



Dienstag, 19. Dezember 2017, 21:00 Uhr
~7 Minuten Lesezeit

Big Data!

Nur ein Hype oder ein bedeutender Fortschritt?

von Wolfgang Romey
Foto: LIPING/Shutterstock.com

Wie in vielen Bereichen der Entwicklung und Anwendung von digitaler Technik wird auch bei Big Data sehr viel versprochen. So schrieb Chris Anderson, der ehemalige Chefredakteur des Technologiema-

Wired, 2008 am Ende seines Aufsatzes „Das Ende der Theorie – Die Datenschwemme macht wissenschaftliche Methoden obsolet“: „Korrelationen machen Kausalitäten überflüssig, und die Wissenschaft kann auch ohne kohärente Modelle, ohne große vereinheitlichenden Theorien Fortschritte machen. Anders ausgedrückt: Sie braucht keine mechanistischen Erklärungen mehr“ (1). Oder: „So konnte etwa die Informatikerin Dr. Carolyn McGregor in den Daten der Vitalfunktionen von Frühgeborenen Muster finden, die mit einer wahrscheinlichen späteren Infektion korrelieren. Damit lässt sich eine mögliche Erkrankung vorhersagen – 24 Stunden vor dem Auftreten der ersten Symptome. Die Analyse großer Datenmengen kann Menschenleben retten, wie bei diesen Frühchen“ (2).

Redaktionelle Vorbemerkung: Die IT-Redaktion möchte in ihrer Kolumne „Schöne neue Welt“ Autoren eine Plattform bieten, die sich kritisch mit der Informationstechnik und den in diesem Feld handelnden Akteuren auseinandersetzen. Wir freuen uns über jeden Artikel, der die technische, aber auch politische, ökonomische, psychologische, physische und soziale Auswirkung und Sicht der IT hinterfragt beziehungsweise behandelt. Bitte kontaktieren Sie uns unter it@rubikon.news.

Big Data – was ist das?

In einer vom Kompetenzzentrum Öffentliche It

[\(https://www.oeffentliche-it.de/\)](https://www.oeffentliche-it.de/) herausgegebenen Schrift findet sich folgende Beschreibung: „Big Data beschäftigt sich mit der intelligenten Auswertung großer Mengen digitaler Daten. Mit intelligenter Auswertung ist hier gemeint, dass Daten aus unterschiedlichsten Quellen gesammelt, aufbereitet, zusammengeführt, analysiert und zu Informationen veredelt werden. Dies erhöht das Wissen der Auswerter signifikant und ermöglicht es, strategische Entscheidungen auf breiter Wissensgrundlage zu treffen. Aus Bereichen wie Wirtschaftsförderung, Energiewende, Sozialhilfe, Bildung, Verkehr oder öffentliche Sicherheit liegen Beispiele für den erfolgreichen Einsatz von Big Data vor. Bei großen Mengen kann es sich um das in heutigen Datenzentren typischerweise vorhandene Datenvolumen von Tera- und Petabytes handeln.

Es kann sich aber auch um das tausend- oder millionenfache Datenvolumen von Exa- und Zettabytes handeln, das die Gesamtmenge der heute überwiegend in sozialen Medien im Internet vorhandenen digitalen Daten beschreibt. Die Bandbreite der betrachteten Datenvolumen umfasst den Bereich von neun Zehnerpotenzen. Benötigt man beispielsweise zur Auswertung eines Terabyte Daten eine Sekunde, so braucht man für die Auswertung eines Zettabyte zweieinhalb Jahre. Damit wird klar, dass man zur Verarbeitung dieser Datenmengen auf das betrachtete Datenvolumen zugeschnittene Werkzeuge benötigt“ (3).

Wo und wie werden Daten gesammelt?

