

Textanalyse mithilfe der Crowd

Ein Werkstattbericht

Pola Lehmann und Malisa Zobel

Summary: In a project on party positions, researchers used the method of crowd coding to collect the data. Instead of well-trained coders, the crowd codes the relevant text sections – making use of an idea known as „swarm intelligence“. Each section of the text is evaluated by multiple individuals from the crowd, and the result is derived from an aggregate of their answers. In this workshop report scientists present the experiences with crowd coding, and outline the advantages of this method – as well as the associated challenges.

Kurz gefasst: In einem Forschungsprojekt zu Parteipositionen wurde zur Datenerhebung auf die Methode des Crowd-Coding zurückgegriffen. Dabei machen sich Wissenschaftler die Schwarmintelligenz zunutze: Anstatt gut ausgebildeter Kodierer bewertet die Crowd die zu untersuchenden Textabschnitte. Dabei wird jeder Textabschnitt von mehreren Personen bewertet, und erst aus dem Aggregat ihrer Antworten ergibt sich das Ergebnis. In diesem Werkstattbericht werden die Erfahrungen mit dem Crowd-Coding vorgestellt und die Vorteile, aber auch die Herausforderungen dieser Methode aufgezeigt.

In den letzten Jahrzehnten sind die methodischen Ansätze der Sozialwissenschaften anspruchsvoller geworden. Gleichzeitig hat die technische Entwicklung, insbesondere die zunehmende Digitalisierung der Gesellschaft, dazu geführt, dass größere Datenmengen leichter und kostengünstiger gesammelt werden können. Die Nutzung von sozialen Medien und die Vernetzung durch das Internet bieten nicht nur die Chance, Daten über individuelle Einstellungen und Präferenzen in zuvor ungekanntem Ausmaß zu erheben. Die neuen Kommunikationskanäle schaffen auch den Zugang zu potenziellen Forschungsteilnehmern, die zuvor nicht einfach zu erreichen waren. Viele Forscherinnen und Forscher wenden sich deswegen dem sogenannten Crowd-Sourcing oder Crowd-Working zu. Allen Crowd-Working-Methoden liegt die Vorstellung der Schwarmintelligenz zugrunde. Die zentrale Idee dieser Vorstellung ist, dass die Masse (auf Englisch *crowd*) gemeinsam Probleme besser löst, als ihre Teile es je einzeln könnten.

In der traditionellen Textanalyse werden Texte von Menschen manuell kodiert, die zuvor dafür ausgebildet wurden. Zunehmend werden einige dieser Aufgaben automatisiert. Die automatisierte Inhaltsanalyse greift auf unterschiedliche Algorithmen zurück – die sich zum Beispiel eines vorher festgelegten Lexikons an Stichwörtern bedienen. Der Algorithmus berechnet dann auf Basis von Worthäufigkeiten oder auch Wortkombinationen, welche Texte sich ähneln und welche nicht. Auf dieser Basis können die Texte zum Beispiel unterschiedlichen Themen zugeordnet werden.

Die neue Methode des Crowd-Coding als eine Form des Crowd-Working ist zwischen der manuellen und automatisierten Herangehensweise an die Textanalyse anzusiedeln. Wie in der klassischen manuellen Textanalyse greift auch das Crowd-Coding auf menschliche Kodierer zurück. Allerdings sind diese nicht zuvor für diese Tätigkeit ausgebildet worden. Sie sind keine Expertinnen. Deshalb eignet sich Crowd-Coding auch am besten für kleinere, klar definierte Aufgaben, die kein spezifisches Fachwissen erfordern. Da nie nur ein Coder einen Text einordnet, sondern stets mehrere, wird die Intelligenz des Schwarms genutzt: Angenommen wird, dass der Mittelwert aller aggregierten Antworten der Crowd sich dem wahren Mittelwert annähert.

Gegenüber der automatisierten Textanalyse hat Crowd-Coding den Vorteil, dass es auch Aufgaben lösen kann, die von Algorithmen nicht oder nur unbefriedigend gelöst werden. Insbesondere wenn es um die Klassifizierung bildlicher Sprache oder andere interpretative Aufgaben geht, sind menschliche Kodierer besser geeignet.

