



Radiotelevision Svizra Rumantscha Big Data – Eine Übersicht

Prof. Dr. Wolfgang Semar

swissuniversities

Diese Unterlagen stehen unter folgender CC-Lizenz:

Namensnennung-Keine kommerzielle Nutzung-Weitergabe unter gleichen Bedingungen 2.0 Deutschland <http://creativecommons.org/licenses/by-nc-sa/2.0/de>

FHO Fachhochschule Ostschweiz



Seite 1

Agenda

1. Was ist Big Data?
2. Was kann Big Data bereits heute?
3. Wie beeinflusst Big Data unser Leben?
4. Wo sind Grenzen von Big Data?
5. Schlussfolgerung



Was ist Big Data?

https://www.youtube.com/watch?v=uH813u7_b0s

- Big Data ist eine neue Perspektive auf die „Wirklichkeit“, die auf der Analyse einer grossen Menge an Datenpunkten fusst.
 - <https://youtu.be/rJC7B-9ZfhE>
 - https://youtu.be/v_PMdAT2SCc
 - <https://youtu.be/y4JJqyTncCA>
- Mit Big Data können wir Muster in grossen Daten-Mengen feststellen und uns somit zu neuen Fragen inspirieren lassen, von denen wir gar nicht wussten, dass wir sie stellen sollten.
- Was ist neu mit Big Data?
 - Früher hat man ein Modell entwickelt und es dann geprüft (Klassisches Big Data).
 - Heute schaut man in die Daten und findet ein Modell (Exploratives Big Data).
 - Zukünftig möchte man automatische Simulationen erhalten.

Was ist Big Data?

- Big Data steht für verschiedenste Methoden, implizit in Daten vorhandene Information mit mathematischen Funktionen explizit zu machen:
 - Klassisches Big Data schätzt den Wert einer Kenngrösse, indem es die Korrelation mit anderen Kenngrössen nutzt.
 - Obama identifizierte so im Wahlkampf die noch unentschiedenen Wähler, auf die er dann seine Kampagne konzentrierte.
 - Exploratives Big Data sucht nach bisher unbekanntem Mustern, die eventuell Bedeutung haben.
 - z.B. weil sie auf Risiken hinweisen oder interessante Gruppen von Kunden identifizieren.
 - Big Data auf „Graphen“-Basis nutzt komplexe semantische Zusammenhänge.
 - z.B. um das Fehlen von Zusammenhängen zu entdecken.
 - Neuere Methoden mit Simulationswerkzeugen.



Was kann Big Data bereits heute?

- Gezielte, personale Werbung für neue „Produkte“:
 - Angenommen Sie besitzen Daten über das Kaufverhalten von als Person identifizierten Kunden und sie möchten ein neues Produkt lancieren und gezielt bewerben.
 - Dann werden sie zuerst ähnliche, bereits existierende Produkte in ihrem Verkaufsportfolio identifizieren und dann jene Kunden herausfiltern, die diese Produkte öfters gekauft haben.
 - Die Wahrscheinlichkeit ist hoch, dass sie sich für das neue Produkt ebenfalls interessieren werden und es macht Sinn, die Marketingkampagne für das neue Produkt auf sie zu konzentrieren.
- Im Customer Relationship Management vieler Banken und Versicherungen wird seit Langem ein ähnliches Verfahren eingesetzt, um neue Finanzprodukte gezielt Kunden zu promoten.

Was kann Big Data bereits heute?

- Gezielte, personale Werbung für neue „Produkte“:
 - Im amerikanischen Präsidentschaftswahlkampf geht es darum, in genügend vielen Staaten die meisten Stimmen zu bekommen.
- In der Schlussphase des Wahlkampfes ist es wichtig, dass das Team eines Kandidaten die Unentschiedenen Wähler identifiziert und seine ganze Energie auf deren Überzeugung konzentrieren.
- Die Kandidaten-Teams konnten mittels Big Data die Unentschiedenen identifizieren.
 - Sie nutzten dabei die Tatsache, dass in den USA über Personen weit mehr Daten einfach beschaffbar sind als in Europa (Cambridge Analytics).

Was kann Big Data bereits heute?

- Gezielte Suche in der Kriminalität:
 - Beziehungskonstellationen oder Transaktionsketten identifiziert, die z. B. auf Geldwäscherei hinweisen.
- Medizinische Diagnostik:
 - 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.

Was kann Big Data bereits heute?

- Herstellen von Produktkombinationen:
 - Eine Zusammenstellung von kundenspezifischen Extras kann besonders attraktiv sein, wenn man diese als Paket verkauft, wobei der Kunde beim Kauf des ganzen Pakets einiges „spart“.
 - Um verlockende Pakete zu schnüren, ist es aber notwendig, zu wissen, welche Kombinationen von Extras für Kunden besonders attraktiv sind.
 - Deshalb wird in Daten über Kundenpräferenzen nach Mustern gesucht, die auf attraktive Paketzusammenstellungen hinweisen.
 - Dieses Vorgehen wird freilich nicht nur im Autohandel praktiziert.
 - Ganz ähnlich lassen sich so auch Produkte zusammensetzen, die am Ende teurer verkauft werden können als ihre Einzelbestandteile – nicht zuletzt im Lebensmittelhandel.

Wie beeinflusst Big Data unser Leben?

