Zur Eignung von Google-Suchaktivität zur Vorhersage von Wahlergebnissen

Sebastian Krause · 24. November 2015
Über Politik aus der Wissenschaft

Von Sebastian Krause

Der Beitrag basiert auf der BA-Arbeit des Autors, die am Lehrstuhl für Empirische Politikwissenschaft durch Florian Rabuza und Achim Goerres betreut worden ist.

In jüngster Vergangenheit wurde in zahlreichen Studien bereits erfolgreich unter Beweis gestellt, dass sich die Suchaktivität von Google-Nutzern eignet, um gegenwärtige Zustände aufzudecken oder gar zukünftige Ereignisse vorherzusagen. Hierzu gehören beispielsweise das regionale Aufkommen von Grippesymptomen, die Handelsaktivität an Börsen, zahlreiche ökonomische Indikatoren, der Sieger von Sportereignissen oder gar das Eintreten terroristischer Anschläge.

Auch in der Politikwissenschaft wurde mit solchen Daten experimentiert, wobei hierfür zum bisherigen Forschungsstand stets auf das frei zugängliche Google *Trends* gesetzt wurde. Hierbei ist das wesentliche Manko der Umstand, dass die Daten zur Suchaktivität nach eingegebenen Begriffen stets im Verhältnis zum Maximum ausgegeben werden, absolute Zahlen auf diesem Wege also nicht zur Verfügung stehen. Während auf solchen Daten fußende Wahlprognosen in den USA bereits länger praktiziert werden, gibt es für deutsche Wahlen nur wenige vergleichbare Studien und die Nachfrage nach konventionellen Umfragedaten, die durch die „Sonntagsfrage“ erhoben werden, dominiert weiterhin den Markt. Dies ist zum Teil dem Umstand geschuldet, dass es für ein Mehrparteiensystem mit koalitionsstrategischen Überlegungen und der 5 %-Klausel

deutlich komplexer ist, zuverlässige Prognosen zu erstellen – sei es mithilfe von Google-Daten oder anderen Prognosemodellen, die z. B. wirtschaftliche Indikatoren oder die Beliebtheitswerte der Kanzlerkandidaten verwenden. Derartige Modelle beschränken sich zumeist darauf, den Wahlsieger oder den Stimmanteil der amtierenden Regierungskoalition vorherzusagen.

Ein Weg zu verwertbareren Daten führt über einen Google *AdWords*-Account. *AdWords* ist ein Angebot, mit dem Werbetreibende Anzeigen schalten können, die nach der Eingabe bestimmter Suchbegriffe als Suchergebnis hervorgehoben neben den Suchergebnissen erscheinen. Mit dem enthaltenen *Keyword-Tool* bietet Google seinen Kunden die Möglichkeit, Suchvolumina für die jeweils werberelevanten Suchbegriffe einzusehen und sie beispielsweise monatsweise oder nach Region, aus der die Anfragen eingespeist wurden, zu filtern. Auf diese Weise erhält man zwar gerundete, aber absolute Zahlen zu den nachgefragten Suchbegriffen, was die Prognose von Stimmanteilen für jede Partei ermöglicht. Jeder betrachteten Partei kann folglich ein absolutes Suchvolumen zugeordnet werden, um es anschließend ins Verhältnis zum Suchvolumen anderer Parteien zu setzen. Dieses Verhältnis korreliert gemäß meiner Überlegungen mit dem späteren Stimmanteil im Wahlergebnis.