Textanalyse von Wahlprogrammen

Wir haben Crowd-Coding in einem Forschungsprojekt angewandt und konnten so Möglichkeiten und Grenzen des Crowd-Coding in der Praxis erfahren. Konkret ging es darum, die Wahlprogramme verschiedener Parteien aus 14 Ländern über mehrere Wahlen hinweg zu analysieren. Ziel war die Erweiterung eines bestehenden Datensatzes: Das Manifesto-Projekt misst schon seit über 35 Jahren, welche Themen von Parteien in ihren Wahlprogrammen besprochen werden. Einzelne Textabschnitte aus den Programmen werden dafür mithilfe von Kodierern jeweils einer von 56 Kategorien zugeordnet. Da das Kategorienschema seit den 1980er Jahren fast unverändert angewendet wird, kommen be-

stimmte Themenbereiche darin nicht vor – eben jene, die zum Zeitpunkt der Erstellung des Kategorienschemas noch nicht relevant oder Gegenstand politischer Debatten waren. Das ist zum Beispiel der Fall beim Thema Zuwanderung und Integration. Wir haben deshalb einen Teil der Wahlprogramme erneut untersucht und haben sie noch einmal nur in Hinblick auf diese Themen einordnen lassen.

Da Crowd-Coder keine Expertinnen in der Einwanderungs- und Integrationspolitik sind, war die zu bearbeitende Aufgabe kurz und einfach. Die Crowd-Coder haben zuerst eine kurze Anleitung bekommen, die die Aufgabenstellung beschreibt und Einwanderungs- und Integrationspolitik definiert. Nach der Lektüre dieser Anleitung wurde sofort mit der Kodierung begonnen. Dabei bestand eine Aufgabe immer aus mehreren zu beurteilenden Textabschnitten. Ein einzelner solcher Textabschnitte hat ungefähr die Länge eines Satzes, deshalb werden sie auch Quasisätze genannt. In der ersten Runde der Datenerhebung mussten die Crowd-Coder entscheiden, ob ein Quasisatz das Thema Einwanderung und/oder Integration behandelt oder nicht.

Um zu verhindern, dass Crowd-Coder absichtlich falsch oder nur zufällig antworten, wurde in jede Aufgabe ein Testsatz eingebaut, der für den Crowd-Coder jedoch nicht als solcher erkennbar war. Der Crowd-Coder musste wie bei anderen Quasisätzen auch eine Zuordnung treffen, allerdings war im Testsatzfall die korrekte Zuordnung schon vorgegeben. Nachdem der Coder alle Quasisätze einer Aufgabe beantwortet hatte, wurde ihm angezeigt, welcher dieser Sätze der Testsatz war und wie die richtige Antwort lautete. Wurden zu viele solcher Testsätze falsch zugeordnet, durfte der Crowd-Coder keine weiteren Aufgaben bearbeiten und seine bisherigen Kodierungen gingen nicht in den finalen Datensatz ein. Mithilfe dieses Tests konnte eine hohe Zuverlässigkeit der Crowd-Coder garantiert werden.

In einer zweiten Runde wurden dann alle Quasisätze, die in der ersten Runde positiv kodiert wurden, mit einem veränderten Kodierschema erneut kodiert. Die Aufgabenstellung war nun, zu entscheiden, ob es sich bei dem vorliegenden Quasisatz um das Thema Einwanderung oder um das Thema Integration handelte. Zudem sollte in einem weiteren Schritt entschieden werden, ob der vorliegende Quasisatz eine skeptische oder eine befürwortende Haltung zum Thema ausdrückte (für Ausnahmefälle wie rein beschreibende oder technisch formulierte Aussagen gab es auch die mögliche Klassifizierung als „neutral“). Wie in der ersten Runde wurden auch jetzt in jede Aufgabe Testsätze eingebaut, durch die unzuverlässige Crowd-Coder ausgeschlossen werden konnten, deren Kodierungen dann gelöscht wurden. Im Nachhinein haben wir evaluiert, wie gut dieser Mechanismus funktioniert hat und ob es uns tatsächlich gelungen ist, Kodierer auszuschließen, die unsere Anleitung nicht befolgt haben.