- Heutige Anwendungsbereiche für Big Data sind:
 - Marketing
 - Verkauf
 - Politikgestaltung
 - Stadtentwicklung
 - Sozial- und Gesundheitspolitik
 - öffentlichen Verwaltung (u.a. Verkehrsmanagement)
 - personalisierte Medizin
 - vorausschauende Wartung, Polizeiarbeit, ...
 - wissenschaftliche Forschung

Wo sind die Grenzen von Big Data?

- Wird ein erdappter Einbrecher zur Waffe greifen?
- Wird der Fanmarsch von Fussballfans in Gewalt ausarten?
 - Die Antwort lässt sich nicht mit Sicherheit aus den verfügbaren Daten ableiten, aber trotzdem kann sie von erfahrenen Polizisten ziemlich zuverlässig gegeben werden.
- Dabei kann man drei Phänomene beobachten:
 1. Die **Zuverlässigkeit** der Analyseergebnisse hängt davon ab, dass die richtigen Daten gesammelt werden.
 2. Der **tatsächliche Ablauf** wird durch Handlungen beeinflusst, die sich aus der Situation ergeben. Teilweise haben wir es also mit selbsterfüllenden Prophezeiungen zu tun.
 3. Sollte die **Zielvorgabe** deshalb nicht lauten möglichst präzise Prognosen zu generieren, sondern **Geschäftsziele** zu erreichen – in unserem Beispiel die Minimierung von Gewalt.

Wo sind die Grenzen von Big Data?

1. Man sollte sich immer bewusst sein, dass Modellannahmen das Ergebnis von Big Data entscheidend beeinflussen, auch dann, wenn wir scheinbar ganz ohne Modelle Daten analysieren.
 - Denn schon bei der Erzeugung von Daten spielen Modelle eine entscheidende Rolle. Es gibt in dieser Hinsicht keine natürlichen Rohdaten (Rohdatenfiktion).
2. Auch grosse Datenmengen können einen klaren Bias (systematischer Fehler, Verzerrung) haben.
 - Nur weil wir viele Daten sammeln, können wir daraus nicht ableiten, dass unsere Daten in irgendeiner Weise repräsentativ sind (Statistikfiktion).
3. Big Data liefert nicht einfach per se gute Resultate (Simplizitätsfiktion)
Es braucht mathematische, technische, fachliche und rechtliche Kompetenzen.

Wo sind die Grenzen von Big Data?

4. Ergebnisse von Big Data Analysen können ohne verständliche Erklärungsmodelle für die zugrunde liegenden Zusammenhänge oft nicht sinnvoll eingesetzt werden (Korrelationsfiktion).
 - Stellen Sie sich einen Polizeieinsatz von Wasserwerfern vor, der damit begründet wird, dass Big Data Massenunruhen prognostiziert hat – und im Nachhinein stellt sich heraus, dass eine überdurchschnittlich hohe Zahl roter Halstücher der Auslöser war.
5. Viele Daten schaffen noch keine Transparenz (Transparenzfiktion).
6. Dauerhafte Anonymisierung ist schwer zu garantieren (Anonymisierungsfiktion).
7. Big Data bringt nicht notwendigerweise Nutzen für alle (Fairnessfiktion).
 - Wer über die Daten von anderen verfügt, kann damit viel Gewinn machen. Die Masse der Datenlieferanten bekommt zwar meist im Tausch kostenlose Online-Dienste, hat aber keinen Anteil an den Milliardengewinnen und gerät im schlimmsten Fall sogar in ein Abhängigkeitsverhältnis.

Wo sind die Grenzen von Big Data?

8. Es ist zu Recht sehr umstritten, ob man hohe Komplexität durch Automatisierung kontrollieren kann (Automatisierungsfiktion).

- China ist einen Schritt weiter:
 - Dort läuft ein Testprojekt, in dem Bürger von zwölf Modellstädten automatisch bewertet werden.
 - Aus der digitalen Vollüberwachung errechnet eine KI, wer fleissig und gesetzestreu ist und wer faul ist und bei Rot über die Straße geht.
 - Der eine wird durch Vergünstigungen bei der Wohnungs- und Schulauswahl belohnt, der andere kommt auf eine öffentliche schwarze Liste.

Schlussfolgerung

- Big Data besitzt ein gewaltiges, derzeit nur in Ansätzen abschätzbares, Nutzenpotential.
- Es wird die Wirtschaft, unser persönliches Leben und das Staatswesen sehr stark verändern.
- Ignoriert die öffentliche Verwaltung das Thema, wird sie in Zukunft ihre Aufgaben nicht mehr zufriedenstellend erfüllen können.
- Gleichzeitig schafft aber Big Data auch neue, grosse Gefahren für die Gesellschaft, die sich verheerend auswirken könn(t)en.

- Es ist deshalb Zeit, dass wir uns ernsthaft mit Big Data auseinandersetzen!
- Wir brauchen eine Ethik im Umgang und in der Verwendung von Daten

Was weiss Google über mich?

<https://myaccount.google.com/dashboard>

Vielen Dank für Ihre Aufmerksamkeit!

Wolfgang.Semar@htwchur.ch

Hochschule für Technik und Wirtschaft
Chur
Schweizerisches Institut für
Informationswissenschaft

www.informationswissenschaft.ch

