

Methodenbericht Wahl Navi

1 Einleitung

Das Wahl Navi, entwickelt von einem Team aus Sozial- und Datenwissenschaftlern von Vox Pop Labs, ist eine Anwendung für bürgerschaftliches Engagement. Das Ziel der Anwendung ist eine Förderung der Wahlkompetenz sowie die Partizipation der Öffentlichkeit während des Wahlkampfes.

Basierend auf den Antworten eines Nutzers auf eine Reihe von Aussagen die wichtige Themen des Wahlkampfes spiegeln, berechnet das Wahl Navi die Übereinstimmung zwischen den persönlichen Ansichten eines Nutzers und den Positionen von politischen Parteien.

Die Positionen der Parteien werden anhand einer Umfassenden Durchsicht von öffentlichen Äußerungen von Politikern zu den Themen im Wahl Navi ermittelt. Darüberhinaus wurde jede Partei, die im Wahl Navi aufgeführt ist, direkt in den Prozess eingebunden und zu mehreren Zeitpunkten eingeladen unsere Erkenntnisse zu begutachten und Feedback zu geben.

Die Ergebnisse des Wahl Navi haben nicht den Anspruch Wahlempfehlungen zu geben und sollten auch nicht so interpretiert werden. Genausowenig sind die Ergebnisse des Wahl Navi Vorhersagen für welche Partei oder welchen Kandidaten ein Nutzer wählen wird. Es ist vielmehr als Einstiegspunkt in eine tiefere Auseinandersetzung darüber, inwieweit sich Parteien in Hinblick auf die relevanten Themen des Wahlkampfes unterscheiden, konzipiert.

2 Die Positionen von Parteien und Kandidaten pro Frage

Die Ausgestaltung des Wahl Navi Fragebogens folgt einem zweistufigen Forschungsansatz. Zuerst wird eine Inhaltsanalyse der am häufigsten auftauchenden Themenbereiche in offiziellen Dokumenten der Parteien und öffentlichen Aussagen von Politikern sowie

in den Medien durchgeführt. Aus einer ersten Liste an potentiellen Fragen suchen wir die Fragen für den endgültigen Fragebogen danach aus, inwieweit sich die Antworten auf die Fragen zwischen Parteien oder Kandidaten und Wählern unterscheiden lassen; wie häufig ein Thema in den Medien auftaucht und wie wichtig eine Frage mit Blick auf die nächsten Wahlen ist.

Zweitens werden die Positionen von Parteien und Kandidaten zu den Fragen im Wahl Navi aus öffentlichen Aussagen der Parteien und Kandidaten abgeleitet. Das wissenschaftliche Team hinter dem Wahl Navi unternimmt dazu eine umfassende Durchsicht von Dokumenten der Parteien und Kandidaten, dazu gehören Wahlprogramme, Internetauftritte, Reden, Pressemitteilungen, Debatten in den Parlamenten und Statements in den Medien, um daraus die Positionen der Parteien und Kandidaten zu den Fragen im Wahl Navi möglichst genau abzuleiten. Dabei werden aktuellere Statments bevorzugt. Im Detail werden öffentliche Äußerungen folgendermaßen nach ihrem Veröffentlichungsdatum priorisiert:

1. Wahlprogramme
2. Offizielle (policy) Dokumente
3. Statements von Kandidaten, Pressemitteilungen der Parteien und (Spitzen-)Kandidaten
4. Statements von Regierungsangehörigen oder Sprecher für die Themenbereiche
5. Statements von gewählten Parteisprechern oder weiteren Kandidaten
6. Gurndsatzprogramme, Parteitagsbeschlüsse

Innerhalb dieser Richtlinien werden vor allem Statements berücksichtigt, die am ehesten dem genauen Wortlaut einer Formulierung im Wahlkompass entsprechen. Auf diesen Kalibrierungsprozess folgt eine Befragung der Parteien und/oder Kandidaten selbst. Diese beiden Schritte werden im Folgenden genauer beschrieben.

Der Kalibrierungsprozess

Auf Grundlage der gesammelten öffentlichen Statements kodieren mehrere Wissenschaftler des Wahl Navi Teams die Positionen der Partei- bzw. Kandidatenaussagen zu jeder der Fragen. Um sicherzustellen, dass die Zuordnung der Positionen unabhängig von einem Wissenschaftler ist (inter-coder reliability), kodiert jeder Wissenschaftler zunächst für sich, dann wird in einem zweiten Schritt die Kodierung auf Konsistenz überprüft. Da alle Antwortkategorien mit einer Likert-Skala oder Bewertungsskala dargestellt werden, gelten folgende Richtlinien für den Kalibrierungsprozess:

- **Stimme sehr stark zu/überhaupt nicht zu, viel weniger/viel mehr, viel stärker/schwächer, viel älter/jünger**

Die Partei oder der Kandidat unterstreicht das jeweilige Thema der Frage klar und formuliert keine Bedingungen, Voraussetzungen oder Einschränkungen für ihre/seine Haltung.

