

Was kann die Sozialwissenschaft von Big Data lernen?

Description

Von Florian Rabuza

Big Data ist heute ein Schlagwort, das immer häufiger in den Medien zu hören ist. Die Einflüsse von Big Data auf unsere Alltagswelt sind vor allem im Bereich des Konsums spürbar. Aber auch die Sozialwissenschaften fangen an, die Potentiale, die die Erschließung größerer, vielfältigerer Daten mit immer größerer Geschwindigkeit freisetzt, zu erkennen. Die Frage, die sich stellt: Kann die datafication unserer Alltagswelt wirklich einen Nutzen für die Sozialwissenschaft bringen?

Alle reden von Big Data. Big Data sind nicht einfach nur große Datenmengen, denn ihre Charakteristika werden allgemein treffender definiert als 3V: *volume, variety, velocity*: Immer größere Datenmengen werden immer schneller verfügbar und stammen dabei aus immer vielfältigeren Bezugsquellen und treten in sehr unterschiedlichen Formaten auf. Diese Entwicklung wird die Welt in der Zukunft nachhaltig verändern und schon heute sind die Einflüsse, die vor allem die kommerzielle Nutzung großer Datenmengen hat, für nahezu jeden im Alltag spürbar. *Recommendation engines* schlagen uns mit teilweise unglaublicher Treffsicherheit Bücher, Reisen, Musik oder Schuhe vor, die wir dann auch tatsächlich häufig kaufen, Google weiß schon nach wenigen Buchstaben, wonach wir mit großer Wahrscheinlichkeit suchen und die Möglichkeit online bestellte Artikel per Lastschrift bezahlen zu können hängt oft davon ab, ob die eigene Wohngegend auf Grundlage vorliegender Daten als zahlungssicher eingestuft ist. Der Beruf des *data scientist*, auch wenn nicht immer ganz klar zu sein scheint, was das denn genau ist, gilt als „sexiest job of the 21st century“ ([Harvard Business Review 2012](#)). Aktuelle Prognosen zufolge benötigt der Arbeitsmarkt der Zukunft Hunderttausende neuer *data scientists*.

Die Auswirkungen der *datafication*, der Prozess der Erfassung und Quantifizierung nahezu jeder Form menschlicher und nichtmenschlicher Aktivität in nahezu jedem Alltagsbereich, sind heute fast ausschließlich im Bereich des Konsums zu sehen. Die Nutzer großer Datenmengen verfolgen also hauptsächlich das Ziel Verkäufe zu steigern, indem sie Kundendaten oder Daten von Konkurrenten auswerten, die sie von ihren eigenen

Websites beziehen oder aus dem Web scrapen, also automatisiert sammeln. Und auch in Zukunft werden große Datenmengen wohl eher zum Zwecke von *business analytics* und nachrichtendienstlichen Erkennungszwecken als zur Analyse von beispielsweise Parteiensystemen eingesetzt werden. Dennoch sind sich viele sicher, dass letztendlich auch die Sozialwissenschaften entscheidend durch die *datafication* geprägt sein werden. Denn soviel scheint sicher: Eine Umkehr der Entwicklungen im Zuge der *datafication* wird es nicht geben.

Big Data in der Sozialwissenschaft

Welchen Nutzen können große Datenmengen, die sehr schnell verfügbar sind und aus vielfältigen Bezugsquellen stammen nun für die Sozialwissenschaft haben? Was ist wissenschaftlich und letztendlich gesellschaftlich gewonnen, wenn Politikwissenschaftler Tweets von Politikern auswerten oder Diskussionsbeiträge von Usern sozialer Netzwerke auswerten? Die Hauptfrage, die sich hier stellt: Für wen oder was sind die Ergebnisse solcher Analysen repräsentativ? Oder anders, was ist eigentlich die Grundgesamtheit, auf die ich rekurriere? Die Bevölkerung eines Landes wie der Bundesrepublik sicherlich nicht. Aber auch nicht die Bevölkerung zwischen 14 und 40 Jahren. Ja, wer denn dann? Man merkt relativ schnell, dass das Problem der Verallgemeinerbarkeit in besonderer Härte auftritt, da die Frage nach Repräsentativität im wörtlichen Sinne sogar von der Uhrzeit abhängt, zu der man die Daten abgreift. Das Potential liegt also sicher nicht in der Verallgemeinerbarkeit. Die Stärke der Analyse von Daten sozialer Netzwerke besteht vor allem in der Flexibilität und Geschwindigkeit, mit der sie analysiert werden können. Die Dynamiken öffentlicher Meinungsbildung können z.B. in Echtzeit abgebildet und beobachtet werden. Und dies sehr kostengünstig. Niemand wird ernsthaft behaupten, dass so etwas mit konventionellen inhaltsanalytischen Methoden in ähnlicher Weise und Geschwindigkeit möglich wäre. Ich denke, es ist kein allzu hoher Preis, wenn dafür Repräsentativität im strengen Sinne geopfert wird, denn die Erkenntnisse, die um Mechanismen von Meinungsbildung, Meinungsführerschaft und Meinungsdynamiken gewonnen werden können, sind äußerst wertvoll.