Im Mittel hat die einzelne Kodiererin in der ersten Runde 94 Prozent der Testsätze richtig eingeordnet. Das ist ein sehr gutes Ergebnis und zeigt, dass die Kodierer tatsächlich ein gutes Verständnis davon hatten, was wir von ihnen erwarteten. In der zweiten Runde sank dieser Wert auf 89 Prozent, was immer noch ein gutes Ergebnis darstellt. Um es allerdings wieder zu einem sehr guten Ergebnis zu machen, kommt uns die Idee der Schwarmintelligenz zugute, die besagt, dass nicht die individuelle Kodiererin immer richtig liegt, sondern dass das gemeinsame Ergebnis der Kodierer im Aggregat korrekt ist. Deshalb haben wir uns angeschaut, wie gut die Kodierer abschneiden, wenn man das mittlere Ergebnis von allen betrachtet, die einen Satz kodiert haben. Und die Annahme der Schwarmintelligenz bestätigt sich tatsächlich: Die Zahl der korrekt eingeordneten Testsätze liegt nun im Mittel wieder bei 95 Prozent.

Es gibt verschiedene Plattformen, die die Vermittlung von sogenannten Crowd-Workern anbieten. Die bekannteste Service-Plattform ist Mechanical Turk (MTurk) des großen Internetversandhauses Amazon. Ein weiterer Anbieter stammt aus den USA und heißt CrowdFlower. Alle Dienste vermitteln über ihre Plattformen Zugang zu einer Reihe von Crowd-Workern in aller Welt. Wir haben uns für CrowdFlower entschieden, da diese Plattform ihre Crowd-Worker über viele unterschiedliche Kanäle rekrutiert. Turker hingegen, wie die Crowd-Work-



Pola Lehmann ist wissenschaftliche Mitarbeiterin in der Abteilung Demokratie und Demokratisierung und im Manifesto-Projekt. In ihrer Forschung beschäftigt sie sich vor allem mit Fragen der programmatischen Repräsentation und untersucht, inwiefern die Präferenzen der Wähler und Wählerinnen von den Parteien aufgenommen und umgesetzt werden.

[Foto: David Ausserhofer]

pola.lehmann@wzb.eu



Malisa Zora Zobel ist akademische Mitarbeiterin am Lehrstuhl für Vergleichende Politikwissenschaft der Europa-Universität Viadrina in Frankfurt (Oder) und Gastwissenschaftlerin der WZB-Abteilung Demokratie und Demokratisierung. Sie beschäftigt sich in vergleichender Perspektive mit Parteien und Parteienwettbewerb. Ihr Forschungsinteresse gilt insbesondere der Frage des Einflusses rechtsradikaler und rechtspopulistischer Parteien auf die Migrations- und Integrationspolitik. *[Foto: David Ausserhofer]*

malisa.zobel@wzb.eu

ker auf MTurk genannt werden, kommen zum Beispiel fast ausschließlich aus den USA und Indien.

Da wir aber Texte in mehreren unterschiedlichen Sprachen analysieren wollten, war es für uns unerlässlich, dass wir auf Crowd-Worker aus vielen unterschiedlichen Ländern zugreifen konnten. CrowdFlower hat es uns zudem erlaubt, spezifische Sprach- und Ländereinstellungen vorzunehmen, sodass wir entscheiden konnten, welche Kenntnisse ein Crowd-Worker mitbringen musste, um eine bestimmte Kodierung für uns zu erledigen. Diese Beschränkungen sind wichtig, um eine hohe Qualität der erhobenen Daten zu erreichen. Als wir mit unserem Forschungsprojekt begonnen haben, gab es auch noch keine großen monetären Unterschiede zwischen den Plattformen, die uns bei der Wahl eingeschränkt hätten. Dies hat sich jedoch mittlerweile geändert. Es ist wichtig darauf hinzuweisen, dass CrowdFlower inzwischen sehr teuer ist: Ein professioneller Account kostet um die 30.000 US-Dollar im Jahr. Es gibt zwar auch eine Option für Forscherinnen und Forscher, jedoch mit Nutzungsvereinbarungen, die für die meisten Universitäten und Geldgeber schwierig sind, da alle Rechte an den Daten unter diesen Nutzungsvereinbarungen an CrowdFlower abgetreten werden.

