

Universität Leipzig
Fakultät für Mathematik und Informatik
Institut für Informatik

**Von Datenverwaltung
und maschinellem Lernen
zu Data Analytics**

Hausarbeit zum Seminarmodul
“Wissen in der modernen Gesellschaft“

Swetlana Shaban, BSc Informatik
Sommersemester 2019

Inhaltsverzeichnis

1. Einleitung.....	3
2. Begriffe.....	3
2.1. Datenverwaltung.....	3
2.2. Maschinelles Lernen.....	4
2.3. Data Analytics.....	5
3. Anwendungsbeispiele.....	6
3.1. Maschinelles Lernen.....	6
3.2. Data Analytics.....	6
4. Soziale Bedeutung.....	7
5. Historischer Überblick.....	7
5.1. Data Science.....	7
5.2. Maschinelles Lernen und Künstliche Intelligenz.....	9
6. Studienangebote zum Thema.....	17
7. Neue Branchen der Informatik.....	18
8. Zusammenfassung.....	19

1. Einleitung

Die digitale Welt hat großen Einfluss auf Menschen und Unternehmen. Zwar ist ein Großteil dieser Veränderungen positiv – man muss sich jedoch zunächst an sie gewöhnen. Das Thema Maschinelles Lernen ist in aller Munde. Doch was verbirgt sich hinter dem Begriff? Maschinelles Lernen wird heute mehr und mehr zum Mainstream. Waren selbstlernende Programme noch bis vor wenigen Jahren ausschließlich ein Thema für Universitäten, Forschungseinrichtungen und einige Technologieunternehmen, finden sie heute zunehmend Eingang in ganz normale Produkte und Lösungen. Unser Alltag und unser Geschäftsleben wird immer mehr von intelligenten Programmen bestimmt, die aus Daten lernen und das Gelernte verallgemeinern.

Der Vortrag liefert Informationen über solche Begriffe wie Datenverwaltung, maschinelles Lernen und Data Analytics und ihre Nutzenanwendung. Der soziale Bedeutungsgehalt wird auch beschrieben. Es gibt auch einen historischen Überblick über das Thema. Die vorliegende Arbeit beschäftigt sich mit den Studienangebote zum Thema und ihrer Bedeutung.

2. Begriffe

Im ersten Abschnitt der Hausarbeit sollen zunächst die wichtigsten theoretischen Begriffe für die späteren Anwendungen erklären werden.

2.1. Datenverwaltung

Zum Datenmanagement gehören das Erfassen von Daten, das Verarbeiten, Verwalten, Archivieren und die Ausgabe. Dazu gehört die Organisation der Daten mit Such- und Abfragefunktionen. Des weiteren gehören dazu die Zuweisung von Speicherplätzen, die Steuerung der Ein- und Ausgabe, die Reorganisation von Datenbeständen bzw. Speichermedien, Funktionen für die Wiedergewinnung der Daten und Präventiv-Verfahren für den Wiederanlauf nach Systemunterbrechungen.

Das moderne Datenmanagement berücksichtigt bei allen Funktionen den Lebenszyklus der Daten. Während dieser Zeit stellen Daten unterschiedliche Anforderungen an ihren Schutz, die Leistungsfähigkeit der Datenträger, die Speichermedien und das Management. Diese Lebenszyklus-Ausprägungen werden teilweise im Hierarchical Storage Management berücksichtigt, umfassender im Information Lifecycle Management. [1]

Data Lifecycle Management (DLM) ist ein regelbasierter Ansatz zur Verwaltung der Daten in einem Informationssystem über den gesamten Lebenszyklus hinweg: Von der Erstellung und ersten Speicherung der Daten bis zu dem Zeitpunkt, an dem sie überflüssig und gelöscht werden. Mehrere Hersteller bieten DLM-Produkte an, aber ein effektives Datenmanagement benötigt neben Anwendungen auch gut durchdachte Vorgehensweisen und das Befolgen von Best-Practice-Empfehlungen.

Es gibt verschiedene Datenmanagementansätze. Master Data Management (MDM) beispielsweise ist eine umfassende Methode, Unternehmen in die Lage zu versetzen, dass sie alle kritischen Daten mit einer Datei, dem so genannten Master File verknüpfen. Diese Datei dient als zentraler Bezugspunkt.

Die Bedeutung des effektiven Managements der Unternehmensdaten gewinnt ständig an Bedeutung, da Unternehmen einer wachsenden Zahl von Compliance-Regularien unterworfen sind. Hinzu kommt, dass die schiere Menge der Daten, die in einem Unternehmen zu verarbeiten sind, derartig gewachsen ist und weiter wächst, dass Big Data für immer mehr Firmen zur Realität wird.

Big-Data-Management ist die Organisation, Verwaltung und Steuerung großer Massen von strukturierten sowie unstrukturierten Daten. Unternehmen, Regierungsbehörden und andere Organisationen implementieren Big-Data-Management-Strategien, um die schnell wachsenden Datenpools zu bewältigen, die typischerweise viele Terabyte oder sogar Petabyte von Daten und zahlreiche unterschiedliche Datenarten beinhalten. [2]

2.2. Maschinelles Lernen

Maschinelles Lernen ist ein Teilbereich der künstlichen Intelligenz. Mithilfe des maschinellen Lernens werden IT-Systeme in die Lage versetzt, auf Basis vorhandener Datenbestände und Algorithmen Muster und Gesetzmäßigkeiten zu erkennen und Lösungen zu entwickeln. Es wird quasi künstliches Wissen aus Erfahrungen generiert. Die aus den Daten gewonnenen Erkenntnisse lassen sich verallgemeinern und für neue Problemlösungen oder für die Analyse von bisher unbekanntem Daten verwenden. [3]

Damit die Software eigenständig lernen und Lösungen finden kann, ist ein vorheriges Handeln von Menschen notwendig. Beispielsweise müssen die Systeme zunächst mit den für das Lernen relevanten Daten und Algorithmen versorgt werden. Zudem sind Regeln für die Analyse des Datenbestands und das Erkennen der Muster aufzustellen. Sind passende Daten vorhanden und Regeln definiert, können Systeme mit maschinellem Lernen unter anderem folgendes:

- ✓ Relevante Daten finden, extrahieren und zusammenfassen,
- ✓ Vorhersagen auf Basis der analysierten Daten treffen,
- ✓ Wahrscheinlichkeiten für bestimmte Ereignisse berechnen,
- ✓ sich an Entwicklungen eigenständig anpassen und
- ✓ Prozesse auf Basis erkannter Muster optimieren. [4]

Die verschiedenen Arten des Maschinellen Lernens

Algorithmen nehmen beim maschinellen Lernen eine zentrale Rolle ein. Sie sind für das Erkennen von Mustern und das Generieren von Lösungen verantwortlich und lassen sich in verschiedene Lernkategorien einteilen.

- ✓ überwachtes Lernen
- ✓ unüberwachtes Lernen
- ✓ teilüberwachtes Lernen
- ✓ bestärkendes Lernen
- ✓ aktives Lernen

Während beim überwachten Lernen im Vorfeld Beispielmuster definiert und spezifiziert werden müssen, um die Informationen passend den Modellgruppen der Algorithmen zuzuordnen, werden die Modellgruppen beim unüberwachten Lernen automatisiert aufgrund eigenständig erkannter Muster gebildet.

Teilüberwachtes Lernen stellt eine Mischung aus beiden Methoden dar. Das bestärkende Lernen basiert auf Belohnungen und Bestrafungen. Dem Algorithmus wird durch diese Interaktion mitgeteilt, wie er auf verschiedene Situationen zu reagieren hat. Diese Lernweise ist dem menschlichen Lernen sehr ähnlich.

Aktives Lernen schließlich bietet dem Algorithmus die Möglichkeit, für bestimmte Eingangsdaten die gewünschten Ergebnisse zu erfragen. Um die Anzahl von Fragen zu minimieren, erfolgt zuvor eine Auswahl relevanter Fragen mit hoher Ergebnisrelevanz durch den Algorithmus selbst.

Abhängig vom jeweiligen System kann die Datenbasis offline oder online vorliegen und wiederholbar oder nur einmalig für das maschinelle Lernen zur Verfügung stehen. Ein weiteres Unterscheidungsmerkmal des maschinellen Lernens ist das gleichzeitige Vorhandensein der Ein- und Ausgabe-Paare oder deren zeitlich versetzte Entwicklung. Je nach Art spricht man vom Batch-Lernen oder vom sequenziellen Lernen.

Wie funktioniert maschinelles Lernen prinzipiell?

Im Prinzip so ähnlich wie menschliches Lernen. Analog wie beispielsweise ein Kind lernt, dass auf Bildern bestimmte Objekte zu sehen sind, kann auch ein Computer "lernen", Objekte zu identifizieren oder Personen zu unterscheiden. Dazu wird die Lernsoftware zunächst mit Daten gefüttert und trainiert. Beispielsweise sagen die Programmierer dem System, dass ein bestimmtes Objekt "ein Hund" und ein anderes "kein Hund" ist.

Im Fortlauf erhält die Lernsoftware ständig Rückmeldungen vom Programmierer, die der Algorithmus nutzt, um das Modell anzupassen und zu optimieren: Mit jedem neuen Datensatz wird das Modell besser und kann schließlich eindeutig Hunde von Nicht-Hunden unterscheiden.

Welche Vorteile bietet maschinelles Lernen?