- **Stimme zu/Stimme nicht zu, stimme weniger/eher zu, etwas stärker/schwächer, etwas älter/jünger**

Die Partei oder der Kandidat formuliert Bedingungen, Voraussetzungen oder Einschränkungen für ihre/seine Haltung, oder spricht nur einen Teil der Wahl Navi These an.

- **Neutral, weder noch, gleich viel wie bisher**

Die Partei oder der Kandidat hat keine konsistente Argumentation für oder gegen sich das Thema, vermeidet eine Positionierung und/oder spricht das Thema nur indirekt an.

Die Kalibrierung der Fragen zu Steuern und Ausgaben erfolgt anhand der Befürwortung nominaler Veränderungen. Für den Fall, dass eine Partei oder ein Kandidat eine Erhöhung/Senkung von Steuern oder Ausgaben befürwortet, die bereits in Plenarsitzungen beschlossen wurde, aber noch nicht in Kraft getreten ist, wird dies ebenfalls als Unterstützung für eine nominale Veränderung gewertet.

Um sicherzustellen, dass die Ergebnisse dieses Kalibrierungsprozesses für die Nutzer transparent sind, werden alle Positionen der Partei und der Kandidaten und die öffentlichen Statements mit URLs im Wahlkompass Tool unter „Du vs. Partei“ und „Partei vs. Partei“ auf der Ergebnisseite zur Verfügung gestellt. Diese Informationen ermöglichen es den Nutzern, die eigenen Antworten mit denen der Parteien oder Kandidaten zu vergleichen und sich näher mit den Parteien- oder Kandidatenplattformen und öffentlichen Dokumente zu beschäftigen.

Befragung von Parteien und Kandidaten

Darüberhinaus befragt das Team von Wahl Navi auch direkt die politischen Parteien oder Kandidaten. Die Parteien/Kandidaten bekommen zunächst eine Kopie des Wahl Navi Fragebogens zugesandt und werden gebeten, sich selbst zu den Fragen zu positionieren. Nach dem Eingang des ausgefüllten Fragebogens stimmt das Team von Wahl Navi die Aussagen der Partei/des Kandidaten mit den Kalibrierungen, die das Kodierer-Team erstellt hat, ab. In den allermeisten Fällen stimmen die jeweiligen Aussagen der Partei/des Kandidaten und des Wahl Navi Forschungsteam überein. Sollten Unterschiede vorhanden sein, schickt Wahl Navi der Partei/dem Kandidaten einen Abweichungsbericht, in dem die übereinstimmenden Aussagen und die abweichenden Aussagen im endgültigen Fragebogen aufgelistet sind. Alle Abweichungen werden gekennzeichnet und mit den öffentlichen Statements der Partei oder des Kandidaten versehen, die vom Forschungsteam gesammelt wurden und Grundlage der Kalibrierung

des Wahl Navi waren.

Die Partei oder der Kandidat kann dann zu jeder abweichenden Kalibrierung Stellung nehmen und ihre/seine Position darstellen sowie alternative öffentliche Statements zur Verfügung stellen, die ihren/seinen Standpunkt zum fraglichen Thema untermauern. In den Fällen, wo die Partei/der Kandidat wichtige Statements vorlegt, die schlüssig seine Selbsteinschätzung dokumentieren, wird Wahl Navi eine Anpassung für die Positionierung vornehmen. Wo Unterschiede so nicht gelöst werden können, werden die strittigen Positionierungen dem wissenschaftlichen Beirat des Wahl Navi zur Begutachtung und zur endgültigen Entscheidung vorgelegt. Der wissenschaftliche Beirat setzt sich aus führenden Politikwissenschaftlern mit einem Fokus auf Wahlen und Wahlkämpfe zusammen. Im Anschluss werden die endgültigen Kalibrierungen an die Partei/den Kandidaten zur Überprüfung geschickt. Parteien und Kandidaten können über den gesamten Zeitraum des Wahl Navis Positionierungen anfechten und öffentliche Statements für eine Anpassung der Position nachreichen. Sollte sich der Standpunkt einer Partei/eines Kandidaten zu einem bestimmten Thema ändern, oder die Partei/der Kandidat zusätzliche offizielle Unterlagen einreichen wollen, die während des Kalibrierungsprozesses nicht berücksichtigt wurden, werden wir die entsprechende Kalibrierung noch einmal daraufhin überprüfen, ob eine Änderung angebracht ist. Wir ermuntern die Parteien und Kandidaten ausdrücklich dazu, während des Wahlkampfes mit uns Kontakt aufzunehmen. Darüber hinaus werden wir während des Wahlkampfes alles dafür tun eine möglichst genaue Kalibrierung der Parteien und Kandidaten auf der Grundlage öffentlicher Statements sicherzustellen.

