



Ethik und künstliche Intelligenz

Michael Mörike, Integrata-Stiftung für humane Nutzung der Informationstechnologie

Künstliche Intelligenz (KI) wird große Veränderungen bringen: spürbare Verbesserungen unserer Lebensqualität und unerwartete Gefahren. Es ist höchste Zeit, sich mit möglichen Folgen vertraut zu machen. Um die Wirkung der KI in eine von uns gewünschte Richtung zu lenken, müssen wir ihr ein ethisches Verhalten beibringen. Damit das gelingen kann, gilt es, unsere eigene Ethik zu analysieren und zu begreifen. Beim Kongress „Ethik und KI“ kann jeder daran mitwirken, der etwas dazu zu sagen hat.

Künstliche Intelligenz (KI) ist aktuell ein Hype. Alle reden davon, dass sie große Veränderungen unserer Arbeits- und Lebenswelt bringen wird. Da der Begriff der Intelligenz nicht klar definiert ist, unterscheidet man in der Technik zwei Arten von künstlicher Intelligenz: schwache KI („weak AI“) und starke KI („strong AI“). Die schwache ist keine Intelligenz in dem Sinn, wie wir sie uns Menschen zuschreiben. Die starke wäre es schon, wenn es sie denn (schon) gäbe. An ihr wird noch

geforscht, ohne dass aktuell klar ist, wie sie einmal aussehen oder funktionieren könnte.

Wenn man von KI spricht, meint man fast immer die schwache Form. Schwache KI wird immer als Software realisiert, die so programmiert ist, dass sie lernen und entscheiden kann. Zunächst wird sie trainiert und lernt beispielsweise, Muster zu erkennen. Einmal trainiert, kann sie neue Muster mit dem gelernten Material vergleichen und daraus Schlüsse ziehen oder Entscheidungen

treffen. Oberflächlich betrachtet sieht das dann so aus, als ob die Maschine intelligent wäre. Daher der technische Begriff „künstliche Intelligenz“.

Künstliche Intelligenz war in den vergangenen Jahrzehnten der Informatik schon mehrfach ein Hype, der nach einiger Zeit aus Enttäuschung über die damals noch schwachen Ergebnisse immer wieder abgeklungen ist. Diesmal aber sieht es anders aus: Vor etwa drei Jahren ist beim Lernen mit neuro-

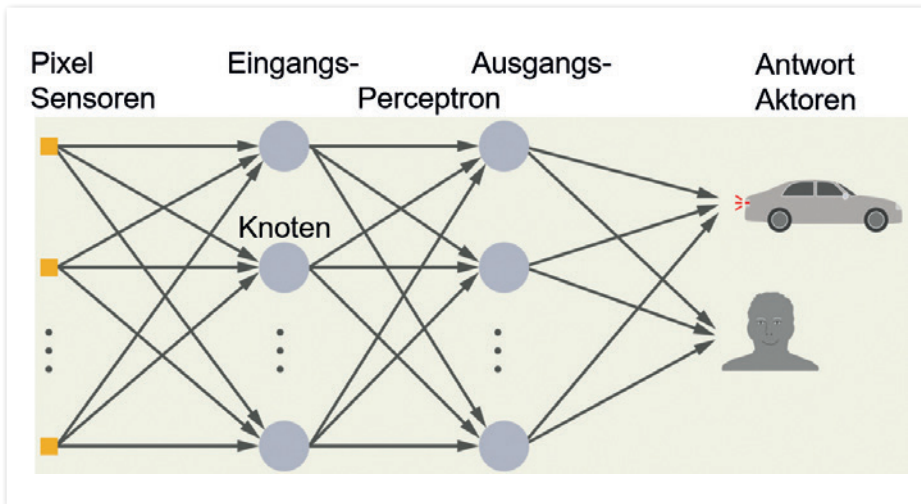


Abbildung 1: Die Berechnung von Entscheidungen ist sehr komplex

nenalen Netzen ein technischer Durchbruch gelungen. Man hat nicht – wie früher - nur einige wenige (etwa ein Dutzend) Schichten eingesetzt, sondern sehr viele (mehrere Hundert). Das machte die seit früher deutlich gestiegene Rechenpower moderner Großrechner möglich. Schließlich sind bei jedem einzelnen kleinen Lernschritt viele Millionen von Datenpunkten zu berechnen (siehe Abbildung 1).