Ethische Bedenken

Im Zusammenhang mit Crowd-Work werden immer wieder ethische Bedenken geäußert. Diese betreffen vor allem die Arbeitsbedingungen in der Crowd. Zum einen handelt es sich nämlich nicht um sozialversicherungspflichtige Jobs, zum anderen fehlt es an Mitbestimmungs- und Mitspracherechten (siehe den Beitrag von Gerber und Krzywdzinski in diesem Heft). Deshalb sollten Beiträge aus der Wissenschaft, die Crowd-Coding enthusiastisch als eine kostengünstige und zeitsparende neue Form der Datengewinnung begrüßen, mit Vorsicht genossen werden. Eine gute Forschungspraxis, die auf eine faire Bezahlung und einen respektvollen Umgang mit Forschungsteilnehmern besteht, ist auch beim Crowd-Coding unerlässlich. Deshalb hat uns auch interessiert, wie die Crowd-Worker im Nachhinein die Arbeit an den von uns gestellten Tasks bewertet haben. Dafür konnten die Crowd-Worker zu unterschiedlichen Fragen ein bis fünf Punkte vergeben. Die durchschnittliche Zufriedenheit mit der Bezahlung lag bei vier Punkten. Ein recht gutes Ergebnis, aber vor allem bei der Bewertung der Testsätze und dem Schwierigkeitsgrad der Aufgabe sank die Zufriedenheit in der zweiten Runde deutlich, sie lag im Schnitt nur noch bei gut drei Punkten.

Hierzu müssen wir selbstkritisch einräumen, dass wir offensichtlich den Aufwand unterschätzt haben. Zum Beispiel hatten die Crowd-Worker die Möglichkeit, nach Aufdecken der richtigen Antworten zu den Testsätzen einen Kommentar zu hinterlassen, wenn sie den Testsatz zu schwer fanden oder mit unserer Antwort nicht einverstanden waren. Wir mussten feststellen, dass wir nicht in der Lage waren, auf diese Kommentare in jedem Fall zeitnah zu reagieren und damit eine größere Wertschätzung der Arbeit auszudrücken. Gute Forschung, die sich an ethische Grundsätze hält, kostet Geld und Zeit – das wird auch in der digitalen Arbeitswelt so bleiben.

Zusammenfassend lässt sich sagen, dass Crowd-Coding eine vielversprechende Methode der inhaltlichen Textanalyse ist. Im Crowd-Coding können Texte auch dann eingeordnet und bewertet werden, wenn sie bildliche Sprache oder Interpretationen und Deutungen sozialer Zusammenhänge enthalten. Allerdings löst Crowd-Coding keinesfalls alle Probleme, die bei der Textanalyse auftreten. So spielen auch beim Crowd-Coding die Sprachbarrieren weiterhin eine große Rolle. Wie in der automatisierten Textanalyse sind auch beim Crowd-Coding englischsprachige Texte weitaus einfacher zu analysieren als andere Sprachen. Denn die Testsätze müssen jeweils in der Sprache gestellt werden, in der auch der zu analysierende Text geschrieben ist, damit sie nicht sofort als Testfragen erkannt werden und ihre Funktion verlieren. Nur wenige Forscherinnen und Forscher sprechen allerdings mehr als drei Sprachen, was insbesondere international vergleichende Untersuchungen vor Herausforderungen stellt. Wir hatten im WZB das Glück, an einem Forschungsinstitut zu arbeiten, das sich durch

eine hohe Internationalität auszeichnet. Ohne die großzügige Hilfe von Kolleginnen und Kollegen wäre das Crowd-Coding so vieler unterschiedlicher Sprachen nicht möglich gewesen. Eine weitere Herausforderung ist die Entwicklung der Crowd-Work-Plattformen selbst. Im Falle von CrowdFlower haben uns der schnelle Preisanstieg und die plötzliche Änderung der Nutzungsrichtlinien vor unerwartete Probleme gestellt. Es bleibt abzuwarten, wie andere Plattformen sich im Druck von Wettbewerb und Rentabilitätsanforderungen entwickeln werden.

Literatur

Benoit, Kenneth/Conway, Drew/Lauderdale, Benjamin E./Laver, Michael/Mikhaylov, Slava: „Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data“. In: American Political Science Review, 2016, Vol. 110, No. 2, pp. 278–295.