prüfung der Suchergebnisse konnte aufgrund der umfangreichen Datenmenge nur stichprobenartig erfolgen. Es ist davon auszugehen, dass Publikationen, Patente und Förderprojekte, die in ihren Titeln und Zusammenfassungen die definierten Suchbegriffe beinhalten, auch einen Bezug zu ML-Technologien besitzen.

Die Anzahl der Publikationen und Patentfamilien wurde auf Ebene der Organisationen berechnet. Dabei wurde die Methode der fraktionalen Zählung (»fractional counting«) angewandt, d.h. jede Publikation bzw. Patentfamilie wurde durch die Anzahl der beteiligten Organisationen geteilt, sodass Doppelzählungen vermieden werden[143].

Auf der Basis der ermittelten Ergebnisse können Aussagen über die Dynamik der Publikationsaktivität, Forschungsförderung sowie der Patentaktivität getroffen werden. Allerdings muss auch auf die Grenzen der Auswertungen hingewiesen werden. Durch die Analyse können keine Rückschlüsse auf Qualität und Kompetenzen der Akteure geschlossen werden, es handelt sich lediglich um eine quantitative Betrachtung des Publikations- oder Patentaufkommens.

Eine weitere Datengrundlage für dieses Kapitel ist die Sekundäranalyse laufender Entwicklungen auf dem internationalen Markt für intelligente Produkte und Dienstleistungen. Hierfür wurden verschiedene Quellen gesichtet, größtenteils Marktstudien und Trendanalysen zu KI-Systemen und kommerziellen ML- und DL-Angeboten. Auch internationale wissenschaftliche Publikationen und Konferenzbeiträge, die sich mit ML-Anwendungen befassen, sowie aktuelle Veröffentlichungen von Unternehmen (z.B. Whitepaper, Berichte, Pressemitteilungen, Web-Beiträge) wurden als Quellen berücksichtigt. Auffällig ist, dass bei den internationalen Marktstudien der letzten Jahre Deep-Learning-Produkte stark im Vordergrund stehen. Am besten beleuchtet ist der US-amerikanische Markt. Für Deutschland sind zum Betrachtungszeitpunkt einzelne umfassende Darstellungen zu ML-basierten Produkten und Dienstleistungen verfügbar[144], Marktdaten liegen lediglich fragmentarisch vor.[145]

[143] Egghe, L., Rousseau, R. Van Hooydonk, G. (2000): Methods for Accrediting Publications to Authors or Countries: Consequences for Evaluation Studies, Journal of the American Society for Information Science. 51, S. 145 – 157.

[144] Z.B. Böttcher et al. 2017.

[145] Z.B. Statista 2017.

3.2 Bibliometische Analyse der ML-Publikationen

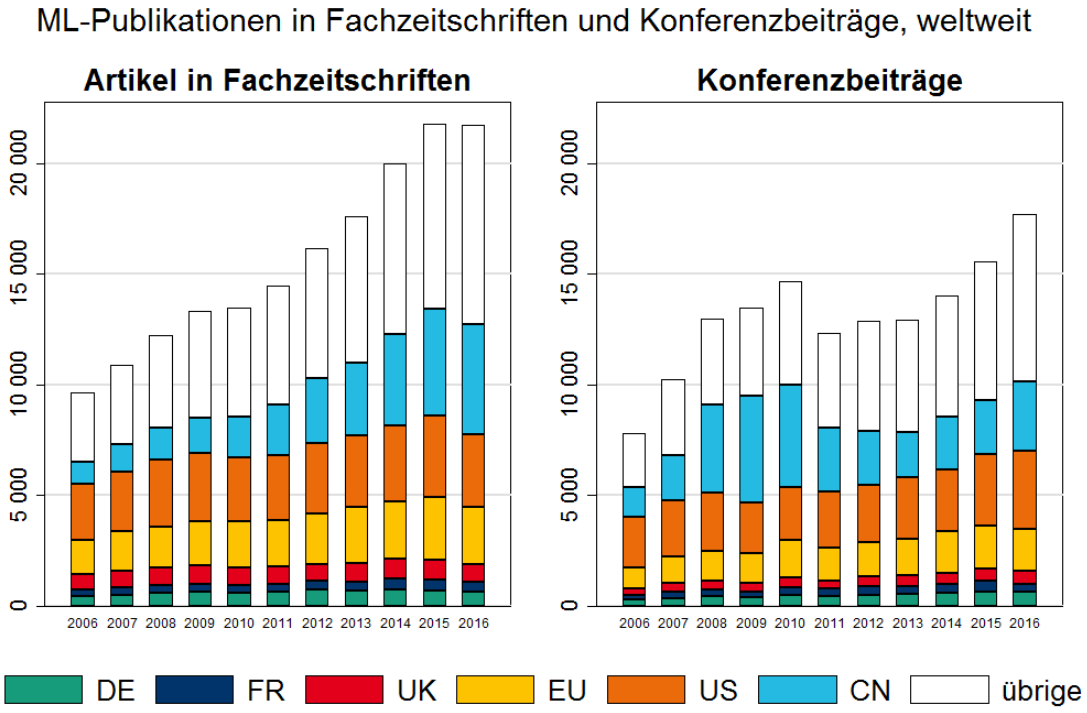
Ein in der Forschung etablierter Indikator für die Betrachtung wissenschaftlicher Aktivitäten sind Publikationen (Artikel in Fachzeitschriften bzw. Fachjournals und Konferenzbeiträge), die mit Hilfe von bibliometrischen Statistiken erfasst und recherchiert werden können. Die Scopus-Datenbank bildet als etablierte Publikationsdatenbank die Datengrundlage für die bibliometrischen Auswertungen der vorliegenden Untersuchung.

Mit Hilfe der definierten Suchbegriffe konnten für das Technologiefeld »Maschinelles Lernen« im Zeitraum zwischen 2006 und 2016 insgesamt 315 817 Publikationen identifiziert werden. Abbildung 48 zeigt, dass das Publikationsaufkommen in wissenschaftlichen Fachzeitschriften bzw. Journals im Zeitraum von 2006 bis 2016 von 9 636 auf 21 724 angestiegen ist. Auch ist die Anzahl der Konferenzbeiträge zu ML von 7 811 in 2006 auf 17 686 in 2016 gestiegen. Im Vergleich hierzu ist das Publikationsaufkommen in den Computerwissenschaften insgesamt im gleichen Zeitraum von 190 583 auf 319 523 zitierbare Publikationen angestiegen[146]. Die ML-Publikationen weisen damit ein relativ stärkeres Wachstum auf als der allgemeine Publikationstrend und lassen auf eine intensivere Forschungstätigkeit schließen.

Mit Blick auf die regionale Verteilung der Publikationsaktivitäten wurden die Ergebnisse nach Ländern aufgeschlüsselt. Es zeigt sich in Abbildung 48, dass die EU, die USA und China in absoluten Zahlen am meisten publizieren. Durchschnittlich entfallen mehr als 63% des weltweiten ML-Publikationsaufkommens auf die oben genannten Länder bzw. Regionen. In absoluten Zahlen gemessen publizieren Forschungseinrichtungen und Unternehmen aus den USA mit 63 436 ML-Publikationen weiterhin etwas mehr als jene in China (61 827 ML-Publikationen). Differenziert nach Konferenzbeiträgen kann ein ähnliches Bild gezeichnet werden. Über den Zeitverlauf betrachtet kann festgestellt werden, dass neben den USA zunehmend weitere Länder besonders viele ML-Publikationen veröffentlichen. Während 2006 noch rund 28% der weltweiten ML-Publikationen aus den USA stammten, waren es 2016 noch 17%. Die Publikationsaufkommen in China hat sich besonders dynamisch entwickelt, so stieg der Anteil an den weltweiten ML-Publikationen von 13% in 2006 auf 21% in 2016.

[146] Vgl. http://www.scimagojr.com/worldreport.php (zuletzt geprüft am 20.02.2018)

Abbildung 48: Weltweite Entwicklung der ML-Publikationen, 2006-2016



Source: Scopus; Fraunhofer IMW; eigene Berechnungen
EU: EU-28 exkl. DE, FR, UK

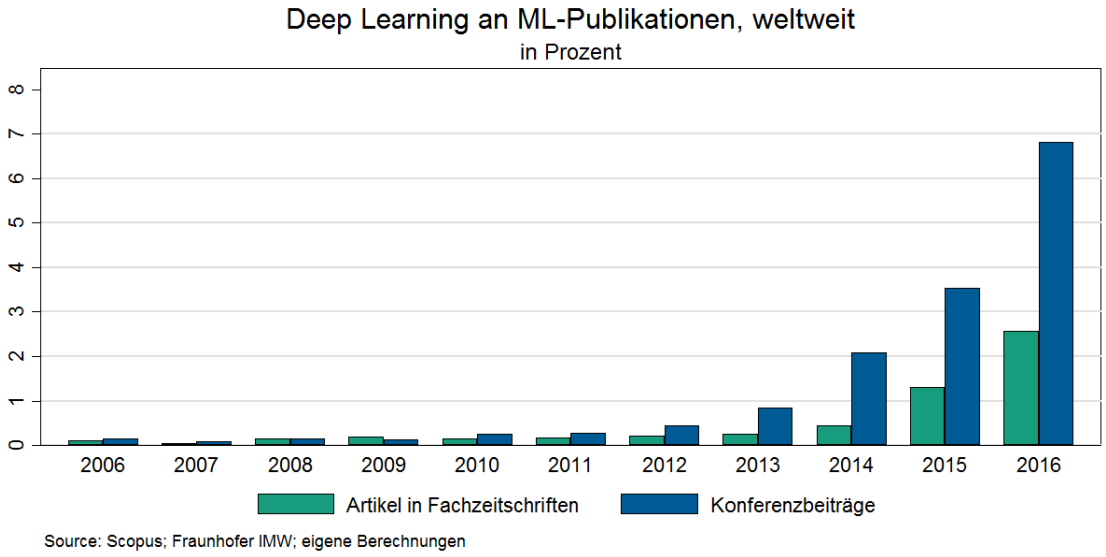
Innerhalb der EU stammen die meisten ML-Publikationen aus Forschungseinrichtungen und Unternehmen aus Großbritannien (13 656 ML-Publikationen) im Gesamtzeitraum von 2006 bis 2016. An zweiter Stelle innerhalb Europas, folgt Deutschland mit insgesamt 11 887 Publikationen im Bereich ML-Technologien, davon 6 674 Aufsätze in Fachzeitschriften und 5 213 Konferenzbeiträge. Die Anzahl der Aufsätze in Fachzeitschriften hat sich von 427 im Jahr 2006 auf 627 in 2016 positiv entwickelt. Die Konferenzbeiträge steigen von 266 in 2006 auf 617 in 2016.

Das Publikationsaufkommen in den USA, Großbritannien und Deutschland entwickelte sich zwischen 2006 und 2016 relativ stetig. Im globalen Vergleich fiel der Anteil der deutschen ML-Publikationen leicht von 4% in 2006 auf 3% in 2016. Dies könnte an einem stärkeren Fokus auf Exzellenz liegen. Dagegen lässt sich in China mit einer jährlichen Zunahme von durchschnittlich 15% ein sehr starkes Wachstum an ML-Publikationen erkennen. Wird jedoch wie in einigen Statistiken nur die reine Anzahl der Veröffentlichungen in einschlägigen Konferenzbeiträgen betrachtet, ohne Berücksichtigung ML-bezogener Such- und Schlüsselbegriffe, fällt auf, dass die Stellung Chinas gegenüber den USA deut-

lich geringer ausfallen kann[147]. Dies kann auf methodische Unterschiede in den Untersuchungen zurückgeführt werden. Zudem könnten insbesondere die USA aufgrund von organisatorischen und räumlichen, sicher aber auch aufgrund von sprachlichen Aspekten dieser Konferenzen hier stärker vertreten sein.

Die Anzahl von Publikationen zu Deep Learning (DL) in wissenschaftlichen Zeitschriften und Konferenzbeiträgen wurde ebenfalls durch unterschiedliche Suchbegriffe erfasst. Seit 2013 kann bei den Fachkonferenzen ein starkes Wachstum konstatiert werden. Hier erreichte der Anteil von Deep Learning 2016 bereits 6,8%. Im gleichen Jahr liegt der Anteil der DL-bezogenen Konferenzbeiträge in den USA mit 10,4% vergleichsweise hoch. In Großbritannien liegt er bei 8,7%, in China bei 8,3% und in Deutschland bei 5,3%. Der Trend scheint in wissenschaftlichen Zeitschriften um zwei Jahre zeitversetzt einzutreten. Es ist davon auszugehen, dass die Dynamik in diesem Bereich weiter anhalten wird, so steht DL im Mittelpunkt vieler Anwendungsszenarien (siehe hierzu Ausführungen in Kapitel 3.6).

Abbildung 49: Anteil der Deep Learning Publikationen an gesamten ML-Publikationen, 2006-2016



3.2.1 Publikationen von Forschungseinrichtungen und Unternehmen

Die zehn publikationsstärksten Forschungseinrichtungen bei wissenschaftlichen Zeitschriften kommen aus der Wissenschaft, wobei acht von zehn der Organisationen aus Asien stammen. Eine Übersicht ist in Tabelle 4 abgetragen. Unter den fünf publikationsstärksten deutschen Einrichtungen finden sich die Max-Planck-Gesellschaft (MPG), Helmholtz-

[147] EFI 2018: Gutachten 2018 https://www.e-fi.de/fileadmin/Gutachten_2018/EFI_Gutachten_2018.pdf

Gemeinschaft (HGF), Technische Universität München (TUM), Universität Bonn, Ludwig-Maximilians-Universität München (LMU) und Universität Tübingen. Die organisatorischen Anreize und Möglichkeiten wissenschaftlich zu publizieren sind in den Hochschulen und großen Forschungseinrichtungen wie der MPG und HGF sicher stärker ausgeprägt als beispielweise in anwendungsorientieren Forschungseinrichtungen wie dem Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) oder den Instituten der Fraunhofer-Gesellschaft (FhG). Die Ergebnisse sind daher nicht ohne weiteres hinsichtlich Qualität oder Kompetenz zu interpretieren, sondern lediglich Aussagen über Quantitäten möglich.

Tabelle 4: Publikationsstärkste Forschungseinrichtungen weltweit, Summe der ML-Publikationen in wissenschaftlichen Zeitschriften zwischen 2006-2016

Artikel in wissenschaftlichen Zeitschriften				
Rang	**Name**	**Organisation**	**Land**	**Pub.**
1	Zhejiang University	Universität	China	816
2	Shanghai Jiao Tong University	Universität	China	704
3	Tsinghua University	Universität	China	672
4	Harbin Institute of Technology	Universität	China	646
5	University of Tehran	Universität	Iran	530
6	Nanyang Technological University	Universität	Singapur	526
7	Chinese Academy of Sciences	Akademie	China	489
8	National University of Singapore	Universität	Singapur	476
9	Max-Planck-Gesellschaft	AuF[148]	Deutschland	462
10	Carnegie Mellon University	Universität	USA	445
24	**Helmholtz-Gemeinschaft**	**AuF**	**Deutschland**	**374**
104	**TU München**	**Universität**	**Deutschland**	**228**
168	**Universität Bonn**	**Universität**	**Deutschland**	**177**
178	**Ludwig-Maximilians-Universität München**	**Universität**	**Deutschland**	**172**
262	**Universität Tübingen**	**Universität**	**Deutschland**	**159**

Die publikationsstärksten Forschungseinrichtungen bei den Konferenzbeiträgen stammen ebenfalls überwiegend aus der Wissenschaft, wobei sieben von zehn Organisationen aus China und drei aus den USA stammen, wie in Tabelle 5 ersichtlich. Im Vergleich zu Publikationen in wissenschaftlichen Zeitschriften ist die relativ größere Anzahl von Unternehmen hervorzuheben. So befinden sich mit IBM und Microsoft zwei US-amerikanische Unternehmen unter den zehn publikationsstärksten Akteuren bei den Konferenzbeiträgen. Unter den fünf publikationsstärksten deutschen Einrichtungen befinden sich HGF, TUM,

[148] AuF = außeruniversitäre Forschungseinrichtung

Fraunhofer-Gesellschaft, MPG und die Rheinisch-Westfälische Technische Hochschule Aachen (RWTH).

Tabelle 5: Publikationsstärkste Forschungseinrichtungen weltweit, Summe der ML-Publikationen in Konferenzbeiträgen zwischen 2006-2016

Artikel in Konferenzbeiträgen				
Rang	**Name**	**Organisation**	**Land**	**Pub.**
1	Carnegie Mellon University	Universität	USA	917
2	Tsinghua University	Universität	China	910
3	Harbin Institute of Technology	Universität	China	613
4	IBM Inc.	Unternehmen	USA	594
5	Beijing University of Posts and Telecommunications	Universität	China	572
6	Chinese Academy of Sciences	Akademie	China	536
7	Zhejiang University	Universität	China	532
8	Microsoft Inc.	Unternehmen	USA	529
9	Huazhong University of Science and Technology	Universität	China	521
10	Beihang University	Universität	China	518
…				
40	**Helmholtz-Gemeinschaft**	**AuF**	**Deutschland**	**292**
47	**TU München**	**Universität**	**Deutschland**	**274**
55	**Fraunhofer-Gesellschaft**	**AuF**	**Deutschland**	**256**
87	**Max-Planck-Gesellschaft**	**AuF**	**Deutschland**	**207**
128	**Rheinisch-Westfälische Technische Hochschule Aachen**	**Universität**	**Deutschland**	**169**

Auch bei den Konferenzbeiträgen liegen Akteure aus China und den USA an der Spitze. Andere Statistiken[149] können jedoch ein anderes Bild insbesondere für China geben, wie unter Punkt 3.2 bereits erläutert wurde.

Im Hinblick auf Unternehmen, entfällt die Hälfte der zehn publikationsstärksten Akteure auf Softwarekonzerne aus den USA, mit Microsoft Research Asia auf Platz sieben. Im Deutschen Raum verzeichnen die Siemens AG, Honda Motor Europe (North) GmbH Honda Motor Europe (North) GmbH (deutsche Niederlassung des japanischen Automobilkonzerns), SAP SE, Robert Bosch GmbH und Daimler AG die meisten Publikationen.

Tabelle 6: Publikationsstärkste Unternehmen weltweit, Summe der ML-Publikationen zwischen 2006-2016

[149] EFI 2018: Gutachten 2018 https://www.e-fi.de/fileadmin/Gutachten_2018/EFI_Gutachten_2018.pdf

Rang	Organisation	Land	Fachzeit-schriften	Konferenz-beiträge	Summe
1	IBM Inc.	USA	197	594	790
2	Microsoft Inc.	USA	156	529	685
3	Google Inc.	USA	74	221	296
4	Yahoo Inc.	USA	54	229	283
5	Microsoft Research Asia Inc.	China	68	164	232
6	Siemens USA Inc.	USA	131	68	199
7	Nippon Telegraph and Tele-phone Corporation	Japan	68	123	191
8	Philips B.V.	Niederlande	111	68	180
9	Hewlett-Packard Inc.	USA	39	106	145
10	Intel Corp.	USA	38	106	144
…					
13	**Siemens AG**	**Deutschland**	**47**	**49**	96
22	**Honda Motor Europe (North) GmbH**	**Deutschland**	**18**	**20**	38
30	**SAP SE**	**Deutschland**	**2**	**29**	32
37	**Robert Bosch GmbH**	**Deutschland**	**6**	**20**	26
48	**Daimler AG**	**Deutschland**	**4**	**14**	18

3.2.2 Publikationsanalyse nach Anwendungsfeldern

3.2.2.1 Definitionen der Anwendungsfelder

Ebenfalls wurden wissenschaftliche und wirtschaftliche Aktivitäten von Akteuren in vier bedeutenden Anwendungsbereichen betrachtet, die ML auf bestimme Datentypen anwenden. Die betrachteten Anwendungsbereiche umfassen:

- Die Verarbeitung visueller Daten (Bild- und Videobearbeitung)
- Die maschinelle Sprachverarbeitung
- Die Analyse von Audiosignalen
- Die maschinelle Signalverarbeitung aus diversen Quellen

Visuelle Daten:

Bei der Bild- und Videoanalyse handelt es sich um die Verarbeitung visueller Daten (z.B. von optischen Sensoren, Kamerasystemen oder Bildern). Dies ermöglicht es einer Maschine, Objekte, Szenen und Aktivitäten in der Umgebung wahrzunehmen und zu identifizieren. Ein Computer-Vision-System erfasst, verarbeitet und analysiert Bilder, um numerische oder symbolische Informationen zu erzeugen. Die traditionelle Herangehensweise, bei der in möglichst kleinen Teilschritten, wie dem Erkennen von Linien und Texturen Bilder berechnet und analysiert werden, um anschließend die Merkmale mit bekannten Objekten

auf wahrscheinliche Übereinstimmungen zu vergleichen[150], wurde in den letzten Jahren durch das tiefe Lernen in KNN überholt. Deep Learning ermöglicht die Verarbeitung visueller Daten mit einer höheren Genauigkeit und Zuverlässigkeit in einem einzigen Schritt (end-to-end). Damit konnten bei Bildklassifikation, Objekterkennung und Bildsegmentierung und Maschinellem Sehen erhebliche Fortschritte erzielt werden.

Neben der Bild- und Videoverarbeitung stellen die maschinelle »Sprachverarbeitung« (18%) und letztlich die »Signalverarbeitung« (11%) wichtige Anwendungsfelder dar, auf denen weltweit geforscht wird. In den Bereichen Sprach- und Signalverarbeitung nimmt die Publikationsdynamik in wissenschaftlichen Zeitschriften seit 2011 deutlich zu. Text- und Sprachverarbeitung (Englisch Natural Language Processing, NLP) sind Technologien, die Computersysteme befähigen, natürliche Sprachen in Wort und Schrift zu verstehen und zu erzeugen.

Sprachverarbeitung:
Text- und Sprachverarbeitung (Natural Language Processing, NLP) umfasst Technologien, die Computersysteme befähigen, natürliche Sprache in Wort und Schrift zu interpretieren und zu erzeugen. Die Algorithmen verarbeiten die menschliche Spracheingabe und wandeln sie in maschinenverständliche Darstellungen um.[151] Dazu gehören unter anderem: Erkennung gesprochener Sprache (speech recognition, automatische Transkription menschlicher Sprache); natural language generation, automatisiertes Schreiben von Texten in stark formalisierten Bereichen wie Sport- oder Finanznachrichten; Sentimentanalyse, die Analyse von Tonalität und Stimmung in Texten; maschinelle Übersetzung sowie das Führen von Unterhaltungen und Dialogen. Die semantische Interpretation ist dabei eine der großen Herausforderungen.

Audioverarbeitung:
Die Analyse von Audiosignalen wird auch in außersprachlichen Bereichen angewandt, beispielsweise in der Produktion, zur Analyse der Geräusche von Maschinen und Anlagen[152], in der Medizin zur Überwachung der Herztöne[153], oder zum Generieren von Musikstücken[154]. Auch Musikstreaming-Dienste wenden Deep Learning[155] an, um personalisierte Playlists zu erstellen. Metainformationen wie Gattung, Stimmung, Tempo bzw. verwendete Instrumente von abgerufenen Audio-Dateien werden erfasst und klassifiziert, um Präferenzen der Kunden vorherzusagen. Das Publikationsaufkommen in wissenschaftlichen Zeitschriften im Bereich »Audioverarbeitung« (2%) fällt deutlich geringer aus als die anderen Anwendungsbereiche und weist insgesamt eine niedrigere Dynamik auf. Dies

150 Gentsch 2018
151 Rao und Verweij 2017
152 https://www.idmt.fraunhofer.de/de/Press_and_Media/press_releases/2017/Acoustic_Quality_Inspection_HMI_2017.html
153 https://www.citi.sinica.edu.tw/papers/yu.tsao/5206-F.pdf, https://arxiv.org/pdf/1707.04642.pdf
154 http://www.asimovinstitute.org/analyzing-deep-learning-tools-music/
155 http://benanne.github.io/2014/08/05/spotify-cnns.html

kann u.a. dadurch erklärt werden, dass die Verarbeitung von Audiosignalen etwa bei der der Prüfung von Maschinen und Produkten noch in den Anfängen steckt und weitere Forschung auf dem Gebiet erforderlich ist.[156]

Signalverarbeitung aus diversen Quellen:
Viele Problemstellungen in datenintensiven Branchen wie Automobilindustrie, Fertigungsindustrie, Energiesektor oder Gesundheitswesen erfordern jedoch eine gleichzeitige Berücksichtigung großer Mengen von Signalen aus verschiedenen Quellen in Echtzeit. Im Rahmen der KI-gestützten Sensordatenfusion werden statistische Interdependenzen zwischen unterschiedlichen Datenquellen unter Verwendung von Bayes-Netzwerken und probabilistischen grafischen Modellen genutzt. Vor allem Deep Learning wird verwendet, um Feeds verschiedener Sensortypen (Beschleunigungsmesser, Gyroskop, Magnetometer, Barometer, Satellitenempfänger) zusammenzuführen[157].

3.2.2.2 Publikationsstatistiken zu Anwendungsfeldern

International nimmt das Publikationsaufkommen im Bereich Sprach- und Signalverarbeitung seit 2011 deutlich zu, während der Bereich »Audioverarbeitung« (2%) deutlich geringer ausfällt und insgesamt eine niedrigere Dynamik aufweist. Die deutsche Forschung folgt diesem Trend, adressiert jedoch überdurchschnittlich stark die »Bild- und Videoverarbeitung« (64%). Die USA und China dagegen zeigen eine vergleichsweise stärkere Fokussierung auf die »Sprachverarbeitung« und »Signalverarbeitung« als Deutschland. In diesem Zusammenhang sei angemerkt, dass bei der Sprach- und Bildverarbeitung bislang die größten Erfolge mit Deep Learning erzielt wurden. In den USA sind die Werte für die »Sprachverarbeitung« (19%) und in China für die »Signalverarbeitung« (12%) hervorzuheben. Des Weiteren sei auf unterschiedlichen Schwerpunkte und Akteure der Artikel in Fachzeitschriften im Vergleich zu Konferenzbeiträgen hingewiesen.

In Abbildung 50 ist die Entwicklung des Publikationsaufkommens der ML-Anwendungsbereiche abgetragen. Im Durchschnitt können für ein Drittel der ML-Publikationen auch ML-Anwendungsbereiche bestimmt werden. Dies kann zum einen methodische Gründe haben, so ist eine eindeutige Abgrenzung der Anwendungsbereiche schwierig. Zum anderen muss davon ausgegangen werden, dass viele Forscherinnen und Forscher zu ML publizieren, aber nur wenige konkrete Anwendungsbereiche adressieren. Auf der anderen Seite kann gezeigt werden, dass Unternehmen stärkere Bezüge zu den definierten Anwendungsbereichen herstellen als etwa Hochschulen oder Forschungseinrichtungen.

156 https://www.idmt.fraunhofer.de/de/Press_and_Media/press_releases/2017/Acoustic_Quality_Inspection_HMI_2017.html
157 Groopman & Kaul 2017.

Abbildung 50: Weltweite Entwicklung ML-Publikationen nach Anwendungsbereichen, 2006-2016

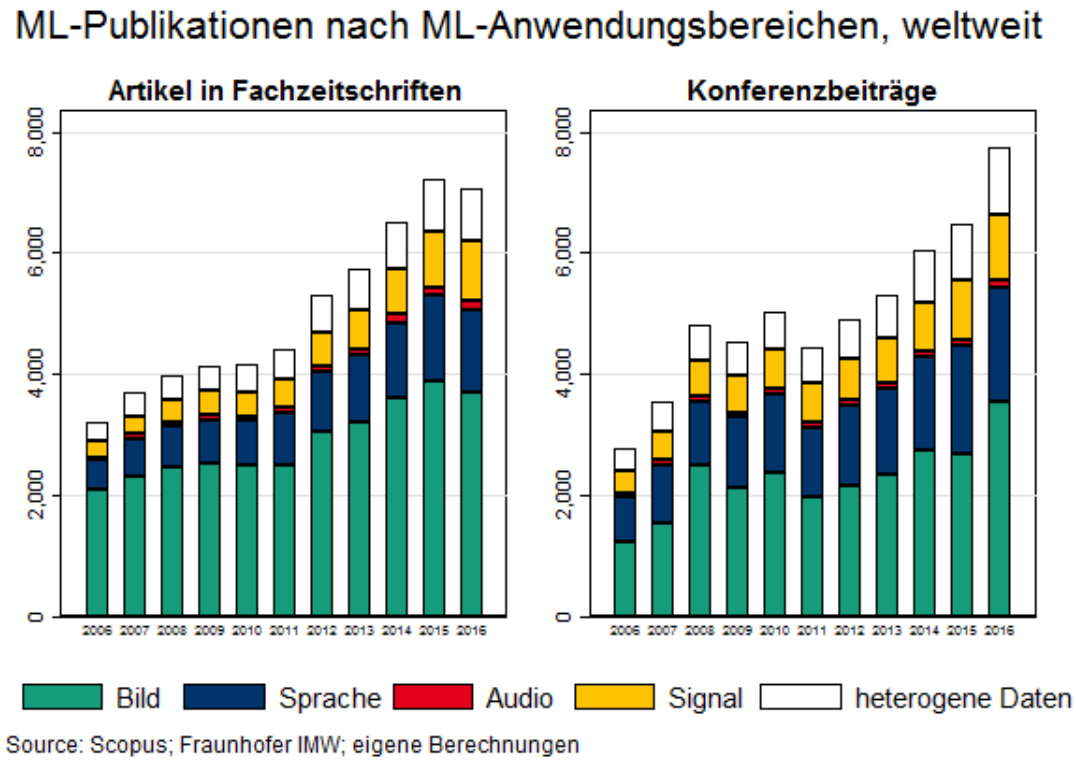


Tabelle 7 zeigt das internationale Publikationsverhalten, gestaffelt nach ML-Anwendungsbereichen. Für Deutschland kann konstatiert werden, dass Forscherinnen und Forscher einen starken Fokus auf den ML-Anwendungsbereich »Bild- und Videoverarbeitung« legen (52%), gefolgt von der »Sprachverarbeitung« (19%) und letztlich »Signalverarbeitung« (13%). Die Daten zeigen ferner, dass die »Audioverarbeitung« (2%) in der deutschen Forschung bislang eine geringe Rolle spielt.

Tabelle 7: ML-Publikationen nach Anwendungsbereichen, 2006-2016

	Bild- und Videoverarbeitung	**Sprachverarbeitung**	**Audioverarbeitung**	**Signalverarbeitung**	**Heterogene Daten**
Artikel in wissenschaftlichen Zeitschriften					
Welt	31 847	10139	1088	6117	6172
	58%	18%	2%	11%	11%
Deutschland	**1.641**	**353**	**47**	**223**	**291**
	64%	**14%**	**2%**	**9%**	**11%**
USA	7.407	2418	189	1173	1325
	59%	19%	2%	9%	11%
China	5062	1497	125	1050	925
	58%	17%	1%	12%	11%

	Bild- und Videoverar-beitung	Sprachver-arbeitung	Audiover-arbeitung	Signalver-arbeitung	Heterogene Daten
Artikel in wissenschaftlichen Zeitschriften					
UK	2030	571	73	275	436
	60%	17%	2%	8%	13%
Konferenzbeiträge					
Welt	25214	14269	1082	7568	7466
	45%	26%	2%	14%	13%
Deutschland	**856**	**573**	**55**	**395**	**396**
	38%	**25%**	**2%**	**17%**	**17%**
USA	4.855	3212	201	1795	1882
	41%	27%	2%	15%	16%
China	5995	2440	124	19	144
	69%	28%	1%	0%	2%
UK	782	488	60	266	287
	42%	26%	3%	14%	15%
Summe					
Welt	57.061	24.408	2.170	13.685	13.638
	51%	22%	2%	12%	12%
Deutschland	**2.497**	**926**	**102**	**618**	**687**
	52%	**19%**	**2%**	**13%**	**14%**
USA	12.262	5.630	390	2.968	3.207
	50%	23%	2%	12%	13%
China	11.057	3.937	249	1.069	1.069
	64%	23%	1%	6%	6%
UK	2.812	1.059	133	541	723
	53%	20%	3%	10%	14%

In den USA sind ebenfalls die »Bild- und Videoverarbeitung« (50%) die wichtigsten Anwendungsbereiche, gefolgt von der »Sprachverarbeitung« (23%) und letztlich »Signalverarbeitung« (12%). Auch in China sind die »Bild- und Videoverarbeitung« (50%), gefolgt von der »Sprachverarbeitung« (23%) die wichtigsten Anwendungsfelder der ML-Forschung.