Was wurde bislang gemacht?

Schaut man sich einmal kurz an, was vor allem in der amerikanischen Forschung gemacht wurde, so wird schnell klar, dass selbst Skeptiker sich vom Nutzen neuer und größerer Datenquellen überzeugen lassen sollten. Hier ein paar Beispiele:

Es gibt sehr interessante Versuche aus dem online erhobenen Datenmaterial auf Eigenschaften der politischen Makro-Ebene zu schließen. Barbera (2013) zeigt mit einer ziemlich komplexen und originellen Analysestrategie, wie man die ideologische Polarisierung und Policy-Positionen, also Positionen bei politischen Sachfragen, in verschiedenen politischen Systemen mit Twitter-Daten messen kann. Es gibt mehrere ähnliche Studien, die sich mit Fragen von Ideologie und Polarisierung befassen. Insgesamt gibt es Hinweise, dass Soziale Netzwerke tatsächlich dazu geeignet sind, ideologische Orientierungen und Policy-Präferenzen zu messen.

Ein weiteres sehr beeindruckendes Beispiel wie eine bedeutsame politikwissenschaftliche Fragestellung mit Daten aus sozialen Netzwerken beantwortet werden konnte ist die Analyse von Gary King, Jennifer Pan und Margaret Roberts (2013) von der Harvard University. Sie analysieren mit eigens hierfür entwickelter Software die Inhalte von über 1400 sozialen Medien in China. Da die eingesetzte Software schneller ist als die chinesische Zensur, können Sie zeigen, welche Inhalte in China im Internet zensiert werden. Die Ergebnisse sind teilweise verblüffend. So werden auch offen kritische und negative politische Äußerungen nur dann mit höherer Wahrscheinlichkeit zensiert, wenn darin zu kollektiven Aktionen aufgerufen wird. Außerdem wird vor allem Pornographie zensiert.

Und in Deutschland?

Die deutsche Sozial- und Politikwissenschaft zeigt zunehmend Interesse für das Thema Big Data. GESIS als wichtigste sozialwissenschaftliche Infrastruktureinrichtung in Deutschland hat auf die zunehmende Bedeutung mit der Einrichtung einer [Abteilung für Computational Social Science](#) reagiert, die sich unter anderem mit der Beantwortung sozialwissenschaftlicher Fragestellungen mit Daten aus dem World Wide Web und der Entwicklung von Methoden für den Umgang mit Big Data beschäftigt. Im Mai 2014 findet an der Uni Duisburg-Essen eine Tagung der Sektion Methoden der Deutschen Vereinigung für Politische Wissenschaft zum Thema Big Data statt. Dennoch steckt die Forschung hierzulande noch in den Kinderschuhen, was sich an der geringen Anzahl an

Publikationen ablesen lässt. Das wird sich aber wohl sehr schnell ändern. Wenn Sie dazu beitragen wollen, sharen, liken und twittern Sie diesen Blogbeitrag.

Verwendete Quellen

[Barbera, Pablo \(2013\): Birds of the same feather tweet together. Working Paper.](#)

[Davenport, Thomas L./ Patil, D.J. \(2012\): Data Scientist: The Sexiest Job of the 21st Century. In: Harvard Business Review October 2012.](#)

[King, Gary/ Pan, Jennifer/ Roberts, Margaret E. \(2013\): How Censorship in China Allows Government Criticism but Silences Collective Expression. In: American Political Science Review 107 \(2\): 326-343.](#)

Date Created

April 9, 2014

Author

politikwissenschaft_h1c5yk