Maschinelles Lernen hilft Menschen, effizienter und kreativer zu arbeiten. Zum Beispiel können sie maschinelles Lernen verwenden, um ihre Bilder schneller zu organisieren und zu bearbeiten. Mit maschinellem Lernen können sie auch langweilige oder aufwändige Arbeiten dem Computer überlassen. Papierdokumente wie Rechnungen kann lernende Software selbständig scannen, speichern und ablegen.

Vor allem sind selbstlernende Maschinen in der Lage, für den Menschen sehr komplexe Aufgaben zu übernehmen - etwa die Erkennung von Fehlermustern oder mögliche Schäden in der Fertigung. Selbst bei der Erkennung von Krebstumoren in der Medizin und bei Therapieempfehlungen helfen inzwischen selbstlernende Programme - und übertreffen dabei oft die besten menschlichen Experten. Diese Fähigkeit, komplexe Zusammenhänge zwischen der Eingabe und der Ausgabe von großen Datenmengen verarbeiten zu können, ist einer der Hauptvorteile von maschinellem Lernen. [3]

2.3. Data Analytics

Data Analytics ist ein wissenschaftliches Vorgehen, Daten aus verschiedenen Datenquellen zu extrahieren und zu untersuchen. Das Ziel ist es, Schlussfolgerungen aus den Daten zu ziehen, die in einem bestimmten Zusammenhang zueinander stehen. Im Vordergrund steht die Auswertung bekannter Daten.

Der Begriff Data Analytics wird häufig im Zusammenhang mit Data Mining genannt oder sogar Synonym verwandt. Der Anwendungsfall von Data Analytics unterscheidet sich zu dem von Data Mining. Durch Data Mining wird versucht, neue und noch unbekannte Zusammenhänge aus Daten zu erschließen.

In der Forschung wird die Data Analytics bereits seit vielen Jahren eingesetzt, um anhand von Daten Modelle und Hypothesen zu verifizieren oder aber zu falsifizieren. Mit dem Aufkommen von Datenanalyse-Software in wirtschaftlichen Unternehmen, hat sich der Begriff Analytics weit verbreitet. Es dient der Ermittlung von Informationen aus Daten, um Entscheidungen in einem Unternehmen (auf unterschiedlichen Ebenen) zu unterstützen.

Meist findet man den Begriff Data Analytics im Zusammenhang mit Business Intelligence, analytischen CRM Systemen, Online Analytical Processing sowie in verschiedenen Unternehmensbereichen, die Datenanalyse verwenden.

Ganz allgemein kann man sagen, dass Data Analytics ein elementarer Bestandteil unternehmerischer Analyse-Methoden geworden ist. Multifunktionale Analyse-Systeme und Dashboards beinhalten die wesentlichen Verfahren der Datenanalyse und können bei Bedarf verwendet werden. [5]

3. Anwendungsbeispiele

3.1. Maschinelles Lernen

Maschinelles Lernen hat ein sehr großes Spektrum an Anwendungsmöglichkeiten. Im Internetumfeld kommt maschinelles Lernen beispielsweise für folgende Funktionen zum Einsatz:

- ✓ selbstständiges Erkennen von Spam-Mails und Entwicklung geeigneter Spam-Filter
- ✓ Sprach- und Texterkennung für digitale Assistenten
- ✓ Bestimmung der Relevanz von Webseiten für Suchbegriffe
- ✓ Erkennung und Unterscheidung der Internetaktivität von natürlichen Personen und Bots
- ✓ die Bild- und Gesichtserkennung
- ✓ automatische Empfehlungsdienste [4]

Maschinelles Lernen findet man bei den Empfehlungsdiensten von Amazon und Netflix ebenso wie bei der Gesichtserkennung von Facebook. Die Möglichkeit, einzelne Mitglieder mit ihren Namen auf Bildern zu markieren, hat bei Facebook zur weltweit größten Sammlung von Gesichtern in einer Datenbank geführt. Diese Daten kann Facebook nutzen, um Maschinen gezielt auf visuelle Erkennung zu schulen.

Auch hinter E-Mail-Anwendungen, die automatisch Spam erkennen, stecken maschinelle Lernverfahren. Der Computer analysiert die Daten, die in der E-Mail enthalten sind, und kategorisiert diese gemäß den erkannten Mustern als Spam oder Nicht-Spam. Wird eine Nachricht als Junk markiert, lernt der Rechner und kann dadurch Junk-Nachrichten noch besser identifizieren. Ebenfalls angewendet werden Lernverfahren bei der Abwehr von Computerattacken, der Bekämpfung von Internet-Kriminalität und dem Suchmaschinen-Ranking. [3]

3.2. Data Analytics

Data Analytics bietet viele Anwendungsmöglichkeiten und ist ein fachlich getriebenes Thema. Die Unternehmens-IT ist hier mehr oder weniger für die Umsetzung der Anforderungen und die Wartung der Systeme und Daten verantwortlich.

Im Bereich Marketing und Vertrieb kann Data Analytics dazu verwendet werden, Zielgruppen zu identifizieren, die mit einer ermittelbaren Wahrscheinlichkeit ein bestimmtes Produkt oder einen Service in Anspruch nehmen werden. Dadurch kann der Aufwand für Marketing-Kampagnen und Vertriebsaktivitäten optimiert werden. Dies führt zu einem besseren Kosten-Nutzen-Verhältnis in diesen Bereichen, was einen großen Beitrag zum Unternehmenserfolg leisten kann.

Im Bankenbereich kann mit Hilfe von Data Analytics eine Einstufung von Kunden und Verträgen durchgeführt werden, um Betrugsfällen vorzubeugen und Ausfallrisiken zu minimieren. Zudem können die Verfahren bei der Bekämpfung von Geldwäsche zum Einsatz kommen, um dies früher zu erkennen und schnellstmöglich reagieren zu können. [5]

4. Soziale Bedeutung

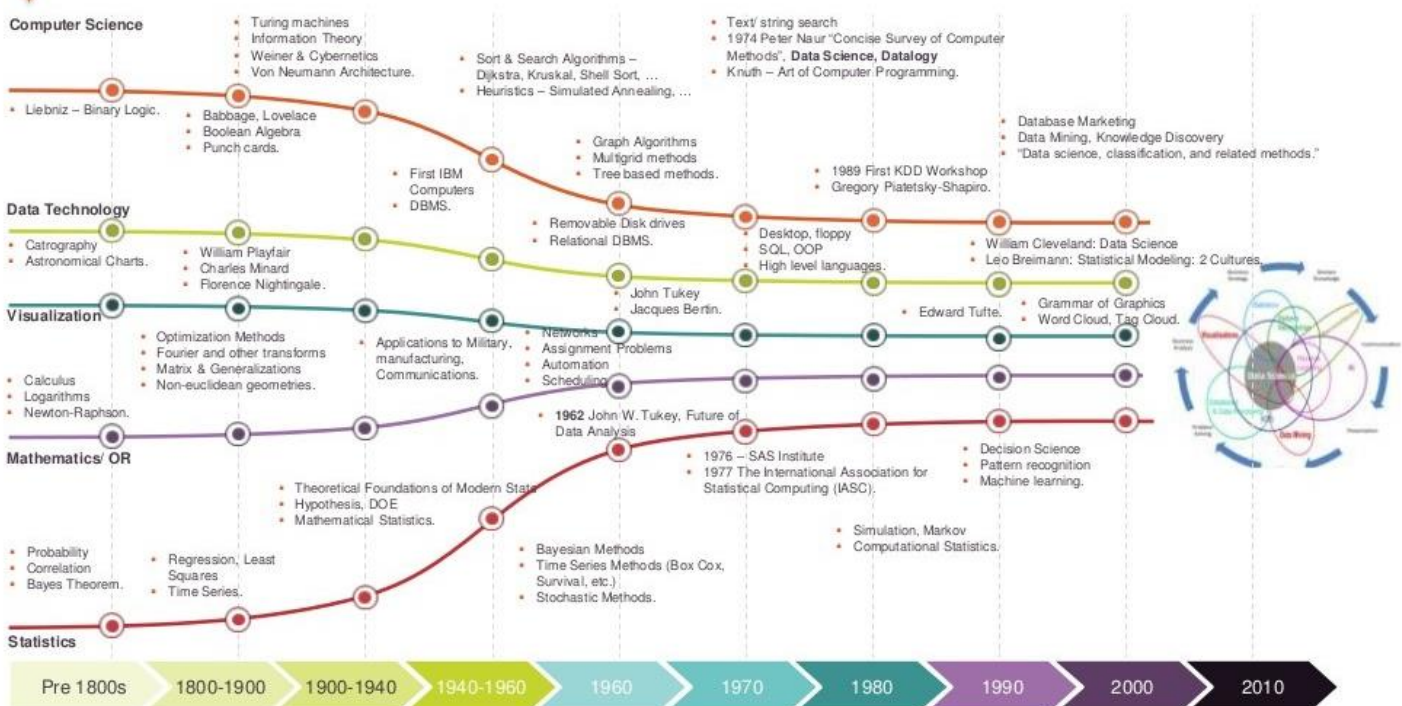
Ob im Rettungsdienst, in der mobilen Pflege, bei der Krankenhausorganisation oder in der Spendenverwaltung: Auch in der Sozial-Branche erkennen immer mehr Akteure, dass Daten ein wertvoller Schlüssel zu noch mehr Service, Effizienz und optimaler Ressourcen-Allokation sind. Gerade im direkten Dienst am Menschen werden positive Effekte professioneller Datenanalyse und -visualisierung ganz unmittelbar wirksam: Denn gewonnene Pflege-Zeit durch optimierte Routenplanung kommt beispielsweise Patienten und Klienten mobiler Pflegedienste zu Gute. Und über die schnelle und pünktliche Anlieferung warmer Mahlzeiten freuen sich Kunden im Menü-Service, der dank Produktions- und Geodaten-Auswertung perfekt abgestimmt ist – von der Großküche bis zum Verzehr daheim.

