

Smart durch Big Data? Plädoyer für eine kritische Betrachtung

Wer wirklich smart sein will, kommt offenbar um Big Data nicht herum. Betrachtet man den Hype etwas näher, kommen Zweifel auf. Was steckt wirklich dahinter? Welches sind die Konsequenzen? Und wird dabei auch die Psychologie überflüssig?

Big Data ist das aktuelle Zauberwort, das sich unmittelbar aus der unausweichlichen Digitalisierung unserer Arbeitswelt ableitet und das als Synonym für Fortschritt gilt. Es gehört zum Standardrepertoire von Keynote-Speakern, Journalisten, Wirtschaftspolitikern, Politikberatern, Hochschulrektoren, Informatikern und letztlich von allen optimistisch querdenkenden, innovativen, progressiven, intelligenten Menschen. *Big Data* sichert angeblich die Wettbewerbsfähigkeit des Standorts Deutschland ebenso wie Arbeitsplätze und Renten, fördert Demokratie auf Augenhöhe, befreit uns von unnötiger Wissenschaft und langatmigen Theorien, bringt uns pragmatisch in die Zukunft und ist also letztlich so etwas wie ein Sechser im Lotto – natürlich mit Zusatzzahl.

Interessant allerdings: Kaum einer derjenigen, die in diesem Chor lautstark über *Big Data* reden, weiß überhaupt, wie *Big Data* funktioniert. Hier verhält es sich ähnlich wie mit dem Zaubertrank bei Asterix, bei dem nur einer im Dorf wirklich weiß, was darin steckt. Der wichtige Unterschied: Im gallischen Dorf versetzt bereits der Glaube an den Zaubertrank Berge, ein Effekt, der bei *Big Data* lediglich beim Einsammeln von Kapital für Projekte mit dem Versprechen *Big Data* nachweisbar ist. Denn Investoren glauben an *Big Data*. Im Kern ist *Big Data* allerdings – so die Argumentation der wenigen Skeptiker – nichts Besonderes, sondern lediglich ein Synonym für das ungesteuerte Zusammentragen von möglichst vielen Daten ungewisser Qualität, die dann ungesteuert auf möglichst viele (auch unerklärliche) Zusammenhänge hin untersucht werden, wobei die Quantität der Daten Unzulänglichkeiten in der Qualität von Daten und Methoden ausgleichen soll.

Kleine, aber nicht unwichtige Vorbemerkung

Der Verfasser dieses Beitrags hat bereits vor über 30 Jahren mit Algorithmen und diversen anderen Modellen experimentiert, die nach heutiger Lesart zu *Big Data* gehören. Neben den später noch in diesem Beitrag

zu behandelnden Markoff-Analysen war dies vor allem die Mustererkennung. Bei dem von dem russischen Mathematiker Andrei Andrejewitsch Markow entwickelten speziellen stochastischen Prozess geht es darum, Wahrscheinlichkeiten für das Eintreten zukünftiger Ereignisse zu berechnen. Ziel ist es, Kombinationen von Merkmalen zu lokalisieren, die mit hohen Effektivitätswerten in Zusammenhang stehen. Das Hauptaugenmerk gilt der Lokalisation von Konfigurationen, also dem gleichzeitigen Auftreten von Merkmalen, die – in Abhängigkeit vom Kontext – Erfolg versprechen.

Die große Schwierigkeit lag (und liegt) allerdings genau in diesem Bezug zum Kontext, also darin, dass es immer von den spezifischen Umständen abhängt, ob Merkmalskombinationen erfolgreich sind. Eine zweite Schwierigkeit lag (und liegt) in der gerade für personalwirtschaftlich-verhaltensorientierte Fragestellungen oft nur unzureichenden Datenqualität. Mit zunehmender Datenverfügbarkeit, aber nur mit wissenschaftlicher Rigorosität, könnten sich diese Schwierigkeiten reduzieren.