3 Wahl Navi Algorithmen & Visualisierungen

Die Darstellung der Ergebnisse des Vote Compass besteht aus zwei Hauptelementen. Jedes dieser Elemente hilft den Nutzern dabei zu erkennen, wie ihre Standpunkte von Parteien oder Kandidaten aufgegriffen wurden, und übergreifend besser zu erkennen, wie diese Standpunkte sich in die politische Landschaft einordnen lassen. Das erste Element ist eine graphische Darstellung der Positionen von Nutzer und Parteien/Kandidaten innerhalb einer abstrakten politischen Landschaft. Zweites Element ist ein Balkendiagramm, das den Grad der Übereinstimmung des Users mit jeder Partei/jedem Kandidaten über mehrere Themen veranschaulicht.

Konsequenz der zwei verschiedenen Maße ist, dass es ab und an Abweichungen zwischen Partei/Kandidat geben wird, die als „am nächsten“ in der politischen Landschaft und „am nächsten“ im Balkendiagramm angezeigt werden. Ein Grund hierfür ist, dass diese Graphen auf unterschiedlichen Konzepten basieren. Außerdem gibt es keine perfekte Messung für politische Differenzen, weder ideologisch noch mit Blick auf einzelne Themen. Bei einem Messwerkzeug dieser Art ist es unabdingbar den Kompromiss zwischen hohem methodischen Anspruch und dem Erfordernis, eine von

jedermann möglichst leicht verständliche Methode anzuwenden, zu akzeptieren. Dem trägt die Nutzung verschiedener Maße Rechnung. Es heißt, die Realität zu akzeptieren, dass die politische Welt – sowohl unter Politikern als auch in der Öffentlichkeit – komplex ist. Es ist aber gerade diese Komplexität, die Politik so lebendig und kontrovers macht und erklärt, warum erfolgreiche politische Arbeit und Strategien oft großartige Vorstellungskraft von Seiten der Öffentlichkeit und auch politischen Vertretern erfordert. Zweck des Wahl Navi ist daher, die Nutzer zu ermutigen, sich mit dieser Komplexität zu beschäftigen und zu erkennen, wie Parteien und Kandidaten sich zu bestimmten Themen positionieren und warum; und auch generell das Niveau und die Qualität der politischen Informationen in der Öffentlichkeit zu verbessern. Aus diesem Grund ermutigen wir alle Nutzer, sich die Begleitdokumentation auf der Ergebnisseite durchzulesen. Dort wird beschrieben, wie und warum Parteien und Kandidaten in der jeweiligen Weise kodiert wurden.

4 Die graphische Darstellung der politischen Landschaft

Die Platzierung von Nutzern und Parteien auf der Graphik zur politischen Landschaft ergibt sich aus den Antworten zu persönlichen Einstellungen und politikbezogenen Fragen. Jede Frage wird der ideologischen Dimension zugeordnet, in die sie am besten passt, sowohl in theoretischer als auch statistischer Hinsicht. Die Dimension, auf der sich jede Frage findet, wird sowohl durch die Theorie als auch durch die Auswertung der Untersuchungsdaten bestimmt, die wir während des Entstehungsprozesses der Anwendung gesammelt haben.