Grundlage für meine Analysen ist die Annahme, dass ein Suchinteresse mit der Intention einhergeht, bei der nächsten Wahl für ebendiese Partei zu stimmen. Die deutlich ansteigende Suchaktivität im Rahmen von Wahlen ist bereits ein Indiz für die Verwendbarkeit der Daten für diese Zwecke (s. Abbildung 1). Die Konzepte der Parteiidentifikation von Angus Campbell (vgl. Campbell et al. 1960) sowie die Ökonomische Theorie der Demokratie Anthony Downs‘ (vgl. Downs 1968) unterstützen die Annahmen auf theoretischer Ebene. Wähler investieren im Zuge von Wahlen Informationskosten, um eine rationale Entscheidung für eine Partei treffen zu können. Im Informationszeitalter nehmen Wähler hierfür natürlich auch den Weg über die Internetsuchmaschinen. Google besitzt in Deutschland eine Quasi-Monopolstellung mit einem Marktanteil von derzeit etwa 94 %. Demnach sollte die Suchaktivität hier das spätere Wahlergebnis widerspiegeln.

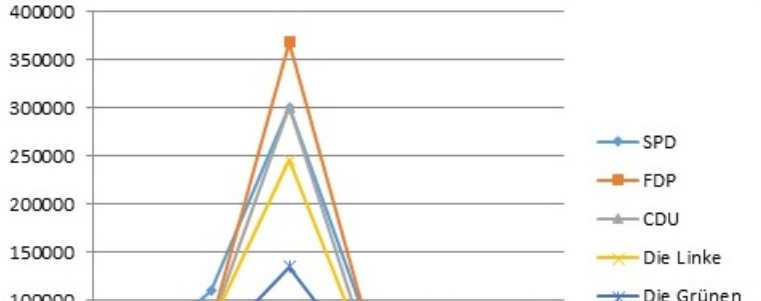


Abbildung 1: Suchvolumen nach Parteinamen im zeitlichen Verlauf der Bundestagswahl 2013. Y-Achse: Suchanfragen

Zum Zeitpunkt der Abfrage können die Daten der vergangenen zwei Jahre abgerufen werden, sodass die Bundestagswahl 2013 und die Europawahl 2014 in die Analyse einflossen. Die Zahlen lassen sich sowohl monateweise als auch nach geografischen Einheiten filtern. Um die Fallzahl für eine Regressionsanalyse zu erhöhen und aussagekräftige Ergebnisse zu erhalten, wurden die in relative Zahlen umgewandelten Suchvolumina nach Parteien aus den 16 Bundesländern jeweils mit dem Zweitstimmenanteil der Parteien (CDU/CSU, SPD, FDP, die Linke und Bündnis 90/die Grünen) auf Länderebene gegenübergestellt. Dies führt zu einer Fallzahl von jeweils 80 für beide untersuchte Wahlen. Eine weitere Vergrößerung der Fallzahl, beispielsweise indem man die geografischen Einheiten verkleinert oder weitere Parteien aufnimmt, wäre zwar wünschenswert, scheitert jedoch daran, dass teils nicht genügend Suchanfragen für einzelne Fälle vorhanden sind.

