

AUSSENANSICHT

Der Arzt kann einpacken

Warum man mit Mathematik Menschen heilen kann, ohne etwas von Medizin zu verstehen – und Big Data besser ist als sein Ruf. *Von Christian Hesse*

Big Data – der Begriff ist verbunden mit dem NSA-Spionageskandal und der Angst vieler Menschen, einer anonymen, bösen Macht ausgeliefert zu sein. Ja: Im weltweiten Netz haben sich auf fast jedem Gebiet riesige Datenmengen angesammelt. Diese Datenmengen können missbraucht werden. Noch viel mehr aber nützen sie. So hat Barack Obama mit Big Data die jüngste US-Präsidentenwahl für sich entschieden. Sein Team aus 150 Datenexperten hatte über jeden Wahlberechtigten mehr als 1000 Datenpunkte gesammelt – Alter, Geschlecht, Wohnbezirk, Clubmitgliedschaften, Restaurantbesuche, medizinische Behandlungen, sowie wer, wann, wo, was mit welcher Kreditkarte bezahlt hat. Die subtile Analyse dieser Daten erlaubte seiner Mannschaft das so genannte microtargeting, die gezielte Ansprache des Wahlvolks auf der entscheidenden Ebene: der des einzelnen Wählers. Statistisch wusste man sehr weitgehend, „wen die Leute wählen würden, noch bevor sie es selber wussten.“ Ist es eine Schande zu wissen, was der Wähler will?

Man kann mit Big Data sehr erfolgreich sein in der Welt – und zwar, ohne die komplexe Welt zu verstehen. Und das ist ein Fortschritt. Nehmen wir Übersetzungen, also die Übertragung eines Textes von einer Sprache in eine andere. Ein menschlicher Übersetzer muss sehr genau Grammatik, Wortschatz und die vielen Feinheiten zweier Sprachen kennen. Das liegt an der gro-

ßen Zahl von Wörtern mit Mehrfachbedeutungen. Bei der Übersetzung eines Wortes kommt es meist nicht nur auf das Wort selbst an, sondern auch auf dessen Umfeld. Der Übersetzer muss also nicht nur die unzähligen Nuancen beider Sprachen beherrschen, sondern auch die inhaltliche Bedeutung des Textes erfassen. Kurzum, er muss diesen Ausschnitt der Welt einigermaßen gut verstanden haben. Sonst klingt, was er fabriziert, so krude wie eine Übersetzung aus dem Internet.

Inzwischen kann man aber auf hohem Niveau rein mit statistischen Methoden übersetzen. Man muss dafür nicht einmal die beteiligten Sprachen sprechen. Auch den zu übersetzenden Text muss man nicht verstehen. Man benötigt für diese Übersetzung keinen Sprachenexperten, der die Übersetzung selbst vornimmt, oder zusammen mit Programmierern einen Computer mit linguistischem Wissen und Sprachregeln füttert. Was man braucht, sind Daten, möglichst viele Daten. Und die gibt es. Benötigt wird ein reichhaltiger Pool mit übersetztem Text von der Aus-

gangs- in die Ziel-Sprache. Ein gutes Beispiel ist die Bibel. Das Neue Testament ist in mehr als 1000 Sprachen übersetzt. Auch die Dokumente internationaler Organisationen wie der EU oder der Uno sind eine reichhaltige Quelle mehrsprachiger Textbausteine. Dann natürlich das Internet: Jede Wortgruppe ist inzwischen von irgendjemandem irgendwo übersetzt worden, meist nicht nur einmal.

Man kann inzwischen auch mit rein statistischen Methoden Texte hervorragend übersetzen

Diese Paralleltexte bilden die Basis für die Übersetzung. Je häufiger eine Wortgruppe im Datenpool in einem bestimmten Umfeld auf eine bestimmte Art übersetzt worden ist, desto größer ist die Wahrscheinlichkeit, dass dies der Übersetzungswahrheit recht nahe kommt. Trifft die Software des Rechners bei einem englischen Text auf das Wort „bank“, so kann das sehr verschieden ins Deutsche übersetzt wer-

den. Ist in der Umgebung des Wortes von Wasser und Gewässern die Rede, steigt die Wahrscheinlichkeit, dass im vorliegenden Text „bank“ „Ufer“ bedeutet. Treten aber in der Nähe Begriffe wie Geld oder Geldanlage auf, wird das Wort wahrscheinlich in der Bedeutung von „Bank“ gebraucht. Das Umfeld jedes Wortes kann von einem Rechner anhand von statistischen Maßzahlen quantitativ erfasst werden.