Doch nicht nur Organisation und Abläufe akuter Geschäftsherausforderungen lassen sich auf der Grundlage von Patienten-, Routen-, Fahrtenbuch- oder Wetterdaten in Diensten und Einrichtungen der Sozialwirtschaft besser aufstellen. Auch wichtige gesellschaftliche Megatrends können mithilfe der stets wachsenden Datengrundlage antizipiert und visualisiert werden. Weisen soziodemografische Daten zum Beispiel darauf hin, dass im Laufe der kommenden Jahre in bestimmten Regionen die Anzahl pflegebedürftiger Menschen steigen wird? Ist diese Region eher geprägt durch Ballungsräume? Oder durch infrastrukturelles schwaches Territorium, wo weite Wege, dünne Besiedelung und wenig Fachkräfte-Nachwuchs die Entwicklung prägen? Entsprechende gesellschaftliche Tendenzen haben – sichtbar und diskutierbar auf der Grundlage von Daten – enorme Bedeutung für die Geschäftsentwicklung von Leistungs-Erbringern in der Sozialwirtschaft. [6]

5. Historischer Überblick

5.1. Data Science

A brief history of Data Science

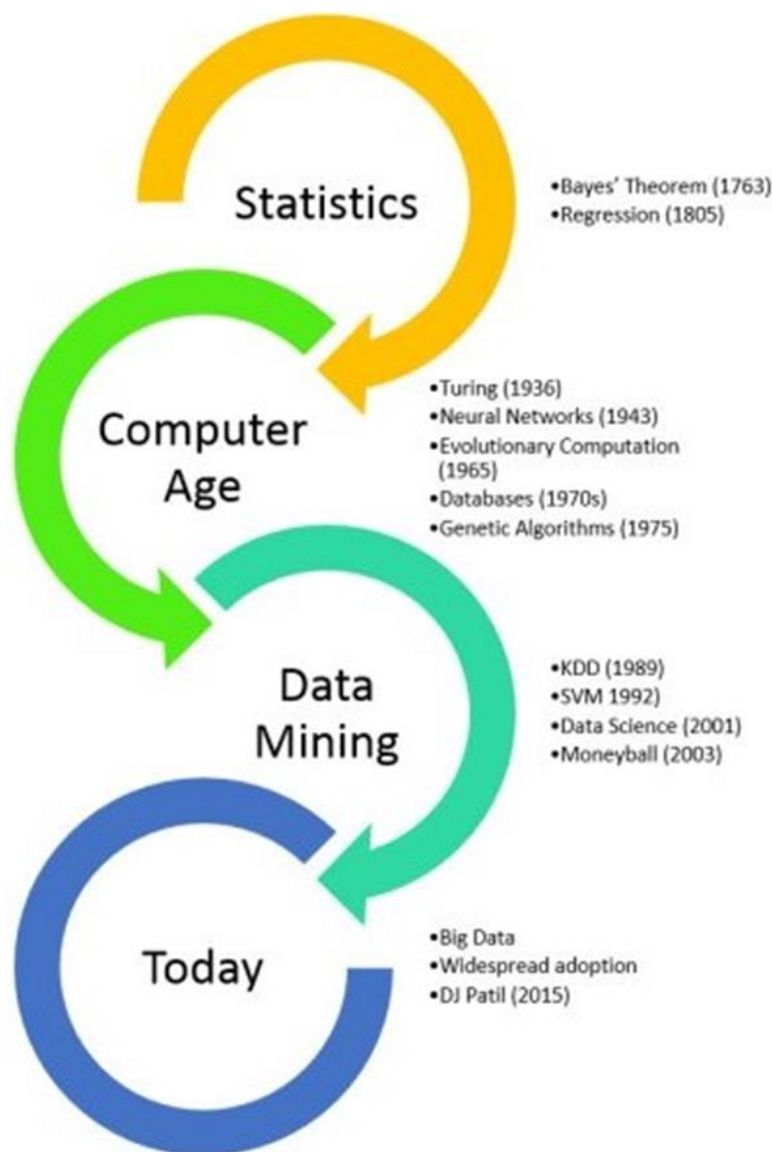


Quelle: [1]

Der Begriff „Data Science“ existiert seit über 40 Jahren und wurde ursprünglich als Ersatz für den Begriff „Informatik“ von Peter Naur im Jahr 1960 verwendet. 1974 veröffentlichte Naur in der Concise Survey of Computer Methods die Umfrage über die zeitgenössische Datenverarbeitung, in welcher der Begriff „Data Science“ frei verwendet wurde.

1996 trafen sich die Mitglieder der International Federation of Classification Societies (IFCS) in Kobe für ihre zweijährliche Konferenz. Bei dieser Konferenz war zum ersten Mal der Begriff "Data Science" im Titel der Konferenz enthalten.

Im November 1997 gab C.F. Jeff Wu den Eröffnungsvortrag mit dem Titel „Statistik = Datenwissenschaft?“ für seine Ernennung zum H. C. Carver Professor of Statistics an der University of Michigan. In diesem Vortrag charakterisierte er die statistische Arbeit als eine Trilogie von Datenerfassung, Datenmodellierung und -analyse und die Entscheidungsfindung. Abschließend rief er den Begriff „Datenwissenschaft“ ins Leben und befürwortete, dass die Statistik in „Datenwissenschaft“ und Statistiker in „Datenwissenschaftler“ umbenannt werden.



Quelle: [2]

Im April 2002 veröffentlichte der internationale Rat für Wissenschaft: Ausschuss für die Daten für Wissenschaft und Theorie, das Data Science Journal, welche sich auf die Problematik, wie die Beschreibung von Datensystemen, ihre Veröffentlichung im Internet, Anwendungen und gesetzlichen Problemen konzentrierte.

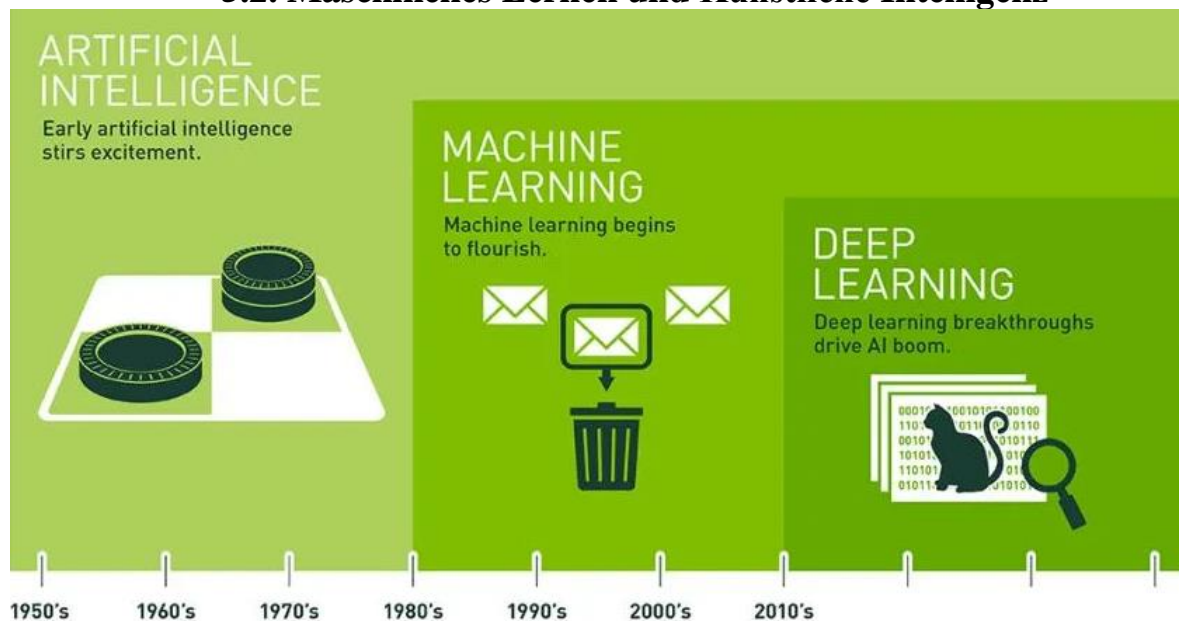
Kurz darauf begann die Columbia University 2003 die Zeitschrift „The Journal of Data Science“ zu veröffentlichen, welche eine Plattform für alle Datenanbieter zur Verfügung stellte um ihre Ansichten und Ideen zum Austausch zu präsentieren. Die Zeitschrift wurde größtenteils der Anwendung von statistischen Methoden und der quantitativen Forschung gewidmet.

2005 veröffentlichte das National Science Board den Bericht „Long-lived Digital Data Collections: Enabling Research and Education in the 21st Century“, in welchem unter dem Begriff Data Scientists verschiedene Experten aufgeführt

werden, die von entscheidender Bedeutung für das erfolgreiche Management digitalen Daten sind. Genannt werden unter anderem Informatiker, Datenbankexperten, Programmierer, Domänenexperten, Bibliothekare, Archivare sowie Experten im Bereich Software Engineering.

