

KÜNSTLICHE INTELLIGENZ Was kann alles schiefgehen, wenn Computer über Menschen entscheiden? Die Informatikerin Katharina Zweig über Algorithmen im Gerichtssaal, programmierte Fairness und durchdrehende Staubsaugerroboter.

In algorithmischer Sippenhaft

Frau Professorin Zweig, in Zukunft werden Computer immer häufiger über Menschen entscheiden.

Wo tun sie das jetzt schon?

Ein Beispiel, mit dem wir alle regelmäßig zu tun haben, ist die Suche im Internet. Wir geben einen Begriff in die Suchmaschine ein, und diese entscheidet, welche Ergebnisse uns am ehesten interessieren. Oder große Onlineshops wie Amazon empfehlen uns Produkte, die zu uns passen.

Welche Technik steckt dahinter?

Die Grundlage solcher Suchmaschinen sind so genannte Algorithmen. Das sind Sequenzen von Handlungsanweisungen für einen Computer. Die Aufgabe, die der Algorithmus lösen soll, lautet: Gib mir die passenden Informationen zu einer bestimmten Frage, und sortiere sie in einer sinnvollen Reihenfolge! Manche dieser Algorithmen sind personalisiert, das heißt, sie entscheiden, welche Informationen für eine bestimmte Person am relevantesten sind. Dafür nimmt der Computer die Daten, die er bisher über sie gesammelt hat – etwa die Suchhistorie oder das bisherige Kaufverhalten –, um daraus eine individuelle Empfehlungsliste zu erzeugen.

Solche Empfehlungen können Nutzer theoretisch auch ignorieren. Gibt es maschinelle Entscheidungen, denen wir nicht so leicht ausweichen können?

Ein Beispiel stammt aus den USA. Dort werden Algorithmen oft in Gerichtsprozessen eingesetzt. Sie sollen etwa beurteilen, ob ein Gefängnisinsasse frühzeitig ent-

lassen werden kann, also wie hoch die Wahrscheinlichkeit ist, dass er rückfällig wird.

Wie tun sie das?

Im ersten Schritt erhält der Computer die Daten von Tausenden Kriminellen, beispielsweise das Alter, die Art der bisherigen Straftaten, Informationen zur Herkunft und zu ihren Einstellungen. Zusätzlich erfährt er, welche dieser Personen innerhalb von zwei Jahren wieder rückfällig wurden. Computer sind deutlich besser als Menschen darin, riesige Datenmengen nach Zusammenhängen zu durchforsten: Welche Eigenschaften treten besonders häufig bei rückfällig gewordenen Straftätern auf und welche bei solchen, die erfolgreich resozialisiert wurden? Aus den Ergebnissen dieser statistischen Berechnungen baut der Algorithmus dann selbst eine Entscheidungsregel.

Das heißt, der Algorithmus zieht Schlüsse über einen ihm bisher nicht bekannten Menschen?

Genau! Das Computerprogramm bestimmt, ob dieser Mensch Eigenschaften hat, die eher denen der Wiederholungstäter ähneln oder eher denen der nicht rückfällig gewordenen Personen. Er erhält schließlich einen Wert zwischen eins und zehn, wobei die größten Werte ein besonders hohes Rückfallrisiko bedeuten. Das Problem ist, dass der Algorithmus die Daten von Herrn Schmidt und Herrn Müller nimmt, um vorherzusagen, ob Herr Meier rückfällig werden wird. Bei Amazon beispielsweise ist das anders. Hier berechnet der Computer



