MERTON
Onlinemagazin des *Stifterverbandes*

INNOVATION

Foto: Jantine Domboos via Unsplash.com

# Auch Algorithmen können diskriminieren

VERÖFFENTLICHT AM 16.01.2018

Je mehr Algorithmen mit Daten gefüttert werden, desto besser können sie Regeln und Muster entwickeln. Diese sind für die menschliche Entscheidungsfindung nützlich. Allerdings sollte man sich nicht bedingungslos auf sie verlassen, warnt unsere Gastautorin.

LESEZEIT: 8 MINUTEN

TEXT:
KATHARINA ZWEIG ›

DER FOLGENDE TEXT STEHT UNTER CC-LIZENZ
CC BY 2.0 DE

Es wirkt so, als wollte uns jeder nur zu gerne das nervtötende Personalmanagement abnehmen: „No more people problems“, so wirbt Cream HR in Großbritannien mit seinem Outsourcing-Service. Und ein **junges Startup** begeisterte in der Fernsehsendung „Die Höhle der Löwen“ die Investoren mit dem Versprechen: „Nur noch halb so viele Bewerbungsgespräche führen!“ Die Gründer hatten eine App entwickelt, die sich dank eines spontan geführten Webvideointerviews sofort einen persönlichen Eindruck von den Kandidaten verschafft; wenn das nicht passt, muss man die Person gar nicht erst einladen. Es scheint, als würde jede Form von Leistungsbewertung – sei es bei Bewerbungen, vor Mitarbeitergesprächen oder bei Beförderungen – Firmen vor Probleme stellen. Eine solche Bewertung darf natürlich nicht diskriminierend sein und sollte gleichzeitig weder zu selektiv noch zu undifferenziert vorgenommen werden. Sind die Experten aus unseren Personalabteilungen und die direkten Vorgesetzten dafür wirklich am besten geeignet?

Wenn man sich die Bestseller der vergangenen Jahre von **Dan Ariely** bis **Daniel Kahnemann** einmal näher ansieht, dann ist wohl klar, dass Menschen vor allen Dingen eins sind: irrationale Wesen mit mangelnder Entscheidungskraft. Könnten es da nicht Computer besser machen? Was hat es auf sich mit diesen Algorithmen, die versprechen, aus großen Datenmengen diejenigen Regeln zu lernen, anhand derer man die erfolgreiche Kandidatin vom Loser trennt oder den aufstrebenden Jungmanager aus einer Reihe von jungen Talenten siebt? Die Algorithmen der künstlichen Intelligenz imitieren das Lernen von Kleinkindern, die versuchen, die Regeln unserer Sprache zu entdecken: Unsere Kinder hören zu, extrahieren daraus Muster, machen Versuche, selbst zu sprechen, und binden das Feedback ihrer Umgebung in ein Update ihrer Regeln ein. Welchen Algorithmus sie dafür verwenden, wissen wir momentan nicht genau. In jedem Fall benötigen sie jede Menge „Datenpunkte“: geduldige Eltern und Großeltern, die mit ihnen in einfachen Sätzen sprechen – dies zeigt ihnen, welche Sätze richtig sind. Spätere Korrekturen der eigenen Sprachversuche verdeutlichen ihnen, welche Sätze noch nicht richtig sind: Sie bekommen ein Feedback zu dem bisher Gelernten. Sie benötigen auch eine Struktur, in der das Gelernte abgespeichert werden kann – das ist ihr Gehirn mit den sich vernetzenden Gehirnzellen.

Die moderne Informatik hat zahlreiche Verfahren entwickelt, die auf diesen vier Aspekten beruhen:

- einer großen Menge an Datenpunkten, zum Beispiel von Bewerbern der vergangenen zehn Jahre,
- einer Information darüber, wer von diesen Personen erfolgreich eingestellt werden konnte (Die große Datenmenge wird also in zwei Gruppen eingeteilt, in „erfolgreiche Bewerber“ und „nicht erfolgreiche Bewerber“. Diese Information stellt das Feedback dar, das dem System sagt, was falsch und was richtig ist.);
- auf der darauf aufbauenden Suche des Algorithmus nach Regeln, welche die erfolgreichen von den nicht erfolgreichen Bewerbern trennen helfen (Meistens beruhen diese Regeln auf reinen Korrelationen, wie etwa: „Unter den in der Vergangenheit erfolgreichen Bewerbern waren überdurchschnittlich viele mit einem BWL-Abschluss, die mindestens vier Monate im Ausland waren.“ Solche Regeln besagen natürlich nicht, dass dies notwendige Bedingungen sind, aber dass eben solche Personen später übermäßig häufig erfolgreich eingestellt werden konnten.);
- der Speicherung der Regeln in unterschiedlicher Form: als mathematische Formel, als Entscheidungsbaum oder in den viel diskutierten „neuronalen Netzen“.

## FEHLERFREIE ROBO-RICHTER?

Die Chancen künstlicher Intelligenz sind ohne Frage vielfältig. Die vergangenen Jahre haben gezeigt, dass Computer damit drastisch besser wurden in der Übersetzung von einer Sprache in die andere, sie besiegten uns erst im Schach und dann im japanischen Brettspiel Go und werden uns bald das chauffeurlose Fahren in autonom gesteuerten Fahrzeugen erlauben. Sie sind – ganz im Gegensatz zu menschlichen Richtern[1] – nicht von der Tagesform abhängig und verrechnen sich so gut wie nie. Ist das ausreichend, um aus ihnen objektive Robo-Richter zu machen, deren Entscheidungen fehlerfrei sind?

Natürlich nicht. Meine Forschungen an sogenannten Rückfälligkeitsvorhersagealgorithmen aus den USA zeigen, dass es eine ganze Reihe von Dingen zu bedenken gibt. Diese Algorithmen wurden entwickelt, um Menschen aus einer Untersuchungshaft zu entlassen, die kaum ein Risiko zeigen, rückfällig zu werden, oder um knappe Ressourcen nach einer Haftentlassung an diejenigen zu verteilen, die am stärksten rückfallgefährdet sind. Die Algorithmen arbeiten genauso wie die oben erwähnten zur Erkennung von erfolgreichen Bewerbern: Sie nehmen Daten aus den letzten Jahren, die Personen beispielsweise anhand eines umfangreichen Fragebogens charakterisieren und zudem die Information enthalten, ob diese innerhalb der zwei Jahre nach Verurteilung wieder rückfällig geworden sind oder nicht. Die Algorithmen extrahieren daraus Regeln, welche die schon bekannten Kriminellen in eine von zehn Kategorien einteilen, wobei die unteren Kategorien – als Gruppe gesehen – für diejenigen stehen, die selten rückfällig wurden, die mittleren Kategorien für Kriminelle, die ungefähr zur Hälfte wieder kriminell wurden, und die obersten Kategorien für all jene, die eine Rückfallquote bis zu 80 Prozent aufwiesen. Verschiedene NGOs haben sich dafür eingesetzt, dass solche algorithmischen Entscheidungssysteme entwickelt wurden, um eine objektive Behandlung aller Kriminellen nach denselben Maßstäben zu gewährleisten.

> Computer schätzen afroamerikanische Kriminelle nachweislich viel zu oft als „hoch rückfallgefährdet“ ein und weiße Kriminelle viel zu selten.

**KATHARINA ZWEIG**

Katharina Zweig (Foto:privat)