Das ist im Prinzip alles, was man für eine gute Übersetzung braucht: Viel bereits übersetzten Text – und filigrane statistische Methoden der Datenanalyse. Linguistische Kenntnisse braucht man dagegen nicht mehr. Die sprachliche Intuition eines Übersetzers ist weitgehend überflüssig geworden.

Dieser Ansatz birgt ungeahnte Möglichkeiten in vielen Bereichen. Zum Beispiel in der Medizin. Um Kranken kompetent zu helfen, muss man nicht mehr genau verstehen, welche Krankheiten sie haben, welche Wirkungen diese Krankheiten auf bestimmte biochemische Abläufe ausüben, welche Störungen diese Wirkungen verur-

sachen, welche Symptome aus diesen Störungen resultieren und welche therapeutischen Maßnahmen bei diesen Symptomen einzusetzen sind. Der klassische Prozess von Diagnose und Therapie ist eine detaillierte, vielschrittige, fehleranfällige Abfolge. Dass es dabei viele Fehlbehandlungen gibt, liegt nicht einfach daran, dass die Ärzte schlecht sind. Es liegt an der hohen Komplexität dieses schrittweisen Prozesses.

Hier hilft „Big Data“, Menschen individuell, gezielt, kompetent und möglichst ohne Fehler zu helfen. Auch hier braucht man einen möglichst großen Datenpool von Informationen über das, was bei Menschen wie passiert ist und funktioniert hat, je nachdem, welchen Geschlechts und Alters sie sind, welche Blut-, Urin- und andere Messwerte bei ihnen erhoben wurden, welche Symptome sie haben. Ist die Datenbasis groß genug, wird der Computer auch hier durch microtargeting eine Therapie vorschlagen, die haargenau auf die persönliche Situation zugeschnitten ist – wahrscheinlich besser als die meisten Ärzte. Deren Datenbasis besteht aus angelesenen Lehrbuch-Informationen über den Durchschnittspatienten, ergänzt durch eigene Erfahrungen; das ist vergleichsweise spärlich und wird oft dem Kranken nicht gerecht, der da vor ihnen steht.

Die wenigsten Patienten dürften sich wünschen, dass ihr Hausarzt durch einen Computer ersetzt wird. Sie sollten sich aber auch nicht wünschen, dass ihr Arzt

auf die Unterstützung gut ausgewerteter Datenpools verzichtet. Der Arzt wird sich zu einem weitgehend computerabhängigen Helfer für Kranke hin entwickeln. Der Doktor herkömmlichen Stils ist dann etwas für die Romantiker unter den Patienten; der Computer wird zum Leibarzt werden. Er hat das entscheidende Know-how. Eine Horrorvision? Nein, im Gegenteil. Es werden weniger Fehler passieren. Es wird mehr Menschen gezielt mit den für sie entworfenen Therapien geholfen werden.

Die verantwortliche Nutzung riesiger Datenmengen schafft ungeahnte neue Möglichkeiten. Man kann Energie effizient nutzen, Verbrechen bekämpfen und verhindern, die Wissenschaft voranbringen. Big Data hat einen schlechten Ruf, doch die gesammelten Datenmengen sind nicht per se schlecht. Der Missbrauch von Daten muss natürlich konsequent bekämpft werden. Doch die Chancen von Big Data überwiegen die Risiken.



Christian Hesse, 53, lehrt Mathematische Stochastik in Stuttgart. Der passionierte Schachspieler hat unter anderem das Buch „Warum Mathematik glücklich macht“ geschrieben.

FOTO: OH