FELIX SCHMITT: WWW.FELIXSCHMITT.COM; MIT FRODL-GEN. VON KATHARINA ZWEIF

KATHARINA ZWEIF

ist Professorin für Informatik und leitet an der Technischen Universität Kaiserslautern das Algorithm Accountability Lab. Nach dem Studium der Biochemie und der Bioinformatik sowie einer Promotion in Tübingen über komplexe Netzwerke ging sie als Postdoc nach Budapest und Heidelberg. 2012 wechselte sie zur TU Kaiserslautern, wo sie den deutschlandweit ersten Studiengang »Sozioinformatik« schuf, der die Auswirkungen der Digitalisierung auf die Gesellschaft untersucht. Sie ist Mitglied der Enquete-Kommission des Bundestags zum Thema »künstliche Intelligenz«. Neben zahlreichen anderen Auszeichnungen erhielt sie den Communicator-Preis 2019 für ihre engagierte Kommunikation zu der Entwicklung, dem Einsatz und den gesellschaftlichen Auswirkungen von Algorithmen.

die Empfehlungen basierend auf Ihrer eigenen Kaufhistorie. Das halte ich für fair. Bei der Rückfallprognose jedoch bestimmt man das persönliche Risiko eines Straftäters aus dem Verhalten anderer Menschen, die ihm in einigen Aspekten ähneln. Ich nenne das daher »algorithmische Sippenhaft«.

Würde ein Mensch anders entscheiden als der Algorithmus?

In Deutschland unterhalten sich Psychologen oder Psychiater mit den Straftätern, um etwa herauszufinden, ob diese in die Sicherheitsverwahrung müssen. Der große Unterschied zu dem algorithmischen Verfahren besteht darin, dass ein menschlicher Gutachter jederzeit

beschließen kann, eine andere Frage zu stellen. Der Computer ist auf einen Pool von beispielsweise 120 Eingaben beschränkt. Er kann nicht plötzlich entscheiden, bei einem bestimmten Punkt noch einmal genauer nachzuhaken. Es fehlt die Flexibilität in der individuellen Prognose.

Ein Psychologe würde den Job also besser machen?

Das kann man so pauschal nicht sagen. Der Vorteil des Algorithmus ist es nämlich, dass er vollkommen objektiv, spricht bei derselben Sachlage immer gleich entscheidet. Das Urteil eines menschlichen Gutachters ist dagegen leicht zu beeinflussen. Es hängt unter anderem davon ab, ob er gerade gegessen, ob er sich über irgendetwas geärgert oder ob seine Ehefrau ihm vor der Arbeit noch einen Abschiedskuss gegeben hat. Das alles spielt beim Computer keine Rolle. Er stellt sicher, dass zwei Personen mit exakt den gleichen Eigenschaften auch gleich behandelt werden.

Gibt es Daten dazu, wie oft sich der Computer irrt?

Ich habe mit meiner Arbeitsgruppe an der TU Kaiserslautern untersucht, wie viele Personen, die in einer Testvorhersage zum Zeitpunkt der Entlassung in der Hochrisikogruppe landeten, nachher wirklich rückfällig wurden. Die Ergebnisse sind leider nicht besonders berauschend. Bei den allgemeinen Straftaten waren es etwas mehr als 70 Prozent, bei den Gewalttaten aber nur etwa jeder Vierte. Das heißt, drei Viertel der auf diese Weise abgestempelten Personen befanden sich zu Unrecht in der Gruppe. Das finde ich bedenklich. Denn das Problem ist: Richter haben keine mathematische Ausbildung. Für viele bedeutet der Begriff »Hochrisikogruppe«, dass jeder darin tatsächlich ein hohes Rückfallrisiko hat.

Was wäre Ihr Lösungsvorschlag?

Wir brauchen einen wissenschaftlichen Prozess der Evaluation. Weder gibt es Daten dazu, wie gut Menschen in solchen Situationen geurteilt haben, noch dazu, wie sich die Entscheidungen verbessern, wenn Maschinen nachhelfen. Solange diese Daten nicht vorliegen, halte ich nicht viel davon, solche Algorithmen im Rechtssystem zu verwenden.

Sollten Maschinen grundsätzlich nicht über die Zukunft von Menschen entscheiden?