Als Teil der Verantwortlichkeiten von Data Scientists wird insbesondere die Entwicklung innovativer Konzepte in den Bereichen Datenbanktechnologie und Informationswissenschaft betont. Hierunter fallen auch Methoden der Informationsvisualisierung, Datenanalyse und Wissensentdeckung in Datenbanken. [7]

5.2. Maschinelles Lernen und Künstliche Intelligenz



Quelle: [3]

- ✓ Erstes Neuronenmodell 1943: McCulloch-Pitts
- ✓ Erstes Lernregel 1949: Donald Hebb
- ✓ Erster (analoge) Neurocomputer 1951: SNARC (Marvin Minsky)
- ✓ Erstes lernfähige Neuronetz 1958: Perceptron (Frank Rosenblatt)
- ✓ ca. 1960 - 2010: Blütezeit der symbolischen KI
- ✓ 1966: Turing-Test ELIZA
- ✓ 1972: Expertensystem MYCIN
- ✓ 1969 - 1981: „Neuronetz-Eiszeit“
- ✓ 1982: Wiedererwachen der Neuronetz-Entwicklung mit Einführung der Backpropagation in KNN
- ✓ 1984: Wissensdatenbank CYC
- ✓ 1996: „Deep Blue“ gewinnt gegen Schachweltmeister Randentwicklungen: Fuzzy Logik sowie evolutionäre Algorithmen

Die Vorstellung einer künstlichen Intelligenz ist keine Erfindung unserer Zeit. Schon im Mittelalter gab es die Idee eines Homunkulus, eines künstlichen Menschen, der natürlich ebenso über künstliche Intelligenz verfügen musste. Versuche zur realen technischen Nachempfindung intelligenter Systeme begannen natürlich erst im vergangenen Jahrhundert – und nahmen recht viel Zeit in Anspruch. Ein Blick in die Geschichte macht klar, wieso KI eher noch am Anfang der Entwicklung steht: Man nutzt heute noch vorwiegend Konzepte, deren Ursprünge schon fast 70 Jahre alt sind. Ihre Entwicklung verlief relativ diskontinuierlich und zeichnet sich durch eine „Bastelphase“ zu Beginn aus. Das ist kaum erstaunlich, lagen die Anfänge doch in einer Zeit, wo auch die Computer erst gerade ihre Entwicklung aufnahmen. Dann kam es durch eine einzige Veröffentlichung – ausgerechnet durch einen großen Anhänger der KI – auch noch zu einer KI-Eiszeit. Sie hielt rund zehn Jahre lang an.

Erschwerend zur Entwicklung der KI kommt hinzu: Abgearbeitet werden entsprechende Algorithmen und Mechanismen mit Maschinen, deren Architekturprinzip ähnlich alt ist: 1945 veröffentlichte der österreichisch-ungarische, später in den USA tätige Mathematiker John von Neumann das seither seinen Namen tragende Konzept. Erst wenige Jahre zuvor – 1941 – hatte der deutsche Ingenieur Konrad Zuse (1910-1995) mit der Z3 den ersten programmierbaren, binär arbeitenden Universalrechner der Welt gebaut. Die Von-Neumann-Computer unserer Tage sind zwar durch die Miniaturisierung der Elektronik unglaublich leistungsfähig geworden. Das diese Steigerung beschreibende Moore'sche Gesetz – eine auf den Mitgründer der Firma Intel, Gordon Moore, zurückgehende Beobachtung und Prognose von 1965 – kommt aber zumindest in seiner bisherigen Auslegung, die im wesentlichen Strukturverkleinerungen berücksichtigt, bald an seine Grenzen.

Logisches Neuron und 1. Lernregel

Einen ersten Schritt dazu machten 1943 die US Mathematiker Warren McCulloch und Walter Pitts: Sie schufen ein „Neuron“ als logisches Schwellwert Element mit mehreren Eingängen und einem einzigen Ausgang. Ihr einfaches Neuronenmodell konnte sowohl Und-, Oder- als auch Nicht-Gatter simulieren. Es gab die logischen Werte true oder false aus, 1 oder 0, schaltete Strom oder nicht, wenn die Summe der Eingangssignale einen Schwellenwert überschritt. Dies entsprach der neurobiologischen Analogie eines Aktionspotentials, das eine Nervenzelle bei einer kritischen Änderung ihres Membranpotentials aussendet. Schon 1894 postulierte der Spanier Santiago Ramon y Cajal, dass das Gedächtnis durch die Stärkung der Verbindung zwischen existierenden Neuronen gebildet werden könnte. 1949 formulierte dann Donald Hebb, ein kanadischer Physiologe, die 1. Lernregel: Gleichzeitig aktive Neuronen verstärken Verbindungen; im Englischen gibt es dafür den eingängigen Merksatz: „What fires together wires together“. Detaillierter ausgedrückt: Wenn eine Nervenzelle A eine Nervenzelle B dauerhaft und wiederholt erregt, wird die Synapse dadurch so verändert, dass die Signalübertragung effizienter wird. Dadurch erhöht sich das Membranpotential im Empfänger-Neuron. Im Jahr 1966 machte der Norweger Terje Lømo die dazu passende Entdeckung. Er führte eine Reihe neurophysiologischer Experimente mit betäubten Kaninchen durch, um die Rolle des Hippocampus in Bezug auf das Kurzzeitgedächtnis zu untersuchen. Er stimulierte dabei einzelne Nervenzellen und stellte fest: Treten die mit den Stimulationen verknüpften Aktionspotentiale häufiger oder schneller oder besser koordiniert auf, so führt dies zur dauerhaften Verstärkung der Signalübermittlung zwischen den Zellen. Zusammen mit seinem Mitarbeiter Timothy Bliss publizierte Lømo 1973 diese Ergebnisse. Die Autoren bezeichneten dieses Phänomen als long-term potentiation (LTP), eine Form der synaptischen Plastizität. Dieser Lernprozess, der wenige Minuten bis zu lebenslang anhalten kann, wurde intensiv im Hippocampus erforscht.

Maus-aus-Labyrinth-Röhrenrechner

Zurück zu den ersten Versuchen, etwas zu entwickeln, was Funktionen ausführte, die man irgendwie mit Intelligenz assoziieren konnte. 1951 baute der erst 24 Jahre alte Mathematiker Marvin Minsky (1927-2016) sogar neuronale Hardware, den SNARC (Stochastic Neural Analog Reinforcement Computer). Er bestand aus 40 Neuronen mit Synapsen, die ihre Gewichte (Maße der synaptischen Permeabilitäten) an den Erfolg einer bestimmten Aufgabe anpassten (Hebbsches Lernen) – vollkommen analog, worauf schon das Akronym hindeutet. Die Maschine bestand aus Röhren, Motoren und Kupplungen, hatte den Charakter einer Bastelarbeit, aber sie modellierte erfolgreich das Verhalten einer Ratte in einem Labyrinth, die nach Nahrung sucht. SNARC konnte, was bis dahin Laborratten vorbehalten war: einen Weg aus einem Labyrinth finden. Damit war SNARC das erste künstliche neuronale Netz, das ein Problem aus der realen Welt löste.

Der Begriff „Künstliche Intelligenz“ taucht auf

Diese Konstruktion löste einen ersten KI-Hype aus, der gleich zur Erfindung des Begriffs „künstliche Intelligenz“ führte. Im Sommer 1956 wurde von John McCarthy, Marvin Minsky, Nathaniel Rochester und Claude Shannon am Dartmouth College in Hanover, New Hampshire, eine Konferenz beantragt, geplant und durchgeführt. Zur Durchführung beantragten die vier Initiatoren bei der Rockefeller Foundation eine Förderung in Höhe von 13500 US-Dollar. In diesem Antrag nutzte John McCarthy (1927-2011) erstmals den Begriff „artificial intelligence“.

Perceptron erkennt Buchstaben

1958 – neun Jahre nach Hebb's Postulat – baute der amerikanische Psychologe Frank Rosenblatt (1928-1971) das Perceptron, ein einfaches künstliches neuronales Netzwerk (KNN oder ANN nach dem englischen „artificial neural net“). Es war lernfähig – der entscheidende Unterschied zu dem McCulloch-Pitts-Modell – und fehlertolerant, konnte Muster klassifizieren und „Erfahrungen“ verallgemeinern, also generalisieren. All das waren Eigenschaften, die man zuvor nur dem Gehirn zuschrieb. Das Mark 1-Perceptron bestand aus einem Raster mit 400 Photozellen, die mit 512 neuronensimulierenden Schaltkreisen verbunden waren. Eine Hebb-ähnliche Lernregel veränderte die Verbindungen (synaptischen Gewichte) zwischen ihnen, und die „Netzhaut“ lernte, Buchstaben zu erkennen. Es war noch Mechanik angesagt bei dem Mark 1-Perceptron: Per Hand wurden die Voreinstellungen vorgenommen, und Elektromotoren verstellten die Potentiometer, die als synaptische Gewichte den simulierten Neuronen vorgeschaltet waren. Diese Neuronen selbst waren eine Kombination aus elektronischem Verstärker, der die Eingabespännungen aufsummierte, und einer Logikschaltung, die nach einer Schwellwertfunktion ein Ausgangssignal erzeugte. Das Lernen erfolgte in vier Schritten:

1. Zufällige Einstellung der Gewichte
2. Berechnen eines Ausgangssignals für ein bekanntes Trainingsbild
3. Veränderung der Gewichte, wenn das Ausgangssignal nicht dem Trainingsbild entsprach
4. Nächster Trainingssatz, Wiederholen der Schritte

Das Perceptron führte zu einem weiteren KI-Hype: Auf einer Pressekonferenz der US-Marine von 1958 gab Rosenblatt Erklärungen über das Perzeptron ab, das eine hitzige Kontroverse unter der jungen KI- Gemeinschaft hervorrief. Die „New York Times“ griff die Erläuterungen mit den Worten auf, dass das Perceptron „der Embryo eines elektronischen Computers ist, von dem man [die Marine] erwartet, dass er in der Lage ist, zu gehen, zu sprechen, zu sehen, zu schreiben, sich selbst zu reproduzieren und sich seiner Existenz bewusst zu sein.“

1969 wurde der KNN-Forschung jedoch ein Tiefschlag versetzt. Marvin Minsky und Seymour Papert (1928-2016, u. a. Erfinder der Programmiersprache Logo) vom Massachusetts Institut of Technology (MIT) wiesen nach, dass das Perceptron bei bestimmten Klassifikationen versagen muss: Ein einlagiges Perzeptron kann den XOR-Operator nicht auflösen (Problem der linearen Separierbarkeit). Daraufhin folgte rund zehn Jahre lang ein „KI-Winter“ in der Erforschung künstlicher Intelligenz, insbesondere solcher mit künstlichen neuronalen Netzen.