Um die große Anzahl der Fragen zu persönlicher Einstellung und politikbezogene Fragen zweidimensional darzustellen, setzen wir eine statistische Technik zur Dimensionenreduzierung ein. Ziel hierbei ist es, den Nutzern aufzuzeigen, was ihre Antworten im Allgemeinen darüber aussagen, wo sie in einer einfach zu verstehenden politischen Landschaft anzusiedeln sind. Wir benutzen hierfür eine statistische Technik, die sich Faktorenanalyse nennt und es uns ermöglicht, die zugrundeliegenden Positionen der Nutzer und Parteien einer kleinen Anzahl abstrakter politischer Dimensionen (Faktoren) zuzuordnen. Man könnte hier an eine Position auf einer Skala denken (z. B. bei sozialen Themen), die wir nicht direkt aufgrund einer einzelnen Antwort ersehen können, sondern erst aus einer Anzahl von Fragen, die sich um ein Thema drehen. Das funktioniert, weil die persönliche Einstellung zu einem Thema oft auch etwas über die Einstellung zu anderen Themen aussagt. Latente Variablenmodelle erlauben Analysten diese Positionen herauszuarbeiten, indem sie viele Fragen gemeinsam betrachten und sich die Beziehung dieser Fragen untereinander anschauen. Antworten, die eng mit der latenten Dimension verknüpft sind, haben daher größeres Gewicht bei der Zuordnung eines Nutzers auf einer Dimension als solche, die nur schwach verknüpft sind. Um

diese Dimensionen zu bestimmen, analysieren wir Daten, die vor dem Start der des Wahl Navi erhoben wurden.

In diesem Absatz legen wir die Annahmen des Modells und die Hauptschritte zur Ermittlung der abstrakten politischen Dimension dar. p ist die Anzahl der Wahl Navi Aussagen. X ist der Vektor der Antworten eines Nutzers auf die p Entscheidungsmöglichkeiten mit $X \in \mathbb{R}^p$. Unter der Annahme, dass wir den Vektor Z von k latenten Variablen (d. h. die Faktoren) finden können, die die Antworten des Users beeinflussen, so dass $Z \in \mathbb{R}^k$. Dann kann die Beziehung zwischen X und Z folgendermaßen ausgedrückt werden:

$$X = \mu + \Lambda Z + \epsilon, \quad (1)$$

mit $Z \sim N(0, I)$, $\epsilon \sim N(0, \Psi)$, und $\Lambda \in \mathbb{R}^{p \times k}$ ist eine Matrix der Faktorladungen. Darüber hinaus nehmen wir ebenfalls an, dass $\text{cov}(X_i, Z_j) = 0$ für $i = 1, \dots, p$ und $j = 1, \dots, k$. Aus Gleichung 1 folgt, dass $X|Z$ verteilt ist als $N(\mu + \Lambda Z; \Psi)$. Aus den Eigenschaften der multivariaten Normalverteilung, folgt die gemeinsame Verteilung von (X, Z) $N(\mu_{xz}, \Sigma)$ mit

$$\mu_{xz} = \begin{pmatrix} \mu \\ 0 \end{pmatrix} \quad \text{und} \quad \Sigma = \begin{pmatrix} \Lambda\Lambda^T + \Psi & \Lambda \\ \Lambda^T & I \end{pmatrix}. \quad (2)$$

Um dieses Modell zu implementieren, nutzen wir das package “factanal” in R. Die Faktorladungen wurden dabei mit der Maximum-Likelihood-Method geschätzt, auf welche dann die Varimax Rotation angewandt wurde. Wir haben jede der theoretischen Dimensionen auf Basis der Güte der Faktorladungen auf jede der Aussagen konstruiert. Nachdem die Gesamtheit der Aussagen für eine spezielle Dimension ausgewählt wurde, haben wir diese Teilmenge an Aussagen induktiv auf eine einzige Dimension projiziert. Diese einzige Dimension für einen Nutzer i wurde mit der folgenden Regressionsmethode erlangt

$$\tilde{z}_i = \hat{\Lambda}^T S^{-1}(x_i - \bar{x})/s_x \quad (3)$$

wo \bar{x} , s_x , und S ein Vektor des Mittelwerts, ein Vektor der Standardabweichungen und eine Stichproben Korrelationsmatrix der Antworten der Nutzer auf die Teilmenge der Aussagen, und $\hat{\Lambda}$ ist ein Vektor der geschätzten Ladungen mit einer Länge gleich der Anzahl an Aussagen in der Teilmenge.

