

Der Heuhaufen sucht mit



ZUKUNFTSPROGNOSEN FASZINIEREN DIE MENSCHHEIT SEIT JEHER. MITTELS PREDICTIVE ANALYTICS WERDEN KOMPLEXE ZUSAMMENHÄNGE BEWERTET UND IN DIE ZUKUNFT PROJIZIERT. DIESE MÄCHTIGE METHODE KANN VIELE PROBLEME LÖSEN UND GLEICHZEITIG NEUE SCHAFFEN. Von Ruth Reitmeier

Eine fiktive Szene in naher Zukunft: „Was hast Du denn da schon wieder bestellt?“, fragt sie ihren Home-Management-Roboter und deutet auf das Paket auf dem Küchentisch. „Das kam heute mit der Post. Und ich habe es nicht bestellt“, antwortet der Roboter. „Der Händler hat dir das Paket wohl in der Annahme geschickt, dass du diese Dinge brauchst. Die Soda-Patronen gehen zur Neige und das Reinigungsmittel für meine Sensoren ist fast leer.“ Sie wirft einen achtlosen Blick auf die Sendung. „Ich brauche dieses Zeug nicht! Und ich will es nicht. Schick das sofort zurück“, befiehlt sie dem Roboter. Der aber antwortet: „Du brauchst es vielleicht heute nicht, doch das System weiß besser als du selbst, dass du es schon morgen brauchen wirst.“ Sie grinst. „Ach so? Nun, das System hat keine Ahnung, denn ich werde morgen nur Leitungswasser trinken und dich, mein Lieber, schalte ich ab. Dann brauchst du auch kein Reinigungsmittel.“

DER ONLINEHÄNDLER KENNT DIE BESTELLUNG VON MORGEN SCHON HEUTE

Vielleicht werden wir uns schon bald durch derart bockige Verweigerung Handlungsraum in Konsumententscheidungen erkämpfen müssen. Der Online-Händler Amazon hat sich nach einem Bericht des Wall Street Journal bereits 2012 ein System des antizipatorischen Paketversands als Patent gesichert. Das Prinzip dahinter: Waren werden verschickt, noch bevor der Kunde sie bestellt hat. Kundendaten zu früheren Bestellungen, das Screening von Wunschzetteln und Warenkörben sowie die Verweildauer des Cursors auf Produkten im Shop liefern die Rohdaten. Diese werden nach Mustern abgesucht und mittels Algorithmen schließlich zu Bestell-

prognosen zusammengeführt. Dieses System wendet Predictive Analytics (PA) an. Während Datamining Muster in Datenbeständen erkennt, liefert PA zukunftsbezogene Auswertungen. Die Trennung zwischen den Begriffen ist unscharf. PA ist jedenfalls so etwas wie das Orakel unserer Zeit. Am Anfang steht klassisch eine Fragestellung. In der Glaskugel spielt sich dann vereinfacht Folgendes ab: Eine Hypothese wird mittels Datamining überprüft und daraus schließlich eine Vorhersage über die Zukunft getroffen. Es wird also nicht nur der Ist-Zustand erhoben, um den Menschen entscheiden zu lassen, sondern weitergerechnet – und erst danach ist der Mensch mit seiner Entscheidungskompetenz an der Reihe. Das Versprechen, das den mächtigen Formeln innewohnt, ist es, komplexe Zusammenhänge effizienter sowie zukunftsbezogen zu überblicken und fundierte, da evidenzbasierte Entscheidungen treffen zu können. Die andere Seite der Medaille ist die Horrorvision eines Datengaugen Mitte des 21. Jahrhunderts in einer Welt, in der sämtliche Informationen gesammelt und Lebensbereiche durchleuchtet werden, in der es keine Anonymität, keine Privatsphäre und keinerlei Freiräume mehr gibt. PA ist keine Zukunftsmusik, sondern findet bereits in zahlreichen Gebieten Anwendung.