Top down: Symbolische KI

Die aufkommende Leistungsfähigkeit digitaler Computer führte dann auch zu einer KI-Entwicklung, die man als „symbolische KI“ beschreibt. Symbolische KI einerseits und neuronale oder auch konnenktionistisch genannte KI andererseits versuchen, sich dem Phänomen Intelligenz von zwei verschiedenen Sichtweisen aus anzunähern. Während die neuronale KI Basisstrukturen für Intelligenz Erscheinungen erschafft – kurz gesagt das Gehirn nachbildet, also von den kleinsten

Einheiten her höhere Leistungen erbringt (bottom up), verfolgt symbolische KI einen top down-Ansatz. Sie nähert sich der Intelligenz von begrifflicher Seite. Symbole sind dabei Objekte, die den Zugriff auf Bedeutungen ermöglichen. Dadurch können dann Schlüsse gezogen und Probleme gelöst werden. Hirnstrukturen spielen also bei der symbolischen KI keine Rolle; es kommt allein auf das Ergebnis an. Typische Verfahren symbolischer KI sind repräsentiert in Expertensystemen. Sie nutzen beispielsweise fallbasiertes Schließen (CBR für case based reasoning). Dabei werden Objekte in Datenbanken abgelegt, mit Attributen gekennzeichnet, miteinander in Beziehung gesetzt. Regelwerke, die auch Ähnlichkeiten berücksichtigen können, erlauben, von bekannten Situationen (case based) auf neue zu schließen. Case-Based Reasoning hat sich besonders in Anwendungssystemen für den Kundendienst, so genannten Help-Desk-Systemen, bewährt, wo man es z. B. zur Diagnose und Therapie von Kundenproblemen nutzt. Ein anderer Ansatz für Expertensysteme sind regelbasierte Systeme. Sie basieren nicht auf konkreten Fallbeschreibungen, sondern auf Regeln der Art „Wenn A, dann B“.

Erste Arbeiten zu Expertensystemen gab es in den 1960er Jahren. Damals beschäftigte man sich auch detaillierter mit der Frage, wann denn ein System als intelligent zu bezeichnen wäre. Dazu hatte bereits 1950 der geniale Mathematiker und Informatiker Alan Turing eine Idee entwickelt.

Der Engländer hatte durch seine Entschlüsselung der deutschen Chiffriermaschine „Enigma“ entscheidenden Einfluss auf den Ausgang des 2. Weltkriegs. Turing war aber auch ein Pionier von Arbeiten zur künstlichen Intelligenz. Grundlegend war sein Werk „Computing Machinery and Intelligence“; darin schlug er einen Test vor, anhand dessen man die Gleichwertigkeit des Denkvermögens einer Maschine verglichen mit dem eines Menschen feststellen könnte. Er nannte den Test, der heute nach ihm benannt ist, „The Imitation Game“ – 2014 der Titel eines Kinofilms über Turing. Er war aber auch Pionier eines Gebiets, das man heute als „künstliches Leben“ bezeichnet. So benutzte er den Ferranti Mark I-Computer der Universität Manchester, um biologisches Wachstum zu modellieren. Von 1952 bis zu seinem Tod 1954 arbeitete Turing an mathematischen Problemen der theoretischen Biologie. Er veröffentlichte 1952 eine Arbeit zum Thema „The Chemical Basis of Morphogenesis“. In diesem Artikel wurde erstmals ein Mechanismus beschrieben, wie Reaktions-Diffusions-Systeme spontan Strukturen entwickeln können. Dieser heute als Turing-Mechanismus bekannte Prozess steht noch heute im Mittelpunkt vieler chemisch-biologischer Strukturbildungstheorien; er erklärt zum Beispiel die Muster auf den Fellen von Zebra oder Giraffe. Turing war überzeugt: „Sage mir exakt, worin deiner Meinung nach der Mensch einer Maschine überlegen sei, und ich werde einen Computer bauen, der deine Meinung widerlegt.“ Und er stand auch dafür ein: „Ich verfechte die Behauptung, dass Maschinen konstruiert werden können, die das Verhalten des menschlichen Geistes weitestgehend simulieren.“ Ein revolutionärer Gedanke zu einer Zeit, was den entsprechend veranlagten und in Großbritannien trotz seiner Verdienste verurteilten Turing 1954 in den Tod trieb.

Kommunikation mit ELIZA, Rat von MYCIN

1966 entwickelte der deutsch-amerikanische Informatiker Joseph Weizenbaum (1923-2008) ein Computerprogramm, das symbolische KI nutzte und sich dem Turing-Test annäherte: ELIZA. Es sollte die Möglichkeiten der Kommunikation zwischen einem Menschen und einem Computer über natürliche Sprache aufzeigen. Es konnte verschiedene Gesprächspartner simulieren, wurde insbesondere bekannt durch die Simulation eines Psychotherapeuten.

Grundlage für ELIZA bildete ein strukturiertes Wörterbuch. Ein eingegebener Satz wurde dahingehend durchsucht, ob er ein Wort enthält, das im Thesaurus vorhanden ist, und dann von diesem Wort ausgehend nach Synonymen, meist aber nach Oberbegriffen gesucht. Das Programm

enthielt weiterhin eine Sammlung von Phrasen zu verschiedenen Themengebieten. Was es damit auf sich hatte – und wie primitiv der Ansatz war – verdeutlicht am einfachsten ein Beispiel:

Person: Ich habe ein Problem mit meinem Vater.

ELIZA: Erzählen Sie mir mehr über Ihre Familie.

Das Programm hatte also zu dem Begriff Vater den Oberbegriff Familie gefunden, dazu passende Phrasen gesucht. Dumm nur, wenn die Zusammenhänge andere waren:

Person: Krieg ist der Vater aller Dinge.

ELIZA: Erzählen Sie mir mehr über Ihre Familie.

Im Grunde erfolgreich war ein medizinisches Expertensystem: MYCIN. Expertensysteme lebten in der Eiszeit der konnektionistischen KI auf. Das Medizin-System wurde seit 1972 an der Stanford University in der Programmiersprache Lisp entwickelt, diente zur Diagnose und Therapie von Infektionskrankheiten durch Antibiotika. Zu der Zeit seiner Entwicklung begann man die übermäßige Anwendung von Antibiotika kritisch zu betrachten und suchte daher nach Methoden, ihre Anwendung in Abhängigkeit vom jeweiligen Krankheitsbild zu optimieren. Hierzu mussten zahlreiche Parameter ermittelt und zueinander in Beziehung gesetzt werden, darunter der Erregertyp, der bisherige Krankheitsverlauf, bestimmte Labordaten usw. Die Komplexität dieses Problems wurde so groß, dass man die Entwicklung eines Expertensystems vorantrieb. MYCIN erreichte in seinen Diagnosen durchaus sehr hohe Trefferquoten. Damals war jedoch die Akzeptanz von Computersystemen noch relativ gering, so dass die Bereitschaft fehlte, sich auf die Diagnose eines schwer durchschaubaren Systems zu verlassen.

Symbolische KI: Schach- und Jeopardy-Erfolg

Herausragende Höhepunkte feierte die symbolische KI 1996 und 2011. 1996 gelang es Deep Blue als erstem Computer, den damals amtierenden Schachweltmeister Garri Kasparow in einer Partie mit regulären Zeitkontrollen zu schlagen. Das Computersystem Watson von IBM schlug im Februar 2011 in der Quizsendung Jeopardy! zwei menschliche Gegner, die in der Show zuvor Rekordsummen gewonnen hatten. Ihre Aufgabe ist es, schneller als die Mitspieler eine passende Frage auf die vorgegebene Antwort zu formulieren. Die Aufgaben sind meist bewusst mehrdeutig formuliert und machen häufig die Verknüpfung mehrerer Fakten erforderlich. Bei ersten Testläufen im Jahr 2006 fand Watson nur für etwa 15 % von 500 Umschreibungen vorangegangener Jeopardy!-Sendungen die korrekte Frage. Die besten Kandidaten von Jeopardy! erreichen im Vergleich dazu etwa 95 % Genauigkeit. Watson wird mittlerweile kommerziell vermarktet. Im Januar 2017 ersetzte eine japanische Versicherung mehr als 30 Mitarbeiter durch die Watson-Plattform. Die KI soll Namen und Daten der Versicherten sowie deren medizinische Vorgeschichte prüfen und Verletzungen bewerten.