5 Die graphische Darstellung der Themen

Die Werte der Nutzer in den Balkendiagrammen wird mit Hilfe der absoluten Distanz zwischen der Position des Nutzers zu einem Thema und der Position der Partei/des Kandidaten berechnet (auch bekannt als Manhattan Distanz). Nehmen wir zum

Beispiel an, dass für eine bestimmte Frage 5 mögliche Antworten zur Verfügung stehen, denen jeweils eine Zahl zugewiesen ist (unten in Klammern):

Stimme überhaupt nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu/teils nicht zu (3)	Stimme eher zu (4)	Stimme voll und ganz zu (5)
	^		^	^
	Nutzer		Partei A	Partei B

Anhand dieser hypothetischen Positionen von Nutzer und Parteien errechnen wir, dass der Nutzer zwei Einheiten von Partei A entfernt ist (d. h. $|2 - 4|$) und drei Einheiten von Partei B entfernt ist (d. h. $|2 - 5|$). Die Summe dieser Abstände wird für jedes Thema errechnet, steht also für die Gesamtentfernung eines Nutzers von einer bestimmten Partei. Da wir es vorziehen, die Entfernungen auf dem Balkendiagramm danach darzustellen, wie *nah* ein Nutzer an einer Partei ist und nicht wie weit entfernt er von einer Partei ist, ziehen wir die ermittelte Entfernung von der Entfernung ab, die einer Partei zugewiesen würde, die so weit wie nur möglich vom Nutzer entfernt ist und teilen dann durch diese Entfernung. Wenn eine Partei eine Position einnimmt, die so weit wie nur möglich vom Nutzer entfernt ist, wird der Balken den Wert 0 anzeigen. Wenn eine Partei eine Position einnimmt, die exakt der des Users entspricht, wird der Balken den Wert 100 anzeigen.

Wir tragen auch der Tatsache Rechnung, dass sich Nutzer für manche Themen mehr interessieren als für andere, indem wir den Nutzer die Möglichkeit geben anzugeben, wie viel mehr er sich für ein Themengebiet interessiert als für andere. Wenn ein Nutzer diese Gewichtung angegeben hat, werden die Entfernungen im Balkendiagramm entsprechend gewichtet: Die Entfernungen eines Nutzers zu einer Partei oder einem Kandidaten bei Themen, für die sich der Nutzer sehr interessiert, werden vergrößert; die, für die er sich nicht so sehr interessiert, verkleinert.

Um genauer zu sein legen wir die Schritte zur Berechnung der Balkendiagramme formal dar. n sei die Anzahl an Aussagen, und $x_i, x_{i,p}$ sind die jeweiligen Positionen eines Nutzers und einer Partei/eines Kandidaten zum Thema i . Somit kann die absolute Distanz zwischen dem Nutzer und der Partei/dem Kandidaten p wie folgt ausgedrückt werden:

$$Differenz_p = \sum_{i=1}^n |x_i - x_{i,p}| . \quad (4)$$

Im nächsten Schritt berechnen wir die maximal mögliche Distanz zwischen Partei/Kandidat und Nutzer, gegeben der Antworten des Nutzers (die Antworten sind auf einer Skala von 1 bis 5 kodiert):

$$\mathit{maxDifferenz} = \sum_{i=1}^n (|x_i - 3| + 2) . \quad (5)$$

Diese Gleichung zentriert die Skala und nimmt den absoluten Wert der Antwort des Nutzers um seine Distanz vom Zentrum zu bestimmen. Zudem wird 2 addiert, der Wert der maximalen Distanz einer Partei/eines Kandidaten vom Zentrum. Zum Beispiel wenn ein Nutzer 1 antwortet, ist eine Partei Position 5 die maximale Distanz (4) vom Nutzer entfernt. Wenn ein Nutzer '2' antwortet, ist eine Partei Position von '5' die maximale Distanz (3) vom Nutzer entfernt. Wenn ein Nutzer '3' antwortet, ist eine Partei Position von '1' oder '5' die maximale Distanz (2) vom Nutzer entfernt, und so weiter. Dadurch findet die Gleichung die Summe der maximalen Distanzen zwischen Partei und Nutzer für jede der Fragen.

Der endgültige Übereinstimmungswert zwischen Partei/Kandidat p und dem Nutzer ist

$$\text{Übereinstimmungswert}_p = 1 - \frac{\mathit{Differenz}_p}{\mathit{maxDifferenz}} . \quad (6)$$

Für den Fall, dass die Themen vom Nutzer nach Wichtigkeit gewichtet wurden, werden die $\mathit{Differenz}_p$ und $\mathit{maxDifferenz}$ wie folgt berechnet:

$$\mathit{Differenz}_p = \sum_{i=1}^n w_i |x_i - x_{i,p}| , \quad (7)$$

$$\mathit{maxDifferenz} = \sum_{i=1}^n w_i (|x_i - 3| + 2) . \quad (8)$$

Der Übereinstimmungswert zwischen Partei/Kandidat p und Nutzer mit gewichteten Themen wird dann wie in Gleichung 6 berechnet.