Im Vergleich zum weitweiten Publikationsverhalten kann festgehalten werden, dass sich die Publikationen in wissenschaftlichen Zeitschriften in Deutschland überdurchschnittlich stark auf die »Bild- und Videoverarbeitung« (65% in Deutschland vs. 58% weltweit) konzentrieren. Die USA und China dagegen zeigen eine relativ stärkere Fokussierung auf »Sprachverarbeitung« und »Signalverarbeitung« als deutsche ML-Publikationen. Besonders auffällig sind die unterschiedlichen Schwerpunkte der Forschung zwischen Artikeln in Fachzeitschriften und Konferenzbeiträgen.

3.2.2.3 Forschungseinrichtungen nach Anwendungsbereichen

In allen ML-Anwendungsbereichen ist China unter den fünf publikationsstärksten Forschungseinrichtungen vertreten. Besonders auffällig ist die relative Stärke auf dem Gebiet der »Bild- und Videoverarbeitung«. Unternehmen, die auf dem Gebiet der »Bild- und Videoverarbeitung« forschen und wissenschaftlich publizieren sind dagegen US-amerikanische und deutsche Konzerne wie Google, Tesla, BMW und Audi. Im Bereich der Audioverarbeitung stechen neben einer chinesischen und einer amerikanischen drei europäischen Universitäte heraus. Darunter die Pompeu Farba in Spanien, die Aristotle University in Griechenland und die Technische Universität München. Im Bereich der Sprachverarbeitung sind drei amerikanische Einrichtungen darunter Microsoft Inc. und IBM Inc. und zwei chinesische Universitäten vertreten. Die fünf ersten Plätze in der Signalverarbeitung teilen sich drei chinesische Universitäten und zwei amerikanische Universitäten auf.

Tabelle 8: Führende Einrichtungen ML-Anwendungsbereiche, 2006-2016

Bild- und Videoverarbeitung	**Audioverarbeitung**	**Sprachverarbeitung**	**Signalverarbeitung**
1. Tsinghua University (China) 2. Carnegie Mellon University (USA) 3. Shanghai Jiao Tong University (China) 4. Huazhong University of Science and Technology (China) 5. Chinese Academy of Sciences (China)	1. Universitat Pompeu Fabra (Spanien) 2. Carnegie Mellon University (USA) 3. Aristotle University of Thessaloniki (Griechenland) 4. Technische Universitat München (Deutschland) 5. Inha University, Incheon (China)	1. Tsinghua University (China) 2. Carnegie Mellon University (USA) 3. Microsoft Inc. (USA) 4. IBM Inc.(USA) 5. Harbin Institute of Technology (China)	1. Zhejing University (China) 2. Tsinghua University (China) 3. Carnegie Mellon University (USA) 4. Harbin Institute of Technology (China) 5. University of Southern California (USA)

3.3 Maschinelles Lernen in Deutschland – Publikationen

In Deutschland ist die Anzahl der ML-Publikationen in Fachzeitschriften von 427 in 2006 auf 627 in 2016 angewachsen. Die Konferenzbeiträge sind im gleichen Zeitraum von 266 auf 617 angewachsen. Zum Vergleich sind die zitierbaren Publikationen aus Deutschland über alle wissenschaftlichen Disziplinen von 115 542 auf 149 645 angestiegen und in den Computerwissenschaften von 11 088 auf 18 691.[158]

Deutsche Forscherinnen und Forscher leisteten insbesondere Pionierarbeit bei der Entwicklung von Support-Vektor-Maschinen (SVM). Weiterhin werden hier auch klassische ML-

[158] Vgl. http://www.scimagojr.com/worldreport.php

Technologien erforscht, verbessert und weiterentwickelt. Ein Anwendungsschwerpunkt in Deutschland sind Datenanalysen für die Industrie 4.0.

3.3.1 Forschungseinrichtungen und Unternehmen

Deutschland verfügt mit etwa 1 000 öffentlich finanzierten Einrichtungen über eine hohe Dichte an Forschungseinrichtungen.[159] Das ML-Publikationsaufkommen ist stark konzentriert. Quantitativ betrachtet verantworten die 20 publikationsstärksten Einrichtungen über 40% der ML-Publikationen in wissenschaftlichen Zeitschriften. In Tabelle 9 sind diese 20 Wissenschaftseinrichtungen abgetragen, gemessen an den für das Forschungsfeld relativ wichtigeren Konferenzbeiträgen, zusammen mit häufig genannten Forschungsthemen.[160] Hierzu zählen u.a. die Max-Planck-Gesellschaft (MPG), zu nennen sind insbesondere das Max-Planck-Institut für Informatik, das Max-Planck-Institut für biologische Kybernetik und das Max-Planck-Institut für Intelligente Systeme und die Helmholtz-Gemeinschaft (HGF), insbesondere das Deutsche Zentrum für Luft- und Raumfahrt (DLR) sowie das Karlsruher Institut für Technologie (KIT). Mit einer starken Anwendungsorientierung befinden sich die Fraunhofer-Gesellschaft (FhG) und Deutsche Forschungszentrum für Künstliche Intelligenz (DFKI) unter den führenden Wissenschaftseinrichtungen. Das DFKI publiziert auch stark in den Bereichen »Computer-Linguistik«, »Mensch-Maschine-interaktion« und »Semantische Techniken«, hinter denen sich neben ML auch andere Techniken verbergen können. Bei der Einordnung und Bewertung der Einrichtungen müssen sowohl Größe als auch Zielrichtung der Organisationen berücksichtigt werden.

[159] https://www.research-in-germany.org/en/research-landscape/facts-and-figures.html

[160] Die jeweiligen Forschungsschwerpunkte wurden anhand der Keywords zu konkreten ML-Lernverfahren und Anwendungsbereichen ermittelt, mit denen die Publikationen verschlagwortet wurden. Generell ist bei den Keywords zu den Publikationen aller betrachteten Akteure auffällig, dass hier sehr viele Begriffe zu den Themen Gesundheit, Medizin und Lebenswissenschaften zu finden sind, z.B.: Alzheimer, Biomarker, DNA, fMRI, EEG, Gehirn, Genetik, Krebs, Neurowissenschaften, Prothetik etc., die hier nicht explizit als Anwendung aufgelistet wurden. Ebenfalls wurden häufig vorkommende allgemeine Begriffe wie bspw. supervised learning, clustering oder data analysis nicht mitgerechnet. Ähnliche Begriffsgruppen wurden hier unter einem verallgemeinernden Oberbegriff zusammengefasst. So können sich bspw. unter »Vorhersagemodelle« unterschiedliche Bereiche verbergen (z.B. Maschinenausfälle, Erdbeben, Klima/Wetter, Finanztransaktionen etc.)

Tabelle 9: Die publikationsstärksten deutschen FuE-Einrichtungen im Bereich Maschinelles Lernen

	Publikationen 2006-2016			**Häufig genannte Forschungsthemen der Publikationen im Zeitraum 2014-2016**	
Forschungseinrichtun[161]	**Fachzeit-schriften**	**Konferenz -beiträge**	**Summe**	**Lernverfahren und Aufgaben**	**Anwendungen**
Deutsches Forschungszentrum für Künstliche Intelligenz DFKI	24	83	107	SVM; Vereinfachung von Repräsentationen, (weitere: Ensemble Learning, Merkmals-extraktion, KNN/Deep Learning)	Mustererkennung; Bildverarbeitung; Text- und Wissensanalyse, semantische Technologien
Fraunhofer-Gesellschaft	127	256	383	KNN/Deep Learning; SVM (weitere: Merkmalsextraktion, PCA; Bayessche Modelle, statistisches ML; Entscheidungs-bäume)	Aktivitäts- und Bewegungsanalyse; Muster- und Objekterkennung; Bildverarbeitung/ Analyse (vision); Vorhersageana-lysen; Text-, Audio, Video- und Geräuschanalyse
Friedrich-Alexander-University Erlangen-Nürnberg (FAU)	116	85	201	KNN/Deep Learning; SVM/Kernmethoden (weitere: Bayessche Netze, Feature Maps, inverse Probleme)	Signalverarbeitung, Industrie 4.0, Sensordatananalyse, Semantik, Stimmerkennung
Gottfried Wilhelm Leibniz Universitat	36	79	114	Neuronale Netze (weitere: active learning, Entscheidungsbäume)	Text- und Semantik-Analysen, linked data, Sensordatan-analysen
Helmholtz-Gemeinschaft (HGF)	374	292	666	KNN/Deep Learning; SVM; Merkmalsextraktion (weitere:: Bayessche Netze, statistisches ML; Active Learning; Entscheidungsbäume)	Hyperspektraldaten-Analyse, Spektroskopie; Fernerkundung (Remote Sensing); Bildverarbeitung
Ludwig-Maximilians-Universität Munchen (LMU)	172	84	256	SVM, Kernmethoden; statistisches ML (weitere: KNN, Ensemble Methoden, Markov-Modelle)	Visualisierungen; Bild Sprach-, und Textverarbeitung
Max-Planck-Gesellschaft (MPG)	462	207	669	KNN; SVM (weitere: Statistische ML-Methoden; Bayessche Methoden)	Gesichts-, Muster- und Objekt-erkennung; Gehirn-Computer-Schnittstellen (BCI); Vorhersagemodelle

[161] Alphabetisch sortiert.

Otto von Guericke Universität Magdeburg	100	122	222	SVM (weitere: KNN; genetische/ evolutionäre Algorithmen; Merkmalsextraktion)	Mustererkennung; (IT)-Forensik; Bildverarbeitung (Vision)
RWTH Aachen	114	169	283	KNN/Deep Learning; (weitere: Markov Modelle; Statistische ML-Methoden)	Sprach- und Textanalyse, semantische Analysen; Bildanalysen; Aktivitätsanalysen; Energiesysteme
Technische Universität Berlin	100	125	225	Entscheidungsbäume; Deep Learning (weitere: SVM/ Kernmethoden; Markov Modelle)	Gehirn-Computer-Schnittstellen (BCI); Mustererkennung; Signalverarbeitung
Technische Universität Darmstadt	66	141	207	Neuronale Netze, markov-Modelle (weitere: Active Learning, SVM)	Aktivitätserkennung, knowledge discovery, Textanalyse, Agentensysteme, ML-Games, akustische Sensoranalyse
Technische Universität Dortmund	53	102	154	Transfer-Learning (weitere: Active Learning, KNN/Deep Learning, SVM)	Ereignisdetektion, Wort- und Geräuscherkennung, Bild-, Form- und Szenenerkennung, Modellierungen (Graphik, Verkehr)
Technische Universität Dresden	111	95	206	Active Learning, SVM	Text- und Inhalts-, Medienanalyse, knowledge discovery, Geräuschanalyse
Technische Universitat München	228	274	502	KNN/Deep Learning; Active Learning; (weitere: SVM; Entscheidungs-bäume; bestärkendes Lernen)	Bildanalyse; Sprachanalyse; Stimmungsanalyse; Emotionserkennung; Aktivitätsvorhersage (activity prediction)
Universität Bonn	177	66	243	SVM (weitere: KNN)	Aktivitätsvorhersage (activity prediction); Bewegungsanalyse
Universität Bremen	65	71	136	Neuronale Netze, SVM	Agentensysteme, Affective Computing/ Emotionsanalyse, Detektionssysteme und Vorhersage-modelle (industrielle Produktion, Fehler, Kopien, Schadstoffe, Erdbeben)

Universität Freiburg im Breisgau	100	101	201	SVM, KNN/Deep Learning (weitere: Active Learning, Randon Forests)	Sentiment-Analyse, Entscheidungsunterstützung, Aktivitätserkennung, autonome Systeme/Roboter
Universität des Saarlandes	64	64	128	SVM, Neuronale Netze	Data-/Text-,Social Media Analysen, Eneriemanagementsysteme
Universität Stuttgart	52	102	154	Neuronale Netze, SVM	Vorhersagemodelle, Optimierung und Steuerung (Produktion), Textanalysen, Wissensintegration
Universität Ulm	77	71	148	SVM/Kernmethoden, KNN/Deep Learning	Affective Computing/ Emotionserkennung, Sprach- und Textanalyse, Signalverarbeitung

Unter den vierzehn publikationsstärksten Hochschulen befinden insgesamt sechs Technische Universitäten, die RWTH Aachen, TU Berlin, TU Darmstadt, TU Dortmund, TU Dresden und TU München.

Unter den außeruniversitären Forschungseinrichtungen sind folgende fünf Institute der Max-Planck-Gesellschaft besonders publikationsstark:

- Max-Plank-Institut für biologische Kybernetik in Tübingen
- Max-Plank-Institut für Informatik in Saarbrücken
- Max-Plank-Institut für Intelligente Systeme in Stuttgart
- Max-Plank-Institut für Kognitions- und Neurowissenschaften in Leipzig
- Max-Planck-Institut für molekulare Genetik in Berlin

Als Teil der Helmholtz-Gemeinschaft Deutscher Forschungszentren sind folgende fünf Institute besonders publikationsstark:

- Karlsruher Institut für Technologie (KIT) in Karlsruhe
- Deutsche-Zentrum für Luft- und Raumfahrt (DLR) in Köln
- Forschungszentrum Jülich (FZJ) in Jülich
- Deutsches Forschungszentrum für Gesundheit und Umwelt (HMGU) in München
- Helmholz-Zentrum für Umweltforschung (UFZ) in Leipzig
-

Die fünf publikationsstärksten Institute der Fraunhofer-Gesellschaft sind:

- Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS in Sankt Augustin
- Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB in Karlsruhe
- Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS in Berlin
- Fraunhofer-Institut für Integrierte Schaltungen IIS in Erlangen
- Fraunhofer-Institut für Hochfrequenzphysik und Radartechnik FHR in Wachtberg

Darüber hinaus sind als weitere außeruniversitäre Forschungseinrichtungen das Deutsche Krebsforschungszentrum (DKFZ) in Heidelberg und das Deutsche Forschungszentrum für Künstliche Intelligenz (DFKI) in Saarbrücken zu nennen.

In Tabelle 10 sind jeweils die fünf publikationsstärksten Einrichtungen nach ML-Anwendungsbereichen in Deutschland abgetragen. Hervorzuheben ist der signifikante Anteil an ML-Publikationen von Universitätskliniken, der auf einen engen Zusammenhang zwischen dem ML-Anwendungsbereich »Bild- und Videoverarbeitung« sowie dem Gesundheitssektor schließen lässt. In der deutschen Forschungslandschaft sind insbesondere die Charité, das Universitätsklinikum Freiburg, das Universitätsklinikum Hamburg-Eppendorf oder das DKFZ zu nennen.

Tabelle 10: Publikationsstärkste Einrichtungen ML-Anwendungsbereiche in Deutschland, 2006-2016

Bild- und Videoverarbeitung	**Audioverarbeitung**	**Sprachverarbeitung**	**Signalverarbeitung**
• TU München • Ludwigs-Maximilians-Universität München • Universität Magdeburg • Universität Heidelberg • Universität Nürnberg-Erlangen	• TU München • Humboldt Universität Berlin • Universität Lübeck • Universität Oldenburg • Universität Heidelberg	• Humboldt Universität Berlin • Technische Universität Dresden • Max-Planck-Institut für Kognitions- und Neurowissenschaften • Universität Bonn • Universität Tübingen	• Technische Universität München • Deutsches Zentrum für Luft- und Raumfahrt (HGF) • Karlsruher Institut für Technologie (HGF) • Rheinisch-Westfälische-Technische Hochschule • Technische Universität Berlin

Bei den fünf publikationsstärksten Unternehmen sind es in Bayern die Siemens AG, in Hessen die Honda Motor Europe (North) GmbH sowie in Baden-Württemberg die Robert Bosch GmbH SAP SE und Daimler AG.

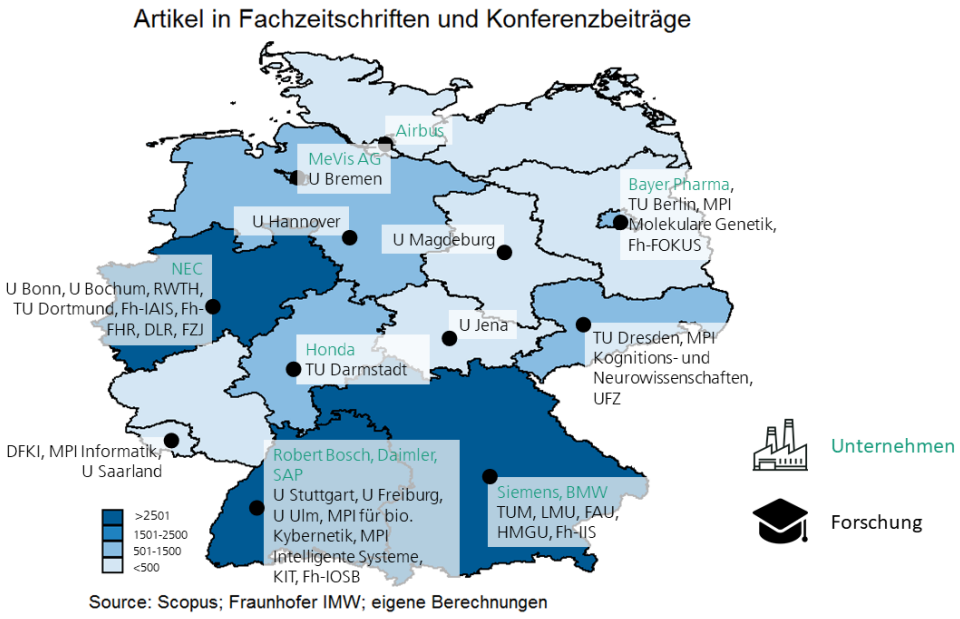
Tabelle 11: Publikationsstärkste Unternehmen in Deutschland, 2006-2016

Organisation	Branche	Fachzeit-schriften	Konferenz-beiträge	Summe
Siemens AG	Technologie	47	49	96
Honda Motor Europe (North) GmbH	Automobil	18	20	38
SAP SE	Software	2	29	32
Robert Bosch GmbH	Technologie	6	20	26
Daimler AG	Automobil	4	14	18
MeVis Medical Solutions AG	Informations- und Kommunikation	12	5	17
BMW AG	Automobil	3	12	15
Airbus Group	Luft- und Raumfahrt	2	12	14
NEC Deutschland GmbH	Informations- und Kommunikation	4	9	12
Bayer AG	Chemie und Pharma	11	1	12

3.3.2 Regionen und deren Akteure

In absoluten Zahlen sind die Bundesländer Baden-Württemberg, Nordrhein-Westfalen und Bayern die publikationsstärksten Bundesländer im Bereich der ML-Technologien. Es folgen Niedersachsen, Berlin, Hessen, und Sachsen. Die Gruppe der publikationsschwächeren Bundesländer bilden Rheinland-Pfalz, Saarland, Thüringen, Sachsen-Anhalt, Hamburg, Bremen, und Schleswig-Holstein, Brandenburg und Mecklenburg-Vorpommern.

Abbildung 51: Publikationen im Bereich Maschinelles Lernen nach Regionen, 2006-2016



Zu den 20 publikationsstärksten Hochschulen und Forschungseinrichtungen, gemessen in Konferenzbeiträgen, zählen in Baden-Württemberg das Karlsruhe Institut für Technologie (KIT), die Universität Freiburg, die Universität Stuttgart, die Universität Ulm sowie das Max-Planck-Institut für biologische Kybernetik und das Fraunhofer- Institut für Optronik, Systemtechnik und Bildauswertung IOSB. In Nordrhein-Westfalen sind unter den Hochschulen und Forschungseinrichtungen die Universität Bonn, die Ruhr-Universität Bochum, die RWTH Aachen, die Technische Universität Dortmund, das Deutsche-Zentrum für Luft- und Raumfahrt (DLR), Forschungszentrum Jülich (FZJ) sowie das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS und das Fraunhofer-Institut für Hochfrequenzphysik und Radartechnik FHR zu nennen. In Bayern sind die TU München (TUM), die Ludwig-Maximilian-Universität München (LMU) sowie Deutsches Forschungszentrum für Gesundheit und Umwelt in München (HMGU) und das Fraunhofer-Institut für Integrierte Schaltungen IIS anzuführen.

In Berlin gehören die TU Berlin sowie das Max-Planck-Institut für molekulare Genetik und das Institut für Offene Kommunikationssysteme FOKUS zu den publikationsstärksten Hochschulen und Forschungseinrichtungen. In Sachsen sind die TU Dresden und die Universität Leipzig sowie das Max-Plank-Institut für Kognitions- und Neurowissenschaften und das Helmholz-Zentrum für Umweltforschung (UFZ) anzuführen. In Hessen gehört die Technische Universität Darmstadt, in Bremen die Universität Bremen sowie die Universität Hannover in Niedersachsen zu den quantitativ publikationsstärksten Hochschulen und Forschungseinrichtungen.

In der Gruppe der verbleibenden Bundesländer befinden sich weitere publikationsstarke Akteure wie das Deutsche Forschungszentrum für Künstliche Intelligenz (DFKI) und das Max-Plank-Institut für Informatik in Saarbrücken, die Friedrich-Schiller-Universität Jena in Thüringen und die Otto-von-Guericke-Universität Magdeburg in Sachsen-Anhalt.

Insgesamt zeigen die Publikationsstatistiken, dass die ML-Forschung der Wissenschaftseinrichtungen in Deutschland auf einem breiten regionalen Fundament steht und im europäischen Vergleich gut aufgestellt ist. Ein ernst zu nehmendes Problem für die Wettbewerbsfähigkeit Deutschlands sind die fehlenden spezialisierten Daten-, ML- und KI-Fachkräfte (siehe hierzu Kapitel 4 und 5).

3.4 Analyse der ML-Projektförderung auf europäischer Ebene

3.4.1 ML-Projektförderung in den EU-Rahmenforschungsprogrammen

Die Europäische Kommission fördert durch die europäische Verbundförderung die Zusammenarbeit der Mitgliedstaaten und begleitet die Koordinierung der nationalen Forschungspolitiken. Im 7. Forschungsrahmenprogramm (FP7) wurden zwischen 2007 und 2014 insgesamt 25 054 Projekte mit einer Summe von rund 44 Mrd. Euro finanziert. Aktuell läuft das 8. Forschungsrahmenprogramm (Horizon 2020), in dem bereits 11 048 Projekte mit einer Summe von über 20 Mrd. Euro gefördert wurden.

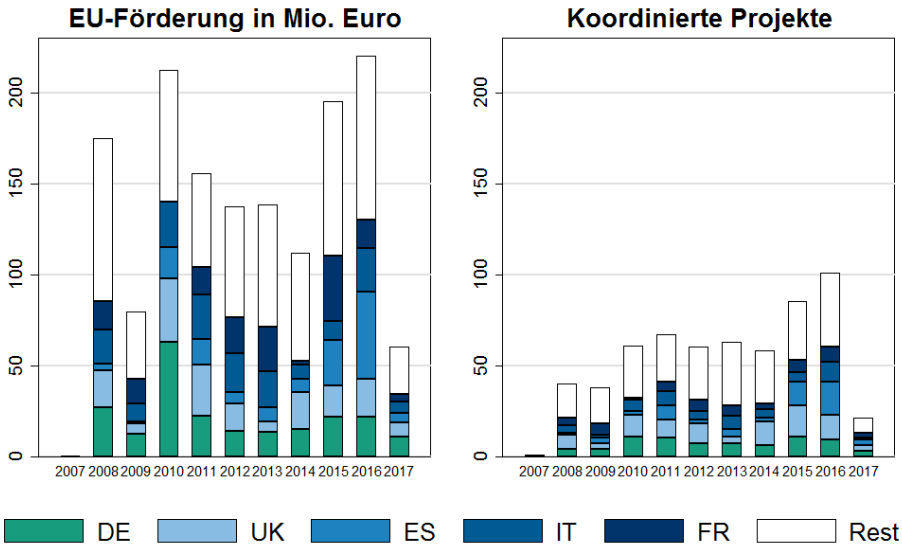
Insgesamt konnten für den Untersuchungszeitraum 595 ML-Forschungsprojekte mit einer EU-Förderung von 1,49 Mrd. Euro identifiziert werden. Die Anzahl der EU-geförderten ML-Forschungsprojekte stieg von 40 in 2008 auf 101 in 2016 stetig an. Die EU- Fördersumme von 175 Mio. Euro aus dem Jahr 2008 fiel im Jahr 2009 auf 80 Mio. Euro und stieg im darauffolgenden Jahr auf 213 Mio. Euro an. Anzumerken ist der Rückgang der EU-Förderung von 213 Mio. Euro in 2010 auf nur noch 112 Mio. Euro in 2014. Erst ab 2015, im Zuge des neuen Forschungsrahmenprogramm H2020, steigt die EU-Förderung für ML-Forschungsprojekte auf 220 (2016) Mio. Euro erneut an. Die Anzahl der Projekte hielt sich in den Jahren von 2010 bis 2014 die Waage und schwankte zwischen 61 (2010) und 58 (2014).

Es fällt auf, dass die durchschnittliche Förderung mit 0,5 Mio. Euro pro Projekt in FP7 (Förderperiode 2007-2014) deutlich kleiner war als im aktuellen Forschungsrahmenprogramm Horizon 2020 (Förderperiode 2014-2020) mit durchschnittlich 1,6 Mio. Euro pro Projekt. Zukünftig ist von einer zunehmenden Forschungsförderung durch die EU auszugehen, da ML-Technologien in nahezu allen Branchen vermehrt zum Einsatz kommen werden.

3.4.2 Projektkoordinatoren im Ländervergleich

Bei den Koordinatoren von EU-geförderten ML-Forschungsprojekten ist die starke Konzentration auf relativ wenige Forschungseinrichtungen auffällig. So koordinieren mit Deutschland, Großbritannien, Frankreich, Italien und Spanien 5 Länder 56% der Gesamtförderung. Großbritannien mit seinen exzellenten Universitäten und der fokussierten Förderung individueller Forscher (Marie Skłodowska-Curie actions) ist der führende Koordinator von Forschungsprojekten auf europäischer Ebene, gefolgt von Deutschland.

Abbildung 52: Entwicklung EU-Förderung ML-Forschungsprojekte in FP7 und H2020 nach Ländern, 2007-2017



Source: EC-Cordis; Fraunhofer IMW; eigene Berechnungen
Stand Oktober 2017

Im Hinblick auf einzelne Akteure, beanspruchen die in Tabelle 11 aufgelisteten 10 Einrichtungen 17% der gesamten ML-Projektförderung der Europäischen Kommission (9% der Projekte). Außeruniversitäre Forschungseinrichtungen (AuF) koordinieren tendenziell weniger (25 vs. 37) Projekte, dafür jedoch größere Projekte als Universitäten. Gemessen an den zur Verfügung stehenden Fördermitteln ist das Karolinska-Institut aus Schweden der führende Koordinator für ML-Forschungsprojekte, bezüglich der Anzahl koordinierter Vorhaben dominiert die Universität Oxford aus Großbritannien.

Tabelle 12: Führende Koordinatoren EU-Förderung ML-Forschungsprojekte in FP7 und H2020 nach Ländern, 2007-2017

Name	Organisation	Land	Koordinierte Projekte	Förderung in Mio. Euro
Karolinska-Institut	Universität	Schweden	2	30,7
Deutsches Forschungszentrum für Künstliche Intelligenz DFKI	Forschungseinrichtung	Deutschland	5	29,3
Max-Planck-Gesellschaft	Forschungseinrichtung	Deutschland	7	27,2
Universität Pierre und Marie Curie	Universität	Frankreich	4	27,2
Vicomtech	Forschungseinrichtung	Spanien	4	26,2
Universität Edinburgh	Universität	UK	9	25,9

Name	Organi-sation	Land	Koordinierte Projekte	Förderung in Mio. Euro
Fraunhofer Gesellschaft	Forschungs-einrichtung	Deutschland	6	24,9
Universität Mailand	Universität	Italien	3	24,1
Kings College	Universität	UK	4	22,8
Universität Oxford	Universität	UK	11	20,3
Summe			**55**	**258,6**

3.4.3 Deutsche Projektkoordinatoren

Im Hinblick auf die von Deutschland koordinierten EU-Projekte, verteilten sich die insgesamt 72 Projekte und über 220 Mio. Euro EU-Fördermittel auf 43 Akteure, davon 8 Unternehmen, 26 Universitäten und 9 außeruniversitäre Forschungseinrichtungen. Damit koordinieren deutsche Forschungseinrichtungen und Unternehmen rund 10% der gesamten ML-Projektförderung der Europäischen Kommission (1,6% der Projekte).

Den Spitzenplatz belegt das Deutsche Forschungszentrum für Künstliche Intelligenz (DFKI), gefolgt von der Max-Planck-Gesellschaft (MPG) und der Fraunhofer-Gesellschaft (FhG). Insgesamt zwei Unternehmen sind unter den führenden zehn Koordinatoren nach Projektförderung.

Tabelle 13: Führende Koordinatoren EU-Förderung ML-Forschungsprojekte in FP7 und H2020 in Deutschland, 2007-2017

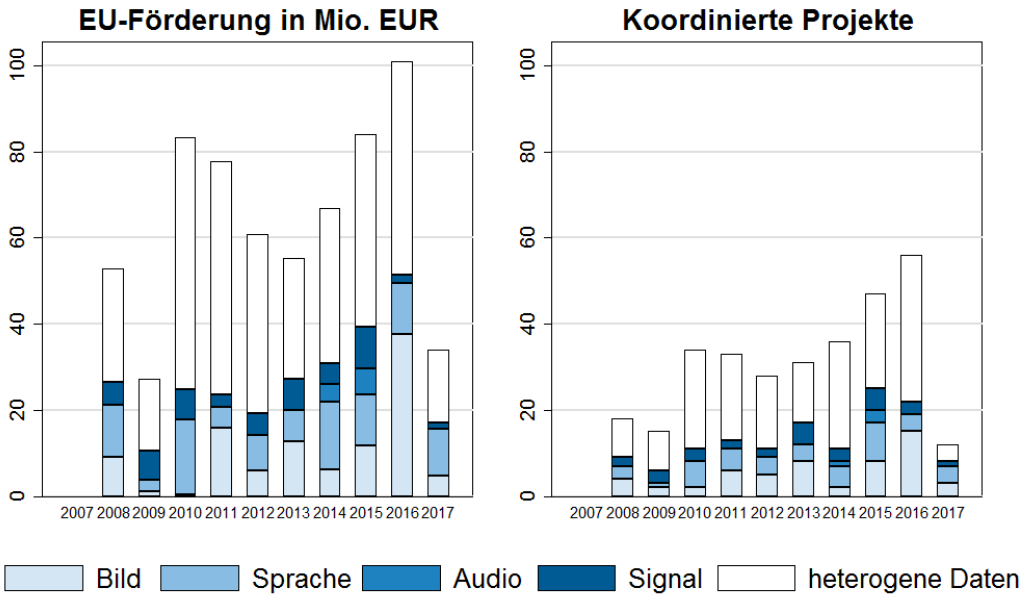
Name	Organi-sation	Ort	Koordinierte Projekte	Förderung in Mio. Euro
Deutsches Forschungszentrum für Künstliche Intelligenz	Forschungsein-richtung	Saarbrücken	5	29,3
Max-Planck-Gesellschaft	Forschungsein-richtung	München	7	27,2
Fraunhofer-Gesellschaft	Forschungsein-richtung	München	6	24,9
SAP SE	Unternehmen	Walldorf	1	12,6
Alacris Theranostics GmbH	Unternehmen	Berlin	1	10,7
Universität Leipzig	Universität	Leipzig	1	9,9
Universität Bielefeld	Universität	Bielefeld	1	9,2
Universität des Saarlandes	Universität	Saarbrücken	3	8,9
Technische Universität München	Universität	München	4	6,3
Albert-Ludwigs-Universität Freiburg	Universität	Freiburg im Breisgau	3	5,5
Summe			**10**	**144,5**

3.4.4 Projektkoordination nach Anwendungsbereichen

Differenziert nach ML-Anwendungsbereichen sind »Bild- und Videoverarbeitung« (13%) sowie »Sprachverarbeitung« (11%) die wichtigsten Anwendungsbereiche der EU-Projektförderung. Im Gegensatz zu wissenschaftlichen Publikationen ist die »Signalverarbeitung« (75%) nicht sehr stark ausgeprägt.