Weitere Hilfen für die KI

Zwischenzeitlich setzte jedoch die neuronale KI zum Überholen an und ist nun auf der Überholspur. Aus der geschichtlichen Entwicklung heraus gibt es zunächst jedoch zwei Anmerkungen. In dem Winter der konnektionistischen KI hatten diese zwei Entwicklungen Auftrieb, die auch heute noch der künstlichen Intelligenz dienen, aber bei der überragenden Beschreibung neuronaler Netze meistens untergehen: Fuzzy-Logik sowie evolutionäre Algorithmen.

Fuzzy-Logik hilft mit Unschärfe umzugehen

Bei der – auf deutsch „unscharfe Logik“ übersetzten Methodik handelt es sich um eine Theorie, welche zunächst in der Mustererkennung zur präzisen Erfassung des Unpräzisen entwickelt wurde, sodann der Modellierung von Unschärfe von umgangssprachlichen Beschreibungen von Systemen diente, heute aber überwiegend in angewandten Bereichen wie etwa der Regelungstechnik eine Rolle spielt. Fuzzy-Logik ermöglicht die logische Verknüpfung unscharfer Begriffe, indem sie die scharfe Unterscheidung von wahr und falsch aufgibt. Anstelle der Booleschen Variablen wahr (1) und falsch (0) treten Wahrheitswerte, die von 0,0 bis 1,0 jeden beliebigen Zahlenwert annehmen können.

Entwickelt wurde die Fuzzy-Set-Theorie, also die unscharfe Mengenlehre, 1965 von dem US-amerikanischen Mathematiker und Elektroingenieur – mit iranisch-ukrainischen Wurzeln – Lotfi Zadeh (1921-2017) an der University of California. Die Fuzzy-Technologie nahm in den 1980er Jahren vor allem in Japan ihren Aufschwung mit der Fuzzy-Welle. Ein Beispiel ist die Regelung der vollautomatischen U-Bahn Sendai, die erste erfolgreiche Großanwendung mit Fuzzylogik. Später fand die Fuzzy-Logik auch in Geräten der Haushaltstechnik breite Anwendung. Fuzzy-Logik kann beispielsweise bei der Steuerung von verfahrenstechnischen Prozessen angewendet werden. Man stelle sich vor, die Brennstoffzufuhr in einer Heizkammer soll als Funktion von Druck und Temperatur in der Kammer geregelt werden. Menschliche Beschreibungen wie etwa „sehr hoch“ und „über dem Normalwert“ sind unscharf.

Evolutionäre Algorithmen optimieren – auch KNN

Die zweite Entwicklung, die in der Eiszeit der konnektionistischen KI aufblühte, waren evolutionäre oder auch genetische Algorithmen (EA oder GA; die aus der historisch erklärbaren Bezeichnungen verschwimmen mehr und mehr in ihren Abgrenzungen). Während bei den Evolutionsstrategien die Mutation und die selbstregulierende Schrittweitenanpassung für Verbesserungen sorgt, ist dies bei den genetischen Algorithmen die genetische Rekombination durch die verschiedenen Crossover-Verfahren. Die beiden Verfahren unterscheiden sich auch durch ihre unterschiedlichen Selektionsprozesse. Beide dienen jedoch u. a. zur Optimierung neuronaler Netzstrukturen und sind somit ein Werkzeug zur Effizienzsteigerung der Neurotechnik.

Schon 1950 schlug der bereits für seine Genialität gerühmte Alan Turing eine „Lernmaschine“ vor, die den Prinzipien der Evolution entsprechen sollte. Computersimulation der Evolution begann bereits 1954 mit der Arbeit von Nils Aall Barricelli, ein norwegisch-italienischer Mathematiker (1912-1993), der den Computer am Institute for Advanced Study in Princeton benutzte. Seine Veröffentlichung von 1954 wurde allerdings nicht allgemein wahrgenommen. Einen Durchbruch erfuhren die evolutionären Algorithmen durch Arbeiten der deutschen Wissenschaftler Ingo Rechenberg (geb. 1934) und Hans-Paul Schwefel (geb. 1940) in den 1960er und beginnenden 1970er Jahren sowie des US-Informatikers John H. Holland (1929-2015) mit seinem Buch über GA von 1975 „Adaptation in Natural and Artificial Systems“. Hans-Paul Schwefel lernte 1963 während seines Studiums an der TU Berlin Ingo Rechenberg kennen. Beide studierten die Luft- und Raumfahrttechnik und beide interessierten sich für Kybernetik und Bionik. Schwefels Diplomarbeit ging dann auch über die Thematik „Kybernetische Evolution als Strategie für experimentelle Strömungsmechanik“.

Die Begeisterung für Evolutionsalgorithmen ist durchaus verständlich: Ein GA findet Antworten auf gestellte Aufgaben, indem eine Vielzahl von Lösungen dieser Aufgabe rückgekoppelt, bewertet und neu berechnet werden. Die Lösungen eines GA werden – in Analogie zum natürlichen Vorbild durch „Chromosome“ repräsentiert, die ihrerseits aus Genen zusammengesetzt sind. In der Natur sind in den Chromosomen die Eigenschaften der Lebewesen verschlüsselt, etwa die Farbe der Flügel von Schmetterlingen. Damit löst jede Art ihre spezifischen

Überlebensaufgaben. Ändert sich die Problemstellung, weil etwa eine Vogelart alle gelbflügeligen Schmetterlinge sofort erkennt und zur Lieblingsspeise erkürt, droht in diesem Beispiel die Schmetterlingsart auszusterben, es sei denn, sie ändert die Farbe ihrer Flügel. Solch eine Änderung kann bei Paarungen passieren, wenn zufällig andersfarbige Liebespartner gefunden werden, oder eine unprogrammgemäße Änderung des genetischen Codes sorgt für eine positive Änderung des Erscheinungsbildes. Genau so arbeiten GAs: Ihre Problemlösungen (Chromosome) werden durch Reproduktion (die ja auch eine Selektion der beteiligten Chromosome bedeutet) und Mutation optimiert. Die Wissenschaftler fügen zu diesen Methoden noch eine spezielle Art der Kreuzung von GA-Chromosomen, das Crossover.

Beim Crossover schneidet man ein ausgewähltes Chromosomenpaar an jeweils gleicher Stelle auseinander und fügt die Teile über Kreuz wieder zusammen. Damit entstehen zwei neue Individuen. Reproduktion bedeutet, dass die Gensätze eines Mutter- und eines Vaterchromosoms kombiniert werden. Mutation schließlich bedeutet, dass ein sehr kleiner Teil des Chromosoms durch Zufall verändert wird. In der Natur geschieht so etwas zum Beispiel durch die Einwirkung starker Strahlung auf die DNA, dem Molekül, das in Chromosome unterteilt die Information über die Erbanlagen enthält. Die Bewertung, ob ein GA-Chromosom eine gute Lösung für ein Problem darstellt, erfolgt durch Kostenfunktionen. Ihr Wert ist bei den technischen Optimierungsproblemen bekannt.

Während man in Deutschland eine Evolutionsstrategie aufbaute, die sich nur auf einfache Vererbung (Selektion) und Mutation stützte, führte man in den USA frühzeitig die Crossover-Operation ein. Mit ihrer Hilfe steigerte sich die Geschwindigkeit der GA, zu guten Lösungen zu kommen, erheblich.

Da diese Algorithmen nicht wissen können, ob nicht eventuell unter den Individuen ein Optimum gefunden wurde, definiert man noch Abbruchkriterien. Grundsätzlich arbeiten evolutionäre Algorithmen metaheuristisch: Sie sind in der Lage, mit vertretbarem Aufwand bei komplexen Problemen gute Lösungen zu finden; die beste Lösung kann aber eine andere sein.

Mit genetischen Algorithmen lassen sich auch Programme erstellen. Das Ziel der genetischen Programmierung ist die Erzeugung von Strukturen, die eine bestimmte Eingabe in eine festgelegte Ausgabe umwandeln sollen (Computerprogramme, Schaltkreise oder mathematische Funktionen). Die Lösungskandidaten werden durch Bäume repräsentiert. Ein Beispiel: Es gibt einen Wert X , der durch bestimmte Operationen – beispielsweise durch einen polynomen Ausdruck – den Wert Y liefert. Vorliegen hat man eine Anzahl von X - und Y -Werten. Der GA findet dann dazu das Polynom.

Wie schon die bisherigen Informationen über KNN zeigen, verfügen diese auch über jede Menge Stellschrauben (Strukturvarianten). Diese Vielfalt – von der im kommenden Artikel noch mehr dargestellt wird – lässt sich durch GA auch für spezielle Problemstellungen optimieren.

Revival der künstlichen Neuronetze

Zurück zur den künstlichen neuronalen Netzen: Sie erlebten ihre Wiedergeburt Ende der 1970er, Anfang der 1980er Jahre. Zum einen hatte sich herausgestellt, dass die von Minsky und Papert beschriebenen Einschränkungen nur für spezielle, sehr einfache Netzwerke ohne „verdeckte Schicht“ gelten; zum anderen ermöglichte die Geschwindigkeitssteigerung konventioneller Computer, immer bessere Simulationen von künstlichen neuronalen Netzen zu konstruieren und zu prüfen.