Das Feld ist breit. Daten werden beispielsweise in industriellen Fertigungsprozessen, bei der Verfolgung von Verkehrsströmen, bei der Wetterbeobachtung gesammelt und dienen in diesen Bereichen zur Optimierung von Prozessen oder zur Verbesserung von Vorhersagen, sie sind damit nützlich.

Gesammelt wird vor allem „bei jedem Klick im Internet, allerdings hinterlassen wir auch außerhalb des Internets ständig Daten – beim Arzt, der Krankenkasse, der Apotheke, beim Finanzamt, beim Arbeitgeber, bei Banken und beim Einkaufen, um nur einige Beispiele zu nennen. Was man oft nicht bedenkt oder schlicht nicht weiß, ist, dass auch an anderen Stellen Daten von uns gesammelt werden. Es gibt Schaufensterpuppen mit eingebauten Kameras, die die Kunden analysieren, Hightech-Mülleimer und Straßenlaternen, die vorbeigehende Passanten anhand der von Handys ausgesendeten Funkwellen verfolgen, smarte Zahnbürsten, die das Zahnputzverhalten aufzeichnen, Autos, die das Fahrverhalten und das Stress- und Müdigkeitslevel auf das genaueste messen oder Fitnessarmbänder, smarte Uhren und jede Menge anderer Geräte, die permanent eine Vielzahl von Vitalwerten messen und damit oft mehr über den Gesundheitszustand Bescheid wissen als der Hausarzt“ (4). Das ist das oben genannte Datenvolumen von Exa- und Zettabytes, das besondere Verfahren zur Auswertung erfordert.

Bei vielen dieser Daten mag die Sammlung harmlos erscheinen. Das ist oftmals aber eine Täuschung. Bekannt ist, dass schon wenige Daten ausreichen, um eine Person zu identifizieren und sie dann komplett zu durchleuchten. „So lassen sich beispielsweise aus den Daten Informationen über Charaktereigenschaften, religiöse und politische Einstellung, sexuelle Orientierung, Intelligenz, Gemütsverfassung, Bildungsniveau, psychopathische Veranlagung oder die Kreditwürdigkeit gewinnen“ (5).

Den gewonnenen Daten wird in der Regel voll vertraut. Nicht bedacht wird dabei das Grundproblem von Big Data: Dass die Gewinnung der Daten, die Verarbeitung der Daten und die Deutung der Ergebnisse nicht losgelöst von menschlichen Interessen und Eingriffen erfolgen. Ein Beispiel: Die Vorhersage von Temperaturen soll durch die umfangreiche Sammlung von Temperaturdaten verbessert werden. Die Daten müssen allerdings aus der Wirklichkeit gewonnen werden und die ist analog und nicht digital;

das heißt, die Werte verteilen sich über ein Kontinuum. Wenn davon ausgegangen wird, dass die physikalische Größe „Temperatur“ den zu untersuchenden Aspekt der Wirklichkeit genau genug abbildet, stellt sich die Frage, wie fein die Daten eingeteilt werden müssen, damit die Auswertung ein brauchbares Ergebnis ergibt. Zehnerschritte (... -20°, -10°, 0°, 10°, 20°, ...) reichen sicherlich nicht. Reichen 1er-, 10tel-, 100tel-, 1.000.000tel-Schritte? Abgebildet ist die Wirklichkeit auch damit nicht. Zwischen jedem 1.000.000tel-Schritt liegen beliebig viele weitere. Jede Abbildung der Wirklichkeit ist eine von Menschen vorgenommene Reduzierung von Wirklichkeit.

Die Verarbeitung der Daten erfolgt durch Rechenvorschriften (**Algorithmen** (<https://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/245571/hat-ein-algorithmus-immer-recht>)), die von Menschen unter Festlegung bestimmter Rahmenbedingungen und Zielstellungen gemacht sind. Wenn die Rechenvorschriften fehlerfrei sind, was die Ausnahme ist, liefern sie Ergebnisse, die wiederum von Menschen gedeutet werden müssen. Vor Big Data stehen also menschliche Entscheidungen und hinter Big Data auch. Die Big-Data-Ideologen sehen davon ab und verkaufen die Daten und die errechneten Ergebnisse als Wirklichkeit.