Wie KI funktioniert

Die erfolgreichsten heutigen KI-Systeme bestehen aus neuronalen Netzen (NN), die in gewisser Hinsicht den Neuronen im menschlichen Hirn nachgebildet sind (nicht wirklich). NN bestehen aus Knoten, die in Schichten angeordnet sind, sogenannten „Perceptrons“. Jeder Knoten multipliziert die Eingangssignale mit einem Gewicht, summiert sie, normiert sie und gibt sie an die nächste Schicht weiter. „Wissen“ ist somit in den Gewichten gespeichert. Einfache NN bestehen aus mindestens drei, meist aus zehn bis zwanzig Schichten, neuerdings auch aus Hunderten von Schichten, die ihrerseits jede aus Hunderten bis Tausenden von Knoten zusammengesetzt sein können.

Lernen bedeutet, die Gewichte in den Schichten zu verändern. Dafür gibt es verschiedene mathematische Methoden und Algorithmen, die mehr oder weniger gut lernen lassen. Sie sorgen dafür, dass die Antwort – bei gegebenem Input-Signal – am Ausgang richtig wird (im Sinne des gewünschten Lern-Ergebnisses). Eine davon heißt „Back-Propagation“, was deutlich macht, dass die Gewichte von hinten, also vom Ausgang her, in kleinen Schritten passend verändert werden. Bis auf vielen Bil-

dern alle Gegenstände richtig erkannt werden, braucht es viele Tausend Durchläufe mit jeweils kleinen Änderungen.

Das „Wissen“ eines NN ist, wie schon genannt, in seinen Gewichten gespeichert; es ist also holografisch: Nicht ein einzelnes Gewicht verursacht eine bestimmte Antwort, sondern alle Gewichte gemeinsam ergeben die richtige Antwort. Es ist nicht mehr einfach nachvollziehbar, wie ein bestimmtes Gewicht zu einer bestimmten Antwort beiträgt, also auch nicht, wie eine bestimmte Antwort zustande kommt. Es kann auch sein, dass die Verteilung der Gewichte sich unterscheidet, je nachdem, in welcher Reihenfolge etwas gelernt wurde, obwohl das in den Gewichten des NN gespeicherte „Wissen“ das Gleiche ist.

Da sehr viele sehr einfache Rechenoperationen parallel ausgeführt werden müssen, eignen sich Grafik-Prozessoren (GPU) von Computern recht gut dafür. Sie werden auch von entsprechender Software eingesetzt, die man als Open-Source von Google, Facebook oder Microsoft kostenlos aus dem Netz herunterladen kann. Jeder begabte junge Mensch kann damit also seine eigene KI bauen.

Bestehende KI-Anwendungen

Wenn man sich heute existierende Anwendungen vor Augen führt, wird deutlich, dass KI vor allem dann stark ist, wenn es um Muster- oder Bild-Erkennung geht, wie der folgende Überblick zeigt:

- Bäume, Blumen, Pflanzen und Pilze bestimmen wir heute per Smartphone-App
- PlantVillage erkennt per Smartphone-Fotos Krankheiten bei Pflanzen

