

Big Data – Risiken und Nebenwirkungen

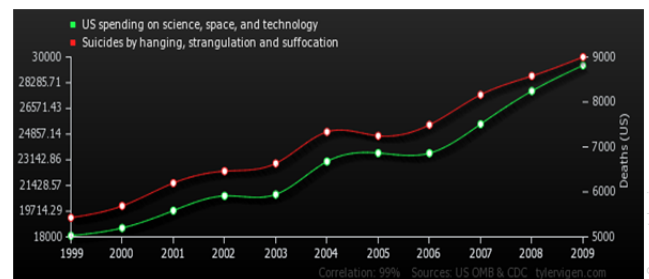
In Kürze

- Big Data steht für die Auswertung großer Datenmengen mittels digitaler Technologien und verspricht neue Einsichten und verbesserte Entscheidungsfindung in vielen Bereichen.
- Potenzielle Anwendungen reichen von Trend-Prognosen bis zu medizinischer Forschung. Um das Potenzial auszuschöpfen braucht es ein tieferes Verständnis über Funktionen und Risiken.
- Big Data kann Komplexität deutlich erhöhen und gepaart mit Automatisierung unerwartete gesellschaftliche Folgen mit sich bringen. Zentrale Anforderungen zur Risikobegrenzung sind Transparenz und Überprüfbarkeit von Big-Data-Analysen.

Worum geht es?

Big Data, also die technologiegestützte Analyse großer Datenmengen, gilt als Wundermittel zur Bewältigung von komplexen Problemen und soll durch die Vorhersage von künftigen Ereignissen Unsicherheiten beseitigen. Potenzielle Anwendungen reichen von Geschäftsprozessoptimierung, bedarfsorientierter Energieversorgung, Trend-Prognosen, medizinischer Forschung, Gesundheitsmanagement bis hin zur vorausschauenden Polizeiarbeit (Predictive Policing). Big Data kann also die Gesellschaft bereichern. Das Erkennen von Korrelationen kann viele Vorteile bringen, insbesondere in der Medizin: etwa durch Erkennen bislang unerkannter Zusammenhänge zwischen Symptomen verschiedener Krankheiten oder Nebenwirkungen von Medikamenten. Auch kann Big Data verborgene Muster zur Unterstützung von Diagnosen aufzeigen, sowie maßgeschneiderte Behand-

lungsformen und Präventivmedizin ermöglichen. Das kann die Behandlung unterstützen und zur Früherkennung beitragen. Analysen von anonymisierten Gesundheitsdaten der Bevölkerung können die Erforschung von Krankheitsentwicklungen (z.B. Krebs) über längere Zeiträume verbessern. Allerdings liegen Erwartungen und Realität oft weit auseinander und deshalb bedarf es hier kritischer Reflexion. Big Data lässt sich durchaus als mutiger Versuch verstehen, Unsicherheit zu reduzieren. Das macht es als Mittel für Zukunftsprognosen attraktiv und wird dementsprechend beworben. Dabei sollte aber nicht vergessen werden, dass seine Visualisierungskraft zum Aufzeigen von Zusammenhängen vor allem auf Wahrscheinlichkeiten und nicht Kausalitäten beruht. Folgendes Bild zeigt etwa, dass die Investitionen der USA für Wissenschaft und Technologie zu fast 100 Prozent mit der Anzahl an Selbstmorden korreliert:

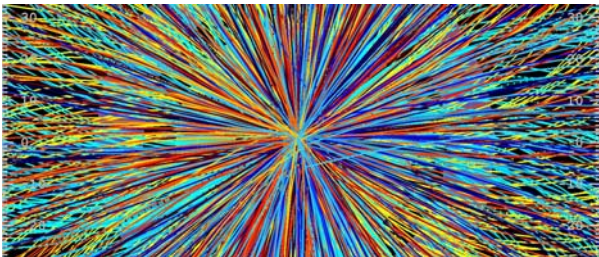


Falle Big Data: fragwürdige Korrelationen

Dieser Zusammenhang ist natürlich völliger Unsinn, zeigt aber die irreführende Kraft von fragwürdigen Korrelationen. Die Interpretation von Big-Data-Ergebnissen ist nicht in jedem Fall so einfach wie hier. Das Vermischen von Korrelation bzw. Wahrscheinlichkeit mit Kausalität kann Unsicherheit drastisch erhöhen. Big-Data-Enthusiasten meinen, dass ungenaue Ergebnisse durch große Datenmengen kompensiert werden können. Dadurch wird suggeriert, Fakten würden mit Quantität entstehen und das Aufspüren von Korrelationen sei der Schlüssel für bessere Entscheidungen. Derartige Ansichten riskieren nicht nur etwas Ungenauigkeit, sondern vor allem Fehlentscheidungen. Denn unabhängig von ihrem Ausmaß sind Daten immer ein zeitliches Konstrukt. Daten in der Gegenwart können daher höchstens solide Information über die Vergangenheit bis zur Gegenwart liefern. Kurzum: Die Zukunft bleibt unberechenbar. Diese simple Tatsache scheint im Daten-Enthusiasmus mitunter unterzugehen. Natürlich kann Big Data sehr hilfreiche Informationen liefern, um Entwicklungen besser zu verstehen. Doch dazu ist Wissen über die Aussagekraft der Analysen und ihre Grenzen nötig. Ein zentraler Faktor ist die Komplexität, die schon mit großen Datenmengen und durch (Teil-)Automatisierung noch weiter zunimmt. Diese Aspekte der Komplexitätssteigerung werden mitunter vernachlässigt. Transparenz ist zentral, um derartige Risiken zu reduzieren, ist aber wirkungslos ohne kompetente Interpretation, die mit steigender Komplexität erschwert wird.