Es gibt auch positive Beispiele dafür, wie man so eine Unterstützung durch Algorithmen transparent machen und auf ihren Effekt hin evaluieren kann, etwa in Österreich. Das dortige Pendant zu unserem Arbeitsamt, der Arbeitsmarktservice AMS, testet zurzeit einen Algorithmus, der Arbeitslose einer von drei Gruppen zuordnen soll. Die eine umfasst Personen, die auch ohne Weiterbildungsmaßnahmen ganz schnell wieder einen Job finden, die andere solche, die schon sehr lan-

TV-Tipp

»scobel – Willkommen in der Arbeitswelt 4.0«:

Katharina Zweig im Gespräch mit Gert Scobel und weiteren Experten im Rahmen des »NeuroForum Frankfurt 2019« der Gemeinnützigen Hertie-Stiftung.
Auf 3sat am 7. 11. 2019 um 21 Uhr

3sat

scobel

ge keine Arbeit mehr hatten und wohl auch in Zukunft so schnell keine Stelle bekommen werden, und eine dritte, mittlere Gruppe alle übrigen Arbeitslosen. Nur Letztere sollen unterstützende Maßnahmen erhalten, in der Hoffnung, dass sie am meisten davon profitieren. Denn das Land hat nur eine bestimmte Menge Geld dafür zur Verfügung und möchte es möglichst sinnvoll einsetzen.

Was zeichnet dieses Verfahren aus?

Die algorithmische Vorhersage in diesem Pilotprojekt wird gleichzeitig evaluiert, denn glücklicherweise kann man recht einfach feststellen, ob jemand wieder eine Stelle gefunden hat. Die Methode des maschinellen Lernens, die dort verwendet wird, ist darüber hinaus sehr simpel, und die vom Computer gelernten Entscheidungsregeln sind leicht nachzuvollziehen. Zudem hat der Entwickler, Michael Wagner-Pinter, eine Reihe von Sozialverträglichkeitsregeln entworfen. So muss ein Urteil der Maschine immer mit den Betroffenen besprochen werden, und die AMS-Mitarbeiter können die Kategorisierung per Hand ändern. Ich bin sehr gespannt, wie das Experiment ausgeht und ob der AMS den Algorithmus dauerhaft einsetzen wird.

Wonach entscheidet der Algorithmus, in welcher Kategorie ein Arbeitsloser landet?

Es sind nur zwei Hand voll, zum Beispiel ob jemand die Matura hat, wie alt er ist und wie oft die Person schon beim AMS vorstellig war. Als ich das hörte, war ich überrascht. Wenn ich mir persönlich meinen Lebenslauf ansehe und mir vorstelle, es gäbe weniger als ein Dutzend Kriterien, nach denen ich beurteilt würde, wäre ich skeptisch. Michael Wagner-Pinter sagte mir aber, die Personen seien vorher im Wesentlichen nach nur einem einzigen Kriterium beurteilt worden, und zwar danach, wie lange sie schon mit dem AMS in Kontakt waren. Natürlich könnten wir uns auch 100 Eigenschaften ausdenken. Doch so viele Arbeitslose gibt es in Österreich gar nicht, um für alle Merkmale genügend Daten zu bekommen, die sich statistisch auswerten lassen.

Sollte es eine Art TÜV für solche Algorithmen geben, die über Menschen entscheiden?

Die Bezeichnung ist irreführend. Denn die Handlungs-vorschriften für die Computer sind selten das Problem. Nehmen wir den AMS-Algorithmus: Man kennt eine Menge Eigenschaften von Zehntausenden von Arbeitslosen und weiß, ob sie wieder einen Job gefunden haben oder nicht. Das Programm sucht in den Daten nach Korrelationen, also danach, welche Merkmale in welcher Gruppe besonders häufig auftreten. Die Methode ist trivial, sie benötigt keinen TÜV. Das Ergebnis dieser Mustersuche ist schließlich eine Formel. Nun gibt es einen zweiten Algorithmus, der die Daten der aktuellen Arbeitslosen in die Formel einfütert und das Ergebnis berechnet. Auch daran ist nichts Geheimnisvolles; das ist einfachste Mathematik und muss nicht durch einen TÜV begutachtet werden. Den gesamten Prozess nennen wir »algorithmisches Entscheidungssystem«. Es besteht aus den Daten, von denen der Computer lernt, der Methode, mit der er lernt, und dem daraus entstehenden statistischen Modell, in diesem Fall der Formel. All das gemeinsam zu überprüfen, macht jedoch durchaus Sinn.