Abbildung 55 zeigt die Ergebnisse der Projektförderung nach ML-Anwendungsbereichen. Insgesamt ist die Abgrenzung der ML-Anwendungsbereiche nach »Bild- und Videoverarbeitung« (13% der Projekte, 12% der EU-Förderung), »Sprachverarbeitung« (11% der Projekte, 12% der EU-Förderung), »Audioverarbeitung« (1% der Projekte, 1% der EU-Förderung) und »Signalverarbeitung« (7% der Projekte, 6% der EU-Förderung) weniger Eindeutig als bei den Publikationen. Viele ML-Forschungsprojekte können der Kategorie »heterogene Daten« (68% der Projekte, 69% der EU-Förderung) zugeordnet werden, sodass für etwas weniger als die Hälfte der 595 geförderten EU-Projekte ein eindeutiger Bezug zu den ML-Anwendungsbereichen herstellt werden kann. Im Gegensatz zu wissenschaftlichen Publikationen ist die »Signalverarbeitung« nicht sehr stark ausgeprägt.

Abbildung 53: Entwicklung der EU-Förderung nach ML-Anwendungsbereichen, 2007-2017



Source: EC-Cordis; eigene Berechnungen
Stand Oktober 2017

Von Deutschland werden insbesondere Projekte zur »Sprachverarbeitung« (22% der Projekte, 26% der EU-Förderung) koordiniert. Mit 65% ist der Anteil der ML-Forschungsprojekte mit »heterogenen Daten« (67% der Projekte, 65% der EU-Förderung) relativ hoch. Großbritannien dagegen koordiniert überwiegend ML-Forschungsprojekte im Bereich »Bild- und Videoverarbeitung« (17% der Projekte, 12% der EU-Förderung), wobei

der Anteil der ML-Forschungsprojekte mit »heterogenen Daten« (70% der Projekte, 61% der EU-Förderung) ebenfalls relativ hoch ist.

3.5 ML-Patentanalyse

3.5.1 Generelle Patententwicklung

Patente sind ein wertvoller Indikator, um die technologische Leistungsfähigkeit von Forschungseinrichtungen, Unternehmen, Regionen oder Ländern nachzuvollziehen und Technologietrends abzuleiten. Untersuchungen zeigen, dass sich die Patentierbarkeit von ML-Technologien in den verschiedenen Patentsystemen stark unterscheidet. Die Espacenet-Patentdatenbank des Europäischen Patentamts bildet die Grundlage für die hier verwendete statistische Analyse von Patentfamilien[162]. Weltweit konnten 7 252 Patentfamilien für ML-Technologien identifiziert werden, die aus 45 386 nationalen und internationalen Patentanmeldungen bestehen. Insgesamt kann für die Entwicklung von Patentanmeldungen ein stetiges Wachstum konstatiert werden. So sind die weltweiten Patentanmeldungen für ML-Technologien von 487 Patentfamilien in 2006 auf 1 258 Patentfamilien in 2015 angestiegen. Die Patentdynamik im Bereich ML folgt dem weltweiten Trend, so stiegen die Patentfamilien über alle Technologien von 901 973 in 2006 auf 1,55 Mio. in 2015 kontinuierlich an.[163]

In den Jahren von 2006 bis 2008 wuchs die Zahl der ML-Patentfamilien kontinuierlich von 487 auf 683 an. Empirische Untersuchungen deuten darauf hin, dass zahlreiche Unternehmen ihre Innovationsausgaben und damit auch die Ausgaben für Patentanmeldungen im Zuge der Finanz- und Wirtschaftskrise konstant gehalten oder zurückgefahren haben.[164] Damit verbunden lässt sich der Rückgang der Patentanmeldungen in 2009 und 2010 in Abbildung 56 erklären. Seit 2010 stiegen die Patentanmeldungen erneut an und erreichten mit 1 258 Patentfamilien in 2015 einen vorläufigen Höhepunkt.

3.5.2 ML-Patente im internationalen Vergleich

Abbildung 57 zeigt, dass die meisten Patentfamilien aus den USA, China und Südkorea stammen. Sie werden von den Unternehmen Microsoft, IBM, Google, Amazon, Cisco, Qualcomm (USA) sowie Samsung, Korea Electronics Telecomm (Südkorea), Huawei und ZTE (China) dominiert. Diese drei Länder vereinen insgesamt über 75% der weltweiten Patentfamilienanmeldungen im Bereich ML. Aus Japan wurden im Untersuchungszeitraum

[162] Patentfamilien setzen sich aus einer oder mehreren nationalen oder internationalen Patentanmeldungen zusammen und schützen die gleiche technologische Erfindung in unterschiedlichen Patentsystemen. Analog zum oben skizzierten Ansatz für die Publikationen wurden auch hier die Felder Titel und Zusammenfassung der Datenbankeinträge semantisch durchsucht. Die Suche beschränkte sich auf Patentfamilien im Zeitraum 2006-2015.

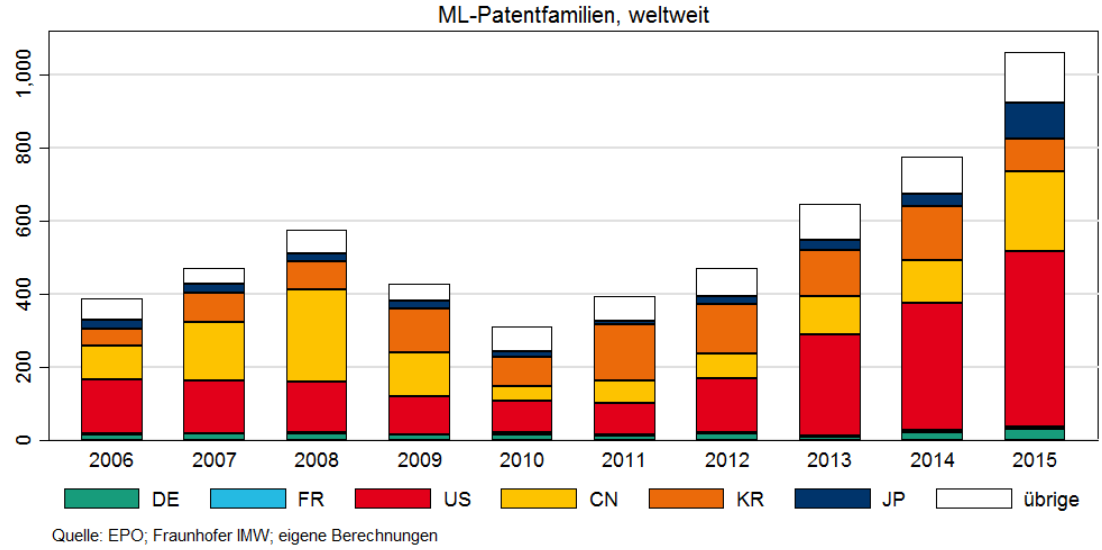
[163] WIPO statistics database. Last updated: December 2017.

[164] http://www.e-fi.de/fileadmin/Inhaltskapitel_2011/2011_C5.pdf

438 ML-Patentfamilien angemeldet und aus Deutschland insgesamt 236. Zusammen stehen die fünf untersuchten Länder für rund 90% aller Patentfamilien für ML-Technologien.

Im internationalen Vergleich auffällig ist, dass die USA in den Jahren von 2006 (189) bis 2010 (170) ein leicht schwankendes Patentverhalten aufweisen, das ab dem Jahr 2010 (242) stark an Dynamik gewinnt. Insbesondere der starke Anstieg zwischen 2010 und 2015 (594) ist hervorzuheben. China dagegen hat in den Jahren 2007 (161) und 2008 (255) relativ viele ML-Technologien zum Patent angemeldet. In den Folgejahren bricht das Patentverhalten zunächst stark ein und gewinnt erst im Jahr 2013 (112) wieder an Dynamik. Das Patentverhalten aus Südkorea zeigt auf hohem Niveau eine konstante Entwicklung mit dem Höhepinkt im Jahr 2011 (171). In den darauffolgenden Jahren schwankt es zwischen 153 und 167 Anmeldungen. Während das Patentverhalten aus Japan in den Jahren 2006 (38) bis 2013 (43) im Vergleich zu den Wettbewerbern gering ausfiel, sind die Patentfamilien dennoch von 37 in 2012 auf 110 in 2015 angestiegen.

Abbildung 54: Entwicklung der Patentfamilien im Bereich Maschinelles Lernen nach Ländern, 2006-2015



Quelle: EPO; Fraunhofer IMW; eigene Berechnungen

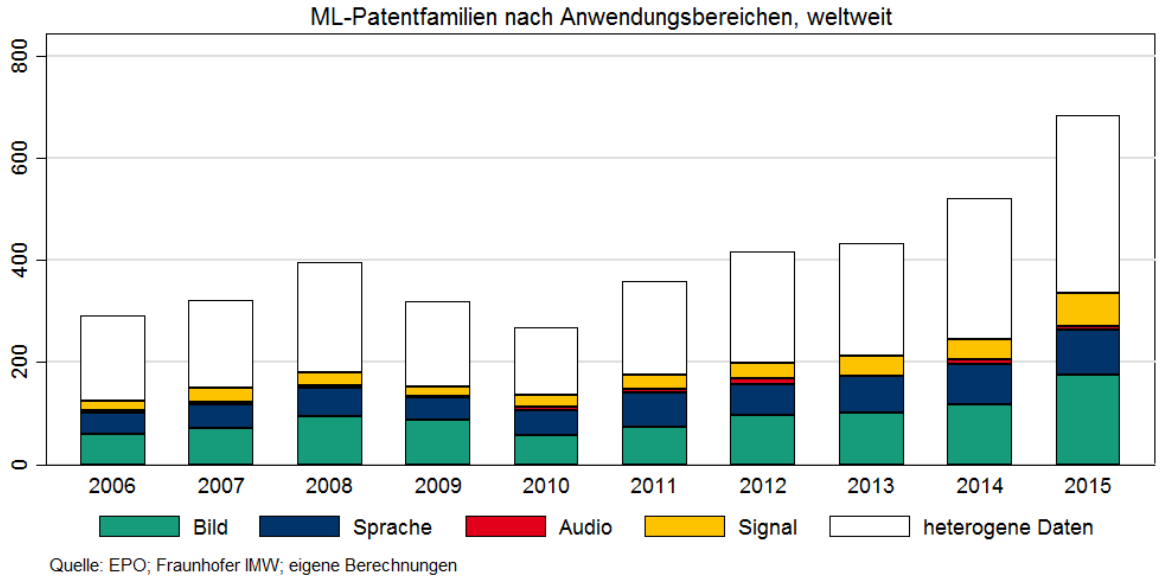
ML-Technologien sind eine bedeutende Grundlage für die Entwicklung kognitiver Maschinen mit sogenannten »kognitiven Computing Fähigkeiten«. Aus diesem Grund investieren Firmen wie bspw. Microsoft, Google, Amazon, Cisco, Yahoo und Facebook in den USA, Samsung in Südkorea oder Huawei bzw. ZTE in China in ML-Technologien bzw. kaufen entsprechende Startup-Unternehmen auf. Auffallend ist, dass in Südkorea und China die Hochschulen und Forschungseinrichtungen zu den aktivsten Patentanmeldern gehören, während in den USA und Deutschland fast ausschließlich Unternehmen in der ML-Patentstatistik vertreten sind. Die Unterscheidung zwischen akademischen und unternehmerischen Patenten erlaub teilweise Rückschlüsse auf die technologische Reife der patentierten Technologien.

Die Patentdynamik deutscher Akteure ist kaum vergleichbar mit den Hauptwettbewerbern USA, China oder Südkorea: Abbildung 54 zeigt, dass die Zahl deutscher Patentfamilien von 21 im Jahr 2006 auf 36 im Jahr 2015 anstieg. Allerdings müssen die Ergebnisse der Patentanalyse im Kontext der Wirtschaftsstruktur der untersuchten Länder interpretiert werden. Die Stärken der deutschen Wirtschaft liegen in Technologiebereichen wie dem Maschinen- und Anlagenbau, Automobil-Industrie sowie der chemischen- und pharmazeutischen Industrie, so dass potentiell weniger deutsche Unternehmen ML-Technologien zum Patent anmelden als die Hauptwettbewerber aus den USA oder aus Ostasien.

3.5.3 Patententwicklung in Anwendungsbereichen

Abbildung 55 zeigt die weltweite Patentdynamik in Bezug auf die relevanten ML-Anwendungsfelder »Bild- und Videoverarbeitung«, »Sprachverarbeitung«, »Audioverarbeitung« und »Signalverarbeitung«. Insgesamt wurden 4 049 Patentfamilien für die ML-Anwendungsbereiche identifiziert, wobei ein stetiger Anstieg von 292 Patentfamilien in 2006 auf 676 in 2015 festzuhalten ist. Ähnlich wie bei der Publikationsanalyse zeigt sich, dass der ML-Anwendungsbereich »Bild- und Sprachverarbeitung« überproportional stark patentiert wird. Nahezu die Hälfte der ML-Anwendungsbereiche entfällt auf die Kategorie »heterogene Daten«, also auf die Kombination von zwei oder mehr Anwendungsbereichen.

Abbildung 55: Entwicklung der Patentfamilien für ML-Anwendungsbereiche, 2006-2015



Quelle: EPO; Fraunhofer IMW; eigene Berechnungen

Ähnlich wie bei der Publikationsanalyse zeigt sich, dass der ML-Anwendungsbereich »Bild- und Sprachverarbeitung« überproportional stark patentiert wird. Außerdem entfällt nahezu die Hälfte der ML-Anwendungsbereiche auf die Kategorie »heterogene Daten«, also auf die Kombination von zwei oder mehr Anwendungsbereichen, die ebenfalls Sprach- und Bilddaten umfassen können.

Tabelle 14: ML-Patentfamilien nach Anwendungsbereichen, 2006-2015

	Bild- und Video-verarbeitung	**Audiover-arbeitung**	**Sprachver-arbeitung**	**Signalver-arbeitung**	**Heterogene Daten**
Welt	932	64	602	310	2141
	23%	2%	15%	8%	53%
Deutschland	**43**	**4**	**9**	**18**	**66**
	31%	**3%**	**6%**	**13%**	**47%**
USA	298	14	241	147	789
	20%	1%	16%	10%	53%
China	140	56	8	20	530
	19%	7%	1%	3%	70%
Südkorea	145	20	143	41	437
	18%	3%	18%	5%	56%
Japan	75	2	35	12	116
	31%	1%	15%	5%	48%

Aus Deutschland werden insbesondere ML-Technologien für die Bild- und Videoverarbeitung (31%), Signalverarbeitung (13%) und Sprachverarbeitung (6%) durch Patente geschützt. Führend sind eindeutig die USA mit 1 486 Patentfamilien. Der Fokus liegt hier ebenfalls auf der Bild- und Videoverarbeitung (20%) und der Sprachverarbeitung (16%). Die Daten für China zeigen, dass die Anbieter einen starken Fokus auf ML-Anwendungsbereiche »Bild- und Videoverarbeitung« (19%) legen, gefolgt von der »Audioverarbeitung« (7%). Die Sprach- und Signalverarbeitung spielt für chinesische Hersteller eine untergeordnete Rolle, wobei hier zu bedenken ist, dass die chinesische und englische Sprache auch kaum vergleichbar sind. Für Südkorea kann eine ähnliche Entwicklung festgehalten werden, so werden dort vorwiegend Erfindungen zur Bild- und Videoverarbeitung (18%) und Audioverarbeitung (18%) zum Patent angemeldet. Aus Japan werden insbesondere ML-Technologien für die Bild- und Videoverarbeitung (31%), Sprachverarbeitung (15%) durch Patente geschützt. Entgegen dem Patentverhalten der asiatischen Länder ist in Deutschland und den USA die Signalverarbeitung noch von Bedeutung.

Tabelle 15: Patentanmeldestärkste Einrichtungen ML-Anwendungsbereiche weltweit, 2006-2015

Bild- und Videoverarbeitung	Audioverarbeitung	Sprachverarbeitung	Signalverarbeitung
1. Microsoft Corp. (USA) 2. Samsung Group Ltd. (Südkorea) 3. Sony Corp. (Japan) 4. Chinese Academy of Sciences (China) 5. Adobe Systems Inc. (USa)	1. University of Korea (Südkorea) 2. Amazon Inc. (USA) 3. Samsung Group (Südkorea) 4. Inha-Industry Partnership Institute (Südkorea) 5. IBM Corp.(USA)	1. Microsoft Corp. (USA) 2. IBM Corp. (USA) 3. Korea Electronics Telecomm (Südkorea) 4. Yahoo Inc. (USA) 5. Google Inc. (USA)	1. Sensormatic Electronics (USA) 2. Microsoft Corp. (USA) 3. Robert Bosch GmbH (Deutschland) 4. Hyundai Motor Co. Ltd (Südkorea) 5. IBM Corp. (USA)

3.5.4 Regionen und deren Akteure - Patentanmeldungen

In Abbildung 56 sind die 20 patentanmeldestärksten Einrichtungen zu erkennen. Die patentstärksten Regionen in Deutschland sind, wie bei den Publikationen Bayern und Baden-Württemberg, gefolgt von Nordrhein-Westfalen und Hessen. Unter den 20 führenden ML-Patentanmeldern finden sich überwiegend Unternehmen.

Insgesamt konnten für 75 Unternehmen, davon 56 Großunternehmen (> 249 Mitarbeiter) allen voran Siemens AG (40 Patentfamilien), gefolgt von Robert Bosch GmbH (15 Patentfamilien), Deutsche Telekom AG (13 Patentfamilien), Daimler AG (9 Patentfamilien), BMW AG (7 Patentfamilien) und SAP SE (6 Patentfamilien), sowie 15 kleine und Kleinstunternehmen (< 49 Mitarbeiter), aber lediglich 4 mittlere Unternehmen (49-249 Mitarbeiter) mit ML-bezogenen Patentfamilien identifiziert werden.

Abbildung 56: Patentfamilien im Bereich Maschinelles Lernen nach Regionen, 2006-2015

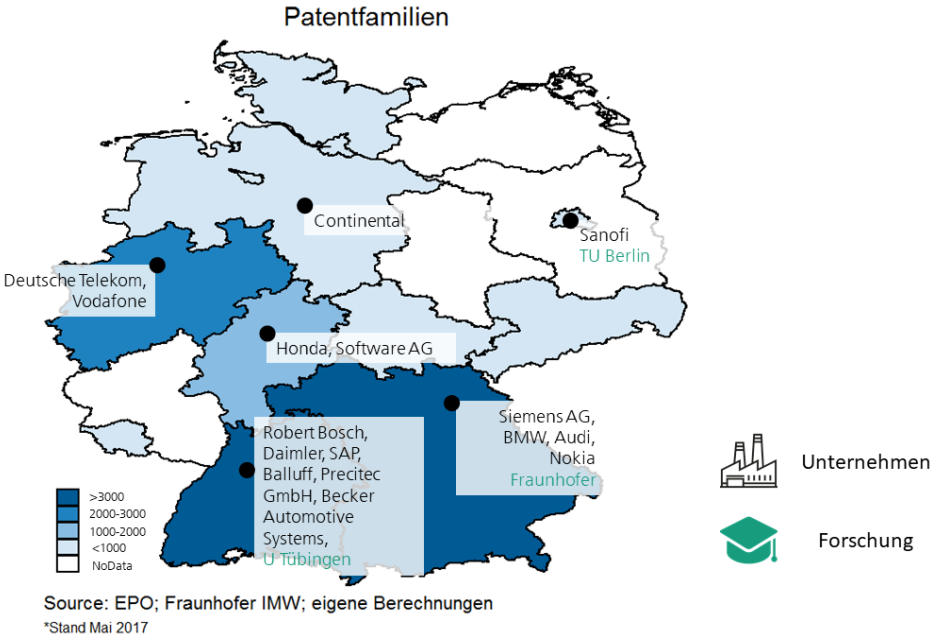


Tabelle 16: Führende Einrichtungen ML-Anwendungsbereiche in Deutschland, 2006-2015

Bild- und Videoverarbeitung	**Audioverarbeitung**	**Sprachverarbeitung**	**Signalverarbeitung**
1. Siemens AG 2. Robert Bosch GmbH 3. Sanofi-Aventis Deutschland GmbH 4. Beckhoff Automation GmbH 5. BASF SE	1. Nero AG 2. Schäffler AG 3. Fraunhofer Gesellschaft 4. Technische Universität Dresden 5. Siemens AG	1. Deutsche Telekom AG 2. Siemens AG 3. Audi AG 4. Becker Automotive Systems GmbH 5. EXB Asset Mgt. GmbH	1. Robert Bosch GmbH 2. Deutsche Telekom AG 3. Siemens AG 4. Brainlab AG 5. Balluff GmbH

Insbesondere stechen die Branchen Gesundheit (Siemens AG, Sanofi- Aventis Deutschland GmbH, Euroimmun), Mobilität (Robert-Bosch GmbH, Daimler AG, BMW AG, Honda Deutschland GmbH, Audi AG, Schäffler AG, Airbus S.A.S.), Industrielle Produktion (Siemens AG, Robert-Bosch GmbH) sowie Informations- und Kommunikationstechnologien (Telekom AG, SAP SE, Vodafone GmbH, Nokia Oyj, Software AG) hervor.

3.6 Maschinelles Lernen: Produkte, Märkte und Wirtschaftsakteure

3.6.1 Generelle Entwicklungen

Mit der Verbreitung der Digitalisierung, dem wachsenden Datenvolumen und steigenden technologischen Anforderungen können herkömmliche technische Ansätze nicht mehr Schritt halten. Führende internationale Hochtechnologie-Unternehmen investieren deshalb

zusehends in die Forschung und Entwicklung von ML-Verfahren, die schnellere, präzisere und effizientere Ergebnisse bei der Analyse großer Mengen heterogener und komplexer Daten liefern und innovative Anwendungen auf den Gebieten der Bild- und Objekterkennung (Computer Vision), Verarbeitung natürlicher Sprache und intelligenter Robotersteuerung ermöglichen. Sie spezialisieren sich dabei auf die Entwicklung von ML-Softwares und -Plattformen, bieten Cloud-basierte Maschine-Learning-as-a-service (ML-aaS) Lösungen an und entwickeln ML-angereicherte Produkte und Dienste für fast alle Branchen[165].

ML-aaS ist ein wachsendes Segment auf dem ML-Markt, das zurzeit von vielen großen internationalen Anbietern[166] (z.B. *Amazon Web Services (AWS), Microsoft - Azure, Google - Cloud Platform, IBM - BlueMix*) vertreten wird. Über die Cloud-Plattformen bieten sie ihren Kunden die Möglichkeit an, bestehende ML-Modelle für spezifische Problemstellungen anzupassen oder neue Modelle zu entwickeln. Mit umfangreichen Speicherdiensten und Tools ermöglichen Cloud-Plattformen Flexibilität beim Verarbeiten der sehr schnell wachsenden Datenmengen. Eine aktuelle Studie von Crisp Research hat neben den bereits genannten[167] Anbietern acht weitere internationale ML-aaS Unternehmen identifiziert, die auf dem deutschen Markt aktuell vertreten sind: *Algorithmia, Algorithms.io, Baidu, BigML, Hewlett Packard Enterprise (HPE), Monkeylearn, wit.ai und yottamine*. Der Markt für ML-aaS befindet sich noch in einer frühen Entwicklungsphase: Die Marktführer sind dabei ihre Nische zu etablieren, die Markt-Analytiker erwarten, dass die Dienstleister sich stärker auf einzelne Anwendungen oder Branchen spezialisieren. [168]

Wie bei vielen Softwaretechnologien fand die Einführung von ML-Produkten in den ersten Jahren überwiegend in Nordamerika statt, bevor diese im Rest der Welt zur Anwendung kamen. Die höchsten Wachstumsraten finden gemäß den Prognosen im asiatisch-pazifischen Raum, insbesondere Ostasien, statt.[169] Die globalen Vorreiter auf dem heutigen ML-Markt sind Hochtechnologieunternehmen aus den USA (*z.B. Amazon, Facebook, Google, IBM, Intel, Microsoft, NVIDIA, Tesla*) und China (z.B. *Baidu, Alibaba, Tencent*). Diese Akteure haben einen begünstigten Zugang zu massiven Mengen an Lerndaten und entsprechender Hardware, bauen Forschungsgruppen für künstliche Intelligenz und Deep Learning auf und setzen für das Training von ML-Algorithmen eigene ML-Plattformen[170] ein.

Die deutsche ML-Akteurslandschaft ist heterogen und fragmentiert. Maschinelles Lernen findet sich vermehrt als Thema in Berichten über Strategien von Unternehmen, die ML- und KI-basierte Technologien in ihre Produkte oder Prozesse integrieren oder in eigenen

[165] Nach Böttcher et al. 2017
[166] Frost & Sullivan 2017b
[167] Böttcher et al. 2017, Crisp Research 2017
[168] Böttcher et al. 2017
[169] Groopman & Kaul 2017
[170] Z.B. Microsoft Azure, IBM Watson, Amazon Web Services

FuE-Abteilungen und Forschungsgruppen weiterentwickeln (u.a. *Siemens, Daimler, Bosch, VW, Audi*).

Deutsche Entwickler von Algorithmen und ML-Dienstleister sind große Softwarehäuser wie *SAP SE* oder *SAS Institute GmbH,* sowie viele Kleinstunternehmen und Start-ups. Letztere verfügen über vergleichsweise geringe Ressourcen und bieten spezialisierte ML-Angebote für bestimmte Problemstellungen, einzelne Branchen oder Kunden an. Beispiele hierfür sind die Verwaltung von Kreditverträgen für die Finanzindustrie, Big Data Analytics für Beratungsunternehmen und Managed-Services Anbieter, Sprachassistenten in Automobilen, Analyse von Pflanzenkrankheiten oder Deep Learning für industrielle Anwendungen. Der deutsche Mittelstand ist unter den ML-Anbietern bzw. -Entwicklern hingegen gering vertreten. Die Fachleute weisen darauf hin, dass die Ursachen hierfür weniger in mangelnder grundsätzlicher Bereitschaft liegen, sondern an einem Defizit an fachlichen Kompetenzen sowie dem unzureichenden Zugang zu umfänglichen Datenerhebungen, der für ML essenziell ist.

Auch bei Start-up-Unternehmen sind die USA führend, gefolgt von Europa, Indien und China. In Europa ist Berlin nach London der zweitgrößte Standort für KI-Start-ups. Von rund 90 deutschen KI-Startups[171] haben hier 18 auf Deep Learning fokussierte Unternehmen[172] ihren Sitz. Die international begrenzte Verfügbarkeit von ML- und KI-Talenten und Know-how in diesem Bereich spiegelt sich in einer hohen Akquisitionsaktivität der Hochtechnologieunternehmen wider. Start-ups werden häufig bereits in einem frühen Entwicklungsstadium aufgekauft, noch bevor sie erfolgreiche Geschäftsmodelle mit ihren Produkten auf dem Markt platzieren können.[173] Einer aktuellen Studie der Unternehmensberatung und Investmentbank Clipperton zufolge belief sich das weltweite Gesamtvolumen an Investitionen in KI-Start-ups im Jahr 2016 auf 1,8 Mrd. US-Dollar, davon wurden 19% in ML-Start-ups und 40% in Deep-Learning-Startups (717 Mio. US-Dollar) investiert.[174]

[171] Gründungs-Update 1/2017: https://www.gruenderwettbewerb.de/service/publikationen/gruenderupdate-1-2017
[172] Valorge et al. 2017
[173] Valorge et al. 2017
[174] Valorge et al. 2017

Abbildung 57: Internationale Investment-Geschäfte in KI, Maschinelles Lernen und Deep Learning seit 2010[175]

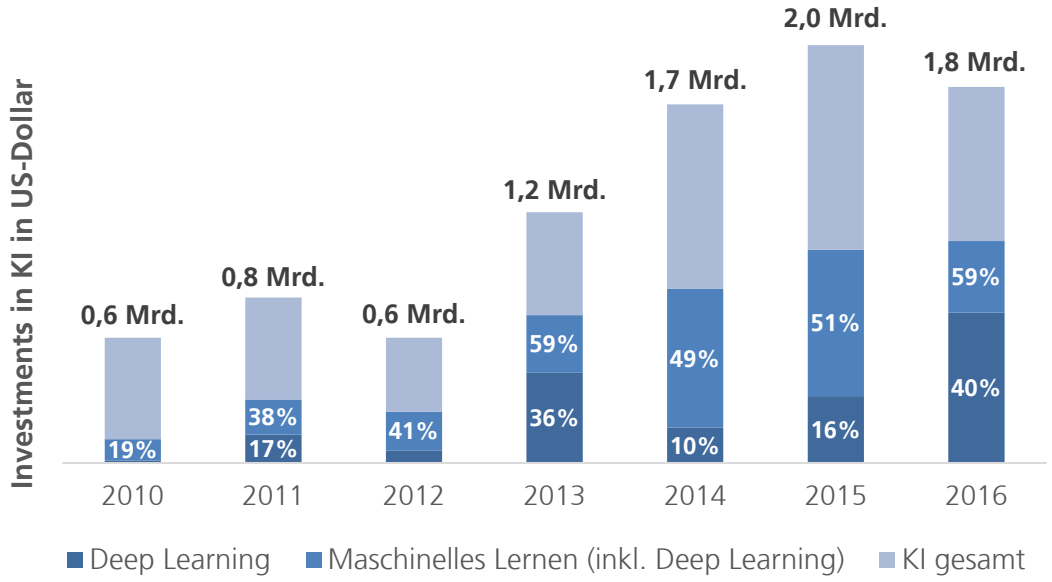


Tabelle 17: Beispielanbieter für ML mit Sitz in Deutschland

Unternehmen	ML-Produkte und Dienstleistungen, Branche	Sitz	Link
Leverton GmbH	Entwickelt und wendet Deep Learning an, um Daten aus Unternehmensdokumenten zu extrahieren, zu strukturieren und zu verwalten, z.B. in der Finanzindustrie.	Berlin	https://leverton.de
Panoratio Database Images GmbH	Big Data Analytics für Beratungsunternehmen und Managed-Services Anbieter.	München	http://www.panoratio.com
Twenty Billion Neurons GmbH	Deep Learning für industrielle Anwendungen (Robotik, Autonomes Fahren, Video Monitoring).	Berlin	https://www.twentybn.com/
PEAT GMBH *(Progressive Environmental & Agricultural Technologies)*	Plantix von PEAT erkennt Pflanzenkrankheiten mittels Machine Learning und Image Recognition via Smartphone und gibt Empfehlungen für Behandlungsmöglichkeiten.	Hannover	https://plantix.net/

[175] Quelle: Valorge et al. 2017

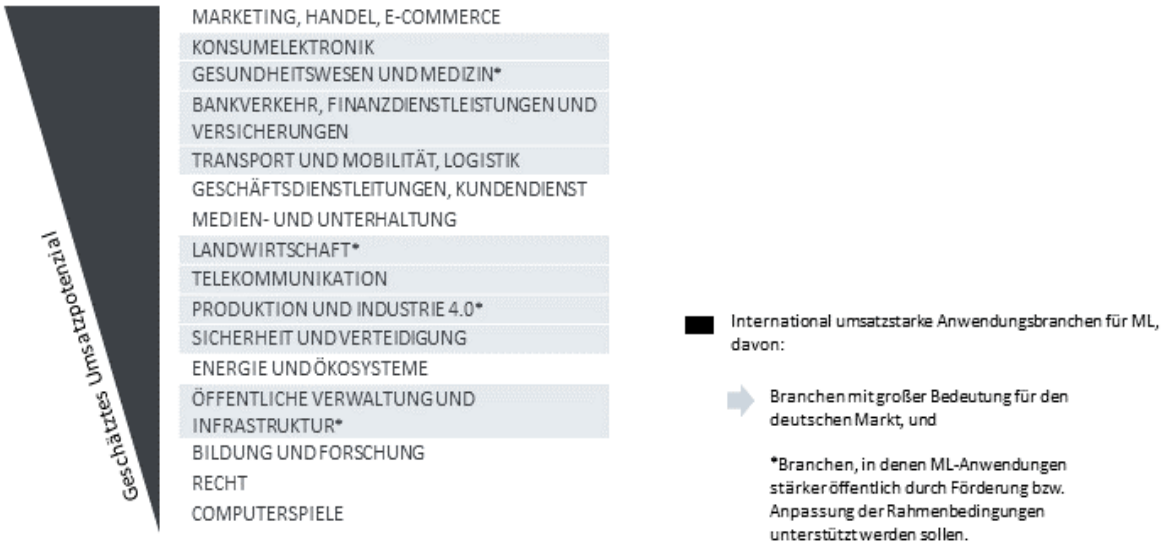
3.6.2 Anwendungsbranchen und Produkte

Maschinelles Lernen hat den Aussagen der befragten Fachleute zufolge das Potenzial, beinahe in allen Branchen erfolgreich kommerziell eingesetzt zu werden. Als international umsatzstärkste Branchen für Maschinelles Lernen und Deep Learning nennen die Marktanalytiker Marketing, E-Commerce und Werbeindustrie, Konsumelektronik und Geschäftsdienstleistungen sowie die Finanzindustrie. Unterschiedlichen Prognosen zufolge werden in den kommenden Jahren ML-basierte Produkte und Dienstleistungen insbesondere datenintensive Bereiche mit hohem Automatisierungspotenzial wie Transport und Mobilität, Gesundheitswesen sowie Landwirtschaft transformieren.[176]

In Deutschland wurden die ersten Produkte bereits Anfang der 1990er Jahre entwickelt, z.B. neuronale Netze für industrielle Anwendungen. Zentrale Einsatzfelder sind heute vor allem Bild- bzw. Videoanalyse sowie die Sprachverarbeitung. Im Gegensatz zur internationalen Entwicklung werden in Deutschland in erster Linie die Automobilindustrie, Maschinen- und Anlagenbau sowie Finanzdienstleistungen zu zentralen Anwendungsbranchen für KI und Maschinelles Lernen. Es folgen Gesundheitswesen, Telekommunikation, Landwirtschaft, Kundendienst, Konsumelektronik und intelligente Umgebungen (Smart Home). Die interviewten Fachleute weisen darauf hin, dass Maschinelles Lernen effektiv zur Effizienzsteigerung im öffentlichen Raum eingesetzt werden kann, etwa für Prozessoptimierungen in der öffentlichen Verwaltung, den Ausbau von Infrastruktur oder in der Verteidigung und der zivilen Sicherheit.