- Fußabdrücke von Tieren im Wald kann man per Smartphone-App bestimmen
- Eine Smartphone-App nennt uns die Namen von Bergen in den Alpen
- Nahrungsmittel auf dem Teller – dank der vielen Fotos von Mahlzeiten im Netz
- Die „VIP-Kamera“ erkennt Prominente in Ladengeschäften und macht das Personal darauf aufmerksam, damit dieses sie bevorzugt bedienen kann
- Die Gesichtserkennung dient als Zugangsberechtigung oder zur Alarmauslösung für die Personensuche an Berliner Plätzen
- SeeingAI oder Alpoly sagt per Smartphone-App einem Blinden beim Gang durch die Stadt, was er da vor sich hat: einen Bekannten, ein falsch parkendes Auto, eine Apotheke, einen Bäcker oder auch spielende Kinder im Park etc.
- Schadensmeldungen von Verkehrsunfällen werden heute teilweise vollautomatisch anhand von Smartphone-Fotos abgewickelt
- In der Medizin werden bestimmte Krebsarten (Beispiel Kehlkopfkrebs) von KI anhand von Aufnahmen besser erkannt als von den meisten Ärzten
- DNA-Sequenzierungen lassen sich per KI im Hinblick auf Krebsarten schneller auswerten
- Erkennung von Autismus ist schon bei Säuglingen aufgrund von Gehirn-Scans möglich
- Aus Langzeit-EKGs erkennt KI, ob Herzprobleme vorliegen und welche
- KI zeichnet die Ränder von Tumoren auf CT-Scans nach und bereitet damit deren Bestrahlung vor
- Per Kinect (einem Modul für die Gestensteuerung von Spielen) kann man Störungen von depressiv Kranken diagnostizieren
- KI überwacht Senioren und lernt deren übliche Gepflogenheiten. Bei Abweichung vom oft sehr eigenwilligen individuellen Muster alarmiert sie Hilfe
- Die Ebay-Bildersuche erlaubt, im Netz per Foto ein Produkt zu suchen (und zu bestellen)
- Durchforsten von Urteilen oder anderen juristischen Dokumenten hilft den Anwälten, schneller einen Prozess vorzubereiten und Gewinnchancen abzuschätzen. Nebenbei macht es den Nachwuchs von Anwälten quasi arbeitslos
- KI kann Plagiate bei Texten automatisch erkennen

- KI beherrscht das Übersetzen von Texten in andere Sprachen inzwischen sehr gut
- Plagiate in der Musikindustrie kann KI schnell aufdecken
- In der Forschung hilft KI bei der Identifizierung von Erdbebenwellen und deren Unterscheidung von Wellen, die von unterirdischen Atomexplosionen ausgelöst wurden
- KI wird zur Erkennung von Gravitationswellen eingesetzt
- Chatbots: Alexa, Cortana, Echo, OK-Google, Siri etc. arbeiten mit KI
- Bots zur glaubhaften Verbreitung von Fake News setzen KI ein
- In Spielen ist KI dem Menschen längst überlegen: Bei Schach schon seit Langem, bei Go seit dem Jahr 2017 und sogar beim Pokern ist KI jedem menschlichen Spieler weit überlegen
- Eine derzeit viel diskutierte Anwendung ist das autonom fahrende Auto. Gleiches gibt es natürlich auch für Lastkraftwagen und Ähnliches für Schiffe
- Eine weitere (fürchterliche) Variante sind autonome Waffensysteme, die selbstständig entscheiden, wen sie umbringen sollen

Die Liste ließe sich weiter fortsetzen. Insbesondere aber verlängert sie sich täglich durch immer neue Anwendungen.

Wie geht es weiter?

Die Muster-Erkennung allein bringt (wie beschrieben) bereits einen deutlichen Nutzen. Sie kann aber auch zur Wahrnehmung der Umgebung eingesetzt werden. So geschieht es zum Beispiel beim autonom fahrenden Auto. Mehrere Erkennungssysteme für die Umwelt lassen sich kombinieren, sodass ein KI-System in einem Roboter (im autonom fahrenden Auto) seine Umwelt im Prinzip beliebig genau erfassen kann. Wenn alle (wichtigen) Objekte aus dem Umfeld erfasst sind, können sie mithilfe einer Ontologie miteinander verknüpft und eingeordnet werden.

Wenn die KI, etwa in einem Roboter dann noch etwas bewirken soll, braucht sie ein klares Ziel und einen Satz von möglichen und erlaubten Handlungen, die sie mithilfe von Aktoren ausführt, um das Ziel zu erreichen oder ihm näherzukommen. Für nützliche Handlungen des ausführenden Roboters benötigt die KI eine (nützliche) Zielsetzung. Das autonom fahrende Auto zum Beispiel soll das ihm genannte Ziel ansteuern. Diese Vorgabe allein reicht aber nicht. Es soll ja