Fehlerrückführung – Schicht für Schicht

Backpropagation meint ausgesprochen „Backpropagation of Error“ also eine Fehlerrückführung. Dieses Lernverfahren von KNN gehört zur Gruppe der überwachten Lernverfahren und wird als Verallgemeinerung der Delta-Regel auf mehrschichtige Netze angewandt. Dazu muss ein externer Lehrer existieren, der zu jedem Zeitpunkt der Eingabe die gewünschte Ausgabe, den Zielwert, kennt.

Der Mechanismus funktioniert wie folgt: Ein Eingabemuster wird angelegt und vorwärts durch das Netz propagiert. Die Ausgabe des Netzes wird mit der gewünschten Ausgabe verglichen. Die Differenz der beiden Werte wird als Fehler des Netzes erachtet. Der Fehler wird nun wieder über die Ausgabe- zur Eingabeschicht zurück propagiert. Dabei werden die Gewichtungen der Neuronenverbindungen abhängig von ihrem Einfluss auf den Fehler geändert. Eine neue Eingabe wird durch die erfolgte Korrektur der Gewichte einen geringeren Fehler aufweisen.

Die Formel des Backpropagation-Verfahrens wird durch Differentiation hergeleitet. Das Ziel ist nun die Minimierung der Fehlerfunktion, wobei aber im Allgemeinen lediglich ein lokales Minimum gefunden wird. Wichtig ist dabei, dass die synaptischen Gewichte in einem angemessenen Maß korrigiert werden. Ist die Lernrate (Korrekturrate) zu groß, kann das Minimum verfehlt werden; ist sie zu klein, wird der Lernprozess sehr langsam. Es gibt Verfahren, die mit variabler, adaptiver Lernrate arbeiten.

Man könnte fragen: Hat nicht auch schon Frank Rosenblatts Perceptron Fehler der Ausgabe rückgekoppelt (siehe die 4 Schritte auf Seite 115)? Das stimmt zwar, aber es wurde beim Trainingsprozess eben nur die Ausgabeschicht mit der korrekten Ausgabe verglichen. Keiner wusste, wie mögliche Zwischenschichten zu korrigieren wären. Der entscheidende Schritt hin zum Backpropagation-Algorithmus war die Einrichtung eines Kettenregel-Mechanismus.

Die Kettenregel ist eine der Grundregeln der Differentialrechnung. Sie trifft Aussagen über die Ableitung einer Funktion, die sich selbst als Verkettung von zwei differenzierbaren Funktionen darstellen lässt. Kernaussage der Kettenregel ist dabei, dass eine solche Funktion selbst wieder differenzierbar ist und man ihre Ableitung erhält, indem man die beiden miteinander verketteten Funktionen separat ableitet und – ausgewertet an den richtigen Stellen – miteinander multipliziert. Die Kettenregel lässt sich verallgemeinern auf Funktionen, die sich als Verkettung von mehr als zwei differenzierbaren Funktionen darstellen lassen. Der Hintergrund für die Anwendung bei KNN ist: Die Aktivierungsfunktion der künstlichen Neuronen, ihr Ansprechen auf Eingangssignale, ist nichtlinear, aber differenzierbar. Mit Anwendung der Kettenregel auf KNN war man also in der Lage, jedem Neuron einen bestimmten Anteil an einem Fehler zuzuschreiben.

Die Rückwärts-Propagierung eines Fehlers erfolgt bei Backpropagation nicht auf einen Schlag von der Ausgabeschicht des Neuronetzes auf die Eingabeschicht; vielmehr geschieht die Fehlerrückführung Schicht für Schicht, von der letzten zur vorletzten usw. bis zur ersten Schicht. Daher kann man ein Backpropagation-KNN nicht ohne besondere Tricks mit besonders vielen Schichten ausführen. Bei mehrschichtigen Neuronetzen mit sehr vielen Schichten (den deep neural nets) versagt der Backpropagation-Algorithmus. Der Einfluss auf einen Fehler wird nämlich mit jeder zusätzlichen Schicht für ein Neuron darin immer geringer, bis es schließlich nahezu nicht mehr aufzuspüren ist, welchen Fehler ein Neuron in einer tiefen Schicht wirklich auf ein Ergebnis hat. Und wenn kein Fehler zu finden ist, lässt sich auch keiner rückkoppeln. Neuerdings gibt es Auswege dazu, wie der nächste Artikel zeigt, der die neuesten Entwicklungen bei KNN vorstellt.

Feedback-Neuronetze

Nur angemerkt sei: Neben den hier bislang beschriebenen Feedforward-Neuronetzen – in ihnen fließt die Information zwischen den Schichten der Prozessorelemente ohne direkte Rückkopplung

(sondern nur von Schicht zu Schicht) von den Eingangsneuronen zu den Ausgangsneuronen – gibt es auch Netze, in denen die Neuronen aller miteinander verkoppelt sind. Weitere Details dazu stehen in dem nächsten Artikel in dem Abschnitt „Verschiedene Lernmechanismen“.

Allgemein gilt: Bei den künstlichen Neuronetzen gibt es etliche Modelle; es lässt sich eine Vielzahl von Parametern ändern. Das macht ihren Einsatz für spezifische Anwendungen so schwierig; es war zumindest bislang eine genaue Abstimmung auf jede Aufgabe nötig. Wissenschaftler und Techniker können Feedforward- oder Feedback-Netze entwerfen, die lernregeln variieren, die Input-, Transfer- sowie die Outputfunktionen der Prozesselemente verändern. Damit wird beispielsweise festgelegt, ob ein Neuron alle Eingangssignale einfach aufsummiert (Input), ob ein Schwellwert zu überwinden ist, damit das Neuron Impulse weitergibt (Transfer), oder ob nur das Neuron einer Schicht mit dem höchsten Output Impulse weitergeben darf (Output), gemäß der Losung: der Gewinner bekommt alles. Solche Optimierungen lösen wie oben schon ausgeführt mittlerweile auch genetische Algorithmen.

Erste kommerzielle Anwendungen

Die bisher hier beschriebenen Neuronetze waren Stand der Technik bis etwa 1995. Ihre Technik galt damals noch als exotisch; dennoch gab es schon einige wenige Anwendungen im geschäftlichen Bereich. Eine der ersten Firmen, die künstliche Neuronetz-Programme vermarkteten, war Nestor in Providence, Rhode Island, USA. Eins ihrer Systeme kam als Decision Learning System für Vorhersagen im Kreditwesen zum Einsatz. Es gab Empfehlungen zur Bausparvertrags-Vergabe. An die Finanzwelt wendeten sich auch andere Firmen. So verkaufte das damalige Unternehmen Siemens Nixdorf (SN) seine Neuronetz-Entwicklungsumgebung „SENN“. Es ging damals schon um Prognosen für Aktien- und Devisenentwicklungen.

Noch ein kleiner Neuro-Winter

Diese ersten Anwendungen änderten jedoch nichts daran, dass die Entwicklung von KNN wieder einmal ins Stocken geriet. Schuld daran war nicht die Theorie – und wichtige Netzarchitekturen, die im nächsten Artikel besprochen werden, nahmen in den 1990er Jahren ihren Anfang – sondern die mangelnde Leistungsfähigkeit der Computer. Dies und dann folgend die jetzigen Erfolge der KNN sollte man im Hinterkopf behalten, wenn man versucht, sich vorzustellen, den Fortschritt abzuschätzen, den neuromorphe Hardware bringen wird. [8]

6. Studienangebote zum Thema

Data Scientists sind in Zeiten von "Big Data" gefragte Experten. Denn die Analyse von großen und komplexen Daten bedeutet in allen Wirtschaftsbereichen einen Wettbewerbsvorsprung.

Wenn in Wissenschaft, Wirtschaft, Verwaltung und Technik neue wertvolle Erkenntnisse gewonnen werden, spielt die Auswertung von großen und komplexen Datenmengen in der Regel eine tragende Rolle – einerlei ob es um den sachgerechten Umgang mit Patienten- oder Kundendaten, um Aktienkursen und Renditen oder Wetter- oder Klimainformationen.

Alle Firmen verfügen über enorme Datensätze, aber nicht die passenden Algorithmen, die die Daten ordnen und nützlich machen. Gefragt sind deshalb Experten, die Instrumente dafür entwickeln. Das Erfassen von Daten und deren Verarbeitung machen nur einen Teil dieses neuen Berufsbilds aus. [9]

1. Was studieren?

In der Data Science spielen mathematische Kenntnisse eine sehr wichtige Rolle. Deshalb muss man mindestens die folgenden Gebiete gut wissen:

- ✓ Lineare Algebra
- ✓ Wahrscheinlichkeitstheorie
- ✓ Mathematische Statistik
- ✓ Maschinelles Lernen Theorie
- ✓ Methoden der Optimierung
- ✓ Algorithmen

2. Online-Angebote

- ✓ Coursera
- ✓ Kaggle
- ✓ Google Academy
- ✓ Bücher

3. Angebote in Deutschland

Bachelor:

- ✓ Universitäten Stuttgart, Marburg, Göttingen
- ✓ Hochschulen Ostwestfalen-Lippe, der Medien Stuttgart, Beuth, Karlsruhe
- ✓ Duales Studium

Master:

- ✓ Universitäten Darmstadt, Mannheim, Marburg, Potsdam
- ✓ TU Chemnitz

4. Universität Leipzig

Es gibt auch ein Masterstudiengang mit Schwerpunkt Big Data an der Leipziger Universität. Man studiert da solche Module, wie Moderne Datenbanktechnologien, Komplexitätstheorie, Künstliche neuronale Netze und Maschinelles Lernen, Computeralgebra, Statistisches Lernen, Neuroinspirierte Informationsverarbeitung.