[176] Austin & Krensky 2017, Gartner 2017

Abbildung 58: Zentrale Anwendungsbranchen für Maschinelles Lernen[177]



Unter den ML-Technologien steht seit einigen Jahren Deep Learning im Mittelpunkt vieler Anwendungsszenarien. Die weltweiten Einnahmen aus Deep-Learning-Software werden den Prognosen von Tractica[178] zufolge von 654,9 Mio. US-Dollar (2016) auf 34,9 Mrd. US-Dollar (bis 2025) ansteigen, was im Jahr 2025 einen Anteil von 57% am gesamten KI-Markt ausmachen wird. Der jährliche Gesamtumsatz für Deep-Learning-Software, -Dienstleistungen und -Hardware soll von 3,3 Mrd. US-Dollar im Jahr 2016 auf 160 Mrd. US-Dollar im Jahr 2025 bei einer jährlichen Wachstumsrate von 54% steigen. Ausgewählte ML-Produktgruppen in den zentralen Branchen werden im Folgenden näher dargestellt.

3.6.2.1 Gesundheitswesen

Das Gesundheitswesen ist eine datenintensive Branche, in der Maschinelles Lernen viele Möglichkeiten für Qualitäts- und Effizienzverbesserungen in der klinischen Medizin, der Krankenhausverwaltung und für die medizinische Forschung eröffnet.

Sowohl weltweit als auch in Deutschland verweisen Fachleute auf ein hohes Zukunftspotenzial bei der Analyse und Verarbeitung von Datensilos, bspw. aus medizinischen Diensten, Geräten, Patientenakten, Aufzeichnungen der Ärzte usw., die derzeit in isolierten Datenbanken gehalten und punktuell auf akute Probleme angewendet werden. Einsatzbeispiele sind die effiziente Verwaltung von Patientendaten, meist noch prototypische Systeme zur Entscheidungsunterstützung bei der klinischen Diagnostik und bildgebende

[177] Eigene Darstellung , basierend auf den Ergebnissen der Sekundäranalyse und Aussagen der befragten Fachleute in den Interviews und dem Validierungsworkshop.

[178] Groopman & Kaul 2017

Verfahren in der Radiologie, Pathologie, Dermatologie[179]. Hier wird vor allem das Deep Learning zur Analyse von medizinischen Bildern (MRT, CT, Röntgen) und anderen komplexen, unstrukturierten Daten eingesetzt. Die Bearbeitung dieser Daten ist besonders zeit- und kostenintensiv, kompliziert und anfällig für menschliche Fehler. Derzeit erlauben ML-Technologien etwa Brustkrebs, Herzerkrankungen, Osteoporose und erste Anzeichen von Hautkrebs zu identifizieren, mit höherer Genauigkeit Prognosen zu erstellen und optimale Anwendung von Medikamenten zu berechnen. In Zukunft werden Systeme zur frühzeitigen Erkennung bzw. Prävention von Pandemien, für die Entwicklung von Medikamenten, in der Präzisionsmedizin und der personalisierten Genetik sowie als kollaborative[180] Chirurgie- und Serviceroboter[181] die Mediziner bei der Arbeit entlasten können.

Fitnessarmbänder, Smart Watches und andere Wearables eröffnen neue Möglichkeiten für gesundheitsrelevante Dienstleistungen[182], die eine aktive Rolle der Patienten voraussetzten. Mit ihrer Verbreitung entsteht ein neuer Markt für »Vor-Grundversorgung«: Fortschrittliche Healthcare-Apps erleichtern die Erfassung von gesundheitsrelevanten Daten und Informationen, ihre Auswertung und gegebenenfalls direkte Übermittlung an den Arzt. So dient der vom Berliner Start-up *ARYA mhealth UG*[183] entwickelte Mood-Tracker der Unterstützung der Therapie von psychischen Erkrankungen und die App »Ada Health« steht dem Nutzer als persönliche Gesundheitshelferin für das Smartphone zur Verfügung. Jedoch ist es umstritten, wie wissenschaftlich präzise und aussagekräftig, sicher und datenschutzkonform die Messungen, Ergebnisse und Ratschläge vieler kommerzieller Wearables und den dazugehörigen Auswertungsalgorithmen und Health-Apps wirklich sind.

Der Markt für intelligente Systeme im Gesundheitswesen ist noch jung, aktuell sind nur wenige Daten verfügbar. Den Einschätzungen der Analytiker von Frost & Sullivan zufolge werden im Jahr 2021 die jährlichen weltweiten Umsätze mit KI-Anwendungen im Bereich Gesundheit rund 6,6 Mrd. US-Dollar erreichen[184]. Für Software-Anwendungen mit Deep Learning prognostiziert Tractica in dieser Branche ein Umsatzvolumen von 3 Mrd. US-Dollar im Jahr 2025[185]. Bedeutende internationale Anbieter sind *IBM (Watson), Google Deepmind, Cognitive Scale*, *Atomwise*, *Nuance*, *Hindsait* und *Radiomics*. In Deutschland

[179] https://www.nature.com/articles/nature21056.epdf?referrer_access_token=gIdA1QINcltmnI2sUPOeF9RgN0jAjWel9jnR3ZoTv0NXpMHRAJy8Qn10ys2O4tuPxB4B9GCLPvFTfGPu3BrO0RvL6Zi_ok9X3NvF_zeHZra2WAHp2WDZLrCkVtI8xa4_39PNUuQbsCAmZKpGj7lHNd79MiOPQ3FmT6znFQ05Cm-_TzsjQ1WUxdFqI2-L7kMH&tracking_referrer=www.bbc.com

[180] Kollaborative Roboter, oder Koboter übernehmen zunehmend komplexere Aufgaben, die in Kooperation mit Menschen durchgeführt werden können.

[181] Rao & Verweij 2017

[182] Deutschland intelligent vernetzt DIV 2017, Statista 2017

[183] https://www.aryaapp.co

[184] Frost & Sullivan 2017c

[185] Groopman & Kaul 2017, Groopman und Wheelock 2017, Sahi und Kaul 2017

forschen und entwickeln in diesem Bereich unter Anderem *SAP*, *Siemens* und *Bayer*, aber auch kleinere Unternehmen, zum Beispiel *ExB Labs*, *Ada Health*, *Heuro Labs und XBird*.

Die Fachleute verweisen auf erhebliche Umsetzungsherausforderungen bei der datenschutzrechtlichen Regulierung bezüglich der Nutzung von Patientendaten, der Transparenz der Systeme (Black-Box Problem) sowie der Akzeptanz seitens der Anwender.

Tabelle 18: Beispielanbieter für ML mit Sitz in Deutschland

Unternehmen	ML-Produkte und Dienstleistungen, Branche	Sitz	Link
Ada Health	Persönliche Gesundheitshelferin für das Smartphone	Berlin	https://ada.com/
ARYA mhealth UG	Mood-Tracker zur Unterstützung der Therapie von psychischen Erkrankungen	Berlin	www.aryaapp.co
ExB Labs GmbH	Cognitive Workbench – Text- und Bildanalyse für die Medizin	München	https://www.exb-health.com/
Heuro Labs	Bild- und Audioerkennung in der Medizin	Berlin	http://heurolabs.com/
MeVis Medical Solutions AG	Medizinische Anwendungssoftware, Dienstleistungen für den Mediziner zur Therapie und Diagnostik oder zur technischen Visualisierung	Bremen	https://www.mevis.de/
xbird GmbH	Digitales Frühwarnsystem für Krankheiten, analysiert die Bewegungsdaten von Handys und Wearables	Berlin	http://xbird.io/
EyeEm	Das Unternehmen nutzt Deep Learning, um qualitativ hochwertige, authentische Fotos für führende Marken und Agenturen zu sammeln und zu kuratieren	Berlin	https://www.eyeem.com

3.6.2.2 Autonomes Fahren

ML ist auch eine Schlüsseltechnologie für das (hoch)automatisierte Fahren, insbesondere für Fahrerassistenzsysteme. Deep-Learning-Architekturen befähigen die Fahrzeuge, aus Erfahrungsdaten zu lernen und sich an Echtzeit-Situationen ohne menschliche Intervention anzupassen. Sie unterstützen die Steuerung und Navigation von Transportmitteln, können auf sich verändernde Umgebungen reagieren, die Trajektorienplanung übernehmen und Kollisionen verhindern. Auch hinsichtlich zukünftiger Sharing-Konzepte mit au-

tonomen Fahrzeugen wird sich ML maßgeblich auswirken.[186] Weitere Funktionalitäten ermöglichen Platooning, prädiktive Wartung, die Überwachung des Fahrzeug- und Fahrerzustandes und diverse Entertainment- und Komfortdienste.

Viele dieser Anwendungsfälle basieren auf dem Erkennen und Identifizieren von Objekten im und außerhalb des Fahrzeugs mittels Bildklassifikation bzw. Segmentierung, realisiert auf Basis von tiefen neuronalen Netzen mit unterschiedlichen Architekturen. Während für die ersten Automatisierungsstufen bereits marktfähige Produkte existieren, steht teil-, hoch- und vollautomatisiertes Fahren im aktuellen Fokus der Forschung und Entwicklung. Vollautonomes Fahren, bei der das Fahrzeug »vollumfänglich auf allen Straßentypen, in allen Geschwindigkeitsbereichen und unter allen äußeren Bedingungen die Fahraufgabe vollständig allein durchführen kann«, wird aktuellen Einschätzungen zufolge gegen 2025 erreicht werden.[187] Aktuellen Prognosen von IHS zufolge, werden bis 2035 ca. 21 Mio. selbstfahrende Autos und etwa 76 Mio. Transportmittel mit unterschiedlichem Automatisierungsgrad weltweit verkauft worden sein.[188] Bis dahin sind, neben den technischen, vor allem rechtlichen Fragestellungen zu beantworten.

Die weltweiten Umsätze mit Deep-Learning-Technologien in der Automobilindustrie in 2016 wurden von Tractica-Fachleuten auf 61,2 Mio. Dollar eingeschätzt, 2025 werden Umsätze bis 1,4 Mrd. Dollar erwartet.[189]

Unternehmen aus den USA, Westeuropa (Deutschland, UK und Schweden) und Asien gelten als stärkste Player auf diesem Markt. Die Mehrheit der weltweiten Patentanmeldungen zum Autonomen Fahren zwischen 2010 und 2016 gehörte den deutschen Unternehmen (Bosch, Audi, Continental), gefolgt von den USA (General Motors, Google, Ford) und Japan (Toyota)[190]. Bis 2025 werden den aktuellen Prognosen zufolge Technologieunternehmen wie *Google, NVIDIA, HERE, QNX, Intel* und *Baidu* auf diesem Markt dominieren. Bereits heute bilden sie strategische Partnerschaften mit führenden OEMs (*AUDI, Daimler, BMW, Toyota, VW, GM, Volvo, Tesla)* und Tier-1-Zulieferern *(ZF & Hella und Bosch)*, um gemeinsam Wettbewerbsvorteile zu erzielen. Die Mehrheit der Start-ups und Pilotprojekte stammen aus den USA: *think, AT&T Quanergy, AdasWork, Five AI, Brain4Cars, DriveAI, Mobileye, Civil Maps, Preferred Networks.*[191]

Weitere bedeutende Anwendungsfelder für autonome Transportmittel sind Logistik, Luft- und Raumfahrt sowie die Landwirtschaft. In der Precision Agriculture (Smart Farming) werden ML-gestützte Datenanalysen und intelligente, autonome Landmaschinen und

[186] PwC 2017
[187] BITKOM 2016
[188] Statista 2017
[189] Burger & Wheelock 2017, Tractica 2017
[190] Statista 2016
[191] Frost & Sullivan 2017a

Drohnen eingesetzt, um Produktivität und Nachhaltigkeit zu steigern. Zu den Aufgaben intelligenter Drohnen (*PrecisionHawk*[192]*, EagleView*[193]*)* gehören beispielsweise die sensorielle Erfassung von Daten für eine optimale Schädlingsbekämpfung oder Düngung. Computer Vision, maschinelle Sensordatenfusion, Robotik und ML befähigen Maschinen darüber hinaus zur intelligenten Identifizierung und Entscheidungsunterstützung bei der Wettervorhersage[194], Landnutzung (*OmniEarth*[195]*)* und Viehzucht (*Cainthus*[196]). In Deutschland sind Akteure wie *John Deere, Horsch, Bosch und Claas* in diesem Bereich tätig, sie integrieren lernende Algorithmen in die Landtechnik, entwickeln Datenanalyse-Apps für Kunden und investieren in KI-Unternehmen.[197] Dem gesamten Markt für Hardware und Software im Prescision Farming sagen Analytiker von Roland Berger einen Anstieg von 2,3 Mrd. Euro in 2014 auf 4,5 Mrd. Euro in 2020 bei einer durchschnittlichen jährlichen Wachstumsrate von 12% voraus.[198] Die jährlichen Umsätze mit Deep-Learning-Software in der Landwirtschaft werden laut Tractica-Prognosen im Jahr 2025 bis zu 2,1 Mrd. Dollar betragen.[199]

[192] http://www.precisionhawk.com/
[193] https://www.eagleview.com/
[194] https://www.hpcwire.com/solution_content/hpe/weather-climate/improving-accuracy-weather-forecasting-deep-learning/
[195] http://www.omniearth.net/
[196] http://www.cainthus.com/
[197] https://www.deere.com/en/our-company/news-and-announcements/news-releases/2017/corporate/2017sep06-blue-river-technology/
[198] Roland Berger Strategy Consultants GmbH 2015
[199] Burger & Wheelock 2017, Tractica 2017

Tabelle 19: Beispielanbieter für ML mit Sitz in Deutschland

Unternehmen	ML-Produkte und Dienstleistungen, Branche	Sitz	Link
Cargonexx	Selbstlernende Algorithmen, die Spotmarktpreise und Ladungsströme vorhersagen, um Schwerlast-LKW-Transporte für Verlader, Speditionen und Transportunternehmen effizienter zu organisieren	Hamburg	https://www.cargonexx.de/de/
German Auto Labs GAL GmbH	Chris - Sprachassistent für Autofahren	Berlin	https://www.hellochris.ai/de
Intelligent Apps GmbH/ Daimler	MyTaxi App für die Bestellung von Taxis auf der Basis von Amazon Web Services (AWS)[200]	Hamburg	https://de.mytaxi.com/impressum.html

3.6.2.3 Sensordatenanalyse in der Industrie und Automatisierung

Auch im Kontext von Industrie 4.0 und Digitalisierung der Produktion bietet ML reichlich Potenzial: sowohl für die Industrierobotik (insb. im Bereich der Bild-/Videoverarbeitung und Handlungsplanung) und automatisierte Produktionsprozesse, ebenso wie für darauf aufbauende Geschäftsmodelle mit der Analyse von Industriedaten und darauf basierenden Dienstleistungen im Service-Bereich (Monitoring, Analysen und Prognosen z.B. für predictive maintenance, Prozess-, Logistik-, und Energie-optimierung sowie Qualitätsmanagement).

Bis 2030 wird eine Vervielfachung der Umsätze durch KI-gestützte Verbesserungen von Produktivität, Qualität und Personalisierung prognostiziert. Es wird erwartet, dass sich dies vor allem in den Bereichen Produktion und Transport niederschlagen wird. Durch Effizienzgewinne wie Automatisierung und Unterstützung von Arbeitskräften wird eine globale Produktivitätssteigerung von 55% im Zeitraum von 2017 bis 2030 vorhergesagt.[201] Der jährliche Umsatz für Anwendungen des Deep Learning in der industriellen Fertigung wird im Jahr 2025 global 762 Mio. US-Dollar erreichen.[202]

In der Robotik besteht eine große Herausforderung derzeit darin, in Robotern Adaptionsmechanismen zu implementieren, damit sie mit ihrer Umgebung und den Menschen in einer sicheren, anpassungsfähigen und flexiblen Art und Weise interagieren können.[203]

[200] https://aws.amazon.com/de/solutions/case-studies/mytaxi/
[201] PwC 2017
[202] Groopman & Kaul 2017
[203] Vgl. Stanford University 2016, S. 9

Die Fortschritte im Bereich des Deep Learning könnten in diesem Kontext entscheidende Vorteile bieten. Dazu bedarf es allerdings i.d.R. einer extrem großen Menge an Beispieldaten für das Trainieren der Maschine.

Bedeutend sind auch die Entwicklungen im Bereich des maschinellen Sehens, der maschinellen Wahrnehmung und Kommunikation. Vorausschauende Wartung, die durch ML verbessert wird, ermöglicht frühere, genauere und präzisere Vorhersagen und damit die Vermeidung von Maschinenfehlern, wofür Daten von modernen IoT-Sensoren (Internet of Things) und Wartungsprotokollen sowie externen Quellen kombiniert werden. Hier wird die Möglichkeit einer Steigerung der Anlagenproduktivität um bis zu 20% ist gesehen, bei gleichzeitig sinkenden Wartungskosten um bis zu 10%.[204]

[204] McKinsey 2017

Tabelle 20: Beispielanbieter für ML mit Sitz in Deutschland

Unternehmen	ML-Produkte und Dienstleistungen, Branche	Sitz	Link
Konux	Predictive-Maintenance-Software lernt Muster in Sensordaten zu erkennen und sowohl den digitalen Fingerabdruck der jeweiligen Maschine als auch den Zeitpunkt des Wartungsbedarfs vorherzusagen und die Instandhaltung somit planbar zu machen	München	www.konux.com/de/
micropsi industries GmbH	Machine Learning in der Robotik und Fertigungsautomatisierung	Berlin	http://www.micropsi-industries.com/
Brick Reply	Manufacturing Operations Platform für Prozessoptierungen und prädiktive Analysen in der vernetzten Produktion	München, London	https://www.reply.com
Bosch Software Innovations GmbH	Predictive Maintenance an Robotern in der industriellen Fertigung	Berlin	https://www.bosch-si.com/
Materna GmbH	Dienstleister für Beratung und Durchführung von Datenerfassung und Analyse zur Prozessoptimierung in Service, Vertrieb und Logistik	Dortmund	https://www.materna.de
FRAMOS GmbH	Kameras, Sensorik und Software für optimierte Diagnose und präventive Wartung in der industriellen Robotik	Taufkirchen	https://www.framos.com/
Bosch Rexroth AG	Zustandsüberwachung und Wartungsoptimierung über das Online Diagnostics Network ODiN	Lohr a.M.	https://www.boschrexroth.com/

3.6.2.4 Einzelhandel und Marketing

In den letzten Jahren wurden ML-basierte Produkte erfolgreich in der Werbebranche, dem Online-Handel sowie im Finanz- und Versicherungswesen umgesetzt. Ein großes Marktsegment bilden professionelle sprachbasierte Assistenzsysteme (Virtual Customer Assistant, VCA), die im Kundendienst und Conversational Commerce[205] zum Einsatz kommen. Sie werden durch die technologischen Entwicklungen bei Sprachtechnologien, ML, Big Data und Real-time Analytics sowie Web Services ermöglicht.[206] Moderne lernende Assistenten werden mit historischen Dialogen trainiert, sind in der Lage den Kontext zu berücksichtigen und sollen ihr Wissen kontinuierlich verbessern. Im Rahmen einer Konversation liefern sie dem Kunden die gesuchten Informationen, die verdichtet und kontextrelevant sein sollen und können in seinem Auftrag handeln, um beispielsweise Transaktionen auszuführen. Die Unternehmen können Kosten sparen und Umsätze erhöhen, indem

[205] Gentsch 2018
[206] Frost & Sullivan 2016

sie beispielsweise die menschlichen Agenten bei Routineaufgaben und einfachen Anfragen entlasten oder Self-Service über IVR und Rund-um-die-Uhr Erreichbarkeit anbieten.[207]

Die Entwicklung im Bereich Service Assistenten wird von Gartner als fortgeschritten eingeschätzt.[208] Der Übergang von statischen, programmierten Assistenten zu lernfähigen, proaktiven Anwendungen wird in den nächsten fünf Jahren erwartet. Viele große Anbieter wie beispielsweise IBM Watson, Microsoft Cortana, Next-IT oder Creative Virtual sind in diesem Bereich unterwegs. Bei der Einschätzung des möglichen Marktvolumens werden Service Assistenten selten von sonstigen digitalen Assistenten getrennt. Nach Einschätzungen von Meisel soll der Markt für diese Anwendungen 2017 ein weltweites Volumen von 9,2 Mrd. US Dollar und 2020 von 80,7 Mrd. US erreichen können[209]. Dabei werden die meisten Umsätze in den Asien-Pazifik Regionen (inkl. Indien) erzielt, gefolgt von China, Europa und Südamerika.

Neben der Verarbeitung gesprochener Sprache wird in der digitalen Wirtschaft auch die automatische Auswertung von Texten nachgefragt. Eine der bedeutenden Technologien ist dabei die Sentiment-Analyse (Englisch für »Stimmungsanalyse«, auch Opinion Mining). Sie basiert auf Textklassifikation und ermöglicht es, automatisch Meinungen und emotionale Einstellungen von Menschen (z.B. negativ/positiv) gegenüber Produkten und Dienstleistungen, Ereignissen oder einzelnen Themen in Texten zu erfassen und sie auszuwerten.[210] Mit den jüngsten Fortschritten im Deep Learning konnte die Fähigkeit von Algorithmen Texte zu analysieren erheblich verbessert werden.

Die rechtzeitige Berücksichtigung von Kundenmeinungen oder Einstellungen der Öffentlichkeit sind sowohl für Unternehmen und Organisationen als auch für politische Entscheidungen strategisch relevant. In den letzten Jahren sind zahlreiche Anwendungen bei der Analyse von Produktbewertungen im Online-Handel, der Antizipation von Kundenentscheidungen im Versicherungswesen oder für die Vorhersage von Börsenbewegungen in der Finanzbranche entstanden. Zentrale Datenquellen für Stimmungsanalysen sind Social-Media-Plattformen (Twitter, Facebook) und Kundenrezensionen (Yelp, Amazon) aber auch Blogs, Artikel, Foren und Umfragen.

Laut einer Marktstudie von Technavio wird der globale Markt für Sentiment-Analyse-Software bis 2021 voraussichtlich ein Umsatzvolumen von 2,14 Mrd. US-Dollar mit einer durchschnittlichen jährlichen Wachstumsrate von mehr als 16% erreichen[211]. Weltweit gibt es hunderte Anbieter in diesem Bereich von Start-ups bis Großunternehmen, darunter

[207] Frost & Sullivan 2016, S.3
[208] Gartner 2016
[209] Meisel 2016
[210] Liu 2015
[211] Technavio 2017

Angoss Software Corporation, Clarabridge, IBM, SAS Institute, Lexalytics, Inc., Microsoft *Cognitive Services* und *ParallelDots, Inc.*

Tabelle 21: Beispielanbieter für ML mit Sitz in Deutschland

Unternehmen	**ML-Produkte und Dienstleistungen, Branche**	**Sitz**	**Link**
ubermetrics	Online-Textanalysen zur Sentimentsbewertung	Berlin	https://www.ubermetrics-technologies.com/de/support/cs_ubermetrics-delta-sentimentsmodell/
Clickworker	Sentiment-Analysen	Essen	https://www.clickworker.de/ueber-uns/
Brandwatch	Sentiment Analyse, Social-Listening-Plattform	Stuttgart, Berlin	https://www.brandwatch.com
Camelot ITLab	Sentiment-Analyse anhand von Produktbewertungen, Social Media Posts und Blogs	Mannheim	https://www.camelot-it-lab.com/de/leistungen/beratungsexzellenz/analytics/sentiment-analysis-with-sap-hana/
talkwalker	Social Media Analyse von Sentiments, Mentions, Performance, Reichweite (Deep Learning)	Frankfurt a.M.	https://www.talkwalker.com/de/social-media-search
Meltwater	Medien-Monitoring, Performance- und Sentiment-Analysen	Berlin	https://www.meltwater.com/de/products/
rapidminer	Text Mining und Sentiment-Analysen anhand von Online-Inhalten	Dortmund	https://rapidminer.com/solutions/text-mining/
Parlamind GmbH	Datenerfassung aus Kundenemails für visuelle Auswertungen	Berlin	https://parlamind.com/de/index
ITyX Solutions AG	Lernfähige Algorithmen zur Inhaltserkennung im Kundenservice	Köln	https://www.ityx.de/
Ambiverse	Text to Knowledge: Ambiverse bietet Lösungen für automatisches Textverständnis und intelligente Textproduktion.	Saarbrücken	https://www.ambiverse.com/
Rasa Technologies GmbH	Virtuelle Assistenten für Bank- und Versicherungswesen sowie Medizin	Berlin	https://rasa.com
Cognigy GmbH	Integration von Sprachsteuerungstechnologie in vernetzte Geräte (IoT) und digitale Anwendungen (Chatbots, Sprachassistenten, Roboter, Web- und Mobilanwendungen oder VR/AR-Applikationen) mit durch ML verbesserter Dialogfähigkeit	Düsseldorf	https://www.cognigy.com/

In weiteren Branchen sind noch viele Einsatzmöglichkeiten für ML enthalten. Als attraktive Anwendungsfelder in der öffentlichen Verwaltung, die in Deutschland aktuell noch nicht im zentralen Blickfeld der Unternehmen stehen, nennen die Fachleute Prozessoptimierung, Predicitive Policing und E-Government. Um die Potenziale des deutschen Forschungs- und Wirtschaftsstandorts im Bereich des Maschinellen Lernens weiter zu stärken, sind innovationsfördernde sozioökonomische, rechtliche und politische Rahmenbedingungen grundlegend erforderlich.

Tabelle 22: ML-basierte Kompetenzen, Forschungsansätze, Einsatzgebiete und Reifegrad

Legende:

Stufe 1: Inzwischen gut etabliert
Stufe 2: Demonstratoren vorhanden, Forschung für komplexere Anwendungen unbedingt erforderlich
Stufe 3: Noch in früher FuE-Phase

Fähigkeit	Nutzen	FuE-Ansätze	Mögliche Anwendungen	Reife-grad
Gruppen ähnlicher Daten bilden	Diese Fähigkeit ermöglicht bspw. das Erkennen von Mustern und hilft, Strukturen in großen Datenmengen zu erkennen.	▪ Clustering ▪ Tiefe Neuronale Netze	▪ Marketing: Kundensegmentierung, Zielgruppenübersicht	1
Objekte klassifizieren	Diese Fähigkeit ermöglicht das Einordnen von Beispielen für die weitere Bearbeitung, um Entscheidungen zu treffen oder Maßnahmen einzuleiten.	▪ Entscheidungsbäume ▪ Stützvektormaschinen ▪ Bayessche Netze ▪ Logistische Regression	▪ Datenfiltersysteme ▪ Sortieraufgaben (z.B. Güteklassifizierung in der Produktion) ▪ Marketing (z.B. Matching von Kunden und Waren)	1
Werte schätzen und vorhersagen	Hier werden lineare oder komplexere Zusammenhänge erkannt und für Vorhersagen über künftige Zustände bzw. Ereignisse genutzt.	▪ Lineare Regression ▪ Regressionsbäume (CART) ▪ Entscheidungsbäume	▪ Generierung von Prognosen (Stau, Angebot und Nachfrage) ▪ Anomalie-Detektion ▪ Maschinen-/Anlagen-Optimierung ▪ Vorausschauende Wartung ▪ Finanz- und Versicherungs- und Rechtswesen ▪ Medizin, Chemie, Materialforschung (Entdeckung neuer Molekularkombinationen etc.) ▪ Steuerungsaufgaben	1
Erfolgversprechende Aktionen für einen Agenten auswählen	Agenten und Roboter können anhand von Feedback lernen, welche Aktionen, Spielzüge etc. die besten Resultate erzielen können. Das ist eine Alternative zum expliziten Planen und Adaptieren von Handlungsfolgen.	▪ Bestärkendes Lernen ▪ Q-Lernen mit tiefen Neuronalen Netzen	▪ Robotik (z.B. um das optimale Greifen unterschiedlicher Objekte zu lernen) ▪ Autonomes Fahren (z.B. um in Simulationen gewünschte von unerwünschten Aktionen unterscheiden zu können) ▪ Spieleindustrie ▪ Konsumelektronik	2

Fähigkeit	Nutzen	FuE-Ansätze	Mögliche Anwendungen	Reife-grad
Bilder erkennen	Auf Bildern werden Objekte lokalisiert, klassifiziert und ggf. Individuen erkannt.	▪ Tiefe Neuronale Netze	Alle Bereiche, in denen Objekterkennung von Nutzen ist, insbesondere ▪ Mobilität (Straßenschild-identifikation etc.) ▪ Industrielle Produktion/ Industrierobotik ▪ Medien (Suche) ▪ Medizin (Radiologische Diagnostik) ▪ Sicherheit (Videoüberwachung)	1
Sprache erkennen	Audiosignale werden als Sprache erkannt und in Text umgewandelt. Der Sprecher und seine Emotionen können identifiziert werden.	▪ Tiefe Neuronale Netze	Alle Bereiche, in denen Spracherkennung von Nutzen ist, insbesondere ▪ Konsumelektronik ▪ Automobilindustrie ▪ Spieleindustrie ▪ Medizin	1
Informationen aus Texten extrahieren und einfachen Aufforderungen nachkommen	Aus Texten werden bestimmte Informationen extrahiert, z.B. Namen, Adressen, Marken. Emotionen können anhand der Wortwahl identifiziert werden. Einfache Befehle oder Fragen werden erkannt und ausgeführt bzw. beantwortet. Äußerungen und Texte werden übersetzt.	▪ Tiefe Neuronale Netze ▪ Textmining ▪ Wissensgraphen	▪ Marktforschung ▪ Medizin (Forschung und Diagnostik) ▪ Informationssysteme ▪ Übersetzungen ▪ Marketing (Sentimentanalyse) ▪ Business Services (Klassifikation von Dokumenten)	1
Sprache und Text verstehen sowie kommunizieren	Hier wird nicht nur ein Wort erkannt, sondern auch eine semantische Einordnung geleistet. Sätze werden in Bezug zu früheren Äußerungen gesetzt und Dialoge fortgesetzt.	▪ Tiefe Neuronale Netze ▪ Wissensgraphen ▪ Kombinationen mit symbolischen Modellen	▪ Kundendienst ▪ E-Commerce, Sales ▪ Wissenschaftliche Arbeit ▪ Konsumelektronik	3
Bild und Video semantisch und im Kontext verstehen	Bei Bildern und Bildfolgen geht es nicht nur darum, einzelne Objekte zu identifizieren, sondern die Szene zu verstehen. Wer interagiert mit wem, was passiert gerade, was könnte als nächstes geschehen?	▪ Tiefe Neuronale Netze ▪ Kombination von Tiefen Neuronalen Netzen und symbolischen Verfahren	Alle Bereiche, in denen ein Lageverständnis von Relevanz ist ▪ Autonomes Fahren (Situationserkennung) ▪ Security-Anwendungen / Überwachung ▪ OP-Unterstützung in der Medizin	2