auch sicher fahren, also alle Verkehrsregeln beachten und dabei auch auf andere Verkehrsteilnehmer Rücksicht nehmen. Kurz: Es braucht einen ganzen Satz von Verhaltensregeln. Davon sind einige sehr klar formulierbar und daher relativ einfach umzusetzen, wenn sie beispielsweise durch die Verkehrsordnung vorgegeben sind. Es gibt aber auch Fälle, in denen die Verkehrsordnung nicht weiterhilft. Beispiel: Das Auto steht im Stau und von hinten kommt offensichtlich ein schnelles Fahrzeug. Es droht ein Auffahrunfall. Soll es nun nach vorne und ganz dicht auffahren, um dem auffahrenden Auto mehr Platz zum Bremsen zu geben? Was ist, wenn das nicht reicht? Soll es ganz schnell nach rechts oder links ausscheren? Dabei eventuell einen hoffentlich kleinen Schaden in Kauf nehmen, um größeren Schaden zu verhindern? Jeder Mensch wünscht, dass es den möglichen Schaden minimiert. Was bedeutet das aber? Kann es bedeuten, dass es die Insassen opfert, um noch größeren Schaden zu vermeiden?

Um dieses und ähnliche Themen lässt sich trefflich streiten. Letztlich läuft es aber immer darauf hinaus, dass der Roboter auch ethische Regeln und nicht nur Verkehrsregeln befolgen soll. Damit die ethischen Regeln nicht zu einer Maschinen-Ethik verkommen, sondern dem Menschen nützen, sollen die Regeln unserer menschlichen Ethik folgen. Damit ist eine Ethik gemeint, die wir Menschen uns gegeben haben, um gut miteinander zusammenleben zu können.

Ethik für die KI

Um die Problematik zu verdeutlichen, ein weiteres Beispiel aus dem Bereich der Pflege, wo bereits heute Roboter eingesetzt werden. Soll der Roboter darauf achten, dass der Patient regelmäßig seine Medikamente einnimmt? Soll er ihn dazu zwingen? Bei lebenswichtigen Medikamenten kann das zweifellos richtig sein. Was aber ist mit Medikamenten, die nur dazu dienen, dass der Patient sich wohlfühlt? Wenn der Patient lieber auch mal auf die Einnahme von weniger wichtigen Medikamenten verzichtet, weil es ihm halbwegs gut geht, wie soll sich der Roboter verhalten? Bei diesem Dilemma wird deutlich, dass die Ethik prinzipiell situationsabhängig ist und dass sie viele Ziele mit unterschiedlicher Priorität zu verfolgen hat.

Die Ethik relativiert (situationsabhängig) die vorgegebenen Ziele und bringt sie in eine Prioritäts-Reihenfolge. Wir Menschen haben Ethik im Laufe unseres Lebens ge-

lernt, ohne sie in klar definierte Regeln zu gießen. Viele große Philosophen haben Jahrhunderte lang darüber nachgedacht, ohne ein sauber definiertes Regelwerk wie etwa die Physik zu hinterlassen. Genau das aber benötigen wir jetzt für die Technik der KI. Noch ist sogar unklar, ob man Ethik überhaupt in ein endliches, sauber definiertes Regelwerk fassen kann. Ist es eine gute Idee, dafür Fuzzy Logic einzusetzen?

Ethik technisch fassbar machen

Kann man ethisches Handeln einer KI durch Lernen beibringen? Schließlich bedeutet es eine Prioritätsreihenfolge von Zielen, die situationsabhängig ist. Welche ethischen Regeln soll die KI lernen? Wie soll sie die Regeln lernen? Soll sie es aus dem Verhalten von Menschen lernen, auch wenn diese immer wieder eine gewisse Schiefelage zeigen? Beispiel: Bei der Begnadigung von Menschen, die eine Gefängnisstrafe in den USA abzusitzen haben, hat sich gezeigt, dass grundsätzlich Weiße besser beurteilt werden als Schwarze. Wie kann man solche Schiefolgen vermeiden?

Diese und ähnliche Themen behandelt der jährliche Kongress Ethik und KI, der von der Integrata-Stiftung für humane Nutzung der IT (*siehe „www.integrata-kongress.de/kongress/2017-euki“*) gemeinsam mit dem Weltethos-Institut Tübingen und der Giordano-Bruno-Stiftung jährlich im Herbst durchgeführt wird. Das Kongresskonzept vertraut auf die Weisheit der Vielen und richtet sich bewusst an die Zivilgesellschaft und nicht an die Wissenschaft. Schließlich müssen die ethischen Regeln für die KI eines Tages von allen Menschen akzeptiert werden. Daher ist es gut, wenn möglichst viele Menschen dabei mitmachen.

*Michael Mörike
michael.moerike@integrata-stiftung.de*