Warum? Was kann dabei schiefgehen?

Es kann zum Beispiel sein, dass das Lernmaterial, also die Daten der Vergangenheit, nicht ausgewogen ist. Wenn etwa bestimmte Bevölkerungsgruppen zu schwach vertreten sind, um statistische Aussagen über sie zu machen.

Sie sagten, bei dem AMS-Projekt urteilen die Algorithmen nicht allein. Ein Mitarbeiter prüft die Ergebnisse und bespricht sie mit dem Arbeitslosen. Mindert das ein wenig das Gefühl, der »Willkür« einer Maschine ausgeliefert zu sein?

Ganz bestimmt. Dennoch ist es auch hier wichtig zu überprüfen, ob der Mensch zusammen mit dem Computer wirklich bessere Entscheidungen trifft. Dafür braucht man eine hohe Transparenz und Methoden, die es Experten erlauben, die Vorgänge nachzuvollziehen. Das sollte insbesondere für Anwendungen des Staates gelten, da dieser eine Monopolstellung hat.

Wer übernimmt die Verantwortung, wenn eine Maschine mal einen falschen Entschluss fasst?

Oft wird von künstlicher Intelligenz geredet, als wäre sie ein Wesen. Aber das ist sie nicht. Wenn jemand einen Algorithmus nutzt, um über andere Menschen zu urteilen, liegt die Verantwortung immer bei dem, der ihn einsetzt. Er muss zudem sicherstellen, dass die Mitarbeiter ausreichend geschult sind, um zu wissen, was die Entscheidung bedeutet.

Kann man bei künstlicher Intelligenz überhaupt von »Entscheidungen« reden? Wie intelligent ist sie wirklich?

Tatsächlich werden in der öffentlichen Diskussion häufig zwei Arten von künstlicher Intelligenz (KI) durcheinandergebracht. Womit wir gerade zunehmend zu tun haben, ist eine so genannte schwache KI. Man baut sie für eine bestimmte Aufgabe, und die erledigt sie oft besser als Menschen. Sie ist aber nicht im eigentlichen Sinn intelligent. Der Begriff entstand in den 1950er Jahren, als Wissenschaftler Geld für ihre Forschung eintreiben wollten. Sie dachten sich, »artificial intelligence« klingt nach etwas, was der Staat gerne fördert. Und jetzt hängen wir an diesem Namen. Die meisten Computerwissenschaftler finden ihn unpassend.

Und die zweite Art von künstlicher Intelligenz?

Eventuell wird man irgendwann einmal eine so genannte starke KI erschaffen. Sie würde sich dann eigene Ziele setzen können und jede Art von Problem in Angriff nehmen. Experten streiten darüber, ob das überhaupt möglich ist.

Was denken Sie?

Ich bin mir da nicht sicher. Mein Kollege Jürgen Schmidhuber glaubt, dass man einer starken KI eine Art Neugier einpflanzen kann – eine Optimierungsfunktion als inneren Kompass. Eine solche Maschine würde dann versuchen herauszukriegen, wie die Welt funktioniert, ähnlich wie Kinder. Sie würde ein Modell bilden mit Regeln darüber, wie sich die Dinge verhalten, etwa dass Gegenstände herunterfallen, wenn man sie loslässt. Auf Grund der Neugier bringt sich die KI immer wieder in neue Situationen, um etwas zu lernen. Wenn sie etwas findet, was ihr Weltmodell bislang nicht enthält, bekommt sie eine »Belohnung«. Und sie ist stets auf der Suche nach neuen Regeln, die allgemeiner sind als die alten.