Fähigkeit	Nutzen	FuE-Ansätze	Mögliche Anwendungen	Reife-grad
Multimodale Inhalte kombinieren	Hier geht es um die Fähigkeit, zusammengehörige Text-, Bild-, Audiodaten in Bezug zu setzen, etwa die Aufnahme von Kamera und Mikrophon oder ein radiologisches Bild mit einem dazugehörigen wissenschaftlichen Report.	▪ Tiefe Neuronalen Netze ▪ Multimodales Lernen	▪ Medizin und Gesundheitswesen ▪ Wissenschaft ▪ Medien und Entertainment ▪ Robotik	2
Neue Inhalte generieren	Maschinen können inzwischen sogar »kreative« Aufgaben durchführen, wie bspw. das Schreiben von Gedichten, das Malen von Bildern und Komponieren von Musikstücken, die Animation von Figuren in digitalen Spielen und die Zusammenfassung von Meldungen zu einem Bericht (RoboJournalismus).	▪ Generative tiefe Neuronale Netze ▪ Probabilistische Modelle	▪ Journalismus ▪ Entertainment, Film, Kunst, Musik, Theater ▪ Spieleindustrie ▪ Simulation	3
Lernen mit zusätzlichem Wissen kombinieren	Wenn physikalische Gesetze, endliche Automaten, logische Regeln und allgemein das formalisierte Wissen von Fachleuten mit ML-Modellen kombiniert werden, können ML-Anwendungen effizienter, verständlicher, verlässlicher und kontrollierbarer gestaltet werden.	▪ Einbeziehung und Generierung von symbolischem Wissen ▪ Grey-Box-Modelle ▪ Regellernen ▪ Lernen mit Wissensgraphen	Überall dort, wo physikalische Gesetze eine bedeutende Rolle spielen oder intrinsisches Expertenwissen von zentraler Bedeutung ist. ▪ Industrielle Produktion ▪ Medizin- und Gesundheitswesen	3
Lernen mit vielen Daten	In vielen Fällen sind die Trainingsdaten so umfangreich, dass ein sequentieller Lernalgorithmus zu lange dauern würde. Hier helfen nur parallelisierte Lernalgorithmen, wie die tiefen neuronalen Netze. Algorithmen, die ohne Speicherung aus vorbeifließenden Daten lernen, sind oft die einzige Möglichkeit, aus sehr großen Datenströmen zu lernen.	▪ Tiefe Neuronale Netze ▪ Verteilte Lernalgorithmen ▪ Lernen aus Datenströmen ▪ Repräsentationslernen	▪ Überwachungsaufgaben (Produktionsprozesse, Finanzaktivitäten, Videoüberwachung, kritische Infrastrukturen) ▪ Empfehlungssysteme	1

Fähigkeit	Nutzen	FuE-Ansätze	Mögliche Anwendungen	Reife-grad
Lernen mit wenigen Daten	Wenn lediglich wenige oder wenig brauchbare Trainingsdaten zur Verfügung stehen, möchte man trotzdem bestmöglich daraus lernen.	▪ Lernen von Labels ▪ Lernen in Simulationen ▪ One-Shot-Lernen ▪ Transfer-Lernen	▪ Medizin- und Gesundheitsanwendungen ▪ Gefahrenabwehr (Produktionsprozesse, Finanzaktivi-täten, Videoüberwachung, kritische Infrastrukturen)	2
Anpassung an eine veränderliche Umgebung	Gelernte Modelle, die für die Lösung einer Aufgabe erfolgreich waren, dienen als Ausgangspunkt für die Lösung neuer, aber ähnlicher Aufgaben mit deutlich reduziertem Trainingsaufwand.	▪ Transfer-Lernen ▪ Lebenslanges Lernen ▪ Multitask-Lernen ▪ Interaktives Lernen	▪ Insbesondere in der Robotik (zur Gestaltung von flexibleren Robotern, bspw. im Service-Bereich)	3
Automatisiertes Lernen	Der Entwicklungsaufwand soll durch Teilautomatisierung in der Modellentwicklung verringert werden. Um ML schneller in die Breite zu tragen, müssen Data Scientists soweit wie möglich entlastet werden.	▪ Auto-ML	Alle praktischen Anwendungen	3
Transparente, nachvollziehbare und robuste Modelle	In viele Fällen möchten oder müssen Menschen verstehen, wie eine ML-basierte Anwendung zu einer Entscheidung gekommen ist und welche Daten und Verknüpfungen hierfür herangezogen worden sind. Hinreichende Robustheit gegen Störungen, das Einhalten von Einschränkungen und das explizite Berücksichtigen von Kompetenzgrenzen stärken das Vertrauen der Nutzer in die ML-Anwendung.	▪ Explainable AI ▪ Adversarial ML ▪ Robuste Lernverfahren ▪ Lernen von Kompetenzgrenzen ▪ Lernen unter Beschränkungen	Insbesondere dort, wo der Gesetzgeber eine Nachvollziehbarkeit verlangt oder es anderweitig sehr wichtig für die Herstellung von Vertrauen in die Anwendung ist, v.a. ▪ Finanz- und Versicherungs- sowie Rechtswesen ▪ Medizin-/Gesundheitswesen ▪ Öffentliche Verwaltung	2

Anhang B: Schlagworte für die Suchanfragen

ML-Technologie	
machine learning	machine-learning learning machine learning-machine learning machines learning-machines learn machine learn-machine learn machines learn-machines
learning algorithm	learning algorithms learning-algorithm learning-algorithms learn algorithm learn-algorithm learn algorithms learn-algorithms
self learning algorithm	self learning algorithms self-learning algorithm self-learning algorithms selflearning algorithm selflearning algorithms self learn algorithm self learn algorithms self-learn algorithm self-learn algorithms selflearn algorithm selflearn algorithms
learning method	learning methods learning-method learning-methods learn method learn methods learn-method learn-methods learn methodology learn methodologies learn-methodology learn-methodologies
reinforcement learning	reinforcement-learning reinforcement learn reinforcement-learn
supervised learning	supervised-learning supervised learn

	supervised-learn
semi supervised learning	semi-supervised learning semi supervised learn semi-supervised learn
unsupervised learning	unsupervised-learning unsupervised learn unsupervised-learn
ensemble learning	ensemble-learning ensemble learn ensemble-learn
pattern recognition	pattern-recognition patterns recognition patterns-recognition
data mining	data-mining
deep learning	deep-learning deep learn deep-learn deep neural network deep neural networks deep neuronal network deep neuronal networks deep neural net deep neural nets deep neuron network deep neuron networks
support vector machine	support-vector-machine support vector machines support-vector-machines
artificial neural network	artificial neural networks artificial neuronal network artificial neuronal networks artificial neural net artificial neural nets artificial neuron network artificial neuron networks
neural network	neural networks neuronal network neuronal networks neural net neural nets neuron network neuron networks

ML-Anwendungsbereiche	
Image	images vision visions video videos visual visuals picture pictures motion picture motion pictures motion-picture motion-pictures audio-visual
speech	speeches natural language natural languages oral orals semantic voice voices
Textverarbeitung	text texts word words document documents narrative narratives
Audio	audios sound sounds
signal processing	signals processing sensor sensors log data signal data signals data data stream data streams

KAPITEL 4

RAHMENBEDINGUNGEN FÜR MASCHINELLES LERNEN

Martin Wegele | Fraunhofer-Zentrale

Inhaltsverzeichnis Kapitel 4

4 Rahmenbedingungen für Maschinelles Lernen 155

4.1 Aus- und Weiterbildung 155

4.2 Transfer in die Praxis 157

4.3 Datenverfügbarkeit und Governance 158

4.4 Rechtliche, ethische und soziale Gestaltung 159

4 Rahmenbedingungen für Maschinelles Lernen

Rahmenbedingungen definieren wir mit übergeordneten rechtlichen, gesellschaftlichen und politischen Aspekten, die im Sinne von Leitplanken die Forschung zu ML und die Anwendungen von ML-basierten Produkten und Dienstleistungen begleiten. Im Folgenden werden jene Aspekte aufgeführt, die von den konsultierten Fachleuten auf einem wissenschaftlichen Validierungsworkshop priorisiert wurden. Diese werden durch Sekundärliteratur etwas detaillierter erklärt und kontextualisiert.

4.1 Aus- und Weiterbildung

Der Bedarf an Fachkräften für digitale Technologien und ML ist sehr groß. In Deutschland reichen die derzeitig ausgebildeten Akademiker und Akademikerinnen nicht aus, um die große Nachfrage abzudecken. Derzeit fehlen rund 85 000 Akademiker mit fortgeschrittenen Datenanalysekenntnissen, wie Mediziner und Ingenieure mit Big-Data Kompetenz. Auch die Breite der ML-Ausbildung wird von den befragten Fachleuten als ein entscheidender Wettbewerbsfaktor angesehen, da oftmals sowohl Ingenieurs- als auch ML-spezifische Kompetenzen benötigt werden, wie z.B. für das in Deutschland bedeutenden Lernen mit »Grey Box«-Modellen.

Daneben fehlen aktuell rund 10 000 IT-Spezialisten in den Bereichen Big Data, Advanced Analytics, Business Analytics und Data Science.[212] Big-Data-Expertinnen und Experten sammeln und strukturieren große Datenbestände, lernen Datenflussarchitekturen und verbessern die Soft- und Hardware der IT-Infrastruktur. Advanced-Analytics/Business-Analytics-Spezialisten analysieren Daten, um Geschäftsentscheidungen zu treffen und Empfehlungen auszusprechen. Data-Science-Experten hingegen generieren Antworten auf analytische Fragestellungen aus (großen) Datenmengen und wenden ebenfalls mathematische und statistische Verfahren zur Wissensgenerierung an.[213] Die Hochschulen in Deutschland reagieren langsam auf den Bedarf auf dem Arbeitsmarkt: von den bundesweit insgesamt 18 467 Studiengängen in 2016/2017 haben lediglich 23 eine explizite Spezialisierung auf Big Data, Data Science und Advanced Analytics/Business Analytics.[214] Der globale Wettbewerb einerseits aber auch die prognostizierte, positive Marktentwicklung für KI- und ML-Produkte und Dienstleistungen andererseits[215] verschärfen den Mangel in Deutschland weiter.[216] Zudem verfügen Universitäten und Forschungseinrichtungen

212 Stifterverband und McKinsey, 2017
213 Stifterverband und McKinsey, 2017
214 Stifterverband und McKinsey, 2017
215 McKinsey Global Institute, 2017
216 Economist, 2016

hierzulande im globalen Vergleich oftmals über weniger Sichtbarkeit und finanzielle Anreize.[217]

Selbstverständlich ist der Einstieg in Datenanalyseberufe nicht ausschließlich über spezialisierte Studiengänge möglich. Auch Absolventinnen und Absolventen anderer Fachrichtungen wie Informatik, Physik oder Mathematik können diese Stellen potenziell besetzen und Big Data/Business Analytics kann durchaus auch in zahlreichen anderen Studiengängen als Vertiefung enthalten sein. Demgegenüber steht, dass für 2017 bereits jetzt eine immense Arbeitskräftelücke i.H.v. 87 0000 Personen aus den MINT-Fächern besteht.[218] Angesichts des bereits vorherrschenden Mangels an IT-Fachkräften ist nicht davon auszugehen, dass sich der Bedarf der Unternehmen an Personen mit Experten- oder vertieften Datenanalysekenntnissen durch interne Weiterbildungsmaßnahmen lösen wird. Hier sehen die befragten Fachleute auch die Unternehmen in der Pflicht, frühzeitig mit Hochschulen zu kooperieren und den Kontakt zu Studierenden suchen. Des Weiteren sollte bestehendes Personal vermehrt zu Datenspezialisten weiterqualifiziert werden.[219]

Zusätzlich zur Förderung der Vermittlung KI-basierter Kompetenzen in der Breite sollten nach Einschätzung der konsultierten Fachleute sowohl mehr IT-Spezialisten in Deutschland ausgebildet als auch global rekrutiert werden, um die Nachfrage zu decken.

Auch Aus- und Weiterbildungsangebote sollte im Hinblick auf den breiten Erwerb von KI-basierten Kompetenzen stärker interdisziplinär orientiert sein. So waren sich die befragten Fachleute einig, dass insbesondere juristische Aspekte in die ML-bezogene Informatikausbildung miteinbezogen werden sollten. Dadurch können gesetzliche Vorgaben beim Entwurf und der Implementierung ML-basierter Anwendungen leichter berücksichtigt werden. In diesem Kontext sind auch Ethik-bezogene Ausbildungsinhalte von Bedeutung, da maschinelles Fehlverhalten der Wahrung ethischer Absichten zuwiderlaufen kann, wie die Diskussion um die sogenannte Dilemmasituation im (voll-)automatisierten Straßenverkehr zeigt.[220]

Aufgrund des disruptiven Potenzials von ML und KI für viele Branchen und Tätigkeitsprofile schlagen die Fachleute vor, ML-Wissen und -Kompetenzen in jedem Aus- und Weiterbildungskurrikulum zu vermitteln. Derartiges Wissen wird bei der betrieblichen Weiterbildung nicht nur durch interne Weiterbildungsprogramme aufgebaut, sondern oft auch extern durch Kooperation und Akquise kreativer Start-ups mit spezifischen ML-Lösungen erworben.[221]

[217] Lakemeyer, G., 2017
[218] IW, 2017
[219] Stifterverband und McKinsey, 2017
[220] The Royal Society, 2017
[221] VDMA, 2017

4.2 Transfer in die Praxis

Wie bereits aus den Patentanalysen ersichtlich, scheinen insbesondere KMU Aufholbedarf beim Einsatz von ML-Technologien zu haben. Nach Einschätzung der konsultierten Fachleute zeigen KMU vielfach Interesse, sobald sie konkrete Einsatzmöglichkeiten mit unternehmerischem Mehrwert aufgezeigt bekommen.

Daher sollten Fördermaßnahmen Führungskräfte und Mitarbeitende allgemein fit machen für den Einsatz von ML, z.B. indem sie mehr über den Einsatz und den Nutzen von ML-basierten Technologien für deren Geschäftsmodelle informieren. Existierende Maßnahmen zur strategischen Unternehmensförderung – die als noch ausbaufähig angesehen werden – sprechen aber eher jüngere Unternehmen und Start-ups und weniger die etablierten KMU an. Der Transfer von Wissen und Erkenntnissen zwischen Akteuren, im Besonderen den KMU, kann des Weiteren auch über Dialogplattformen, wie z.B. der »Plattform lernende Systeme«, erfolgen.[222]

Die allgemein als gut angesehenen Kompetenzen Deutschlands bei der Entwicklung von Machine Learning Methoden durch Forschungseinrichtungen bieten einen guten Ausgangspunkt.[223] Angesichts der Position Deutschlands in der theoretischen sowie der ML-Grundlagenforschung[224] sollte jedoch der Anwendungsbezug in der (Grundlagen-)Forschung gestärkt werden, um Ideen, Ansätze und Methoden zu fördern, die in Anwendungen münden können. Dies könne durch gezielte Transferprojekte in Kooperation zwischen Grundlagenforschung und Anwendung erfolgen, so die Experten. Auch in öffentlich finanzierten Forschungsprojekten sollte der Marktbedarf bzw. die Umsetzung am Markt stärker herausgearbeitet werden, so die konsultierten Fachleute. Vor dem Hintergrund der Größenordnungen der derzeit in China, Japan oder USA anlaufenden KI-Fördermaßnahmen[225] kommt es zu einem globalen Wettlauf, wenn Kunden künftig vermehrt auf US-amerikanische oder chinesische Produkte und Lösungen zurückgreifen. Dies gefährdet die langfristige Wettbewerbsfähigkeit besonders in deutschen Schlüsselbranchen wie Medizin, industrieller Produktion und Maschinenbau. Die verstärkte Zuwendung zu ausländischen Herstellern könnte mit einer Abwanderung der Daten einhergehen, was in Folge dessen zum Verlust der Datenhoheit führen könnte.

[222] EFI, 2018
[223] VDMA, 2017
[224] EFI, 2018.
[225] EFI, 2018.

4.3 Datenverfügbarkeit und Governance

Die Verfügbarkeit großer Datenmengen ist für das Maschinelle Lernen essenziell, besonders für die tiefen Lernverfahren. Große US-amerikanische und chinesische B2C Unternehmen wie Google, Facebook, Amazon, Baidu oder Alibaba verfügen über großen Datenmengen durch ihre große, globale Nutzerschaft. In Deutschland hingegen stehen aufgrund der stärkeren produktionsorientierten Industriestruktur einerseits, sowie strengerer Regelungen zum Datenschutz andererseits, weniger bzw. andersartige Daten zur Verfügung.

Den konsultierten Fachleuten zufolge mangelt es in Deutschland allgemein an der Datenverfügbarkeit. Hier besteht Handlungsbedarf, sowohl bei öffentlich-geförderten Forschungsprojekten als auch entsprechenden privatwirtschaftlichen Aktivitäten.

In diesem Zusammenhang sollte die Bereitstellung von »Open Data«, also offen zugänglichen Daten, die von jedem frei genutzt, weiterverbreitet und weiterverwendet werden dürfen, laut den befragten Experten stärker gefördert werden. Prinzipiell werden öffentliche Einrichtungen als wichtige Lieferanten von »Open Data« gesehen. Außerdem könnte in öffentlich geförderten Forschungsprojekten das Teilen der gewonnenen Daten explizit eingefordert werden, wie es die Pilotinitiative »Open Research Data« der EU-Kommission vormacht.[226]

Neben »Open Data« spielt die Veröffentlichung neuartiger Datensätze eine große Rolle bei der Förderung von ML-basierten Anwendungen in den Markt. Hierfür ist besonders der Zugang zu neuartigen Datenbeständen essenziell zur Sicherung des Wettbewerbsvorsprungs bei ML-basierten Systemen. Denn werden neu gewonnene Datensätze publiziert, entstehen innerhalb kurzer Zeit (im Schnitt innerhalb von 3 Jahren) neue Anwendungen. Im Gegensatz dazu münden Veröffentlichungen einer neuen ML-Methodik statistisch gesehen wesentlich langsamer in neue Anwendungen (im Schnitt 18 Jahre).[227]

Die branchenübergreifende Nutzung von Daten aus den Unternehmen bietet große Potenziale für neue ML-basierte Geschäftsmodelle. Nach Einschätzung der befragten Fachleute werden besonders Daten zum Kunden- und Maschinenverhalten von den Unternehmen ungern geteilt, weil sie dem Ausbau der eigenen Wettbewerbsfähigkeit dienen. Dies benachteiligt KMU, die zumeist aufgrund ihrer Größe und Struktur über weniger Daten verfügen[228]. Die Schaffung entsprechender Regelungen zum institutionenübergreifenden Austausch der Daten werden daher von den Fachleuten empfohlen.

Derartige Governance-Modelle für den Datenaustausch sollten nach Einschätzung der befragten Experten die Sicherheit und Kontrollierbarkeit des Dateneigentümers gewähr-

[226] https://www.openaire.eu/what-is-the-open-research-data-pilot (zuletzt geprüft am 19.03.2018)
[227] https://www.kdnuggets.com/2016/05/datasets-over-algorithms.html (zuletzt geprüft am 19. 03.2018)
[228] VDMA, 2017

leisten können. Oftmals können Governance-Modelle nur sektorspezifisch gestaltet werden, insbesondere im Gesundheits- und Produktionsbereich.[229] Unabhängige Instanzen als Treuhänder, die den Zugang zu den Daten kontrollieren und gewährleisten, werden zunehmend nachgefragt. Modelle wie der Industrial Data Space[230] wo die Unternehmen, die ihre Daten für digitale Dienste bereitstellen möchten, stets die Kontrolle über ihre Daten behalten und ihre Datenschutzvorgaben durchsetzen können, sind laut Einschätzung der Experten ein gutes Beispiel für die Sicherheit und Kontrollierbarkeit des Datenaustausches. Speziell für das Gesundheitswesen könnte ein »Datenspendeausweis« eine Idee sein. Dieser könnte festhalten, welche Daten eine Person für eingrenzbare wissenschaftliche oder medizinische Zwecke freigeben möchte. Diese freigegebenen Daten könnten wiederum zur Beschleunigung der medizinischen Forschung und Therapie beitragen.

Neben der Quantität an Daten ist auch die Qualität, speziell die Vollständigkeit, Korrektheit, und Dokumentation der Herkunft, von hoher Bedeutung für ML-Anwendungen. Dies erfordert eine gute Auf- und Vorbereitung der Daten. Einige Fachleute haben darauf hingewiesen, dass die sehr ressourcenintensive Tätigkeit der Datenkuration im wissenschaftlichen Umfeld kaum honoriert wird. Es wird vorgeschlagen, die Datenarbeit explizit in Publikationen und akademischen Arbeiten zu würdigen, z.B. in dem Zitiervorschlag zu einer Publikation. Als weitere Maßnahme fördern Standardisierungen die Vergleichbarkeit und Validierung von Modellen. Sie können sich auf Datenformate, aber auch auf Anforderungen an trainierte Modelle oder die Güteklasse von Resultaten beziehen. So können anwendungsbezogene Mindeststandards für Zuverlässigkeit, Robustheit, Performanz und Repräsentativität festgelegt werden. Qualitätskriterien für Datenbestände sollten über Metadaten definiert werden, um so den Austausch und (kommerziellen) Handel mit Datenbeständen erleichtern. Dokumentierte Trainingsdatenbestände und zertifizierte Validierungsdatenbestände dienen der Nachvollziehbarkeit der Modelle im Einsatz,[231] erleichtern den Nachweis, dass automatisierte Entscheidungen in Einklang mit der Datenschutz-Grundverordnung getroffen werden[232] und können das Vertrauen in ML-basierte Produkte und Dienstleistungen erhöhen.[233] [234]

4.4 Rechtliche, ethische und soziale Gestaltung

ML-Systeme werfen rechtliche Fragen auf, die in der Gesellschaft diskutiert werden sollten mit dem Ziel, einen akzeptierten Rechtsrahmen für zukünftige Entwicklungen zu schaffen. Zugleich sind die sozialen und ethischen Fragen des Einsatzes von ML entscheidend für

[229] The Royal Society, 2017
[230] PwC, 2017; Fraunhofer-Gesellschaft, 2017
[231] Fachforum Autonome Systeme im Hightech-Forum, 2017
[232] Vgl. §13f, EU Datenschutz-Grundverordnung (DSGVO)
[233] National Science and Technology Council, 2016a
[234] Campolo, A.; Sanfilippo, M.; Whittaker, M.; Crawford, K., 2017

das Vertrauen in ML-basierten Lösungen beim Nutzer. Maßstäbe für den verantwortungsvollen Einsatz von ML- Technologien müssen in vielen Fällen anwendungsspezifisch getroffen werden.[235]

Die Rechtssicherheit in Bezug auf die Datenhoheit ist von großer Bedeutung für ML.[236] Aktuell stehen alle Unternehmen und Forschungseinrichtungen vor der Herausforderung, die Frage der Datenhoheit zu klären. Dies umfasst zum einen die Daten, welche bereits heute erfasst werden, bei denen die Datenhoheit jedoch nicht ausreichend geklärt ist. Zum anderen geht es auch um die zukünftig erfassten Daten, beispielsweise durch die Integration weiterer Sensoren in die Produkte.[237] Der Rechtsrahmen sollte auch die Hoheit über zukünftig erfasste, neuartige Daten regeln, und zwar in Bezug auf die Erhebung, die Zugriffsrechte und die sichere Verarbeitung, Speicherung und Verteilung der Daten. Hier können Zielkonflikte bestehen zwischen dem Schutz personenbezogener oder personenbeziehbarer Daten und den Nutzungsmöglichkeiten für ML, welche sich aus dem organisationsübergreifenden Datenaustausch ergeben.[238]

Eine Weiterentwicklung des Rechtsrahmens ist ebenfalls erforderlich für das Haftungsrecht, in Bezug auf maschinelle Entscheidungsprozesse. Dabei sollte geklärt werden, in welchem Umfang Entscheidungskompetenzen an automatisierte ML-Anwendungen abgegeben werden können und sollten.[239] ML-Anwendungen werden dann verstärkt den Weg in den Markt finden, wenn haftungsrechtliche Verantwortungen der Hersteller, Betreiber und Nutzer ML-basierter Systeme geklärt sind und das Risiko ML-basierter Anwendungen versicherbar wird. Insbesondere bei menschennahen Dienstleistungen, wie im Gesundheitswesen, sollte die Rechtssicherheit für alle Beteiligten sichergestellt werden.[240] Im Gewährleistungsrecht[241] sind Verantwortlichkeiten für die korrekte Ausführung von Sicherheits- und Funktionsupdates zu klären, ebenso wie Fragen der Verantwortung im Falle von Fehlern.[242] Dies umfasst nach Einschätzung der befragten Expertinnen und Experten auch die (kartell-)rechtliche Beurteilung für die Verantwortung ML-basierter Markthandlungen, wie z.B. bei online Buchungsplattformen und ihren dahinter liegenden Algorithmen.

Die Wahrung ethischer Prinzipien und Normen von ML-basierten Produkten und Dienstleistungen sind für die Anwendung am Markt zentral. Ethische Prinzipien und Normen müssen dabei immer kontextspezifisch, d.h. in Abhängigkeit von dem im Einsatzland vor-

[235] The Royal Society, 2017
[236] VDMA, 2017
[237] VDMA, 2017
[238] Fachforum Autonome Systeme im Hightech-Forum, 2017
[239] VDMA, 2017
[240] Burgess, M., 2017
[241] Fachforum Autonome Systeme im Hightech-Forum, 2017
[242] Heckmann, D.; Schmid, A., 2017

herrschenden Wertekanon betrachtet werden. Die Möglichkeit, dass ML-basierte vollautomatisierte Systeme mit moralischen Vorstellungen unvereinbar sind, weil ihre Entscheidungen auf einem - dem in Deutschland vorherrschenden gegensätzlichen – Wertesystemen basieren, wird als Gefahr gesehen.[243] In Deutschland entwickelte ML-basierte Produkte und Dienstleistungen sollten die allgemein anerkannten ethischen Prinzipien nicht verletzen. Auch in Bezug auf den Einsatz ML-basierter Technologien spielen derartige Aspekte in Abhängigkeit Ihres jeweiligen, gesellschaftlichen Kontexts eine Rolle, wie das Beispiel des in China modellhaft erprobten Sozialkreditsystems zeigt. Entwicklungsprinzipien, die eine Kontrolle und Anpassung KI-basierter Systeme auf der Grundlage gesellschaftlich anerkannter ethischer Prinzipien gewährleisten, sollten in einem breiten Dialog, der auch in der Politik verankert ist, erarbeitet werden.[244]

Auch Aspekte wie beispielsweise die Auswirkungen des zunehmenden Einsatzes von ML-Verfahren auf den Arbeitsmarkt, nicht nur für routinebasierte Tätigkeitsprofile sondern auch für Hochqualifizierte und in der Wissensarbeit Tätige[245], ebenso wie weitere soziale Aspekte, sollten hinsichtlich der Chancen und Risiken gesellschaftlich breit diskutiert werden.[246] ML-Anwendungen können die soziale Teilhabe der Bürger beeinflussen, beispielsweise durch die zunehmend KI-gestützte Bewerberauswahl für Vorstellungsgespräche oder auch durch die gestiegenen Möglichkeiten in der Arbeitsunterstützung von Schwerbehinderten. Öffentliche Foren für die Auseinandersetzung mit möglichen Vorbehalten oder Sorgen vor einem Verlust der menschlichen Entscheidungshoheit durch ML und KI können den Anfang bilden.[247] Sie sollten helfen, mögliche Vorbehalte der Endverbraucher von Beginn an zu berücksichtigen und so langfristig zur Akzeptanz ML-basierter Produkte und Dienstleistungen beitragen.

[243] National Science and Technology Council, 2016b
[244] EFI, 2018.
[245] OECD, 2017.
[246] Vieth, Kilian, Wagner, Ben, 2017 im Auftrag der Bertelsmann Stiftung
[247] The Royal Society, 2017

KAPITEL 5

EXPERTENKONSULTATION

Inga Döbel (Fraunhofer IMW)
Dr. Juliane Welz (Fraunhofer IMW)
Dr. Henning Petzka (Fraunhofer IAIS)
Frieder Schmelzle (Fraunhofer IMW)

Inhaltsverzeichnis Kapitel 5

5 Expertenkonsultation 164

5.1 Einleitung 164

5.2 Maschinelles Lernen – Methoden und Lerndaten 165
5.2.1 (Lern-)Methoden 165
5.2.2 Deep Learning 166
5.2.3 Nutzung von Frameworks 167
5.2.4 Lerndaten und offene Beispielmengen 167

5.3 Forschungsthemen und Forschungsbedarf 168
5.3.1 Entwicklung des Maschinellen Lernens in den letzten 8 Jahren (Wissenschaft und Praxis) 169
5.3.2 Alleinstellungsmerkmale der deutschen Forschung 169
5.3.3 Zukünftige Relevanz ausgewählter Forschungsthemen 170

5.4 Empfehlungen für die Forschungspolitik 171

5.5 Kompetenzlandschaft in Deutschland 172
5.5.1 Wichtige Akteure in der Wissenschaft 173
5.5.2 Akteure und Kompetenzen in der Wirtschaft 174
5.5.3 Anwendungspotenziale und Branchen 176

5.6 Sozioökonomische, rechtliche und politische Rahmenbedingungen 177
5.6.1 Sozioökonomische Rahmenbedingungen 177
5.6.2 Rechtliche Rahmenbedingungen 179
5.6.3 Politische Rahmenbedingungen 179

5.7 Ausblick 180

5 Expertenkonsultation

5.1 Einleitung

Im Rahmen des Projektes »Maschinelles Lernen – Kompetenzen, Anwendungen und Forschungsbedarf« wurde eine Bestandsaufnahme der deutschen Forschungs- und Kompetenzlandschaft im Themenbereich des Maschinellen Lernens und den damit verbundenen Technologien erarbeitet. Ein wesentlicher Bestandteil dieser Bestandsaufnahme war die Erfassung und Integration von Expertenwissen, um die wissenschaftlichen Analysen durch Praxiswissen zu validieren.

Vor diesem Hintergrund wurden Expertenkonsultationen mit relevanten Akteuren aus der Wissenschaft und Wirtschaft durchgeführt. Diese zielten darauf ab, **aktuelle und zukünftige Forschungsthemen**, entsprechende **wissenschaftlich-technische und ausbildungsbezogene Kompetenzen** sowie die **wichtigsten Einsatzgebiete in Deutschland** zu identifizieren.

Die Expertenkonsultationen umfasste die Konzeption eines Interviewleitfadens in Abstimmung mit Inhalten der anderen Arbeitspakete (AP), die Koordination, Durchführung und Auswertung von Telefoninterviews mit Fachleuten aus der unternehmerischen Praxis und Face-to-Face-Interviews mit Fachleuten aus der Forschung. Die Ergebnisse mündeten in einer deskriptiven, qualitativen Studie mit wichtigen Interviewpassagen, welche wesentliche Erkenntnisse der jeweils anderen AP validieren und in einen praxisbezogenen Kontext einordnen.

Insgesamt wurden 15 Interviews mit 18 Fachleuten vom 1. September bis 30. Oktober 2017 durchgeführt.[248] Es wurden 8 Vertreter der Grundlagen- und anwendungsorientierten Forschung, darunter Professorinnen und Professoren und wissenschaftliche Mitarbeitende der deutschen Universitäten und außeruniversitären Forschungseinrichtungen befragt. Zehn weitere interviewte Fachleute kamen aus Großunternehmen mit eigenen ML-Forschungsgruppen, von Lösungsanbietern, KMU bzw. Start-ups. Die Tiefeninterviews hatten eine Dauer von 45 bis 60 Minuten und wurden unter Zustimmung der Interviewten audiotechnisch aufgezeichnet und transkribiert. Die Datenreduktion und Kategorienbildung bei der Auswertung erfolgte auf Basis einer zusammenfassenden qualitativen Inhaltsanalyse. Für die jeweiligen Teilnehmergruppen („Wissenschaft“ und „Branchenvertreter“) wurden die Interviewleitfragen angepasst. Die Interviews wurden anfänglich von einem Experten für Maschinelles Lernen aus dem Projektteam begleitet. Einzelne Aussagen, insbesondere hinsichtlich Kapitel 1 und 2, wurden ebenfalls in Abstimmung mit den

[248] Über 40 Fachleute wurden dafür angeschrieben und zu Interviews eingeladen.