Ein methodischer, aber wichtiger Hinweis

Was ist eigentlich *Big Data*? Zum einen bezieht sich *Big Data* auf die Datenbasis. Da geht es um die Nutzung von allen Daten jeglicher Art: Bestellungen in der Kantine, über Amazon gekaufte Bücher, Unterlagen der Krankenkasse, Kreditkartenabrechnungen, Reisedaten im Navigationssystem, Dokumente auf Servern sowie in der Cloud, persönliche E-Mails, Flickr- oder Snapchat-Fotos, Facebook-Einträge, WhatsApp- sowie Twitter-Nachrichten und vieles mehr, was man sich lieber nicht vorstellen

*Professor Dr. Christian Scholz,
Inhaber des Lehrstuhls für Betriebs-
wirtschaftslehre, insbesondere Orga-
nisation, Personal- und Informations-
management an der Universität des
Saarlandes in Saarbrücken,
scholz@orga.uni-sb.de*



New Work

möchte. Zum anderen bezieht sich *Big Data* auf Algorithmen, die diese Daten auf Zusammenhänge untersuchen. Das Ergebnis sind im Regelfall Koinzidenzen als gleichzeitiges Auftreten von bestimmten Merkmalen, wobei Koinzidenz aber nichts zur Kausalität aussagt.

Allerdings sind beide Punkte problematisch: Zum einen wirft die grenzenlose Datenbasis Probleme auf. Ist es zulässig, wenn man einen Kandidaten googelt oder sein Verhalten auf Facebook automatisch analysiert und einen entsprechenden „Background-Check“ in Auftrag gibt? Oder will man sogar eine DNA-Analyse? Bereits nach dem Datenschutz- und Mitbestimmungsgesetz scheidet vieles aus, was zurzeit propagiert und auch praktiziert wird. Zum anderen sind die Verarbeitungsverfahren problematisch. Gerade die induktiv lokalisierten „Zusammenhänge“ zwischen scheinbaren Ursachen und scheinbaren Wirkungen können – falsch interpretiert – fatale Konsequenzen für Betroffene haben.

Koinzidenz bedeutet nicht Kausalität. So sinnvoll Koinzidenz und auch *Big Data* im Gesundheitsmanagement sein können (bestimmte Symptome führen zu einem spezifischen Krankheitsverdacht und leiten die weitere Diagnose), so problematisch sind sie im beruflichen Alltag. Stellt man beispielsweise fest, dass kreative Mitarbeiter während der Kreativphase Kaugummi kauen, kann man daraus kaum schließen, dass Kaugummi eine Ursache für Kreativität ist. Es ist also weder sinnvoll, alle Mitarbeiter zum Kaugummigenuß zu zwingen, noch sollte man Einstellungsentscheidungen davon abhängig machen, ob Kandidaten beim Bewerbungsgespräch einen Kaugummi im Mund haben.

Ein kleines, aber interessantes Beispiel

Wie man aber trotzdem *Big Data* einsetzen kann, zeigt das Anwendungsbeispiel

Recruiting. Hier wird massiv gearbeitet: Wer hat sich zu einem bestimmten Thema geäußert? Wer taucht auf welchen Teilnehmerlisten auf? Wer ist auf welchen Fotos „markiert“? Auch liefert *Big Data* Hinweise auf Stärken und Schwächen potenzieller Kandidaten. Hierzu lassen sich Bewegungsprofile quer durch die sozialen Netze erstellen, woraus man dann auf Persönlichkeitsstrukturen schließt. Man kann es aber noch weitertreiben und über Stimmanalysen von Bewerbern Persönlichkeitsprofile erstellen.

Auch wenn in vielen dieser Fälle Zweifel an der Wissenschaftlichkeit angebracht sind, wirkt das alles vielleicht noch relativ harmlos. Nur: Was ist, wenn man aufgrund einer vielleicht sogar extrem fragwürdigen Analyse vermutet, dass Frauen mit einer Vorliebe für Rosamunde Pilcher viel mehr Elternzeit beanspruchen als solche, die Filme vom Typ „Erin Brockovich“ bevorzugen. Dann hat eine harmlos scheinende Frage im Bewerbungsgespräch nach Filmen für die Kandidatin weitreichende Konsequenzen.

Erstes, aber schon ernüchterndes Zwischenfazit

Big Data arbeitet oft mit trivialer Statistik. Auf diese Weise die Ungenauigkeit, Unschärfe und Unsicherheit in Daten ausgleichen zu wollen, grenzt an unrealistischen Machbarkeitswahn. Weder verfügt *Big Data* damit über zuverlässige Daten noch über aussagefähige Algorithmen. Dementsprechend gibt es nur wenige Beiträge, die wirklich zeigen, wo ein solider personalwirtschaftlicher Erkenntnisgewinn durch *Big Data* realisiert wird. Hier verhält es sich ähnlich wie mit dem nebulösen Data Mining, das auch nach lautstarker Eröffnung wenig Verwertbares zutage gebracht hat. Meist werden lediglich Personen, von denen kaum einer je ernsthaft mit *Big Data* gearbeitet hat, befragt, wo sie die Entwicklungspotenziale für *Big Data* sehen. Aus solchen Aussagen wird dann auf die Zukunft von *Big Data* geschlossen.