Im Vergleich zu Daten, die aus sozialen Prozessen oder aus dem Verhalten von Menschen gewonnen werden, sind die beschriebenen Prozesse ungleich komplexer. Ungenauigkeiten und Fehler werden deshalb eher die Regel als die Ausnahme sein. Problematisch ist das insbesondere, wenn Menschen aufgrund der gewonnenen Daten Eigenschaften zugeschrieben werden und ihnen beispielsweise ein Kredit verweigert wird, eine negative Gesundheitsprognose zugeschrieben wird, sie einen Arbeitsplatz nicht erhalten oder ein hohes Rückfallrisiko nach einer Straftat prognostiziert wird.

Big Data und soziale Netzwerke

Nicht ganz so problematisch mag es erscheinen, wenn aus der Analyse der Daten Medienvorschläge oder Werbebotschaften erzeugt werden, die auf den mit der Hilfe von Big Data erkannten Vorlieben der Menschen basieren. Gefährlich wird es aber, wenn zum Beispiel bei Facebook auf der Grundlage der Nutzerdatenanalyse über die Anzeige von Inhalten entschieden und damit zur selektiven Wahrnehmung der Welt beigetragen wird. „Hier übernehmen Big Data Analysen eigentlich journalistische Funktionen, ohne sich journalistischen Standards zu unterwerfen. Auch Aussagen aus der Wissenschaft, die auf Big Data Analysen fußen, werden im öffentlichen Diskurs gerne als Wahrheit aufgenommen und als normativ gesetzt. Die Öffentlichkeit fordert von Wissenschaft eine Erklärung der Welt, eine Analyse der Zusammenhänge, die Big Data Analysen nicht zu liefern im Stande sind. Hier kann der Erfolg und die große Verbreitung von Big Data Technologien langfristig zum Problem werden, wenn den Datenanalysespielereien und Visualisierungen keine Modellbildung folgt.“ (5)

Selbstverständlich ist Facebook nicht die Ausnahme. Wer beispielsweise die meisten der Google-Dienste nutzt, hat seine Privatsphäre aufgegeben. Das umfasst die Auswertung der über Gmail verschickten und erhaltenen E-Mails, der über Google gestellten Suchanfragen, der von dem Smartphone für Google bereitgestellten Kontakte, der Vorlieben bei YouTube-Videos, des vom Smartphone gelieferten aktuellen Bewegungsprofils, der Gesundheitsdaten, der genutzten Spiele, der gelesenen Texte, ... Gefährlich ist das, weil sich die Nutzer meist nicht bewusst sind, dass diese ungeheure Fülle von Daten erhoben wird; leider oftmals auch unter deren aktiver Beteiligung. Mit den Verfahren von Big Data ist eine umfassende Profilbildung möglich, die sich hervorragend für die Beeinflussung und Überwachung der Nutzer

eignet. Das betrifft nicht nur den einzelnen Nutzer. Da auch die Kontakte bekannt sind und das Kommunikationsverhalten ausspioniert wird, können sich auch Nutzer, die Google keine Daten direkt geben, der Datensammlung und den Folgen nicht vollständig entziehen.

Ist Big Data also nur ein Hype?

„Insgesamt hat sich der Begriff Big Data heute weitgehend in neuen Hype-Begriffen der Technologiebranche aufgelöst: ‚Predictive X‘ (zum Beispiel als ‚Predictive Policing‘), ‚Machine Learning‘ oder auch der alle paar Dekaden wieder populäre Begriff der ‚künstlichen Intelligenz‘“ (4).