## VERMEINTLICHE OBJEKTIVITÄT

Aber der Stand der Dinge ist, dass sie trotz ihrer vermeintlichen Objektivität auch diskriminierend sein können, wie der gemeinnützige journalistische Thinktank ProPublica zeigte: Die Computer schätzen afroamerikanische Kriminelle nachweislich viel zu oft als „hoch rückfallgefährdet“ ein und weiße Kriminelle viel zu selten. Damit wird die Last der „falschen Verurteilung“ ungleichmäßig auf die verschiedenen Volksgruppen verschoben. Auch wenn die Frage nach der Fairness hoch kompliziert und spannend ist, zeigt sie doch nur eine Facette der Risiken von algorithmischen Entscheidungssystemen: Im Verlauf ihrer Entwicklung muss eine Vielzahl von kritischen Entscheidungen von einer großen Menge unterschiedlicher Entscheider getroffen werden. Diese Verteilung auf viele Schultern in einem noch nicht hinreichend qualitätsgesicherten Prozess kann leicht dazu führen, dass die vom „objektiven“ System getroffenen Entscheidungen suboptimal sind: Aus alten Daten können Diskriminierungen mitgelernt werden, fehlerhafte Daten können zu fehlerhaften Entscheidungsregeln führen, der gewählte Algorithmus und die Art, in der er die Regeln extrahiert, können aus mathematischen Gründen unpassend sein – und nicht zuletzt kann die Verwendung des Algorithmus im betrieblichen Alltag zu ungewollten Nebeneffekten führen. Wer traut sich schon, bei anderslautendem Computerbefehl seiner Intuition zu folgen und eventuelle Fehlentscheidungen zu verantworten, wenn man sich stattdessen bequem hinter Kollege Chip verstecken kann? Es zeigt sich also schnell, dass wir Computer mindestens so sorgfältig ausbilden müssen wie menschliche Entscheider; sie sind keine magischen Wunderwaffen, die eine komplexe Situation in einer einzigen Zahl abbilden können.

Doch selbst wenn sie optimal gebaut wurden, muss klar sein, dass ein algorithmisches Entscheidungssystem am besten die Eigenschaften erfolgreicher Bewerber und Arbeitnehmer entdeckt, die häufig sind. Wollen Sie in einem Betrieb leben, in dem alle Mitarbeiter denselben „optimalen“ Lebenswegen gefolgt sind, bevor sie angestellt wurden? Was ist mit dem schrägen Vogel im Vertrieb, der vielleicht nicht so viel arbeitet, aber die ganze Abteilung aufheitert und jeden Geburtstag kennt? Was ist mit der Japanologin, die heute Ihre beste Produktmanagerin ist, oder mit dem Handwerker, der sich über Meisterschule und Teilzeitstudium im Alter von 45 mit Herzblut einbringt? Der Algorithmus wird alle diese Exoten nicht entdecken können.

Es gibt einen goldenen Mittelweg zwischen der menschlichen, subjektiven Entscheidung und der datengetriebenen Computerentscheidung, der vielversprechend erscheint: die konsequente, datengefütterte Entscheidung von menschlichen Experten. Dazu müssen Sie zuerst festlegen, woran man eine gute Entscheidung erkennt, und messen, ob momentan überhaupt schlechte Entscheidungen getroffen werden.

Katharina Zweig (Foto:privat)

Katharina Zweig ist Professorin am Fachbereich Informatik der Technischen Universität

Gibt es gemessen an den erfolgreich Eingestellten Personalmanager, die ein goldenes Händchen bei der Bewerberauswahl haben? Chefinnen, die massenweise junge Talente hochziehen, die dann gestärkt ins globale Firmengeschehen eingreifen? Was macht diese Personalentscheider erfolgreich? Auch das kann ein Algorithmus lernen und damit den menschlichen Entscheidern den Rücken stärken. Und das, ohne das falsche Versprechen zu geben, die Komplexität der Welt auf eine einzige Zahl zu reduzieren.

*[1] Danziger, S.; Levav, J.; Avnaim-Pesso, L.: Extraneous factors in judicial decisions, in: Proceedings of the National Academy of the Sciences, 2011, 108, S. 6889–6892.*

Kaiserslautern. Sie ist Mitgründerin der NGO **AlgorithmWatch** und Studiengangskoordinatorin des deutschlandweit einmaligen Sozioinformatik-Studiengangs an der Technischen Universität Kaiserslautern. 2017 erhielt sie den vom Stifterverband und vom Dachverband der Fakultätentage der Ingenieurwissenschaften und der Informatik 4ING vergebenen **Ars legendi-Fakultätenpreis** für exzellente Hochschullehre in den Ingenieurwissenschaften und der Informatik.

**QUELLE: HTTPS://MERTON-MAGAZIN.DE/AUCH-ALGORITHMEN-KOENNEN-DISKRIMINIEREN**