Je weniger Personen es gibt, die sich professionell mit Personalmanagement auseinandersetzen können, umso mehr wächst der Glaube an *Big Data* als personalwirtschaftliche Wunderwaffe. Daher überrascht es nicht, wenn sich gerade diejenigen, die für ein Abschaffen von professionellen Personalabteilungen trommeln, lautstark für *Big Data* einsetzen. Denn jetzt kann man als angeblich „smarter“ Human-Resource-Manager ohne verhaltensorientierte Theorien, ohne komplexe Modelle und ohne dicke Lehrbücher beziehungsweise aufwendige Studiengänge ganz einfach vom Smartphone ablesen, welche Mitarbeiter man einstellen, fördern oder freisetzen soll.

Eine offenkundige, aber hochproblematische Beobachtung

Februar 2013: In einem interessanten Interview mit der „Zeit“ zählt der im österreichischen Zell am See geborene und in höchste international-akademische Ebenen aufgestiegene Viktor Mayer-Schönberger auf, in welchen Sprachen sein zuerst in Englisch erschienenes Buch zu *Big Data* gerade publiziert wird: Portugiesisch, Französisch, Koreanisch, Spanisch, Italienisch, Niederländisch, Japanisch, Taiwanesisch und Türkisch. Nur auf Deutsch erscheine es nicht. Keiner der großen Verlage hätte Interesse an einer „ausbalancierten“ Analyse von *Big Data*, bei der es auch um die Risiken geht, erklärte der Jurist, der als Professor für Internet Governance und Regulation am Oxford Internet Institute tätig ist. Gerade in Deutschland mit seiner extrem ausgeprägten Nähe zu MINT-Themen (Mathematik, Informatik, Naturwissenschaft, Technik) gibt es wenig Bereitschaft, mit *Big Data* kritisch umzugehen, obwohl nur eine derartige Auseinandersetzung Fortschritt ermöglicht und Missbrauch eindämmt. Vor allem bei personalwirtschaftlichen Fragestellungen, die sich zwangsläufig auf Menschen beziehen, wäre genau dies erforderlich, unterbleibt aber bei Forschungsgipfeln ebenso wie auf Personalkongressen.

New Work



neal/istockphoto

„Zauberkraft von Big Data“: Beeinflussung von Bürgern, Mitarbeitern und Konsumenten

Eine problematische, aber aufschlussreiche Gegenüberstellung

Big Data basiert auf der These, wonach sich Qualitätsdefizite von Daten durch Steigerung der Quantität von Daten kompensieren lassen. Also: Man braucht nur extrem viele (auch schlechte) Daten zu sammeln, dann mutieren diese irgendwie automatisch zu „guten“ Daten, die jede Realität perfekt abbilden. Nur gibt es für diese These bisher keinen zwingenden Beweis. Im Gegenteil spricht vieles dafür, dass immer noch die Müllbehandlungstheese „Garbage in, Garbage out“ gilt.

Auch wenn es vielleicht problematisch, weil historisch belastet ist, soll an das DDR-Ministerium für Staatssicherheit (MfS) erinnert werden. Es wurde 1950 gegründet und sammelte so ziemlich alle Daten, derer es irgendwie habhaft werden konnte. So gesehen könnte man Walter Ulbricht und Erich Mielke als Erfinder von *Big Data* betrachten. Im Westen existierte keine vergleichbare Institution. Stattdessen gab es 1970 ein erstes Datenschutzgesetz und 1983 ein Volkszählungsurteil, vor allem aber wirtschaftlichen Aufschwung – trotz

weitgehenden Verzichts auf Datensammlung über Bürger. Letztlich hat im „Kampf der Systeme“ das MfS mit seinem Datensammeltrieb doch nicht gewonnen.