ML-Experten im Projektteam ausgewertet. Ausgewählte Aspekte wurden darüber hinaus im Rahmen des Validierungsworkshops mit weiteren Fachleuten aus der Wissenschaft, Politik und Praxis diskutiert.

5.2 Maschinelles Lernen – Methoden und Lerndaten

Die Beherrschung und Anwendung der einschlägigen ML-Methoden bildet den Kern jeder Disziplin. In diesem Zusammenhang wurden folgende Fragen an die Fachleute gestellt:

a) Welche (Lern-)Methoden finden in Ihrem Forschungsfeld Anwendung und welchen Stellenwert nehmen Deep Learning und probabilistische Modelle ein?
b) Welche Fragestellungen stehen dabei im Mittelpunkt?
c) Mit welchen ML-Tools, Plattformen oder Bibliotheken arbeiten Sie?

Darüber hinaus wird im Allgemeinen die Verfügbarkeit von Lerndaten und offenen Beispielmengen als eine wesentliche Voraussetzung für ML gesehen, die jedoch bis heute vor großen Herausforderungen steht. Diesbezüglich wurden die Experten gefragt:

d) Welche Rolle spielen »offene Beispielmengen« und »Lerndaten« in Ihrem Anwendungsgebiet (Art der Daten, Quelle, Umfang und Aufwand für die Datenaufbereitung und Qualitätskontrolle der Daten)? (s. Abschnitt 5.2.4).
e) Welche Art von Daten sollten in Zukunft verfügbar gemacht werden und welche Rahmenbedingungen müssten dafür erfüllt sein?

Kapitel 5 fasst die wesentlichen Aussagen aus diesen Fragestellungen zusammen und nimmt zunächst Bezug auf die genutzten (Lern-)Methoden im Allgemeinen und auf Deep Learning im Speziellen. Es folgen Aussagen zu den relevanten Frameworks sowie zur Nutzung offener Beispielmengen und Lerndaten.

5.2.1 (Lern-)Methoden

Insgesamt verweisen die Fachleute auf eine große Bandbreite an Methoden, die sie in ihrer Tätigkeit – meist als Wissenschaftler – anwenden. Zu den genannten Methoden zählen Reinforcement Learning, Support Vector Machines, Entscheidungsbäume, Gauß-Prozesse, Regellernen und Repräsentationslernen. Weitere Themen sind die Multi Label Klassifikation, Kernel-basierte Ansätze und verteiltes Lernen. Die Anwendung dieser Themen zeigt auch, dass eine ganze Reihe von Ansätzen Anwendung finden und diese nach wie vor aktuell und relevant sind.

Im Interview wurde gezielt nach Deep Learning und probabilistischen Modellen in der Wissenschaft und in der Anwendung gefragt. Die Expertenmeinungen zum Thema Deep Learning werden in Abschnitt 5.2.2im Detail beschrieben.

Die **probabilistischen Modelle** werden sowohl aktuell als auch in Zukunft als ein sehr wichtiges Thema eingeschätzt, insbesondere, weil in den letzten Jahren sehr viel Wissen in

diesem Bereich aufgebaut wurde (u.a. Modellierung von Rankings und Präferenzdaten). Dessen ungeachtet, ergänzen einige Fachleute, dass auch in diesem Bereich noch nicht alle Potenziale ausgeschöpft sind und deshalb auch mit einer stetig hohen Bedeutung von probabilistischen Modellen in Zukunft zu rechnen sein wird. Verbesserungen werden in den probabilistischen Eigenschaften der Methoden erwartet, denn noch immer haben die Fragestellungen, die probabilistischen Modellen zugrunde liegen, eine hohe Relevanz. Es werden darüber hinaus Potenziale in Ansätzen gesehen, die über die klassischen probabilistischen Methoden hinausgehen. Dies könnten u.a. probabilistische Modelle als Teilkomponenten von tiefen Lernverfahren sein, vor allem dort, wo komplexe neue Modelle mit einer großen Menge von Parametern oder vorklassifizierten Trainingsbeispielen vorliegen. Probabilistisch-logische Modelle könnten hier als strukturelles Element dienen, indem das grafische Element der Wissensgraphen mit einer präzisen Numerik und Modellierung verbunden wird.

5.2.2 Deep Learning

Tiefes Lernen bedeutet das Lernen in künstlichen neuronalen Netzen mit mehreren bis sehr vielen inneren Schichten. Das Verfahren konnte in den letzten Jahren herausragende Erfolge insbesondere bei Sprach- und Text-, Bild- und Videoverarbeitung verzeichnen. In diesem Abschnitt wird der Stellenwert, den Deep Learning im aktuellen Forschungsfeld der befragten Experten einnimmt, skizziert.

Aus Sicht der Fachleute hat Deep Learning einen Durchbruch in der Technologie und damit die „**Renaissance der künstlichen Intelligenz“** eingeleitet. Von diesem Erfolg haben auch **andere Verfahren**, wie u.a. Wissensgraphen und natürliche Sprachverarbeitung und Multi Label Klassifikation profitiert. Deep Learning hat in vielen Bereichen den **Übergang in die Praxis/Anwendung** gefunden, insbesondere, weil sich die **Fehlerraten halbiert** haben und »Quantensprünge« in der Genauigkeit der Methode erzielt werden konnten. Die Marktrelevanz von Lernverfahren hat sich dadurch erhöht und neue Anwendungsklassen konnten erschlossen werden.

Deep Learning wurde in den letzten Jahren besonders populär und erfährt viel Aufmerksamkeit, ist jedoch laut Aussagen der Fachleute **nur eines von vielen Verfahren**, die sich je nach Anwendungsfall unter Umständen besser eignen können. Besonders erfolgreich ist Deep Learning dort, **wo sehr große Datenmengen und die notwendigen Rechnerleistungen zur Verfügung stehen,** wie zum Beispiel in der Bild- und Spracherkennung. In anderen Bereichen hingegen, wo lediglich kleine Datenmengen vorhanden oder komplexe Aufgabenstellungen zu bearbeiten sind, werden bislang andere Methoden als geeigneter angesehen.

Den Fachleuten zufolge sind **noch nicht alle Potenziale des Deep Learnings ausgeschöpft**, so dass es aus Sicht der meisten Befragten noch längerfristig eine wichtige Rolle für Maschinelles Lernen spielen wird, sowohl in der Grundlagenforschung als auch in der anwendungsorientierten Forschung. Vereinzelt wurde allerdings auch erwähnt, dass es sich beim Deep Learning möglicherweise um einen kurzfristigen Hype handelt, der wieder

abflachen könnte. Insbesondere im Zusammenhang mit Sicherheitsaspekten könnten sich in manchen Anwendungen aktuelle Defizite hinsichtlich der Interpretier- und Nachvollziehbarkeit als problematisch erweisen. Auch aus diesem Grund empfehlen einige Fachleute, dass es in Zukunft auch wichtig sein wird, andere Methoden weiter zu stärken und wissenschaftlich weiter zu entwickeln.

5.2.3 Nutzung von Frameworks

Die Fachleute wurden zu den von ihnen verwendeten »ML-Tools, Plattformen oder Bibliotheken« befragt. Mit Abstand am häufigsten wurde **TensorFlow** (Google) genannt, gefolgt von **Keras, scikit-learn, Weka, RapidMiner** (jeweils entwickelt von Forschungsgruppen oder Einzelpersonen) und **Matlab** (einzige kommerzielle Software).
Hinsichtlich der verwendeten Tools und Bibliotheken herrscht innerhalb der Forschung eine **offene Policy**, insbesondere die intensive Nutzung von **Open-Source-Lösungen** ist weit verbreitet. Für besonders spezifische Anwendungsfelder werden von Unternehmen oft **eigene Tools** entwickelt **und interne Datenbanken** verwendet.

Die verwendeten Lernmethoden sind sowohl in der Forschung als auch in der Anwendung **äußerst vielfältig**. Als vermutlich am weitesten verbreitete Programmiersprache gilt **Python**, gefolgt von C++.

Die Antworten der Befragten zeigen weiterhin, dass wichtige Frameworks **häufig von großen internationalen Unternehmen entwickelt und zur Verfügung gestellt** werden (z.B. TensorFlow von Google), was sich als problematisch erweisen könnte, sollten diese ihre Angebote nicht mehr frei zur Verfügung stellen. Auch die **Verfügbarkeit von Trainingsdatensätzen ist von hoher Bedeutung**. Klassische Datawarehouse-Lösungen werden wegen verlässlicher Performanz und gut strukturierter Daten teilweise bevorzugt.

5.2.4 Lerndaten und offene Beispielmengen

Die **Verfügbarkeit von Lerndaten** ist aus Sicht der befragten Fachleute noch immer eine große Barriere für die Weiterentwicklung des Maschinellen Lernens. Obgleich im Bereich der Bild- oder Sprachklassifikation internationale Anbieter große Datenbanken aufgebaut haben, besteht in vielen Anwendungen ein Bedarf an relevanten Lerndaten. Dabei ist die Qualität der Lernmengen entscheidend für die Qualität von ML-Lösungen. Dies wird aktuell als eine der größten Herausforderungen gesehen. Während in der Forschung und Wissenschaft eine Kultur von »Open Data« für etablierte Probleme besteht, d.h. es steht eine größere Menge von Testdaten/Benchmark-Daten zur Verfügung, ist die Bereitstellung von **Daten aus dem wirtschaftlichen Umfeld** noch immer limitiert. Insbesondere Datenströme und Ranking-Daten (bei denen mindestens eines der Merkmale eine Rangordnung der Einträge darstellt), bzw. Daten für Präferenzlernen, haben einen hohen kommerziellen Wert und werden oft als unternehmensstrategische Geschäftsgeheimnisse gehütet. Aus wissenschaftlicher Sicht schränkt dieser Umstand die Nachvollziehbarkeit und Reproduzierbarkeit veröffentlichter Forschungsergebnisse ein. Aus diesem Grund wird insbesondere von den Befragten aus der Wissenschaft darauf hingewiesen, dass für die Grundlagenforschung ein Schwerpunkt auf die Bereitstellung von Daten sowie die Anerkennung und

Honorierung für die Bereinigung, Veröffentlichung und Standardisierung von Daten gelegt werden sollte.

Die Experten unterstreichen aber auch, dass es in Zukunft nicht darum gehen wird, lediglich Massendaten zu sammeln, sondern vielmehr **tatsächliche Realweltdaten** (insbesondere aus Unternehmen), die auch **Anomalien** enthalten. Es muss **Wissen um die „richtigen" Daten** generiert werden, wofür auch entsprechendes Verständnis und aktives Mitwirken seitens der Daten sammelnden (zukünftigen) ML-Anwender nötig ist. Neben dieser Kooperationsbereitschaft ist für qualitativ möglichst hochwertige Datensätze den Befragten zufolge auch das **gewissenhafte Kuratieren** von zentraler Bedeutung. Entsprechend sollte nicht nur die Veröffentlichung von Daten, sondern auch deren Aufbereitung, Dokumentation und Pflege stärker honoriert werden.

Aus Sicht der Experten sind vor allem die folgenden Fragen relevant:

- Woher bekommt man Daten?
- Wie können Daten verdichtet werden?
- Wie schaffe ich es, dass ich die richtigen Daten behalte und die falschen aussortiere?
- Wie können Datenmassen reduziert werden?

Unabhängig von Strategien und Programmen zur Datengenerierung, heben die Fachleute die Notwendigkeit hervor, dateneffiziente Verfahren (Transferlernen, Lernen mit wenigen Beispielen) in Forschungsprogrammen stärker zu adressieren. Im Kern sollte in Zukunft das **Lernen mit weniger großen Datenmengen** stärker in den Mittelpunkt der Forschung und Entwicklung gestellt werden.

5.3 Forschungsthemen und Forschungsbedarf

Mit Blick auf die aktuelle internationale Forschungslandschaft (z.B. USA, Südkorea, China) ist die **Identifizierung von zentralen und zukunftsrelevanten Forschungsthemen für den Wissenschaftsstandort Deutschland** von besonderem Interesse. In diesem Zusammenhang wurden folgende Fragen an die Fachleute gestellt:

a) In welchen Themenfeldern ist Deutschland im Bereich des ML gut aufgestellt und was sind aus Ihrer Sicht die Alleinstellungsmerkmale der deutschen Forschung?
b) Welche Forschungsthemen sind aus Ihrer Sicht für Deutschland in den nächsten 4 bzw. 8 Jahren wichtig und welche Ideen und Ansätze sollten dazu besonders verfolgt werden?

Darüber hinaus war die zukünftige Relevanz von zehn vorgegebenen Forschungsthemen bzw. -ansätzen für die Entwicklung von ML durch die Fachleute einzuschätzen. Die Befragten konnten auf einer Skala von „sehr relevant" bis „nicht relevant" ihre Einschät-

zung einordnen und diese begründen. Die Frage wurde nur Fachleuten aus der Wissenschaft gestellt.

Dieses Kapitel fasst die wesentlichen Aussagen aus diesen Fragestellungen zusammen, beginnt mit einem Rückblick auf die Entwicklung der Forschungsthemen in den letzten 8 Jahren und nimmt anschließend Bezug auf Forschungsthemen, die Deutschland als Wissenschaftsstandort im Bereich Maschinellen Lernens auszeichnen. Daran anschließend werden Zukunftsthemen und -schwerpunkte vorgestellt und Empfehlungen an die Forschungspolitik formuliert.

5.3.1 Entwicklung des Maschinellen Lernens in den letzten 8 Jahren (Wissenschaft und Praxis)

Das Thema Maschinelles Lernen zeigt unterschiedliche Entwicklungspfade in der Wissenschaft und in der Praxis. Die Fachleute aus der **Wissenschaft** können auf eine lange Forschungstradition zurückblicken, teilweise bis in die 1980er Jahre, und stellen die Kontinuität in der Forschung des Maschinellen Lernens heraus. In den letzten Jahren verzeichnen sie einen steigenden und stetigen Anstieg sowohl in der Anzahl der Forschungsprojekte als auch in der Studentenzahl in Vorlesungen und Seminaren.

In der **Wirtschaft** hingegen war Maschinelles Lernen bis vor fünf Jahren noch ein Nischenthema, welches nur von einigen wenigen Akteuren aufgegriffen wurde. In den letzten Jahren ist es jedoch immer stärker in den Vordergrund gerückt. Einige Industriebereiche haben begonnen, eigene ML-Einheiten aufzubauen und stellen zunehmend Expertinnen und Experten für Deep Learning, Signalverarbeitung oder Zeitreihenanalyse ein.

Insgesamt hat Maschinelles Lernen **an Relevanz gewonnen**. Immer mehr Wissenschaftsakteure und Unternehmen interessieren sich für dieses Thema, sowohl aus anderen Teildisziplinen der Informatik als auch aus fachfremden Bereichen oder Wissenschaftsgebieten kommend.

5.3.2 Alleinstellungsmerkmale der deutschen Forschung

Die deutsche Forschungslandschaft des Maschinellen Lernens ist laut den befragten Fachleuten international sehr gut positioniert. Ende der 1990er Jahre sind die Themenschwerpunkte **Support Vector Machines (SVM)** und generell **Kernel-basierte Methoden** international durch die deutschen Forschungsakteure geprägt worden. Diese sind auch heute noch von Relevanz und gelten als Stärken der deutschen Forschung. Darüber hinaus zeigt die deutsche Forschungslandschaft Stärken u.a. in den Grundfragen der **Algorithmik** (große Zahl von Publikationen) sowie **strukturiertes Lernen und Lernverfahren** (ganze Paradigmen sind stark in Deutschland geprägt worden, wie z.B. die **Subgruppenentdeckung**).

Im internationalen Vergleich hat die deutsche KI-Forschung relativ lange an klassischen Ansätzen logik- und regelbasierter Verfahren festgehalten. Mitunter aus diesem Grund weniger gut aufgestellt ist die Forschungslandschaft laut Experten **im Bereich Deep Learning** und bei der **Internetdatenanalyse** (Suchmaschinen u.Ä.). Hier sind andere

Länder, insbesondere die USA, federführend. Bei Studenten und der jungen wissenschaftlichen Generation ist das Interesse an dem Thema Deep Learning heute allerdings groß, sie schafft eine zukünftige Nachfrage nach entsprechenden Lehrangeboten an den deutschen Hochschulen.

Als deutsche bzw. europäische Besonderheit wurde weiterhin festgehalten, dass Maschinelles Lernen hier enger mit dem **Bereich Data Mining** verknüpft ist als etwa in den USA, wo sich die Forschung an maschinellen Lernmethoden früher bzw. stärker in Richtung Statistik entwickelt hat.

5.3.3 Zukünftige Relevanz ausgewählter Forschungsthemen

Die Fachleute aus der Forschung wurden zu verschiedenen Themen des maschinellen Lernens explizit befragt und jeweils um eine **Einschätzung der zukünftigen Relevanz** gebeten.

Auffällige Ergebnisse der abgegebenen Einschätzungen sind:

- die einhellig sehr hohe Relevanz von nachvollziehbaren Modellen und der Verbindung maschinellen Lernens mit bestehenden Wissensvorräten,
- die Einschätzungen, dass kontinuierliches Lernen und das Lernen von Algorithmen derzeit noch nicht vor dem Durchbruch stehen,
- Effizienzsteigerung von Lernalgorithmen (mit gleicher Rechenleistung weniger Ressourcen verbrauchen, schneller lernen, weniger Daten zum Lernen nutzen),
- kontroverse Anmerkungen und Einschätzungen zum Lernen in verteilten Geräten (pro: Skalierbarkeit und cloudbasierte Lösungen, dezentraler Datenaustausch / kontra: Datenschutzproblematik, „sehr gefährlich“), insbesondere Uneinigkeit bzgl. zukünftigem Machine Learning auf Smartphones.

Insbesondere im ersten Punkt, der Verknüpfung von Lernprozessen mit Hintergrund-, Experten- bzw. Vorwissen, sehen einige der interviewten Wissenschaftler umfangreiches Potenzial für zukünftige Durchbrüche. Eine Möglichkeit bestünde im Einbinden explizit formulierter Funktionsblöcke, die das Lernen bestimmter Teilmechanismen ersetzen. Ferner existieren Ansätze, welche die Menge der erforderlichen Daten reduzieren, oder synthetische Daten zum Vortrainieren neuronaler Netze generieren könnten.

Weitere inhaltliche Themenschwerpunkte (neben Sprach- und Bildverarbeitung und häufig mit Anwendungsbezug), die von den Fachleuten angesprochen wurden, sind:

- Deep Learning für andere Arten von Anwendungen, z.B. aus dem Chemie- oder Biologiebereich;
- in der Medizin für Auswertung von Patientendaten, Patientenakten, um dann darauf basierend den Gesundheits- oder Krankheitszustand von einer Person zu charakterisieren;

- im Pharmabereich zur Identifizierung von Medikamenten für bestimmte Patientengruppen;
- in der individualisierten Landwirtschaft;
- Vorhersagemodelle für z.B. Marketing-Bereiche oder Schätzmodelle für Verkäufe.

»Wie würden Sie die zukünftige Relevanz der nachfolgenden Forschungsthemen bzw. Ansätze für die Entwicklung von ML einschätzen (für die nächsten 4-8 Jahre)?« (Experten 1, 2, 3a, 3b, 4, 9, 11, 12)						
(Einschätzung der Wissenschaftsexperten)		sehr relevant	Relevant	weniger relevant	nicht relevant	kann ich nicht einschätzen
1	Nachvollziehbare, vertrauenswürdige Modelle, Qualitäts- und Leistungszusicherungen	7	1			
2	Verbindung von Lernen mit Experten-, Hintergrund- und Alltagswissen	7	1			
3	Lernen in Mensch-Maschine-Teams (Modellentwicklung, -evaluation, Instruktion, Delegation von Antworten, …)	4	3	1		
4	Kontinuierliches Lernen, lebenslanges Lernen	2	5	1		
5	Lernen mit wenigen Daten (one-shot, zero-shot, automatisches Labeln, ...)	5	1	2		
6	Lernen von Algorithmen (z.B. Neural Computing)		4	2		2
7	Performantes Training (z.B. durch neuromorphic computing, Quantum Learning, Spezial-Hardware, verteilte Algorithmen)	2	5	1		
8	Lernen in verteilten Geräten, Lernen mit beschränkten Ressourcen	3	4		1	

5.4 Empfehlungen für die Forschungspolitik

Die Fachleute haben basierend auf ihren Erfahrungen Empfehlungen für die Forschungspolitik formuliert, die sich in zwei übergeordnete Themenfelder untergliedern lassen: thematische und organisatorische Empfehlungen.

Nach den Beobachtungen der Fachleute wird Maschinelles Lernen in der deutschen Forschung, anders als in bspw. den USA, als **kleines Spezialgebiet** behandelt. Dabei ist es ein großer, zukunftsträchtiger Forschungsbereich, welcher viel mehr Ressourcen und Aufmerksamkeit sowohl in der Informatik als auch im interdisziplinären Kontext erhalten sollte. In Bezug auf die **thematischen** Schwerpunktsetzungen der Förderprogramme wünschen sich die Befragten eine stärkere **Förderung der Grundlagenforschung** im Bereich des Maschinellen Lernens. Sie sehen u.a. noch Potenziale im Bereich der **Algorithmik**, insbesondere zum Thema komplexe Datenstrukturen, Hochgeschwindigkeitsanforderungen auf Datenströmen sowie Veröffentlichung in standardisierten Formaten. Darüber hinaus wird angeregt, Anreize für das Kuratieren, die Standardisierung, die Erstellung und Veröffentlichung von Datensätzen zu schaffen. Damit könnte das **Transferler-**

nen von einem System auf das andere besser ermöglicht werden. Diese Themen könnten in **Modell- oder Testfabriken** erforscht werden. Bestimmte Algorithmen könnten hier basierend auf öffentlichen Datensätzen getestet oder Demonstrationen analysiert werden. Algorithmen und Standard-Datensätze könnten so besser verglichen und eingesetzt werden. Im Bereich der anwendungsorientierten Forschung werden Potenziale von ML für **Industrie 4.0** gesehen. Neben der Stärkung der Anwendungen in den „klassischen" Branchen, wie industrielle Produktion und Maschinenbau, Automobilindustrie und Medizin, sollten auch weitere Anwendungsgebiete mit hohem Innovationspotenzial, wie z.B. öffentliche Verwaltung (Nahverkehr), Rechtswissenschaften, Chemie, Biologie, Landwirtschaft, Versicherungswirtschaft und das Thema „Process Mining", gefördert werden.

Die Empfehlungen **für die organisatorische Ausrichtung** der Forschungsförderung weisen in folgenden Richtungen:

1. Stärkung der **Grundlagenforschung**, um neue Methoden, Ansätze und Ideen zu entwickeln, die u.a. in Anwendungen münden;
2. Sicherung der **Langfristigkeit** von Forschungsförderung (Förderung von Langzeitprojekten).
3. **Interdisziplinarität** adressieren und den Dialog zwischen Naturwissenschaften und Geisteswissenschaften stärker fördern;
4. Förderung von **Transferprojekten zur gezielten Stärkung der Kooperation zwischen Grundlagenforschung und Anwendung.**

5.5 Kompetenzlandschaft in Deutschland

Die Wahrnehmung der Kompetenzlandschaft in Deutschland lässt Rückschlüsse auf wichtige Schwerpunkte des Maschinellen Lernens schließen und ist dementsprechend einer der zentralen Bausteine dieser Untersuchung. In diesem Zusammenhang wurden folgende Fragen an die Fachleute gestellt:

a) Wer sind aus Ihrer Sicht die wichtigsten Akteure bzw. Organisationen im Bereich des ML in der deutschen Wissenschaft/Wirtschaft?
b) Mit Blick auf Ihr Kooperationsnetzwerk, mit welchen der genannten Akteure stehen Sie im engsten Austausch (z.B. durch gemeinsame Projektverbünde)? Mit wem würden Sie darüber hinaus in Zukunft kooperieren wollen (ML-Forschungsprojekt in den nächsten 4 Jahren)?
c) Welche deutschen Wirtschaftsakteure (Lösungsanbieter bzw. Anwender) sind für Ihren ML-Fachbereich besonders relevant? Welche internationalen Unternehmen könnten Sie sich für zukünftige Kooperationen als „Wunschpartner" vorstellen?

Die Einschätzung der Fachleute zu möglichen Anwendungspotenzialen erfassen den derzeitigen Stand sowie zukünftige Dynamiken im Feld des Maschinellen Lernens. Folgende Fragen standen im Mittelpunkt:

d) Im Hinblick auf den Einsatz der für Sie relevanten ML-Verfahren, welche Anwendungen, Dienstleistungen und Produkte sind aus Ihrer Sicht bereits gut erschlossen?
e) Welche Anwendungsgebiete bzw. Branchen werden in den nächsten 8 Jahren Ihrer Meinung nach einen hohen Stellenwert einnehmen?

Abschnitt 4 fasst im Rahmen einer Auswertung die wesentlichen Aussagen zu diesen Fragestellungen zusammen und nimmt zunächst Bezug auf die wichtigsten Akteure in Wissenschaft und Wirtschaft. Im Anschluss daran werden Anwendungspotenziale und Branchen näher beschrieben.

5.5.1 Wichtige Akteure in der Wissenschaft

Die Einschätzungen der Fachleute zu den wichtigsten Akteuren in der deutschen Forschungslandschaft sind bemerkenswert **konsistent**. Die Antworten bezogen sich dabei teils auf wissenschaftliche Institutionen und teils auf einzelne Forschende.
Insgesamt konstatieren die Befragten, dass **Deutschland in der ML-Forschung gut aufgestellt** ist. Als eines der führenden Länder in Europa verliert Deutschland allerdings an Boden gegenüber den USA und China. Spezielle Themen (kernel-basierte ML Verfahren und SVM, statistical relational learning) sind dabei für die deutsche Forschung charakteristisch (vgl. Abschnitt. 3.2) und mit den Aktivitäten führender Forschungsgruppen verbunden.

Die Nennung einzelner Personen spiegelt ein sehr ähnliches Bild, wie das der genannten Institutionen, wider. Als wichtigste und auch international sichtbare Forschende wurden **Bernhard Schölkopf** (MPI Tübingen) und **Klaus-Robert Müller** (TU Berlin) genannt.

Als wichtigste Forschungsgruppen sehen die Experten:

Technische Universität Berlin	Gruppe Maschinelles Lernen am Institut für Softwaretechnik und Theoretische Informatik	Leitung Klaus-Robert Müller
Max-Planck-Institut für Intelligente Systeme Tübingen	neben weiteren, insbesondere die Abteilung Empirische Inferenz	Leitung Bernhard Schölkopf
Technische Universität Dortmund	Lehrstuhl für Künstliche Intelligenz an der Fakultät für Informatik	Leitung Katharina Morik, bis April 2017 mit Kristian Kersting
Universität Bonn	Institut für Informatik III – Intelligent Systems / Fraunhofer IAIS	Leitung Stefan Wrobel
Technische Universität Darmstadt	Professur für Intelligente Autonome Systeme im Fachbereich Informatik	Leitung Jan Peters

Als weitere bedeutende Forschungseinrichtungen (1 Erwähnung) wurden genannt: DFKI (allgemein), DFKI Saarbrücken, Fraunhofer IOSB (Ettlingen/ Karlsruhe), Fraunhofer ITWM (Kaiserslautern), KIT, TU München, DFKI Bremen, DLR, Uni Bielefeld, Uni Bremen, LMU München, Uni Potsdam, Uni Tübingen, RWTH Aachen, Uni Duisburg-Essen, Uni Trier.
Als **Partner in bestehenden Forschungskooperationen** wurden darüber hinaus weitere Forschungseinrichtungen (z.B. Uni Mannheim, Uni Paderborn, Uni Duisburg-Essen) genannt.

Als mögliche »Wunschpartner« für zukünftige Kooperationsprojekte wurden abhängig von der thematischen Fragestellung und weiteren Faktoren, u.a. die Fraunhofer Institute (z.B. ITWM, IAIS, IOSB), das DFKI, das Max-Planck-Institut in Tübingen, die TU Dortmund und die TU München genannt.

5.5.2 Akteure und Kompetenzen in der Wirtschaft

Die Angaben der Fachleute zu relevanten Wirtschaftsakteuren in der kommerziellen ML-Anwendung sind im Vergleich **zurückhaltend und weniger einheitlich**. Mögliche Gründe könnten hierfür sein, dass die Wirtschaftsexperten sich hier bewusst bedeckt halten, und dass den Wissenschaftsexperten nur beschränkte Einblicke in spezifische Anwendungsbereiche zur Verfügung stehen. Weiterhin sind Einheitlichkeit und Vergleichbarkeit durch das qualitative Befragungsdesign und kontextabhängige Angaben nur sehr eingeschränkt gegeben. Die Experten weisen auch darauf hin, dass es **wenig transparent ist, welche Anbieter tatsächlich ML in ihren Systemen bzw. Produkten verwenden**. Dadurch wird die Einschätzung der Wettbewerbssituation erschwert.
Im Vergleich zur ML-Forschung **hängt das Angebot an ML-Produkten und Dienstleistungen in Deutschland noch weit zurück**. Während im Vergleich zu anderen europäischen Ländern die Positionen (insbesondere großer Anbieter) noch stark sind, gelten im internationalen Vergleich die Anbieter aus den USA (z.B. IBM, Google, Amazon, Face-

book, Tesla) als führend und wirtschaftlich erfolgreich. Diese Unternehmen bündeln starke ML-Kompetenzen, verfügen über große Mengen an Trainingsdaten bzw. Infrastruktur und ziehen die »besten Köpfe« an. Einige Themengebiete, wie bspw. »Intelligente Automatisierung« sind in den USA im Vergleich zu Deutschland jedoch weniger gut vertreten. Darüber hinaus kommen starke Player aus Asien (z.B. Baidu, Weibo).

Aus den Aussagen ergibt sich das Bild einer **fragmentierten und heterogenen Akteurslandschaft**, die seitens der Software-, bzw. Toolanbieter und Dienstleister in große Player wie SAP und Kleinstunternehmen/Startups, die einzelne industriespezifische Nischen adressieren, aufgeteilt ist. Maschinelles Lernen findet sich darüber hinaus als Thema in den Unternehmensstrategien großer Technologieunternehmen, die die Technologie in eigenen FuE-Abteilungen und Forschungsgruppen weiterentwickeln und in Produkte sowie Prozesse integrieren (u.a. Siemens, Daimler, Bosch, VW, Audi). Der deutsche Mittelstand ist unter den Anbietern bzw. Entwicklern hingegen nur bedingt vertreten. Es wurde mehrfach angesprochen, dass es dabei weniger an grundsätzlicher Bereitschaft, sondern an fachlichen Kompetenzen sowie dem für Maschinelles Lernen essentiellen Vermögen umfänglicher Datenerhebungen fehlt.

Genannte Wirtschaftsakteure:

Softwareanbieter:	**Großunternehmen:**	SAP
	Klein(st)unternehmen und Start-ups:	RapidMiner (Dortmund), Point8 (Dortmund), arconsis IT-Solutions GmbH (Karlsruhe), IPS Engineers (Dortmund)
Technologieunternehmen und OEMs	Amazon, Bosch, VW, Siemens, Continental, Audi, Daimler, Hella, BWM	

Die befragten Fachleute beobachten innerhalb der letzten Jahre ein **erheblich gestiegenes Interesse am Thema Maschinelles Lernen** bei den deutschen Wirtschaftsakteuren. Viele der Befragten berichten von engem Austausch zwischen Wissenschaft und Praxis im Rahmen transdisziplinärer Forschungsprojekte, die sowohl die Unternehmen (national und EU-weit) als auch die Wissenschaftlerinnen und Wissenschaftler betreiben (nationale und internationale Anwendungs- und Industrieprojekte). Diese Projekte adressieren **verschiedene Anwendungsfelder und Branchen**, darunter industrielle Fertigung, Stahlbau, Maschinenbau, Medizin, Energiewirtschaft, Agrarwirtschaft, Raumfahrt und Logistik.