EIN SUPERCOMPUTER SCHLÄGT MEDIZINISCHE DIAGNOSEN UND THERAPIEN VOR

Die sprichwörtliche Suche nach der Nadel im Heuhaufen hat im digitalen Zeitalter eine neue Dimension erreicht. „Der Heuhafen sucht mit“, sagt Datenschutz-Experte Christof Tschohl. Und das kann Leben retten. Der IBM-Supercomputer Watson etwa durchforstet Fachliteratur und klinische Daten, um etwa optimale, persona-

lisierte Behandlungsstrategien für Krebspatienten zu finden. Medizinische Suchmaschinen helfen beim Wissensmanagement, doch dies ist erst der Anfang. Mittels PA macht die Maschine auf der Basis der durchforsteten Literatur samt Arztbriefen und klinischen Daten konkrete Diagnose- und Behandlungsvorschläge. „Mittlerweile erscheinen rund eine Million medizinische Fachartikel pro Jahr“, betont Allan Hanbury vom Institute of Software Technology and Interactive Systems an der Technischen Universität Wien. Dass hier kein Mensch den Überblick behalten kann, versteht sich von selbst. Ein von Hanbury geleitetes Projektteam hat solch ein Suchsystem für die Radiologie entwickelt. Der Prototyp ist fertig und funktioniert – und zwar so: Zeigt die Röntgenaufnahme etwas, das der Radiologe noch nie gesehen hat, speist er das Bild in die Suchmaschine ein. Über Bilderkennung werden ähnliche Röntgenaufnahmen blitzschnell aus dem Archiv gehoben, mittels Textanalyse werden zusätzlich relevante Informationen aus Arzt-Berichten sowie klinischen Daten gefiltert und sodann die wichtigen Textstellen und Werte bereits markiert präsentiert. „Das System ist ein Tool für den Arzt, es kann den Arzt aber nicht ersetzen“, betont Hanbury, „Nur ein ausgebildeter Mediziner verfügt über genug Fachwissen, das Ergebnis zu interpretieren und gegebenenfalls zu verwerfen.“

GEN-SCREENING BEDEUTET ZU WISSEN, WAS EINMAL IN DER EIGENEN KRANKENGESCHICHTE STEHEN WIRD

Nicht nur in der Krebsforschung gewinnt das Genetic Sequencing – ein Screening der Gene – an Bedeutung. Nachdem allein ein einziges Gen enorme Datenmengen enthält, war dies

bisher sehr aufwendig und teuer. Mittlerweile liefern Hochleistungsrechner Ergebnisse deutlich schneller und billiger ab. Dies ist zweifellos ein wichtiger Fortschritt in der Behandlung von Krankheiten wie Krebs, die einen Lauf gegen die Zeit bedeuten. Die Methode eröffnet aber auch ganz andere Möglichkeiten: Etwa jene, bereits ab der Geburt zu wissen, welche Krankheiten einem Menschen bevorstehen könnten. In den USA wird aktuell in einem Modellversuch mit insgesamt 480 Babys Genetic Sequencing eingesetzt und dabei dessen Folgen erforscht. Das Projekt soll zeigen, welche Chancen, aber vor allem auch welche Risiken dieses Wissen birgt. In Boston entschlüsseln Ärzte dabei erstmals auch das Genom gesunder Neugeborener. Ihr Motiv: schneller zu sein als der Markt. Denn die Medizintechnik ist seit einiger Zeit in der Lage, solcherart Prognosen zu erstellen, man nutzte diese aber bisher nicht, weil ihr Einsatz naturgemäß höchst umstritten ist. Es stellen sich jede Menge Probleme, allen voran jenes, dass ja die Babys selbst nicht darüber entscheiden können, ob sie eine Analyse ihres medizinischen Schicksals wünschen oder nicht. Und man weiß freilich auch noch nicht, wie die Eltern mit dem Wissen über mögliche künftige Krankheiten ihrer Kinder umgehen werden.