Eine hochinteressante, aber selten praktizierte Offenlegung

Sucht man nach Anwendungsfällen, die erklären, wie ein Unternehmen bei der Nutzung von *Big Data* vorgegangen ist, so wird man zwar nicht unbedingt in personalwirtschaftlichen Quellen fündig, wohl aber an ganz anderer Stelle. So erklärt der Produktchef von Netflix, Neil Hunt, in einem Interview in der österreichischen Zeitschrift „Profil“ den Erfolg seines Unternehmens mit seinem Wissen um die Sehgewohnheiten seiner Konsumenten. Dazu verarbeitet der Streamingdienst und Filmanbieter Netflix die Daten von 69 Millionen Kunden, um unter anderem herauszufinden, was der Kunde als Nächstes sehen will. Dazu nutzt Neil Hunt – und jetzt wird es gleichermaßen interessant wie konkret – explizit Markoff-Ketten, also Übergangswahrscheinlichkeiten, die Auskunft darüber geben, was ein Zuschauer in einer bestimmten Situation gemacht

hat: also zum Beispiel abschalten oder umschalten – und im zweiten Fall, wohin. Da Netflix ausschließlich über das Internet empfangen werden kann und der Zuschauer somit 1:1 mit dem „Sender“ verbunden ist, bekommt Netflix riesige Datenmengen. Sie dienen dann sogar dazu, die Akzeptanz bestimmter Schauspieler in bestimmten Situationen zu ermitteln. Und auch wenn es Neil Hunt nicht explizit erwähnt, letztlich lassen sich auf diese Weise sogar ganze Spielfilme als optimierte Übergangsfolgen „komponieren“.

Hier sieht man, wie *Big Data* funktioniert, und auch, dass dies durchaus auf Personalplanung übertragbar ist. Derartiges hat im Übrigen schon Andrew Forbes 1971 mit seinen Arbeiten über Markoff-Modelle bewiesen und damit vor langer Zeit Eingang in anspruchsvolle Lehrbücher zu Personalmanagement gefunden. Allerdings wurden damals derartige Vorgehensweisen noch belächelt und als akademisch-theoretisch abgetan. Vielleicht liegt hier eine interessante Chance: nämlich plausible Modellansätze mit dem Etikett *Big Data* zu versehen und damit nicht nur die Filmwirtschaft auf neue Ebenen zu heben.

Eine behavioristische, aber erneut thematisierbare Annahme

Es geht jedoch noch wesentlich weiter: „Nichts treibt die Utopie einer perfekt zu managenden Gesellschaft derzeit stärker an als der Glaube an die Zauberkraft von *Big Data*.“ Mit diesem Satz fasst die „FAZ“ das „Nudging-Prinzip“ des Verhaltensökonom Richard Thaler und des Juristen Cass Sunstein zusammen, das auf psychologische Verhaltenssteuerung durch sanftes „Anstupsen“ hinausläuft. Aber anders als beim klassischen Behaviorismus von B. F. Skinner liefern jetzt die Menschen selbst Daten. In dieser von dem Informatiker Alex Pentland als „Social Physics“ bezeichneten Welt muss man dann nur noch „wie ein Ingenieur an den entsprechenden Reglern drehen“, und Menschen

New Work

ändern ihr Verhalten. Experimente und Theorien spielen keine Rolle mehr: Die schiere Datenmenge liefert in Echtzeit Verhaltensdaten. Beeinflussung und Verstärkung erfolgen nicht mehr durch Käsestückchen, sondern durch Informationsfilter. Wir bekommen also einen digitalen Neobehaviorismus.

Das Ergebnis ist dann ein *Big Nudging* als Verbindung von *Big Data* und Nudging: Hier werden basierend auf großen Datenmengen und konsequenter Mustererkennung Mitarbeiter, Konsumenten und Bürger, ohne es zu merken, in bestimmte Richtungen „gestupst“. Das interessiert nicht nur Unternehmen, sondern plötzlich auch die Politik, die auf diese Weise den unmündigen und irrationalen Bürger sanft in die richtige Richtung glaubt schubsen zu müssen und zu können. *Big Nudging* hat Konsequenzen für die Forschung. Denn durch die Logik von *Big Data* mit seiner konsequenten Theorielosigkeit wird – folgt man seinen Verfechtern – als Nebeneffekt auch die Psychologie als Grundlagen- und Anwendungsdisziplin überflüssig, da wir weder eine theoriebasierte Organisations- und Personalforschung noch eine Konsumentenforschung mehr brauchen. Es genügt Informatik als reine Massendatenanalyse.

Eine naheliegende, aber zweischneidige Überlegung

In der Wissenschaft gab es schon immer ethische und methodische Einwände gegen den Behaviorismus, die alle verstärkt auf *Big Nudging* zutreffen. Denn dahinter steckt das identische Menschenbild von fremdgesteuerten Wesen, die sich nicht einmal der Fremdsteuerung bewusst sind. Da es zudem nicht nur darum geht, den Einzelnen zu steuern, sondern ganze Gruppen von Menschen, ergeben sich auch gesellschaftspolitische Probleme – worauf unter anderem der Soziologie-Professor Dirk Helbing in einem warnenden Artikel zum *Big Nudging* hinweist. Deshalb ist hier Vorsicht angebracht.