Die große Komplexitätsfalle

Google Flu Trends wurde etwa als ein "großer Erfolg" angepriesen – die Verbreitung von Grippe wurde letztlich mit über 50 Prozent fehlgeschätzt. Deutlich gravierender sind z.B. Fehlberechnungen der (Google-nahen) US-Firma *23andme*, die Risiken in Gen-Daten analysiert. Ein Kunde wurde über zwei Mutationen in seiner DNS informiert, die typisch sind für Gliedergürteldystrophie – eine seltene, meist tödliche Generkrankung. Der Kunde entdeckte erhebliche Fehler in der Analyse, konfrontierte *23andme* damit und erhielt nur eine kurze Bestätigung und Entschuldigung per E-Mail. Noch schwieriger ist ein Fall zur Unterstützung einer medizinischen Diagnose: Big Data entdeckte Tuberkulose-Symptome sowie eine mögliche Neigung zu Blutgerinnseln, der Patient erhielt blutverdünnende Medikation. Es traten keinerlei Gerinnsel auf, was jedoch keineswegs belegt, dass überhaupt ein reales Risiko bestand. Es ist schlichtweg nicht mehr feststellbar, ob Big Data hier erfolgreich war oder die Analyse völlig falsch. Solche Fälle verdeutlichen, wie Big Data Komplexität und Fehleranfälligkeit steigern kann.



Big Data Visualisierungen suggerieren fundierte Erkenntnisse

Richtiges Interpretieren von Ergebnissen und ihrer Aussagekraft ist also keineswegs trivial. Fehldiagnosen bedeuten nicht nur eine potenziell hohe emotionale Belastung. Es drängt sich die Frage auf: Was geschieht, wenn Fehler unerkant bleiben? Big Data ist mitunter schwer überprüfbar. Insbesondere dann, wenn ein rein prognostiziertes Ereignis mit Präventivmaßnahmen verhindert werden soll. Kann aber ein noch nicht geschehenes Verbrechen überhaupt verhindert, oder eine vermutete Krankheit wirksam bekämpft werden? Big Data birgt also auch Risiken bezüglich selbsterfüllender Prophezeiungen. Automatisierte Entscheidungsfindung kann sogar zu Situationen führen, in denen die sozialen und wirtschaftlichen Kosten zur Fehlerbehebung höher ausfallen als jene Kosten dafür, Fehler in Kauf zu nehmen. Zudem können sich neue Technik-Abhängigkeiten, verstärkte Quantifizierung des Menschen, Überwachbarkeit etc. negativ auf Machtverteilung, Autonomie und Privatsphäre auswirken. Automatisierte Vorschläge könnten es erheblich erschweren, die Grenze zwischen sinnvoller Intervention und übertriebener Prävention überhaupt zu erkennen. Sogar genbasierte Diskriminierung ist möglich – Menschen mit risikoreichen Genen hätten dann geringere Chancen auf soziale Entwicklung und Wohlstand.

Was tun?

Um diese Gefahren einzudämmen, ist es angemessen, den schmalen Grat zwischen (Leicht-)Gläubigkeit an automatisierte Technik und sinnvollem Technikeinsatz besser auszuloten. Das erfordert einen breiteren gesellschaftlichen Diskurs über Anspruch, Wirklichkeit und Gefahren von Big Data. Zentrale Herausforderungen für alle Akteure sind u.a.:

- Erhöhte Anforderungen an die Datenqualität und Interpretation: Big Data hat Einfluss auf die Aussagekraft von Informationen. Fundierte analytische Fähigkeiten sind daher essentiell, da letztlich Menschen betroffen sind. Daten müssen richtig interpretiert, Fehler erkannt und Ergebnisse verantwortungsvoll genutzt werden.
- Benutzerfreundlichkeit ist wesentlich, um Mensch-Maschine-Schnittstellen so zu gestalten, dass sie die Nutzung und Interpretation von komplexen Daten vereinfachen, ohne den Informationsgehalt zu stark zu verkürzen.
- Aufbau und Funktionsweise von Big-Data-Anwendungen müssen nachvollziehbar sein, um Risiken durch automatisierte (Fehl-)Entscheidungen zu verringern. Relevante Fragen sind etwa: Welche Information wird in welcher Qualität, von welchen Quellen und für welchen Zweck verarbeitet.
- Es braucht Transparenz, Verantwortlichkeit und Überprüfbarkeit von Big-Data-Analysen, besonders um Risiken von automatisierten Einzelentscheidungen und Autonomieverlusten frühzeitig entgegenzusteuern. Aus diesen Aspekten können bei steigendem Anwendungsgrad neue Regulierungsfragen entstehen.

Zum Weiterlesen

Strauß, S. (2015): Datafication and the Seductive Power of Uncertainty – A Critical Exploration of Big Data Enthusiasm. *Information*, 2015 (6), 836-847. mdpi.com/2078-2489/6/4/836

Kontakt

Stefan Strauß

E-mail: tamail@oeaw.ac.at

Telefon: +43(1)51581-6582