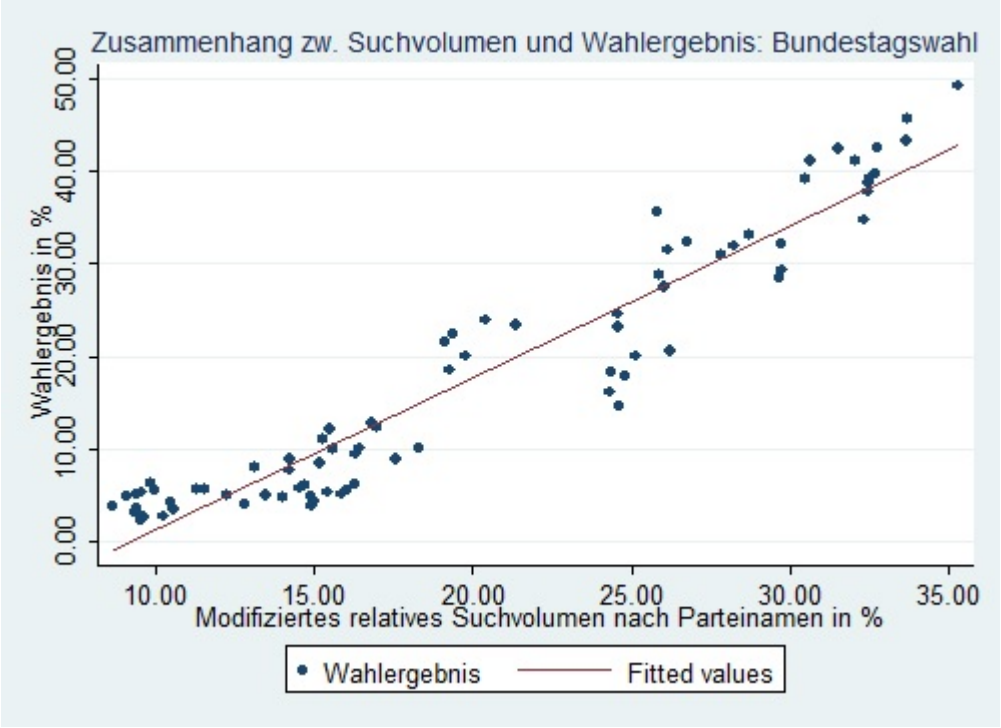


Abbildung2: Streudiagramm mit Regressionsgerade. Suche nach Parteinamen unter Verknüpfung mit programmatischen Suchbegriffen und Wahlergebnisse der Bundestagswahl 2013

Die Suchaktivität nach bloßen Parteinamen zeigt im Zuge beider Wahlen eine eher geringe Korrelation mit den Wahlergebnissen. Insgesamt kann für die Europawahl 2014

resümiert werden, dass die Suchaktivität einen zu hohen Teil der Varianz in den Wahlergebnissen unerklärt lässt. Sowohl das Suchvolumen als auch die Wahlbeteiligung sind hier deutlich geringer, was sich negativ auf die Qualität der Vorhersage niederschlägt.

Die Prognosen werden umso erfolgreicher, je mehr die Suchbegriffe mit programmatischen Aspekten verknüpft werden (z. B. „SPD Wahlprogramm“ anstelle von „SPD“). So wurden Keywordlisten für jede der fünf Parteien erstellt, für die das Suchvolumen bei *AdWords* in einem Schritt aufsummiert abgerufen werden kann. Das auf diese Weise aufgestellte Regressionsmodell für die Bundestagswahl 2013 erklärt die Varianz der Stimmanteile zu 91 %. Die durch die Regressionsgerade vorhergesagten Stimmanteile weisen im Schnitt eine nur geringe Abweichung zu den tatsächlichen Wahlergebnissen auf (s. Abbildung 2). Die experimentelle Erweiterung des Modells um die unabhängige Variable „Zweitstimmenergebnis bei der vergangenen Bundestagswahl“ erklärte die Varianz sogar zu 98 %.

In den Ergebnissen sind systematische Schwankungen der Prognosegenauigkeit nach den Parteien zu erkennen (s. Abbildung 3). So liegt z. B. die relative Suchaktivität nach CDU-Begriffen vermutlich wegen der im Schnitt älteren Parteiklientel und ihres geringeren Informationsbedarfs aufgrund der Regierungstätigkeit auffällig häufig unter dem späteren Wahlergebnis.