Vorsicht ist aber nicht gleichzusetzen mit Verzicht. Denn wenn man sich anschaut, auf welch wackligen Beinen viele psychologische Verfahren, Erkenntnisse und Tests stehen, bei denen ganz häufig (wenn überhaupt) relativ kleine Gruppen von Studenten als Probanden auftreten, so ist die Hinwendung zu größeren Datenmengen durchaus plausibel. Trotzdem bedeutet richtiges empirisches Arbeiten nicht, möglichst alle verfügbaren Daten zu nutzen. Gut ausgewählte Daten, reliabel erhoben und valide genutzt, sind der bessere Weg – ohne dass man sich blind auf „pseudoanonymisierte Daten“, „Metadaten“ oder „Record Linkage“ einlässt. Gleichzeitig könnte theoriebasiert die kontextbezogene Mustererkennung in den Mittelpunkt rücken, allerdings basierend auf strenger Wissenschaftlichkeit sowie Freiwilligkeit und informationeller Selbstbestimmung.

Eine unbequeme, aber nötige Schlussfolgerung

Big Data wird in doppelter Hinsicht unterschätzt: Einerseits sind die Analysemöglichkeiten bereits mit heutigen Techniken wesentlich vielfältiger als landläufig gedacht, andererseits sind die damit verbundenen Probleme deutlich größer, als es die aktuelle Diskussion vermuten lässt.

Damit ergibt sich eine zentrale Schlussfolgerung: *Big Data* gehört raus aus der reinen Informatik mit dem Streben nach Massendatenverarbeitung. *Big Data* gehört hinein in eine strikt verhaltensorientierte und am Menschen ausgerichtete Wissenschaft. Dazu brauchen wir aber ein völliges Umdenken bei Politikern, Journalisten, Beratern, Hochschulrektoren, Politikberatern und – als ersten Schritt – in der Forschungsförderung.

Weiterführende Literatur

Brodnig, I. (2015, 11. Dezember). Netflix-Produktchef Neil Hunt: „Ich weiß das alles über Sie“. *Profil.at*. Verfügbar unter:

<http://www.profil.at/gesellschaft/netflix-produktchef-neil-hunt-ich-sie-6141400> [28.4.2016].

Fischermann, T. & Hamann, G. (2013, 21. Februar). Den TÜV fürs Netz, bitte: Eine neue Expertenkaste soll Google und andere kontrollieren: Ein Treffen mit Computerforscher Viktor Mayer-Schönberger. *Zeit.de*. Verfügbar unter: <http://www.zeit.de/2013/09/Internet-Tuev-Viktor-Mayer-Schoenberger-Big-Data> [28.4.2016].

Forbes, A. F. (1971). Markov chain models for manpower systems. In D. J. Bartholomew and B. R. Morris (Eds.), *Aspects of manpower planning: A volume of papers* (pp. 93–113). London: English University Press.

Helbing, D. (2015, 12. November). „Big Nudging“ – zur Problemlösung wenig geeignet [Kommentar]. *Spektrum.de*. Verfügbar unter: <http://www.spektrum.de/news/big-nudging-zur-problemloesung-wenig-geeignet/1375930> [28.4.2016].

Pentland, A. (2014). *Social physics: How good ideas spread. The lessons from a new science*. New York: Penguin.

Scholz, C. (1981). Bildschirmorientierte Personalplanung mit SIMIPOC. *OR-Spektrum*, 3, 161–174.

Scholz, C. (2013, 12. August). Aktuell, aber fiktiv: Die Personalabteilung als Geheimdienstzentrale. *Per-anhalter-durch-die-arbeitswelt.de*. Verfügbar unter: <http://per-anhalter-durch-die-arbeitswelt.de/personalabteilung-als-geheimdienstzentrale/> [28.4.2016].

Scholz, C. (2014, 26. August). Big Data: Große gläserne Mitarbeiter. *Per-anhalter-durch-die-arbeitswelt.de*. Verfügbar unter: <http://per-anhalter-durch-die-arbeitswelt.de/grosse-glaeserne-mitarbeiter/> [28.4.2016].

Staun, H. (2014, 7. September). Wie wir gern leben sollen. *Faz.net*. Verfügbar unter: <http://www.faz.net/aktuell/feuilleton/big-data-social-physics-wie-wir-gern-leben-sollen-13126401-p2.html> [28.4.2016].

Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.