7. Neue Branchen der Informatik

Heutzutage gibt es nicht nur Künstliche Intelligenz, sondern auch andere neue Branchen sind immer mehr populär. Rund um die Informatik haben sich einige interdisziplinäre Teilgebiete und Forschungsansätze entwickelt, teilweise zu eigenen Wissenschaften. Zum Beispiel:

- ✓ Die **Bioinformatik** ist eine recht junge Teildisziplin der Informatik, die sich vorwiegend mit der Auswertung biologischer Daten, sowie deren Speicherung und Strukturierung beschäftigt. Zu Beginn waren es meist die DNA-Sequenzen von Genen bzw. Genomen, die in großen Datenbanken gespeichert und analysiert wurden. Inzwischen gehören zur Bioinformatik u.a. aber auch die Vorhersage von Protein- und RNA-Strukturen, Erstellung von phylogenetischen Bäumen und die Untersuchung von Wechselwirkungen zwischen Substanzen in der Pharmazie. Deshalb spielt die Bioinformatik inzwischen auch in den traditionellen Naturwissenschaften wie z.B. der Molekularbiologie, Biochemie, Pharmazie und Genetik eine immer größer werdende Rolle. [10]
- ✓ **Chemoinformatik**, Cheminformatik oder Chemieinformatik bezeichnet einen Wissenschaftszweig, der das Gebiet der Chemie mit Methoden der Informatik verbindet mit dem Ziel, Methoden zur Berechnung von Moleküleigenschaften zu entwickeln und anzuwenden. Der Begriff „Chemoinformatik“ ist relativ jung, während die älteren Termini Computerchemie und chemische Graphentheorie das gleiche Gebiet bezeichnen.

- Computerchemie wird heutzutage eher als ein Teilgebiet der Theoretischen Chemie und der Quantenchemie begriffen. [11]
- ✓ Die **Wirtschaftsinformatik** ist Wissenschaft von dem Entwurf, der Entwicklung und der Anwendung computergestützter Informations- und Kommunikationssysteme und -techniken in Unternehmungen und Verwaltungen zur Unterstützung ihrer Geschäftsprozesse. Die Wirtschaftsinformatik nimmt eine Schnittstellenfunktion zwischen der oft technisch ausgerichteten Informatik und der anwendungsorientierten Betriebswirtschaftslehre wahr. Ihr interdisziplinärer Charakter wird durch den Einbezug arbeitswissenschaftlicher, psychologischer, soziologischer und mathematisch-technischer Aspekte noch verstärkt. [12]
 - ✓ Die **Sozioinformatik** ist ein junges Teilgebiet der Informatik, das die Wechselwirkung zwischen sozialen Gruppen und Softwaresystemen untersucht und Gestaltungsprinzipien und Realisierungsmethoden für Software mit großen Nutzergruppen entwickelt. Dazu bedient sie sich Methoden aus der Informatik, der Soziologie, den Wirtschaftswissenschaften und der Psychologie. [13]
 - ✓ Unter **Sozialinformatik** versteht man in erster Linie die systematische Betrachtung des IT-Einsatzes in der sozialen Arbeit und ihren Handlungsfeldern. Sozialinformatik beleuchtet verschiedene Aspekte der IT-gestützten Informationsverarbeitung in sozialen Organisationen. Neben den Grundlagen sozialer Arbeit gehören auch technologisches Wissen und Sozialmanagement-Aspekte zu den Inhalten dieser Disziplin. [14]
 - ✓ Die **Medieninformatik** ist eine angewandte Informatik, die sich mit der zunehmend digitalen Informationsverarbeitung in den Medien und dem hierdurch ausgelösten Zusammenwachsen von modernen Konzepten der Informatik, digitalen Kommunikationstechniken und Multimedia beschäftigt. [15]
 - ✓ Das Fach **Computerlinguistik** ist eine eher junge Disziplin, die sehr stark mit den Entwicklungen in der Informationstechnologie verbunden ist: flexible Suchmaschinen, automatisierte Dialogsysteme, durch Sprache bedienbare Büromaschinen und Werkzeuge, e-learning Unterrichtsmaterial. Die Entwicklung und Verbesserung dieser Anwendung hängt in hohem Grad ab von der gelungenen Integration sprachlicher bzw. linguistischer Datenstrukturen. Und genau darin liegt das Hauptinteresse der Computerlinguistik, nämlich linguistische Datenstrukturen so zu modellieren, dass menschliche Sprache(n) von Computern verarbeitet und verstanden werden. [16]
 - ✓ Der Begriff „**Umwelt-Informatik**“ wird erstmals 1986 durch J. Seggelke und Bernd Page verwendet. Inzwischen hat sich ein daraus eigenes Fachgebiet der Umweltinformatik entwickelt. Im Allgemeinen versteht man unter der Umweltinformatik umweltbezogene Informatik-Anwendungen mit deren Hilfe ein Beitrag zur Bewältigung der Umwelt- und Ressourcenproblematik geleistet werden soll. Dies wird in erster Linie dadurch erreicht, indem man Methoden und Verfahren der Informatik für Institutionen der Umweltforschung, für Umweltbehörden und auch für umweltorientierte Unternehmen bereitstellt. [17]

8. Zusammenfassung

Schließlich kann man sagen, dass heute sich die Informatik sehr schnell entwickelt und jede 5-10 Jahre entstehen neue. Informatik hat einen Einfluss auf alle Bereiche unseres Lebens. Heute die Künstliche Intelligenz ist sehr wichtige Disziplin der Informatik. Jeden Tag wird es immer mehr in neuen Branchen eingeführt. Maschinelles Lernen ist ein Teilbereich der künstlichen Intelligenz. Der Anwendungsbereich des maschinellen Lernens ist sehr weit: von Gesichtserkennung bis Chatbots. Diese Themen kann man nicht nur laut besprechen, sondern auch gibt es viele Studiengänge an den deutschen Universitäten und Online-Angebote, wo man die Möglichkeit das zu studieren und dann selbst anzuwenden hat. Dazu braucht man gründliche mathematische

Kenntnisse. Auch man muss erwähnen, dass nicht nur die Künstliche Intelligenz eine neue perspektivische Disziplin der Informatik ist. Solche interdisziplinären Teilgebiete, wie Bioinformatik, Wirtschaftsinformatik, Computerlinguistik und so weiter, sind heutzutage auch sehr gefragt.

Literatur

- [1] Datenverwaltung, <https://www.itwissen.info/Datenverwaltung-data-management-DM.html>
- [2] Margaret Rouse, Datenmanagement, 2015, <https://www.computerweekly.com/de/definition/Daten-Management>
- [3] Dr. Klaus Manhart, Was Sie über Maschinelles Lernen wissen müssen, 2018, <https://www.computerwoche.de/a/was-sie-ueber-maschinelles-lernen-wissen-muessen,3329560>
- [4] Stefan Luber, Was ist Machine Learning?, 2016, <https://www.bigdata-insider.de/was-ist-machine-learning-a-592092/>
- [5] Data Analytics – Was ist Data Analytics?, <http://www.datenbanken-verstehen.de/business-intelligence/data-analytics-grundlagen/data-analytics/>
- [6] Robert Schmitz, Menschlich, sozial – und digital, 2018, <https://www.bi-scout.com/menschlich-sozial-und-digital>
- [7] Data Science, https://de.wikipedia.org/wiki/Data_Science
- [8] Analytik News, Das Online-Labormagazin, 2018, http://www.clb.de/Probeartikel_files/CLB3-4-2018kuenstlicheIntelligenz.pdf
- [9] Ulrike Hartmann, Data Science: Analytische Fähigkeiten für den Umgang mit Big Data, <https://studieren.de/data-science.0.html>
- [10] Was ist Bioinformatik?, <https://www.imise.uni-leipzig.de/Lehre/BioInf/Vorstellung.jsp>
- [11] Chemoinformatik, <https://de.wikipedia.org/wiki/Chemoinformatik>
- [12] Prof. Dr. Richard Lackes, Technische Universität Dortmund, Wirtschaftsinformatik, <https://wirtschaftslexikon.gabler.de/definition/wirtschaftsinformatik-50673>
- [13] Sozioinformatik, <https://de.wikipedia.org/wiki/Sozioinformatik>
- [14] Sozialinformatik, <http://www.sozialinformatik.info/>
- [15] Medieninformatik, <https://www.uni-muenchen.de/studium/studienangebot/...>
- [16] Was ist Computerlinguistik?, <https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/seminar-fuer-sprachwissenschaft/studium-lehre/studiengaenge/faq/was-ist-computerlinguistik/>
- [17] Michaela Runge, Umweltinformatik als Teildisziplin der angewandten Informatik, <https://www.grin.com/document/25871>

Bilder:

- [1] Dörte Böhner, [Infografik] Eine kurze Geschichte der Wissenschaft der Daten, 2018, <http://blog.bibliothekarisches.de/blog/2018/08/21/infografik-eine-kurze-geschichte-der-wissenschaft-der-daten/>
- [2] History of Data Mining, <https://www.kdnuggets.com/2016/06/rayli-history-data-mining.html>
- [3] Dr. Klaus Manhart, Was Sie über Maschinelles Lernen wissen müssen, 2018, <https://www.computerwoche.de/a/was-sie-ueber-maschinelles-lernen-wissen-muessen,3329560>