ES GILT: WAS MÖGLICH IST, WIRD AUCH GEMACHT

Beim Baby-Seq-Projekt geht es um Technologiefolgenabschätzung, um die Frage also, was das Einsetzen dieser Technologie mit den Menschen macht. Allfällige verheerende Folgeschäden sollen verhindert werden, bevor es zu spät ist und diese Untersuchung im Kreissaal zur Routine wird. Experten sind im Übrigen der Ansicht, dass die Genom-Sequenzierung von Kindern nicht aufzuhalten ist. Aber nicht nur die Ergebnisse, sondern der bloße Einsatz von Dataming und PA müssen auf ihre Folgewirkun-

gen hin überprüft werden. Wird in Österreich etwa Predictive Policing – das heißt Vorhersagen im Bereich der Kriminalistik zu treffen – erstmals im Hinblick auf das Einbruchrisiko getestet, wird es in den USA bereits viel weitreichender angewandt: In Florida hat nun das Department of Juvenile Justice beschlossen, künftig PA einzusetzen, um das Rückfallrisiko jugendlicher Straftäter zu ermitteln und gegebenenfalls gegenzulenken. Das Programm nennt sich „The Positive Achievement Change Tool“. Die Risiko-Bewertungs-Software screent Daten zur Vorgeschichte der jugendlichen Straftäter aus Polizeiberichten, Gerichts- und Gefängnisakten, durchstöbert die Herkunft der Jugendlichen und ihr soziales Umfeld sowie Risikofaktoren wie etwaige Traumatisierungen oder Abhängigkeiten.

VORHERSAGEN IN DER KRIMINALISTIK SIND EIGENTLICH EIN SCHUTZMECHANISMUS

Das Ziel ist ein hehres: Gefährdete Jugendliche sollen vor erneuter Straffälligkeit und Inhaftierung geschützt werden – vor sich selbst also. Fast reflexartig stellt sich hier die Frage, was das wohl mit einem jungen Menschen macht, wenn er von vornherein als Wiederholungstäter abgestempelt wird. Ergibt sich daraus nicht zwangsläufig eine selbsterfüllende Prophezeiung, die noch dazu nicht auf bloßen Annahmen beruht, sondern Schlussfolgerungen auf Basis harter Daten darstellen? Nach Definition des österreichischen Psychotherapeuten Paul Watzlawick handelt es sich bei der selbsterfüllenden Prophezeiung um eine Voraussage, die rein aus der Tatsache heraus, dass sie gemacht wurde, zur Wirklichkeit wird und damit ihre eigene Richtigkeit bestätigt. Wer also den Teufel an die Wand malt, verstärkt demnach die Wahrscheinlichkeit, dass es tatsächlich so kommt.

VORHERSAGEN SIND NUR SO GUT WIE DIE DATEN, AUS DENEN SIE STAMMEN

Themenwechsel zu einem der Praxisprobleme des Dataming: die Datenqualität. Grundsätzlich gilt der Leitsatz: „Garbage in, garbage out“ (kommt Müll hinein, kommt Müll heraus), wie es Mobilitätsforscher Martin Köhler vom Austrian Institute of Technology auf den Punkt bringt. Ist also die Datenbasis schlecht, kann bei deren Analyse nicht viel herauskommen. Im hektischen Spitalsalltag kommt es immer wieder zu schlampigen Aufzeichnungen, bei Verkehrsprognosen ist das Problem der ungenauen Verortung noch nicht gelöst. Ist etwa gerade keine Satellit in der Nähe, können GPS-Daten um bis zu 100 Meter unscharf sein. Mobilität ist jedenfalls bereits heute eines der großen Anwendungsbereiche von PA. Reisezeiten können immer genauer berechnet, Staubildungen vorausgesagt werden. Mussten bisher Autos gezählt und Umfragen durchgeführt werden, bieten heute Sensoren deutlich effizienter Möglichkeiten der Datenerfassung. „Smart Survey“ (schlaue Umfrage) heißt diese neue Art der Verkehrsdatenerhebung. Freiwillige Probanden, die eine entsprechende App auf ihren Handys installiert haben, liefern aktuelle Verkehrsinformationen. Die Systeme können sogar unterschiedliche Verkehrsmittel erkennen. Es geht längst nicht nur um den Autoverkehr, sondern genauso um Zugverbindungen, um den Umstieg von einem auf ein anderes Verkehrsmittel, der Nachfrage nach Car-Sharing an Verkehrsknotenpunkten. Neben den Handy-Probanden sind zudem mit Sensoren ausgestattete Taxis unterwegs und

WAHRSCHEINLICHKEIT, NICHT SICHERHEIT

liefern Daten direkt von der Straße. Diese Methoden weisen zugleich in die Zukunft der Verkehrsforschung,