Geblichen sind leistungsstarke Verfahren zur Analyse riesiger Datenmengen, die in einer Reihe von Bereichen nützliche Anwendungen ermöglichen. Sicherlich sind auch im Bereich der Medizin nützliche Anwendungen entstanden. Dass von den Apologeten von Big Data ausgerechnet mit der Gesundheit von Frühchen argumentiert wird, Katzenbilder passen da wohl nicht, ist für mich ein Indiz dafür, dass die Bürger den gegen sie gerichteten Einsatz von Big Data nicht hinterfragen sollen. Denn gefährlich ist Big Data insbesondere bei der Nutzung sozialer Netzwerke, da hier leistungsstarke Verfahren zur Durchleuchtung von Menschen, zur Unterstützung von Überwachung und zur politischen Lenkung von Menschen eingesetzt werden.

Quellen und Anmerkungen:

(1) <http://www.feinschwarz.net/big-data-anspruch-und->

ideologie/ (<http://www.feinschwarz.net/big-data-anspruch-und-ideologie/>)

(2) <http://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/228949/was-ist-big-data>

(<http://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/228949/was-ist-big-data>)

(3) Big Data ungehobene Schätze oder digitaler Albtraum

[https://www.oeffentliche-it.de/publikationen?](https://www.oeffentliche-it.de/publikationen?doc=14702&title=Big+Data+ungehobene+Sch%C3%A4tze+oder+digitaler+Albtraum)

[doc=14702&title=Big+Data+ungehobene+Sch%C3%A4tze+oder+digitaler+Albtraum](https://www.oeffentliche-it.de/publikationen?doc=14702&title=Big+Data+ungehobene+Sch%C3%A4tze+oder+digitaler+Albtraum)

([https://www.oeffentliche-it.de/publikationen?](https://www.oeffentliche-it.de/publikationen?doc=14702&title=Big+Data+ungehobene+Sch%C3%A4tze+oder+digitaler+Albtraum)

[doc=14702&title=Big+Data+ungehobene+Sch%C3%A4tze+oder+digitaler+Albtraum](https://www.oeffentliche-it.de/publikationen?doc=14702&title=Big+Data+ungehobene+Sch%C3%A4tze+oder+digitaler+Albtraum))

(4) <https://www.bpb.de/dialog/netzdebatte/238966/big-data-eine-zustandsbeschreibung>

(<https://www.bpb.de/dialog/netzdebatte/238966/big-data-eine-zustandsbeschreibung>)

(5) [http://www.bpb.de/lernen/digitale-](http://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/228951/die-wahre-macht-der-datensammler)

[bildung/medienpaedagogik/228951/die-wahre-macht-der-](http://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/228951/die-wahre-macht-der-datensammler)

[datensammler](http://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/228951/die-wahre-macht-der-datensammler)

(<http://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/228951/die-wahre-macht-der-datensammler>)

Dieser Artikel erschien bereits auf www.rubikon.news.



Wolfgang Romey arbeitete nach dem Studium der Theoretischen Elektrotechnik als Lehrer für Mathematik, Elektrotechnik und Digitaltechnik im Berufsbildenden Bereich, später als Lehrerausbilder im Vorbereitungsdienst, dem Referendariat. Dann folgte ein Wechsel in die Bezirksregierung Düsseldorf als Dezernent

für Lehrerausbildung und später auch -fortbildung. Er verfügt über etwa 20 Jahre Erfahrung darin, angehende Lehrerinnen und Lehrer auf die Bildungsarbeit mit Digitalen Medien vorzubereiten und deren Urteilskraft in diesem Feld zur Entfaltung zu verhelfen. Die kritische Auseinandersetzung mit den dramatischen Folgen der Digitaltechnik, die ihm extrem unterentwickelt scheint, ist bis heute sein Thema.

Dieses Werk ist unter einer **Creative Commons-Lizenz (Namensnennung - Nicht kommerziell - Keine Bearbeitungen 4.0 International (<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>))** lizenziert. Unter Einhaltung der Lizenzbedingungen dürfen Sie es verbreiten und vervielfältigen.