Große Technologieunternehmen (z.B. Continental) und Softwareanbieter (z.B. SAP) sind die meistgenannten Partner in bestehenden (Forschungs-)Kooperationen. Als mögliche »Wunschpartner« für zukünftige Kooperationsprojekte wurden große Wirtschaftsunternehmen, Softwaresystemhäuser sowie kleine, innovative Unternehmen genannt.

5.5.3 Anwendungspotenziale und Branchen

Maschinelles Lernen hat den Aussagen der befragten Fachleute zufolge das Potenzial, beinahe in allen Branchen erfolgreichen kommerziellen Einsatz zu finden. In Deutschland wurden die ersten Anwendungen bereits Anfang der 1990er Jahre entwickelt, z. B. neuronale Netze für industrielle Anwendungen. Zentrale Anwendungsfelder heute sind vor allem Bild- bzw. Videoanalyse sowie die Sprachverarbeitung, aber auch autonome Systeme (bspw. Robotik). Als einen bedeutenden Trend nennen die Experten den Einsatz von Wissensgraphen (z. B. in der Medizin).
ML ist allerdings **keine Nischentechnologie mehr**, sie ist sehr umfassend und branchenübergreifend. In Deutschland kommt ML schwerpunktmäßig in »klassischen« Branchen zum Einsatz. Deutschland ist ein starker Industrie- und Maschinenbaustandort, *Autonomes Fahren* ist ein bedeutendes Anwendungsfeld in der Automobilindustrie. In diesem Bereich ist die Verbindung von Deep Learning und Bildverarbeitung besonders erfolgreich gewesen.

Industrielle Produktion, insbesondere im Kontext von Industrie 4.0 und Digitalisierung, bietet viel Potenzial sowohl für die *Industrierobotik* (insb. im Bereich Bildverarbeitung/Videoverarbeitung und Handlungsplanung) als auch für die Analyse von Industriedaten und darauf basierenden *Dienstleistungen im Service-Bereich* (Analysen und Prognosen z.B. für Predictive maintenance, Prozess-, Logistik-, und Energieoptimierung sowie Qualitätsmanagement).

Das **Gesundheitswesen** wird ebenfalls als großes Innovationsfeld für ML sowohl in medizinischen als auch in organisatorischen Aspekten *(z.B. Unterstützung bei ärztlichen Entscheidungen, Diagnostik Endpunktvorhersage, Individualisierte Medizin)* betrachtet.

Ferner ist das **Finanz- und Versicherungswesen** ein bedeutender Sektor für *automatisierte Datenanalyse* (z.B. bei Versicherungsfällen). Die Experten beobachten mit Interesse die Entwicklungen im Bereich Blockchain.

In weiteren Anwendungsbranchen, die in Deutschland aktuell nicht im zentralen Blickfeld der Anbieter stehen, in anderen Ländern aber bereits gut erschlossen sind, sind noch viele Marktpotenziale für ML enthalten. Die Befragten nennen die **öffentliche Verwaltung** (*z.B. für Prozessoptimierung, predictive policing, e-government*), den **Sicherheitssektor** (*z.B. Trendmonitoring in sozialen Medien, Personenerkennung*), die **Chemie- und Pharmaindustrie** (Medikamentenentwicklung) und **intelligentes Gebäudemanagement** im gewerblichen Bereich.

ML-Lösungen werden darüber hinaus für Problemstellungen in anderen wissenschaftlichen Disziplinen entwickelt, z.B. in der Astrophysik, Sprachwissenschaft, Physik, Sozialwissenschaften oder Chemie.

Ein bedeutender Trend bei internationalen Anbietern ist das Angebot von **ML-Dienstleistungen.** Deutsche Unternehmen sind in diesem Segment noch gering vertreten. Die Experten weisen auf den Bedarf an ML-Dienstleistern hin, die bspw. Anpassung

von ML-Systemen wie IBMs »Watson« an Lösung spezifischer Probleme für Endkunden vornehmen.

Bei der Entwicklung von ML-Lösungen ist aus der Sicht der Anbieter die **Verfügbarkeit und Qualität der Lerndaten** von essentieller Bedeutung. Das Bewusstsein dafür, dass das Sammeln von qualitativ hochwertigen Lerndaten für den eigenen Bedarf kritisch für die erfolgreiche und effiziente Einführung von ML-Systemen ist, ist allerdings noch bei wenigen Anwendern vorhanden. Der Bedarf an rechtlicher Regulierung, insbesondere für medizinische Daten, wird akut gesehen. Ein weiterer zentraler Aspekt, der im Zusammenhang mit der Datenerfassung und den Optimierungsprozessen einer Maschine genannt wurde, ist die Anonymisierung und Systematisierung von Daten. Hier bestehen derzeit noch keine einheitlichen Richtlinien bzw. Vorgehensweisen.

5.6 Sozioökonomische, rechtliche und politische Rahmenbedingungen

Um die Potenziale des deutschen Forschungs- und Wirtschaftsstandorts im Bereich Maschinellen Lernens weiter zu stärken, sind innovationsfördernde sozioökonomische, rechtliche und politische Rahmenbedingungen grundlegend. In diesem Zusammenhang wurden folgende übergreifende Frage an die Fachleute gestellt, und ad-hoc während des Interviews detaillierter diskutiert:

> Welche politischen, rechtlichen und sozioökonomischen Rahmenbedingungen sind aus Ihrer Sicht entscheidend für die stärkere Verbreitung von ML in der deutschen Wirtschaft?

Abschnitt 5 fasst die wesentlichen Aussagen aus diesen Fragestellungen zusammen und gliedert sich entlang der einzelnen Themenschwerpunkte.

5.6.1 Sozioökonomische Rahmenbedingungen

Die Verbreitung von Maschinellem Lernen wird unsere Gesellschaft in Zukunft verändern und u.a. **Änderungen in der Arbeitsorganisation** hervorrufen. Dies betrifft auf der einen Seite die **Verfügbarkeit von Fachkräften** mit ML-Kenntnissen. Es werden den Befragten zufolge in Zukunft sowohl »Hyperspezialisten«, als auch **interdisziplinär aufgestellte und anwendungsorientierte Fachkräfte** gebraucht. Obgleich die Anzahl der Studentinnen und Studenten in den Vorlesungen in den letzten Jahren stetig gewachsen ist, betrachten die befragten Fachleute die aktuelle Fachkräftesituation mit Sorge und sehen hier einen wichtigen Handlungsbedarf.

Ausgehend von dem wachsenden Interesse an der Thematik gehen die Befragten davon aus, dass in naher Zukunft (in fünf Jahren) eine echte Maschinenlernexpertise in Deutschland vorhanden sein sollte. Das bedeutet aber auch, dass die **Bildungspläne und die Hochschulcurricula** bereits jetzt ML in der Breite und der Tiefe aufnehmen sollten. Ma-

schinelles Lernen kann als eine Brücke zwischen den Disziplinen dienen. Es ist jedoch noch offen und zu klären, wie viele Fachkräfte in welchen Bereichen gebraucht werden. Die Befragten sind sich einig, dass nicht alle zukünftigen Fachkräfte ausschließlich in Computer Science ausgebildet werden sollten. ML sollte einerseits in der **Grundlagenausbildung und -forschung** langfristig gestärkt werden und andererseits ist eine **interdisziplinäre Integration** zu stärken. Dies bedeutet, dass Kenntnisse und Fähigkeiten in unterschiedlichen Bereichen erwartet werden und dementsprechend ML ein fester Bestandteil der Curricula u.a. in BWL, in Naturwissenschaften/Mathematik, in Wirtschaftswissenschaften, im Werkzeug-/Maschinenbau, Produktionstechnik, Produktionsplanung etc. sein sollte.

Dementsprechend sollte auch die Aus- und Weiterbildungsstruktur an Hochschulen und in Unternehmen angepasst und eine **Flexibilität im Ausbildungssystem** erreicht werden, um auf mögliche Schwankungen in dem Bereich reagieren zu können. Die befragten Fachleute bemängeln, dass **zu wenig neue Stellen an Hochschulen für dieses Fach** geschaffen werden, da dieses Fach innerhalb der deutschen Informatik nicht den notwendigen Rückhalt erfährt. In Deutschland wird dieses Fach nicht als zentral angesehen und entsprechend wenig Professuren werden besetzt bzw. die Aussicht auf einen unbefristeten Arbeitsvertrag in der Wissenschaft ist derzeit noch gering, was in der Folge nicht nur einen Nachwuchsmangel erzeugt, sondern auch teilweise zur Abwanderung ins Ausland führt.

Andererseits wird davon ausgegangen, dass insbesondere im Bereich der einfachen manuellen Tätigkeiten bis zu 80% des Personals durch intelligente, autonome und lernende Maschinen in den nächsten Jahren substituiert werden könnten. Dies wirft die grundlegende Frage auf, wie unsere Gesellschaft der Zukunft gestaltet sein soll. Hier wünschen sich die Befragten eine **stärkere Auseinandersetzung und öffentlichen Diskurs**.

Vor diesem Hintergrund sehen die Fachleute einen Schwerpunkt in der **„Awareness Creation"**, d.h. in der Generierung von Aufmerksamkeit, sowohl in der Gesellschaft als auch in deutschen Unternehmen. Der **Kompetenzaufbau in Unternehmen** wird von den Befragten besonders hervorgehoben, d.h. das Aufzeigen von Anwendungsmöglichkeiten und Potenzialen von Maschinellem Lernen. Dies kann durch Bündelung von fachlicher Kompetenz mit Netzwerkbildung zu Universitäten, Spezialanbietern oder anderen Wissenschaftsorganisationen erfolgen. Es thematisiert auch die geforderte Interdisziplinarität zwischen den verschiedenen Branchen. Deutsche Unternehmen haben zwar das Know-how und die Fähigkeiten, Maschinelles Lernen umzusetzen, „doch das auch zu tun ist noch nicht genügend ausgeprägt". Dementsprechend sollten die Adressaten gezielt angesprochen werden und Möglichkeiten der Anwendung hinsichtlich der Grenzen der Technologie, der Machbarkeit, der Verfügbarkeit von Fachkräften und Kooperationsnetzwerken diskutiert werden. Insgesamt stehen die Unternehmen vor grundsätzlichen Strukturveränderungen. Herkömmliche Aufgabenstrukturen und Denkweisen werden sich komplett in den Unternehmen ändern müssen, eingefahrene Prozesse, Handlungsweisen und Überzeugungen müssen aufgebrochen und neu strukturiert werden. Hier sind die oberen

Führungsebenen angesprochen, welche diese Umstrukturierungsprozesse klar gestalten müssen und ihren Arbeitsnehmer in multidisziplinären Teams Raum und Zeit bieten, um sich dieser Aufgabe zu widmen.

5.6.2 Rechtliche Rahmenbedingungen

Im Zusammenhang mit der Verbreitung Maschinellen Lernens in der Industrie und in der Gesellschaft nannten die befragten Fachleute noch Herausforderungen in den rechtlichen Rahmenbedingungen. Dazu zählen die folgenden Aspekte:

- **Datenschutz**, um die Persönlichkeitsrechte zu wahren
- **Die wirtschaftlichen Interessen** zu erhalten (u.a. Exportkontrolle - „Daten aus einem Land in das andere zu schaffen"),
- **Anonymisierung von Daten**, soweit, dass sie für Testprojekte und Demonstratoren ausreichend Aussagen treffen, aber nicht die Persönlichkeitsrechte einschränken,
- **Urheberrecht**, insbesondere beim Zugriff auf Internetdaten und Analyse von News-Schnipseln aus dem Internet,
- **Verbraucherschutzrecht**, Umgang mit persönlichen Daten, Rechte und Pflichten beim Umgang mit persönlichen Daten,
- **Haftungs- und Versicherungsrecht**, ein selbstlernendes System muss dafür Sorge tragen, dass Fehler in einem abgesteckten Sicherheitsrahmen bleiben bzw. nachvollziehbar oder begründbar sind,
- Schutz vor **Cyberangriffen**, Schutz von Assets insbesondere bei industriellen Anwendungen (es werden Garantien benötigt) - Hacking-Attacken lösen große Ängste aus.

5.6.3 Politische Rahmenbedingungen

Bei den politischen Rahmenbedingungen, die eine stärkere Verbreitung des Maschinellen Lernens unterstützen könnten, stellen die Fachleute insbesondere zwei große Themenfelder heraus: (a) stärkere Förderung der Unternehmen und (b) Bewusstsein und Transparenz schaffen. Deutschland ist aufgefordert diese **Transformation zu gestalten** und **gezielt Akzente zu setzen**.

In Bezug auf die **Förderung von Unternehmen** wünschen sich die Befragten gezielte Anreize in der Wirtschaftspolitik, u.a. in der Analyse von Industriedaten, Forschungsprogramme für deutsche Unternehmen, Erarbeitung passgenauer Lösungen für KMUs sowie Förderung von Innovationen und Startups. Insgesamt sprechen diese Maßnahmen die strategische Förderung von Unternehmen an, die relativ jung am Markt sind und Themen des Maschinellen Lernens in die Anwendung bringen möchten. Auch **kooperierendes Lernen** wird von den Experten thematisiert, d.h. gezielt junge Unternehmen oder Startups mit der etablierten Industrie zusammenzubringen. Dies wäre aus Sicht der Experten eine *„Win-win-Situation"*: Junge Unternehmen würden die Möglichkeit erhalten, Referenzen und Netzwerke aufzubauen sowie innovative Probleme zu lösen, und große und mitt-

lere Unternehmen könnten diese Kooperationen nutzen, um diese innovativen Technologien schnell umzusetzen und auf dem Markt zu etablieren. Insgesamt fordern die Experten auch, dass **Technologien greifbar gemacht werden**. Dies könnte durch die Sammlung von Anwendungs- und Problembeschreibungen, Lösungen, *best practices* oder auch *templates* auf einer Open-Source Web-Plattform erfolgen.

Die **öffentliche Verwaltung** (Bund, Land und Kommunen) könnte hier eine **Vorreiterrolle** einnehmen. Hier besteht ein großes Potenzial beim Ausbau der Digitalisierung im Allgemeinen, aber auch in der Optimierung von Prozessen durch Maschinelles Lernen. Aus Sicht der Experten könnte das Thema *Process Mining* interessant sein, um z.B. bestehende Prozesse deutlich effizienter zu gestalten, um Daten zu analysieren, Probleme und Verzögerungen besser zu antizipieren.

Das zweite große Themenfeld – **Schaffung von Bewusstsein und Transparenz** – wird von vielen Fachleuten als zentrales politisches Aufgabenfeld gesehen. Insgesamt sollte die Kommunikation mit der Öffentlichkeit stärker in den Fokus rücken, denn oft sind Ängste relativ unbegründet und es verbreitet sich eine allgemeine Ungewissheit, wie die gesellschaftliche Zukunft aussehen wird. Die Befragten konstatieren, dass es einer öffentlichen Debatte darüber bedarf, wie Maschinelles Lernen z.B. den Alltag und die Arbeitswelt verändern wird, aber auch darüber, wie sich die Kriegsführung verändern könnte (hin zu Cyber War). Eine öffentliche Debatte zu den komplexen Veränderungen und Auswirkungen durch Maschinelles Lernen trägt zur Aufklärung der Bevölkerung bei.

5.7 Ausblick

Den befragten Fachleuten zufolge wird das Thema des Maschinellen Lernens auch in den nächsten Jahren weiter aktuell bleiben, evtl. wird der Hype etwas abflachen und die eine oder andere Methode stärker in den Vordergrund rücken. Aus Expertensicht wird es eine stärkere Verschiebung von derzeit sehr forschungsintensiven Aktivitäten hin zu mehr Anwendungen in der Industrie geben. Insgesamt werden in Zukunft große Schübe erwartet, jedoch ist es auch heute noch schwer abzuschätzen, in welche Richtung sich dieser zukunftsrelevante Forschungs- und Wirtschaftsbereich entwickeln wird und was das im Einzelnen für die Unternehmen und die Gesellschaft bedeutet. Viel Potenzial wird weiterhin Deep Learning und neuronalen Netzen zugesprochen, aber Klassifizierungsalgorithmen und sonstige einfache Maschinenlernverfahren werden bei Themen wie Predictive Maintenance nicht an Bedeutung verlieren.

Die bisherigen substanziellen Veränderungen in der Bild- und Sprachverarbeitung, die Steuerung von Systemen oder auch in Spielen prägen schon heute unsere Gesellschaft und werden dies auch in Zukunft nachhaltig tun. Es wird sehr stark davon abhängen, wie die zukünftige Forschungsförderpolitik auf diese Trends und Herausforderungen reagieren wird und in welche Richtung sie die Transformation gestalten möchte.

6 Fazit

Maschinelles Lernen und Künstliche Intelligenz sind inzwischen zu einem wirtschaftlich, gesellschaftlich und strategisch hochrelevanten Thema geworden. Ihr Einsatz wird in immer mehr Anwendungsgebieten wettbewerbsentscheidend sein. Insbesondere die USA und China gehören zu den stärksten Wettbewerbern. Somit rücken Fragen zu Forschung und Entwicklung, geeignetem Personal, verfügbaren Daten und kommerziellen Anwendungen immer stärker in den Fokus der Politik, Wissenschaft und Wirtschaft in Deutschland.

Die Forschung der Wissenschaftseinrichtungen zu ML ist in Deutschland gut aufgestellt. Allerdings bestehen Defizite bei der wertschöpfenden Umsetzung der wissenschaftlichen Errungenschaften im Markt. Dies spiegelt sich in den vergleichsweise geringen Patentanmeldungen Deutschlands wider, ebenso wie in der geringen Präsenz von KMU in dem Bereich. Auch von den Fachleuten wurde betont, dass der Transfer in die Praxis unbedingt gestärkt werden muss. Hierzu wurden gezielte Transferprojekte in Kooperation zwischen Grundlagenforschung und Anwendung als Möglichkeit vorgeschlagen.

Ein ernst zu nehmendes Problem für die Wettbewerbsfähigkeit Deutschlands sind die fehlenden Datenspezialisten, also ML- und KI-Fachkräfte. Hierbei geht es nicht allein um die benötigte Menge an Fachleuten, sondern auch um die Breite ihrer Ausbildung, um ML-Technologie für die Bedarfe der Industrien einzusetzen.

Der Zugang zu Daten und ihr kontrollierter Austausch sind essenziell in einem Kontext, wo datengetriebenes Vorgehen wettbewerbsentscheidend ist. Die Datenverfügbarkeit ist momentan eine Herausforderung für Deutschland. Dies erfordert Governance-Modelle, die Anreize schaffen, Daten mit anderen Akteuren für wissenschaftliche und wirtschaftliche Zwecke auszutauschen, ohne, dass hier Kontrollverlust und Datenschutzbedenken auftreten.

Bei neuen, von vielen als disruptiv erachteten Technologien wie ML und KI müssen frühzeitig Chancen und mögliche Vorbehalte berücksichtigt und breit diskutiert werden. Denn Anwendungen, die rechtliche Herausforderungen aber auch ethische Vorstellungen und das Bedürfnis nach Sicherheit und Transparenz adressieren, tragen zum Erfolg auf dem heimischen Markt bei und so mittelbar auch zur globalen Wettbewerbsfähigkeit Deutschlands.

Danksagungen

Wir bedanken uns bei allen Expertinnen und Experten, die für ein Experteninterview zur Verfügung gestanden oder an dem wissenschaftlichen Validierungsworkshop am 29.11.2017 in Berlin teilgenommen haben, sowie den Mitwirkenden aus den Bundesministerien.

Interview mit Expertinnen und Experten:

- Prof. Dr.-Ing. habil. Jürgen Beyerer, Fraunhofer IOSB und KIT
- Alexander Fabisch, DFKI GmbH
- Prof. Dr. Johannes Fürnkranz, Technische Universität Darmstadt
- Eberhard Hechler, IBM Deutschland Research & Development GmbH
- Prof. Dr. Eyke Hüllermeier, Universität Paderborn
- Ralf Klinkenberg, RapidMiner GmbH
- Dr. Melanie Knapp, ITyX Solutions AG
- Prof. Dr. Katharina Morik, Technische Universität Dortmund
- Dr. Michael May, Siemens AG
- Prof. Claus Oetter, VDMA
- Marc Otto, AG Robotik, Universität Bremen
- Prof. Dr. Martin Riedmiller, DeepMind
- Manuel Pereira Remelhe, Bayer AG
- Prof. Dr. Stefan Wrobel, Fraunhofer IAIS

Validierungsworkshop:

- Felix Assion, neurocat GmbH
- Dr. Tarek Besold, City, University of London
- Prof. Dr.-Ing. habil. Jürgen Beyerer, Fraunhofer IOSB und KIT
- Dr. Daniel Büscher, Universität Freiburg
- Peter Deussen, Microsoft Deutschland GmbH
- Florens Greßner, neurocat GmbH
- Prof. Dr. Horst-Karl Hahn, Fraunhofer MEVIS und Jacobs University Bremen
- Matthias Himmer, Salesforce, Einstein Analytics EMEA Central
- Dr. Christoph Kehl, Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag/TAB

- Dr. Anna-Lena Kranz-Stöckle, Bundesministerium für Bildung und Forschung
- Andrea Martin, IBM Deutschland GmbH
- Dr. Christian Reisswig, SAP SE, Deep Learning Center of Excellence
- Catrin Schaefer, Bundesministerium für Gesundheit
- Dr. Matthias Schulz, Projektträger des DLR
- Dr. Michael Weber, Bundesministerium für Bildung und Forschung
- Dr. Stefan Wess, Empolis Information Management GmbH
- Dr. Steffen Wischmann, Institut für Innovation und Technik in der VDI/VDE-IT GmbH

Unser Dank gilt auch den Expertinnen und Experten, die nicht namentlich erwähnt werden wollten.

Anlagen: Verzeichnisse

Literaturverzeichnisse nach Kapiteln

Kapitel 1 und 2

- Amazon (Hrsg.) (2017a): Amazon aws. https://s3-ap-south-1.amazonaws.com/av-blog-media/wp-content/uploads/2017/03/06100746/grad.png. Zuletzt geprüft am 23.10.2017.
- Amazon (Hrsg.) (2017b): Amazon aws. http://docs.aws.amazon.com/machine-learning/latest/dg/images/mlconcepts_image5.png. Zuletzt geprüft am 23.10.2017.
- Arulkumaran,K., Deisenroth, M.P., Brundage,M., Bharath, A. A., (2017): Deep Reinforcement Learning: A Brief Survey. IEEE Signal Processing Magazine 34(6)
- Audibert, J., Munos, R., (2011): Tutorial: Introduction to Bandits: Algorithms and Theory ICML 2011 https://sites.google.com/site/banditstutorial/ Zuletzt geprüft am 08.01.2018
- Bader, S., Hitzler, P. (2014): Dimensions of Neural-symbolic Integration — A Structured Survey. http://daselab.cs.wright.edu/resources/publications/pdf/nesy_survey_05.pdf. Zuletzt geprüft am 08.01.2018.
- Barocas, S., Hardt, M., (2017): Fairness in Machine Learning NIPS 2017 Tutorial http://mrtz.org/nips17/#/ Zuletzt geprüft am 08.01.2018
- Bengio, Y. et al. (2007): Greedy layer-wise training of deep networks, in: Advances in neural information processing systems Vol. 19, 153.
- Bennett, K., Parrado-Hernandez, E. (2006): The Interplay of Optimization and Machine Learning Research. http://www.jmlr.org/papers/volume7/MLOPT-intro06a/MLOPT-intro06a.pdf. Zuletzt geprüft am 23.10.2017.
- Binder, A. et al. (2016): Analyzing and Validating Neural Networks Predictions. https://www.hhi.fraunhofer.de/fileadmin/News/2016/Auszeichnung_fuer_Wissenschaftler_des_Fraunhofer_HHI__der_TU_Berlin_und_der_SUTD/Analyzing_and_Validating_Neural_Networks_Predictions.pdf. Zuletzt geprüft am 23.10.2017.
- Bitkom (Hrsg.) (2017): Künstliche Intelligenz verstehen als Automation des Entscheidens, Berlin.
- Bolukbasi, T. et al. (2016): Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings https://arxiv.org/abs/1607.06520 Zuletzt geprüft am 08.01.2018
- Boulard, H., Kamp, Y. (1988): Auto-Association by Multilayer Perceptrons and Singular Value Decomposition.

https://pdfs.semanticscholar.org/f582/1548720901c89b3b7481f7500d7cd64e99bd.pdf. Zuletzt geprüft am 24.10.2017.

- Bousmalis, K., Levine, S., (2017): Closing the Simulation-to-Reality Gap for Deep Robotic Learning https://research.googleblog.com/2017/10/closing-simulation-to-reality-gap-for.html Zuletzt geprüft am 08.01.2018
- Breiman, L. et al. (1983): CART: Classification and Regression Trees, Belmont.
- Breslow, L., Aha, D. (2006): Simplifying Decision Trees. http://ce.aut.ac.ir/~shiry/lecture/machine-learning/hw/OLD/AIC-96-014.pdf. Zuletzt geprüft am 08.01.2018.
- Britz, D., (2016): Attention and Memory in Deep Learning and NLP http://www.wildml.com/2016/01/attention-and-memory-in-deep-learning-and-nlp/ Zuletzt geprüft am 08.01.2018
- Brownlee (2017): Classification and Regression Trees for Machine Learning. https://machinelearningmastery.com/classification-and-regression-trees-for-machine-learning/. Zuletzt geprüft am 23.10.2017.
- Bullinaria, J. (2015): Recurrent Neural Networks. http://www.cs.bham.ac.uk/~jxb/INC/l12.pdf. Zuletzt geprüft am 23.10.2017.
- Business Insider (Hrsg.) (2016): Chinese search giant Baidu has launched a medical chatbot. http://www.businessinsider.de/chinese-search-giant-baidu-has-launched-a-medical-chatbot-2016-10. Zuletzt geprüft am 23.10.2017.
- Chowdrhy, A. (2014): Facebook's DeepFace Software Can Match Faces With 97.25% Accuracy. https://www.forbes.com/sites/amitchowdhry/2014/03/18/facebooks-deepface-software-can-match-faces-with-97-25-accuracy/#3ddfb84254fc. Zuletzt geprüft am 23.10.2017.
- Clark, J. (2015): Google Turning Its Lucrative Web Search Over to AI Machines. https://www.bloomberg.com/news/articles/2015-10-26/google-turning-its-lucrative-web-search-over-to-ai-machines. Zuletzt geprüft am 23.10.2017.
- Clickworker (2017): Datasets for Machine Learning & Artificial Intelligence (AI) training. https://www.clickworker.com/machine-learning-ai-artificial-intelligence/. Zuletzt geprüft am 23.10.2017.
- Constine, J. (2016): Facebook’s new DeepText AI categorizes everything you write. https://techcrunch.com/2016/06/01/facebook-deep-text/. Zuletzt geprüft am 23.10.2017.
- Cortes, C., Vapnik, V. (1995): Support-vector networks, in: Machine learning, Vol. 20, 273-297.
- Cycorp (Hrsg.) (2017): Knowledge Base. http://www.cyc.com/kb/. Zuletzt geprüft am 17.10.2017.

- DARPA (2016): Explainable Artificial Intelligence (XAI). https://www.darpa.mil/program/explainable-artificial-intelligence. Zuletzt geprüft am 23.10.2017.
- Dasgupta, S. (2017): A neural algorithm for a fundamental computing problem. https://www.ncbi.nlm.nih.gov/pubmed/29123069. Zuletzt geprüft am 08.01.2018
- Decker, A. (2015): Is Skype Translator a Threat to the Translation and Interpreting Sector? https://info.moravia.com/blog/is-skype-translator-a-threat-to-the-translation-and-interpreting-sector. Zuletzt geprüft am 23.10.2017.
- Deshpande, M. Kuramochi, M., Wale, N., Karypis, G. (2005): Frequent substructure-based approaches for classifying chemical compunds, Transactions on Knowledge and Data Engineering 17(8), S. 1036-1050
- Domingos, P. (2016): The Five Tribes of Machine Learning- And What You Can Take from Each. https://learning.acm.org/webinar_pdfs/PedroDomingos_FTFML_WebinarSlides.pdf. Zuletzt geprüft am 17.10.2017.
- Doshi-Velez, F., et al. (2017): Accountability of AI Under the Law: The Role of Explanation https://arxiv.org/abs/1711.01134 Zuletzt geprüft am 08.01.2018
- Elman, J.L. (1990): Finding structure in time, in: Cognitive science Vol. 14, 179-211.
- Evtimov, I. et al. (2017): Robust Physical-World Attacks on Deep Learning Models. https://arxiv.org/abs/1707.08945. Zuletzt geprüft am 08.01.2018.
- Fan, Y., Yang, Z., Cohen, W. (2017): Differentiable Learning of Logical Rules for Knowledge Base Completion. arXiv preprint arXiv:1702.08367
- Fernando, C. et al. (2017): PathNet: Evolution Channels Gradient Descent in Super Neural Networks. https://arxiv.org/pdf/1701.08734.pdf. Zuletzt geprüft am 08.01.2018.
- Forbes (Hrsg.) (2017): Why 2017 Is The Year Of Artificial Intelligence. https://www.forbes.com/sites/forbestechcouncil/2017/02/27/why-2017-is-the-year-of-artificial-intelligence/#3003774f57a1. Zuletzt geprüft am 17.10.2017.
- Fortune (Hrsg.) (2016): 2017 Will Be the Year of AI. http://fortune.com/2016/12/30/the-year-of-artificial-intelligence/. Zuletzt geprüft am 17.10.2017.
- Fürnkranz, J., (1999): Separate-and-Conquer Rule Learning. J. Artificial Intelligence Review 13: 3
- Futurezone (Hrsg.) (2016): Google-KI entwickelt Sprache, die niemand sonst versteht. https://futurezone.at/science/google-ki-entwickelt-sprache-die-niemand-sonst-versteht/232.882.618. Zuletzt geprüft am 23.10.2017.
- Gama, J., A survey on learning from data streams: current and future trends (2012): Prog Artif Intell (2012) 1: 45. https://doi.org/10.1007/s13748-011-0002-6 Zuletzt geprüft am 08.01.2018.

- Gal, Y., Ghahramani, Z. (2016): Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. Proceedings of the 33rd International Conference on Machine Learning, PMLR, No. 48. S.1050-1059.
- Gartner (Hrsg.) (2016): Gartner's 2016 Hype Cycle for Emerging Technologies Identifies Three Key Trends That Organizations Must Track to Gain Competitive Advantage. https://www.gartner.com/newsroom/id/3412017. Zuletzt geprüft am 17.10.2017.
- Gartner (Hrsg.) (2017): Gartner Identifies Three Megatrends That Will Drive Digital Business Into the Next Decade. https://www.gartner.com/newsroom/id/3784363. Zuletzt geprüft am 17.10.2017.
- Gärtner, T., Lloyd,J., Flach, P.A.. (2002) Kernels for structured data. International Conference on Inductive Logic Programming
- Geitgey, A. (2016): Machine Learning is Fun! Part 3: Deep Learning and Convolutional Neural Networks. https://medium.com/@ageitgey/machine-learning-is-fun-part-3-deep-learning-and-convolutional-neural-networks-f40359318721. Zuletzt geprüft am 23.10.2017.
- Goodfellow, I. et al. (2014): Generative adversarial nets. https://arxiv.org/abs/1406.2661. Zuletzt geprüft am 23.10.2017.
- Goodfellow, I., Bengio, Y., Courville, A. (2017): Deep Learning, Cambridge.
- Goodman, N., Tenenbaum, J. (2016): Patterns of inference. https://probmods.org/chapters/04-patterns-of-inference.html. Zuletzt geprüft am 17.10.2017.
- Graves, A. et al. (2016): Hybrid computing using a neural network with dynamic external memory. Nature, 538. S.471-476.
- Graves, A., Wayne, G., Danihelka, I. (2014): Neural Turing Machines. https://arxiv.org/pdf/1410.5401.pdf. Zuletzt geprüft am 08.01.2018.
- Grosse, R. (2017): Deep learning. https://metacademy.org/roadmaps/rgrosse/deep_learning/version/25. Zuletzt geprüft am 23.10.2017.
- Gunning, D., (2016): https://www.darpa.mil/program/explainable-artificial-intelligence Zuletzt geprüft am 08.01.2018
- Guttenberg, N. (2017): Stability of Generative Adversarial Networks. http://www.araya.org/archives/1183. Zuletzt geprüft am 23.10.2017.
- Hayon, D. (2017): Google verrät sein Geheimnis: Darum ist Google Maps besser als jedes Navi. http://www.chip.de/news/Google-verraet-sein-Geheimnis-Darum-ist-Google-Maps-besser-als-jedes-Navi_114029678.html. Zuletzt geprüft am 23.10.2017.
- He, K. et al. (2015): Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. https://arxiv.org/pdf/1502.01852v1.pdf. Zuletzt geprüft am 23.10.2017.