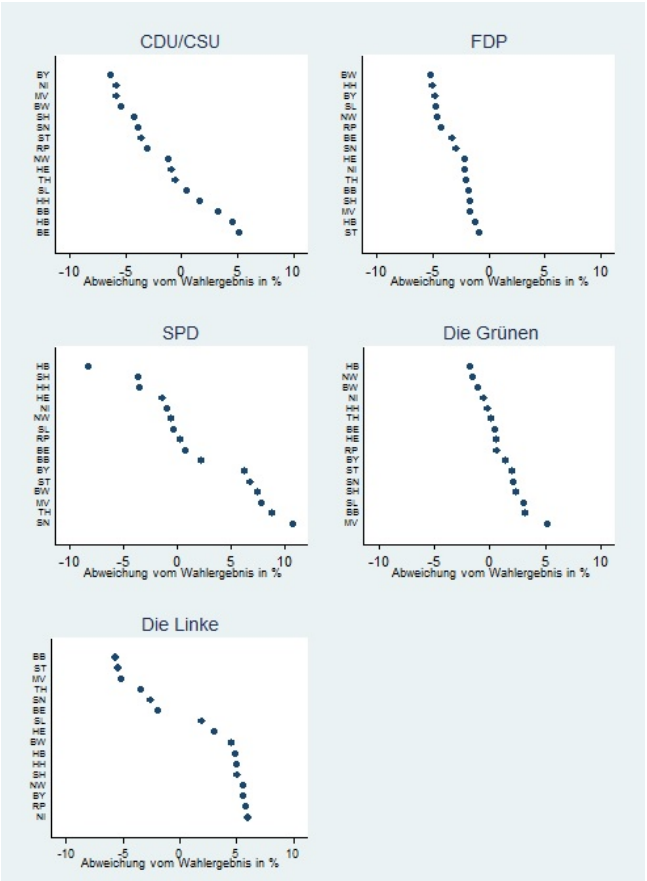


Abbildung 3: Abweichung der Vorhersage vom Wahlergebnis nach Parteien. Y-Achse: Kürzel der Bundesländer (Bundestagswahl 2013)

Es besteht eine teilweise starke Korrelation zwischen der oben operationalisierten Suchaktivität im Wahlmonat, die einer Partei zugeordnet werden kann, und dem Wahlergebnis. Bei Wahlvorhersagen – sei es in Deutschland oder im Ausland – blieb ein Teil der Wahlergebnisse stets unvorhersehbar. Es ist daher notwendig und wünschenswert in Zukunft neue und bessere Datenquellen aufzuspüren, die die Prognosegenauigkeit erhöhen. Zusammenfassend kann die Verfügbarkeit derartiger Internetnutzerdaten in Zukunft zu neuen und vielversprechenden Ansätzen im Rahmen von Wahlprognosen abseits der etablierten Umfrageforschung führen. Der Zugang zu solchen Daten ist deutlich einfacher, erfolgt schneller und verursacht im Gegensatz zu repräsentativen Umfragen nur geringe Kosten.

Verwendete Quellen:

AT Internet (2015): Suchmaschinenbarometer. Dezember 2014. Online verfügbar unter http://www.atinternet.com/wp-content/uploads/2015/01/Suchmaschinenbarometer-Dezember-2014.pdf, zuletzt geprüft am 12.09.2015.

Bank, Matthias; Larch, Martin; Peter, Georg (2011): Google search volume and its influence on liquidity and returns of German stocks. In: Financial Markets and Portfolio Management 25 (3), S. 239–264.

Campbell, A., Converse, P., Miller, W. E., & Stokes, D. (1960). The American voter. New York: John Wiley

Choi, Hyunyoung; Varian, H. A. L. (2012): Predicting the Present with Google Trends. In: Economic Record 88, S. 2–9.

Downs, Anthony (1968): Ökonomische Theorie der Demokratie. Tübingen: J. C. B. Mohr (Paul Siebeck) (Die Einheit der Gesellschaftswissenschaften, 8).

Groß, Jochen (2010): Die Prognose von Wahlergebnissen. Ansätze und empirische Leistungsfähigkeit. Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage, Wiesbaden (VS research. Forschung und Entwicklung in der analytischen Soziologie).

Preis, Tobias; Moat, Helen Susannah; Stanley, H. Eugene; Bishop, Steven (2012): Quantifying the Advantage of Looking Forward. Scientific Reports 2, 350

Sebastian Krause (2015): Zur Eignung von Google-Suchaktivität zur Vorhersage von Wahlergebnissen. Empirische Politik UDE.
https://www.politik-wissenschaft.org/2015/11/24/zur-eignung-von-google-suchaktivitaet-zur-vorhersage-von-wahlergebnissen/
Lesefassung aus dem Onlinebeitrag. Interaktive Inhalte sind in der Onlinefassung verfügbar.