Auf Basis der Vergangenheit wird die Zukunft vorausgesagt: Durch gesammelte Daten, etwa über das Einkaufsverhalten einer Person, berechnet ein Algorithmus, welche Produkte bald in den Einkaufswagen wandern werden. Hier geht es nicht nur um den täglichen Einkauf, sondern auch um neue Artikel wie zum Beispiel ein Buch oder ein Fahrrad. Aber auch im Bereich der Mobilität liefert das Verhalten von gestern Material für Prognosen aus der Glaskugel. So können zum Beispiel Staus genau vorausgesagt oder Reisezeiten sehr genau berechnet werden.

zumal die Daten dichter und präziser werden, und die Systeme immer genauer arbeiten und laufend dazulernen. Dennoch, betont Köhler, bilden selbst die ausgereiftesten Analysemethoden immer nur eine Wahrscheinlichkeit ab. Selbst bei einer Trefferquote von 97 Prozent bleibt ein Restrisiko, dass die Vorhersage dennoch nicht eintrifft. Die Ziele von Dataming und PA in der Verkehrsforschung sind der optimale Verkehrsdurchfluss, exakte Verkehrsplanung sowie das Vermeiden von Unfällen. „Den Verkehr ohne Risiko wird es in naher Zukunft noch nicht geben“, betont Köhler.

..... DATENSICHERHEIT ALS UNVERZICHTBARER BESTANDTEIL DES DATAMININGS

Ein zentrales Thema im engen Zusammenhang mit Dataming, das immer stärker ins Bewusstsein rückt, ist die Datensicherheit. Denn wurden personenbezogene Daten erst einmal erhoben, lassen sich Ergebnisse mit hoher Treffersicherheit exakt rückverfolgen.

„Dann ist es mit der Anonymität vorbei“, betont Tschohl. Er leitet das Research Institute in Wien – ein Forschungszentrum an der Schnittstelle von Technik, Recht und Gesellschaft –, ist Nachrichtentechniker und Jurist und war als Mitglied der Initiative AK Vorrat mitverantwortlich dafür, dass die Vorratsdatenspeicherung aufgehoben wurde. Denn Sicherheit bedeutet angesichts der zunehmenden Durchdringung vieler Lebensbereiche mit Informations- und Kommunikationstechnologie nicht nur Schutz vor Viren, Trojanern und Hackerangriffen, sondern es geht um rechtliche Sicherheit beim Erheben von und im Umgang mit Daten. Konkret: Datenanalysen müssen im Einklang mit Bürger- und Menschenrechten gestaltet werden. Zahlreiche Beispiele zeigen, dass dies möglich ist. Die Weichen müssen rechtzeitig gestellt werden. Tschohl arbeitete zuletzt an einer noch nicht veröffentlichten Roadmap für IKT-Sicherheit in Österreich mit. Eine wichtige Erkenntnis der Studie: Gutes Privacy Design in der Datenverarbeitung gewinnt zunehmend an Bedeutung und ist folglich keine Bremse für

den Wirtschaftsstandort, sondern ein Innovationsmotor.

..... PERSONENBEZOGENE DATEN SIND OFT GAR NICHT NÖTIG

Die Technologie hat dem Menschen zu dienen und nicht umgekehrt. Deshalb muss vor jeder Umsetzung, vor jeder Datenerhebung, die Frage nach dem konkreten Nutzen gestellt und beantwortet werden. Aus Sicht des Datenschutzes gilt: Je weniger in die Grundrechte eingegriffen wird, desto besser. Der legitime Zweck einer Erhebung und Verarbeitung von Daten muss dabei jedesmal genau definiert werden. „Abstrakte Formulierungen wie ‚im Interesse der öffentlichen Sicherheit‘ sind viel zu schwammig“, betont Tschohl. Zudem heilige der Zweck eben nicht die Mittel. Die Maßnahmen müssen tauglich sein, dürfen aber nicht übers Ziel hinausschießen. In den meisten Fällen ist die Erhebung personenbezogener Daten nicht notwendig, zumal auch aggregierte, also zusammengefasste Daten zu soliden Ergebnissen führen. ✱