- Hke, S. (2017): Interpretability of Neural Networks. https://datawarrior.wordpress.com/2017/10/31/interpretability-of-neural-networks/. Zuletzt geprüft am 04.12.2017.
- Hochreiter, S., Schmidhuber, S. (1997): Long short-term memory, in: Neural computation Vol. 9, 1735-1780.
- Horaczek, S. (2017): Amazon Echo Look uses a camera and AI to judge your outfit, sell you stuff. https://www.popsci.com/amazon-echo-look-uses-alexas-ai-to-judge-your-outfit-sell-you-stuff. Zuletzt geprüft am 23.10.2017.
- Hornik, K. (1991): Approximation capabilities of multilayer feedforward networks, in: Neural Networks Vol. 4, 2.
- Horvath, T., Gärtner, T., Wrobel, S. (2004): Cyclic pattern kernels for predictive graph mining, Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. S. 158-167
- ICML (2018): ICML 2018 – Thirty-fifth International Conference on Machine Learning. https://icml.cc/. Zuletzt geprüft am 08.01.2018.
- INWT Statistics (2017): Einfache Lineare Regression. https://www.inwt-statistics.de/blog-artikel-lesen/Einfache_lineare_Regression.html. Zuletzt geprüft am 23.10.2017.
- Ishibuchi, H., and Yamamoto , T. (2005): Rule weight specification in fuzzy rule-based classification systems, IEEE Trans.FuzzySyst.13(4)
- Ivakhnenko, A. (1965): Cybernetic Predicting Devices, Kiev.
- Jagannath, V. (2017): Random Forest Template for TIBCO Spotfire® - Wiki page. https://community.tibco.com/wiki/random-forest-template-tibco-spotfirer-wiki-page. Zuletzt geprüft am 23.10.2017.
- Kaggle (2017): The State of Data Science & Machine Learning 2017. https://www.kaggle.com/surveys/2017. Zuletzt geprüft am 30.01.2018.Kühn, S. (2017): Wie lernen Maschinen? https://data-science-blog.com/blog/2016/01/20/wie-lernen-maschinen-3/. Zuletzt geprüft am 23.10.2017.
- Kirkpatrick, J. et al. (2017): Overcoming catastrophic forgetting in neural networks. http://www.pnas.org/content/114/13/3521.full. Zuletzt geprüft am 08.01.2018.
- Kusner, M. et al., (2017): Counterfactual Fairness, arXiv:1703.06856v2 Zuletzt geprüft am 08.01.2018
- Lake, B., Salakhutdinov, R., Tenenbaum, J. (2015): Human-level concept learning through probabilistic program induction. https://www.cs.cmu.edu/~rsalakhu/papers/LakeEtAl2015Science.pdf. Zuletzt geprüft am 08.01.2018.
- LeCun, Y. et al. (1998): Gradient-based learning applied to document recognition, in: Proceedings of the IEEE 86.11, 2278-2324.

- LeCunn, Y., Cortes, C., Burges, C. (2015): The MNIST Database of handwritten digitis. https://github.com/ShinAsakawa/2015corona/blob/master/MNIST%20handwritten%20digit%20database,%20Yann%20LeCun,%20Corinna%20Cortes%20and%20Chris%20Burges.pdf. Zuletzt geprüft am 17.10.2017.
- Levy, S. (2016): The iBrain Is Here—and It's Already Inside Your Phone. https://www.wired.com/2016/08/an-exclusive-look-at-how-ai-and-machine-learning-work-at-apple/. Zuletzt geprüft am 23.10.2017.
- Lowe, R. et al. (2017): Learning to Cooperate, Compete, and Communicate. https://blog.openai.com/learning-to-cooperate-compete-and-communicate/#. Zuletzt abgerufen am 08.01.2018.
- Lundberg, S., Su-In, L.(2017): A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems
- Maron, O., Moore, A. (1997): The Racing Algorithm: Model Selection for Lazy Learners. Artificial Intelligence Review, No.11. S.193-225.
- McDonald (2017): Correlation and linear regression. http://www.biostathandbook.com/linearregression.html. Zuletzt geprüft am 17.10.2017.
- McMahan, B., /Ramage, D., (2017): Federated Learning: Collaborative Machine Learning without Centralized Training Data https://research.googleblog.com/2017/04/federated-learning-collaborative.html Zuletzt geprüft am 08.01.2018
- Merkert, J. (2016): Google: Translate-KI übersetzt dank selbst erlernter Sprache. https://www.heise.de/newsticker/meldung/Google-Translate-KI-uebersetzt-dank-selbst-erlernter-Sprache-3502351.html. Zuletzt geprüft am 23.10.2017.
- Metz, C. (2015): Google Says Its AI Catches 99.9 Percent of Gmail Spam. https://www.wired.com/2015/07/google-says-ai-catches-99-9-percent-gmail-spam/. Zuletzt geprüft am 23.10.2017.
- Mnih, V. et al (2013): Playing Atari with Deep Reinforcement Learning. https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf. Zuletzt geprüft am 23.10.2017.
- Mnih, V. et al. (2015): Human-level control through deep reinforcement learning. Nature. 518, 529- 541.
- Mnih, V., Kavukcuoglu, K.: Methods and apparatus for reinforcement learning. US 14/097,862. 5. Dez. 2013.
- Neapolitan, R. (2003): Learning Bayesian Networks, Chicago.
- Ng, A., Stanford University (2017): Machine Learning. https://www.coursera.org/learn/machine-learning. Zuletzt geprüft am 08.01.2018.
- NIPS (2018): NIPS 2018. https://nips.cc/. Zuletzt geprüft am 08.01.2018.

- Nickel, M., et al. (2016): A review of relational machine learning for knowledge graphs. Proceedings of the IEEE 104.1 S. 11-33
- N-tv (2013): Neues Raubtier in den Anden entdeckt. https://www.n-tv.de/wissen/fundsache/Neue-Tierart-entdeckt-Olinguito-lebt-in-den-Nebelwaeldern-der-Anden-article11178231.html. Zuletzt geprüft am 08.01.2018.
- Olah/Carter (2016): Attention and Augmented Recurrent Neural Networks. https://distill.pub/2016/augmented-rnns/. Zuletzt abgerufen am 08.01.2018.
- Open CV (Hrsg.) (2017): Introduction to Support Vector Machines. https://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html. Zuletzt geprüft am 23.10.2017.
- Pan, S.J., Yang Q. (2010): A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering archive 22(10) S. 1345-1359
- Quantitatitve Journey (Hrsg.) (2015): Learning Gridworld with Q-learning. http://outlace.com/rlpart3.html. Zuletzt geprüft am 23.10.2017.
- Quantup (Hrsg.) (2016): Case study: Electric power load forecasting — a comparison of three approaches. http://quantup.eu/2016/09/02/bez-kategorii-en/case-study-electric-power-load-forecasting-a-comparison-of-three-approaches/. Zuletzt geprüft am 23.10.2017.
- Quinlan, J.R. (1986): Induction of Decision Trees, in: Machine Learning Vol. 1 (1), 81-106.
- Radford, A., Metz, L., Chintala, S. (2016): Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. https://arxiv.org/abs/1511.06434. Zuletzt geprüft am 23.10.2017.
- Rao, V. (2013): Introduction to Classification & Regression Trees (CART). https://www.datasciencecentral.com/profiles/blogs/introduction-to-classification-regression-trees-cart. Zuletzt geprüft am 23.10.2017.
- Rezende, D., Mohamed, S., Danihelka, I., Gregor K., Wierstra, D. (2016): One-Shot Generalization in Deep Generative Models , Proceedings of The 33rd International Conference on Machine Learning, PMLR 48:1521-1529
- Ribeiro, M., Singh, S., Guestrin, C. (2016): »Why Should I Trust You?" Explaining the Predictions of Any Classifier KDD 16 Proceedings of the 22 ACM SIGKDD Intern. Conf. on Knowledge Disc. And Data Mining, S. 1135-1144
- Rimstar (Hrsg.) (2017): Backpropagation neural network software (3 layer). https://rimstar.org/science_electronics_projects/backpropagation_neural_network_software_3_layer.htm. Zuletzt geprüft am 23.10.2017.
- Rosenblatt, F. (1958): The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain. Psychological Review. 65 (6), 386–408.
- Ruder, S, (2017): An Overview of Multi-Task Learning in Deep Neural Networks. https://arxiv.org/pdf/1706.05098.pdf. Zuletzt geprüft am 08.01.2018.

- Ruder, S. (2017b): Transfer Learning - Machine Learning's Next Frontier http://ruder.io/transfer-learning/ Zuletzt geprüft am 08.01.2018
- Ryerson University (Hrsg.) (2017):Department of Computer Science. http://www.scs.ryerson.ca/. Zuletzt geprüft am 17.10.2017.
- Salk Institute (Hrsg.) (2017): Fruit fly brains inform search engines of the future. https://phys.org/news/2017-11-fruit-brains-future.html. Zuletzt geprüft am 08.01.2018.
- Samuel, A. (1959): Some Studies in Machine Learning Using the Game of Checkers. 2-Recent Progress. http://researcher.watson.ibm.com/researcher/files/us-beygel/samuel-checkers.pdf. Zuletzt geprüft am 23.10.2017.
- Schlafer, M. (2017): Facebook: AI kann Fotos nach Inhalten durchsuchen. https://www.computerbase.de/2017-02/facebook-ai-fotos-inhalte/. Zuletzt geprüft am 23.10.2017.
- Schölkopf, B., Smola, A., Müller, K.-R. (1998): Nonlinear Component Analysis as a Kernel Eigenvalue Problem, in: Neural Computation, Vol. 10 (5), 1299-1319
- Shallue, C. (2016): Show and Tell: image captioning open sourced in TensorFlow. https://research.googleblog.com/2016/09/show-and-tell-image-captioning-open.html. Zuletzt geprüft am 23.10.2017.
- Shavlik, J.W. (1994): Combining symbolic and neural learning. Mach Learn 14: 321
- Shu, Kai, et al. (2017): Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter 19.1 S. 22-36
- Siegelmann, H., Sontag, E. (1995): On the computational power of neural nets. Journal of Computer System and Science, 50(1). S.132-150.
- Silver, D., Yang, Q., Li, L. (2013): Lifelong Machine Learning Systems: Beyond Learning Algorithms. AAAI Spring Symposium - Technical Report.
- Smirnov, E., Timoshenko, D., Andrianov, S. (2014): Comparison of Regularization Methods for ImageNet Classification with Deep Convolutional Neural Networks. https://www.researchgate.net/figure/262678168_fig1_Fig-1-Deep-Convolutional-Neural-Network-architecture. Zuletzt geprüft am 23.10.2017.
- Snider, D. (2017): Introduction to the SQL Server Analysis Services Logistic Regression Data Mining Algorithm. https://www.mssqltips.com/sqlservertip/3471/introduction-to-the-sql-server-analysis-services-logistic-regression-data-mining-algorithm/. Zuletzt geprüft am 23.10.2017.
- Stolpe, M., Morik, K. (2011): Learning from Label Proportions by Optimizing Cluster Model Selection. In: Gunopulos D., Hofmann T., Malerba D., Vazirgiannis M. (eds) Ma chine Learning and Knowledge Discovery in Databases ECML PKDD 2011 Lecture Notes in Computer Science 6913

- Tata, S. (2017): Quick Access in Drive: Using Machine Learning to Save You Time. https://research.googleblog.com/2017/03/quick-access-in-drive-using-machine.html. Zuletzt geprüft am 23.10.2017.
- Technologie Review (Hrsg.) (2017a): Forget AlphaGo—DeepMind Has a More Interesting Step Toward General AI. https://www.technologyreview.com/s/608108/forget-alphago-deepminds-has-a-more-interesting-step-towards-general-ai/. Zuletzt geprüft am 17.10.2017.
- Technologie Review (Hrsg.) (2017b): Forget AlphaGo—DeepMind Has a More Interesting Step Toward General AI. Google's AI Guru Says That Great Artificial Intelligence Must Build on Neuroscience. Zuletzt geprüft am 17.10.2017.
- Tensorflow (2017a): An open-source software library for Machine Intelligence. https://www.tensorflow.org/. Zuletzt geprüft am 23.10.2017.
- Tensorflow (2017b): Tinker With a Neural Network in Your Browser. Don't Worry, You Can't Break It. We Promise. http://playground.tensorflow.org/#activation=tanh&batchSize=10&dataset=circle&regDataset=reg-plane&learningRate=0.03&regularizationRate=0&noise=0&networkShape=4,2&seed=0.31795&showTestData=false&discretize=false&percTrainData=50&x=true&y=true&xTimesY=false&xSquared=false&ySquared=false&cosX=false&sinX=false&cosY=false&sinY=false&collectStats=false&problem=classification&initZero=false&hideText=false. Zuletzt geprüft am 23.10.2017.
- Torres, T. (2015): Deep Style: Inferring the Unknown to Predict the Future of Fashion. http://multithreaded.stitchfix.com/blog/2015/09/17/deep-style/. Zuletzt geprüft am 23.10.2017.
- Vanschore, J., Brazdil, P., Hoos, H., Hutter, F., (2017): Auto-ML ECML Tutorial 2017 https://sites.google.com/site/automl2017ecmlpkdd/tutorial Zuletzt geprüft am 08.01.2018
- Van Veen, F. (2016): The Neural Network Zoo. http://www.asimovinstitute.org/neural-network-zoo/. Zuletzt geprüft am 23.10.2017.
- Varshney, K. (2016): Engineering Safety in Machine Learning. https://arxiv.org/abs/1601.04126. Zuletzt abgerufen am 08.01.2018.
- W3C (Hrsg.) (2014): Resource Description Framework (RDF). https://www.w3.org/RDF/. Zuletzt abgerufen am 08.01.2018.
- Watkins, C., Dayan, P. (1992): Q,-Learning, in: Machine Learning Vol.8, 279-292.
- Wikipedia (2017b): Entscheidungsbaum. https://de.wikipedia.org/wiki/Entscheidungsbaum. Zuletzt geprüft am 17.10.2017.
- Wikipedia (2017c): Random Forest. https://de.wikipedia.org/wiki/Random_Forest. Zuletzt geprüft am 23.10.2017.

- Wikipedia (2017d): k-means clustering. https://en.wikipedia.org/wiki/K-means_clustering. Zuletzt geprüft am 23.10.2017.
- Wikipedia (2017e): Long short-term memory. https://en.wikipedia.org/wiki/Long_short-term_memory. Zuletzt geprüft am 23.10.2017.
- Wikipedia (2017f): Fluch der Dimensionalität. https://de.wikipedia.org/wiki/Fluch_der_Dimensionalit%C3%A4t. Zuletzt geprüft am 23.10.2017.
- Wikipedia (Hrsg.) (2017a): Perceptrons. https://en.wikipedia.org/wiki/Perceptrons_%28book%29. Zuletzt geprüft am 17.10.2017.
- Wilcke, X., Bloem, P., de Boer, V. (2017): The knowledge graph as the default data model for learning on heterogeneous knowledge. Data Science. 1-19. 10.3233/DS-170007
- Xian Y., Lampert, C.H., Schiele, B., Akata, Z., (2017): Zero-Shot Learning - A Comprehensive Evaluation of the Good, the Bad and the Ugly. https://arxiv.org/abs/1707.00600 Zuletzt geprüft am 08.01.2018.
- Xing E, Ho Q., Xie P., Wei, D. (2016): Strategies and Principles of Distributed Machine Learning on Big Data. Engineering 2(2), S. 179-195
- Zhang, C. et al. (2016): Understanding deep learning requires rethinking generalization. https://arxiv.org/abs/1611.03530. Zuletzt geprüft am 23.10.2017.
- Zhang, Y., Yang, Q., (2017): A Survey on Multi-Task Learning. https://arxiv.org/abs/1707.08114 Zuletzt geprüft am 08.01.2018.

Kapitel 3

- Austin, Tom; Krensky, Peter (2017): Machine Learning/AI: Hard Facts, Conclusions and Actions. Webinar. Gartner, 2017.
- BITKOM (Hg.) (2016): Germany - Excellence in Big Data. Unter Mitarbeit von Mathias et al. Weber. Berlin, zuletzt geprüft am 14.11.2016.
- Böttcher, Björn; Schwalm, Anna-Lena; Velten, Carlo (2017): MACHINE LEARNING ANBIETER & DIENSTLEISTER IM VERGLEICH. Hg. v. Crisp Research.
- Burger, Rudy; Wheelock, Clint (2017): Computer Vision Report. Market Trends, M&A Transactions, Private Placement Financings, and Public and Private Company Profiles. Hg. v. Woodside Captial Partners und Tractica.
- Crisp Research (Hg.) (2017): Machine Learning im Unternehmenseinsatz. Künstliche Intelligenz als Grundlage digitaler Transformationsprozesse.
- Deutschland intelligent vernetzt DIV (Hg.) (2017): Fitness- und Healthcare-Wearables. Einfluss auf Gesundheit und Gesellschaft. Expertengruppe Smart Wearables.
- EFI (Hg.) (2018): Gutachten 2018 https://www.e-fi.de/fileadmin/Gutachten_2018/EFI_Gutachten_2018.pdf
- Frost & Sullivan (Hg.) (2017a): Impact of Artificial Intelligence on Autonomous Driving Development. OEMs to Have Ai-incorporated Autonomous Driving Software by 2022 but to be Focused on Object and Road Furniture Detection Rather than on Core Decision Engine Software ((Keine Angabe), K1B1-18).
- Frost & Sullivan (Hg.) (2017b): Machine Learning Goes Hyperscale (Digital Transformation Beats).
- Frost & Sullivan (Hg.) (2017c): PowerPoint Presentation. Recent Innovations in Healthcare, Electric Vehicles, Artificial Intelligence, and Food Packaging. Inside R&D TechVision Opportunity Engine.
- Gartner (Hg.) (2017): Hype Cycle for Data Science and Machine Learning. Unter Mitarbeit von Peter Krensky und Jim Hare.
- Gentsch, Peter (2018): Künstliche Intelligenz für Sales, Marketing und Service. Mit AI und Bots zu einem Algorithmic Business - Konzepte, Technologien und Best Practices. Wiesbaden: Springer Gabler.
- Groopman, Jessica; Kaul, Aditya (2017): Deep Learning. Enterprise, Consumer, and Government Applications for Deep Learning Software, Hardware, and Services: Market Analysis and Forecasts for 112 Use Cases. Research Report. Hg. v. Tractica.
- Groopman, Jessica; Wheelock, Clint (2017): Artificial Intelligence: 10 Key Themes Across Use Cases. White Paper. Hg. v. Tractica.
- McKinsey (Hg.) (2017): Smartening up with Artificial Intelligence (AI). What's in it for Germany and its Industrial Sector? (Digital).
- Moor Insights & Strategy (Hg.) (2017): A Machine Learning Application Landscape. And Appropriate Hardware Alternatives.
- Rao, Anand S.; Verweij, Gerard (2017): Sizing the prize. What's the real value of AI for your business and how can you capitalise? Hg. v. PWC.

- Sahi, Manoj K.; Kaul, Aditya (2017): Customer Service Robots. Humanoid and Non-Humanoid Robots for Retail, Travel and Hospitality, Financial Services, Restaurants, Healthcare, and Other Customer-Facing Applications: Global Market Analysis and Forecasts. Executive Summary. Hg. v. Tractica.
- Stanford University (Hg.) (2016): Artificial Intelligence and Life in 2030. One Hundred Year study on Artificial Intelligence. Report of the 2015-2016 Study Panel. Unter Mitarbeit von Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee Saxenian, Julie Shah, Milind Tambe, and Astro Teller. Online verfügbar unter http://ai100.stanford.edu/2016-report, zuletzt geprüft am 06.10.2016.
- Statista (Hg.) (2016): Künstliche Intelligenz. Statista-Dossier, zuletzt geprüft am 28.11.2016.
- Statista (Hg.) (2017a): Artificial Intelligence (Statista Report 2017).
- Statista (Hg.) (2017b): Digital Economy Compass.
- Tractica (Hg.) (2017): Deep Learning. Enterprise, Consumer, and Government Applications for Deep Learning Software, Hardware, and Services: Market Analysis and Forecasts for 112 Use Cases.
- Valorge, Stéphane; Combaudou, Olivier; Chastel, Théodore; Lyet, Mathilde; Vasco, Alexandre (2017): ARTIFICIAL INTELLIGENCE. FROM HYPE TO MATURITY? Hg. v. Clipperton.

Kapitel 4

- Burgess, M. (2017): NHS DeepMind deal broke data protection law, regulator rules. In: Wired UK. Online verfügbar unter http://www.wired.co.uk/article/google-deepmind-nhs-royal-free-ico-ruling, zuletzt geprüft am 20.02.2018.
- Campolo, A.; Sanfilippo, M.; Whittaker, M.; Crawford, K. (2017): AI Now 2017 Report. Hg.: AI NOW.
- Economist (2016): Million-dollar babies. As Silicon Valley fights for talent, universities struggle to hold on to their stars. Artificial intelligence. Online verfügbar unter https://www.economist.com/news/business/21695908-silicon-valley-fights-talent-universities-struggle-hold-their, zuletzt geprüft am 20.02.2018.
- EFI-Expertenkommission Forschung und Innovation (2018): Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands 2018, Berlin: EFI. S. 75 ff.
- Evtimov, I. et al. (2017): Robust Physical-World Attacks on Deep Learning Models. https://arxiv.org/abs/1707.08945. Zuletzt geprüft am 20.02.2018.
- Fachforum Autonome Systeme im Hightech-Forum (2017): Autonome Systeme – Chancen und Risiken für Wirtschaft, Wissenschaft und Gesellschaft. Kurzversion, Abschlussbericht, Berlin, März 2017. S. 22.
- Fraunhofer-Gesellschaft (2017): Industrial Data Space – White Paper.
- Heckmann, D.; Schmid, A. (2017): Rechtliche Aspekte automatisierter Systeme. Rechtskonforme Gestaltung unserer Zukunft. In: Informatik Spektrum (5/2017).
- IW (2017): Gutachten für BDA, BDI, MINT Zukunft schaffen und Gesamtmetall.
- KDnuggets.com (Hg.) (2016): Datasets Over Algorithms https://www.kdnuggets.com/2016/05/datasets-over-algorithms.html zuletzt geprüft am 19.03.2018.
- Lakemeyer, G. (2017): Künstliche Intelligenz. Analysen & Argumente - Digitale Gesellschaft, Ausgabe 261, Hg.: Konrad Adenauer Stiftung. Berlin.
- McKinsey Global Institute, 2017: Artificial Intelligence, the next digital frontier? Discussion Paper, June 2017.
- National Science and Technology Council (2016a): Preparing for the future of artificial intelligence. Online verfügbar unter: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf, zuletzt geprüft am 19.03.2018.
- National Science and Technology Council (2016b): The National Artificial Intelligence Research and Development Strategic Plan. Online verfügbar unter https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf, zuletzt geprüft am 19.03.2018.
- OECD (2017): Digital Economy Outlook 2017, S. 306 ff.
- OpenAire: What is the Open Research Data Pilot? https://www.openaire.eu/what-is-the-open-research-data-pilot, zuletzt geprüft am 19.03.2018.
- PwC (Hg.) (2017): Datenaustausch als wesentlicher Bestandteil der Digitalisierung.

- Stifterverband, McKinsey (Hg.) (2017): Hochschulbildungsreport 2020 - Höhere Chancen durch höhere Bildung. Jahresbericht 2017/18 – Halbzeitbilanz 2010 bis 2015: 70-74.
- The Royal Society (2017): Machine learning: the power and promise of computers that learn by example. S. 89 ff.
- VDMA (2017): Machine Learning 2030. Zukunftsbilder für den Maschinen- und Anlagenbau / Band 1. S. 24 ff.
- Vieth, K., Wagner, B. im Auftrag der Bertelsmann Stiftung (2017): Teilhabe, ausgerechnet - Wie algorithmische Prozesse Teilhabechancen beeinflussen können.

Weiterführende Literatur zum Thema Maschinelles Lernen

- Bishop, Christopher M. (2007): Pattern Recognition and Machine Learning (Information Science and Statistics). Springer
- Downey, Allen B. (2011): Think Stats: Probability and Statistics for Programmers. Green Tea Press Needham, Massachusetts http://greenteapress.com/thinkstats/thinkstats.pdf
- Ertel, Wolfgang (2016): Grundkurs Künstliche Intelligenz. Springer
- Géron, Aurélien (2017): Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow: Konzepte, Tools und Techniken für intelligente Systeme (Animals). O'Reilly
- Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016): Deep Learning. MIT Press
- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2009): The Elements of Statistical Learning. Springer
- Ng, Andrew (2016): Machine Learning Yearning.
- Rashid, Tariq; Langenau, Frank (2017): Neuronale Netze selbst programmieren: Ein verständlicher Einstieg mit Python. O'Reilly
- Shai, Shalev-Shwartz; Shai, Ben-David (2014): Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press
- Wartala, Ramon (2017): Praxiseinstieg Deep Learning: Mit Python, Caffe, TensorFlow und Spark eigene Deep-Learning-Anwendungen erstellen. O'Reilly

Abbildungsverzeichnis

Abbildung 1: Gartner Hype Cycle für Emerging Technologies 2017 10

Abbildung 2: Beispielhafte Variationen handschriftlicher »7« 12

Abbildung 3: Automatische Ziffernerkennung mit ML 14

Abbildung 4: Bayessches Netz 18

Abbildung 5: Lineare Regression 19

Abbildung 6: KNN, eigene Darstellung 20

Abbildung 7: Entscheidungsbaum zur Vorhersage der Apfelernte 22

Abbildung 8: Überblick über die wichtigsten Phasen im Zusammenspiel von KI und ML, eigene Darstellung 25

Abbildung 9: Intentionen bei überwachtem und unüberwachtem ML 27

Abbildung 10: Lineare Regression 29

Abbildung 11: Logistische Regression 30

Abbildung 12: Iterativer Dichotomiser 3 (ID3) 30

Abbildung 13: Klassifikationsbaum 31

Abbildung 14: Random Forest 31

Abbildung 15: k-means Clustering 32

Abbildung 16: Stützvektormaschine (SVM) 33

Abbildung 17: Kernel Principal Component Analsis (PCA) 33

Abbildung 18: Feed-forward Network, eigene Darstellung 34

Abbildung 19: Bayessches Netz (BN) 35

Abbildung 20: Q-Lernen nach 35

Abbildung 21: Verwendete Methoden der von Kaggle befragten Data Scientists und ML-Fachleute 36

Abbildung 22: Tiefes KNN zur Vorhersage der Stromnachfrage 37

Abbildung 23: Aufbau eines Künstlichen Neuronalen Netzes in TensorFlow 38

Abbildung 24: Seite zum Testen von Parametern eines KNN 38

Abbildung 25: Fortschritte durch Maschinelles Lernen, adaptiert 40

Abbildung 26: Autoencoder 41

Abbildung 27: Generative Adversarial Networks 41

Abbildung 28: Künstlich generierte Bilder 42

Abbildung 29: Deep feedforward network (DFF) 43

Abbildung 30: Convolutional neural network (CNN) 44

Abbildung 31: Recurrent network (RNN) 44

Abbildung 32: Long short-term memory (LSTM) 45

Abbildung 33: Autoencoder 45

Abbildung 34: Deep belief network (DBN) 46

Abbildung 35: Generative adversarial network (GAN) 46

Abbildung 36: Unteranpassung (links) und Überanpassung (rechts) 49

Abbildung 37: Minimierung der Kostenfunktion 50

Abbildung 38: Anzahl der Expertenmeinungen zur ML-Technologie, auf die fokussiert werden sollte 57

Abbildung 39: Olinguito 64

Abbildung 40: Ansätze zur Wiederverwendung von tiefen Netzen 67

Abbildung 41: Integration von Lernen und Wissen Bildquelle 73

Abbildung 42: Übersicht zu Methoden zum Verkleinern von Entscheidungsbäumen 76

Abbildung 43: Teil des bAbI Datensatzes 78

Abbildung 44: Netzarchitekturen zum Lernen von Algorithmen 79

Abbildung 45: Heatmap für die Visualisierung der Parameter 82

Abbildung 46: Binäre Klassifikationsfunktion mit Ausgabe eines Konfidenzwertes 86

Abbildung 47: Analyserahmen ML-Akteure und Forschungslandschaft 98

Abbildung 48: Weltweite Entwicklung der ML-Publikationen, 2006-2016 104

Abbildung 49: Anteil der Deep Learning Publikationen an gesamten ML-Publikationen, 2006-2016 105

Abbildung 50: Weltweite Entwicklung ML-Publikationen nach Anwendungsbereichen, 2006-2016 111

Abbildung 51: Publikationen im Bereich Maschinelles Lernen nach Regionen, 2006-2016 119

Abbildung 52: Entwicklung EU-Förderung ML-Forschungsprojekte in FP7 und H2020 nach Ländern, 2007-2017 122

Abbildung 53: Entwicklung der EU-Förderung nach ML-Anwendungsbereichen, 2007-2017 124

Abbildung 54: Entwicklung der Patentfamilien im Bereich Maschinelles Lernen nach Ländern, 2006-2015 126

Abbildung 55: Entwicklung der Patentfamilien für ML-Anwendungsbereiche, 2006-2015 127

Abbildung 56: Patentfamilien im Bereich Maschinelles Lernen nach Regionen, 2006-2015 130

Abbildung 57: Internationale Investment-Geschäfte in KI, Maschinelles Lernen und Deep Learning seit 2010 133

Abbildung 58: Zentrale Anwendungsbranchen für Maschinelles Lernen 135

Tabellenverzeichnis

Tabelle 1: Überblick zu ausgewählten Meilensteinen im Einsatz von Maschinellem Lernen 15
Tabelle 2: Zukünftige Forschungsthemen und die Bewertung der Fachleute 55
Tabelle 3: Anforderungen aus der Praxis und Forschungsthemen, die diese Anforderungen aufnehmen 87
Tabelle 4: Publikationsstärkste Forschungseinrichtungen weltweit, Summe der ML-Publikationen in wissenschaftlichen Zeitschriften zwischen 2006-2016 106
Tabelle 5: Publikationsstärkste Forschungseinrichtungen weltweit, Summe der ML-Publikationen in Konferenzbeiträgen zwischen 2006-2016 107
Tabelle 6: Publikationsstärkste Unternehmen weltweit, Summe der ML-Publikationen zwischen 2006-2016 107
Tabelle 7: ML-Publikationen nach Anwendungsbereichen, 2006-2016 111
Tabelle 8: Führende Einrichtungen ML-Anwendungsbereiche, 2006-2016 113
Tabelle 9: Die publikationsstärksten deutschen FuE-Einrichtungen im Bereich Maschinelles Lernen 115
Tabelle 10: Publikationsstärkste Einrichtungen ML-Anwendungsbereiche in Deutschland, 2006-2016 118
Tabelle 11: Publikationsstärkste Unternehmen in Deutschland, 2006-2016 119
Tabelle 12: Führende Koordinatoren EU-Förderung ML-Forschungsprojekte in FP7 und H2020 nach Ländern, 2007-2017 122
Tabelle 13: Führende Koordinatoren EU-Förderung ML-Forschungsprojekte in FP7 und H2020 in Deutschland, 2007-2017 123
Tabelle 13: ML-Patentfamilien nach Anwendungsbereichen, 2006-2015 128
Tabelle 15: Patentanmeldestärkste Einrichtungen ML-Anwendungsbereiche weltweit, 2006-2015 129
Tabelle 16: Führende Einrichtungen ML-Anwendungsbereiche in Deutschland, 2006-2015 130
Tabelle 17: Beispielanbieter für ML mit Sitz in Deutschland 133
Tabelle 18: Beispielanbieter für ML mit Sitz in Deutschland 137
Tabelle 19: Beispielanbieter für ML mit Sitz in Deutschland 140
Tabelle 20: Beispielanbieter für ML mit Sitz in Deutschland 142
Tabelle 21: Beispielanbieter für ML mit Sitz in Deutschland 144
Tabelle 22: ML-basierte Kompetenzen, Forschungsansätze, Einsatzgebiete und Reifegrad 146