Tun wir Menschen das auch?

Ja, wir versuchen immer, das grundlegende Konzept zu entdecken. Kleine Kinder merken beispielsweise durch Ausprobieren, dass ein Ball herunterfällt, aber ebenso ein Apfel und eine Banane. Irgendwann bilden sie die allgemeine Regel: Alle Dinge fallen auf den Boden, wenn ich sie loslasse. Und dann kommt die Neugier hinzu; vielleicht gibt es ja Dinge, bei denen das anders ist. Und siehe da, ein heliumgefüllter Ballon steigt nach oben. Das Weltmodell wird nun um eine Regel ergänzt. Nach weiterem Experimentieren kommt irgendwann die allgemeine Erkenntnis, dass es von der Dichte des

Gegenstands abhängt, ob er nach oben steigt oder herunterfällt. Könnte man so etwas nachbilden, wäre das nach Schmidhuber ein Weg zu einer starken KI.

Sollte man das überhaupt versuchen? Immer wieder ist von den Gefahren einer unkontrollierbaren KI die Rede.

Die Optimierungsfunktionen von Menschen sind unter anderem ihre Emotionen. Schmerz, Freude, Trauer et cetera lehren uns, was richtig und was falsch ist, was wir suchen und was wir vermeiden sollten. So etwas in Maschinen umzusetzen, ist nicht so einfach und kann leicht schiefgehen.

Was kann etwa passieren?

Dazu gibt es eine schöne Anekdote: Jemand hat versucht, seinen Staubsaugerroboter mit Hilfe eines künstlichen neuronalen Netzes zu tunen. Auch dieses besitzt eine Optimierungsfunktion, die bewirkt, dass gewünschtes Verhalten »verstärkt«, also wiederholt, und unerwünschtes »bestraft« und so mit der Zeit abgeschaltet wird. Der Roboter sollte lernen, schneller zu fahren, ohne dabei in die Möbel zu rauschen. Das neuronale Netz gab somit eine Strafe, wenn der Sensor eine Kollision mit einem Gegenstand bemerkte, und einen Bonus, wenn der Staubsauger schön schnell fuhr. Während der Besitzer den Tag über bei der Arbeit war, sollte die künstliche Intelligenz das neue Verhalten lernen. Doch als er abends zurückkam, raste der Staubsauger in der Wohnung herum und rammte dabei gegen alles, was ihm im Weg stand. Was war passiert? Der Roboter kann rückwärtsfahren, hat hinten aber keinen Sensor. So hat er gelernt, rückwärts besonders schnell zu fahren, ohne jemals bestraft zu werden. Das ist eine sehr wichtige Lektion.

Was lernen wir aus solchen Beispielen?

Für alles, was wir modellieren, brauchen wir meistens mehrere Versuche. Bei einer starken künstlichen Intelligenz jedoch hätte man unter Umständen nur einen. Deshalb sollten wir gar nicht erst versuchen, eine solche zu entwickeln. Die Wahrscheinlichkeit, dass wir diese globale Optimierungsfunktion auf Anhieb so hinkriegen, dass die KI die gleichen Dinge für gut befindet wie wir Menschen, halte ich für äußerst gering. ★

Die Fragen stellte »Gehirn&Geist«-Redakteurin Anna von Hopffgarten.

LITERATURTIPP

Zweig, K.: Ein Algorithmus kennt kein Taktgefühl. Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können. *Heyne*, 2019.
Mit viel Witz und anhand einfacher Beispiele erklärt die Autorin, wie Algorithmen funktionieren, welche völlig harmlos sind und welche uns tatsächlich Sorgen bereiten sollten.

Dieses Interview im Internet: www.spektrum.de/artikel/1678242