

Contents

| | |
|--|----------|
| 1 A scaling Model for Estimating Time-Series Party Positions from Texts | 1 |
| 1.1 Introduction | 1 |
| 1.2 Current Methods for Estimating Party Positions | 1 |
| 1.2.1 Expert Surveys | 2 |
| 1.2.2 Hand Coding: Comparative Manifestos Project | 2 |
| 1.2.3 Computer-Based Content Analysis | 2 |
| 1.3 A Scaling Approach to Party Positions | 3 |

1 A scaling Model for Estimating Time-Series Party Positions from Texts

1.1 Introduction

- Textanalyse mittels *Wordfish* zur Schätzung von deutschen Parteipositionen auf Rechts-Links Skala zw 1990 u 2005
 - Parteipositionen sind abstrakte Konzepte die nicht direkt beobachtbar/messbar sind
 - * bislang Messversuche via Surveys, Handkodierung von Parteiprogrammen, computergestützte Kodierung von Parteiprogrammen
 - Text will mit statistischen Modell, das auf existierenden Methoden aufbaut + erweitert, Beitrag leisten und Parteipositionen & verbundene Unsicherheiten über die Zeit hinweg mittels Worthäufigkeiten in Parteiprogrammen untersuchen
 - * Einführung eines neuen Modells und Vergleich mit Existierenden

1.2 Current Methods for Estimating Party Positions

- Parteipositionen sind nicht beobachtbar → *latente Variable*
 - werden indirekt durch Aktivitäten preisgegeben, zB Parteiprogramme
- drei Parteiposition Analysemethoden

1.2.1 Expert Surveys

- in idealer Welt beste/genaueste Methode
- Experten können große Mengen aus verschiedensten Quellen analysieren
- allerdings teuer, aufwendig und eingeschränkte Vergleichbarkeit über Studien/Surveys hinweg

1.2.2 Hand Coding: Comparative Manifestos Project

- CMP Gruppe kodiert Parteiprogramme von Hand (umfangreiches Projekt)
- Score für die Parteien ist der jeweilige Prozentsatz von vorhandenen (Quasi-)Sätzen die in eine von 56 Issue Kategorien fallen
- um davon Links-Rechts Positionierung abzuleiten zB Spezifizierung von Issues die insbed. links/rechts relevant ist und somit Berechnung eines Links/Rechts Scores
- the hand-coding approach provides the only cross- sectional time-series database on party positions to date

1.2.3 Computer-Based Content Analysis

1. Laver, Benoit & Garry: Verwendung von Referenztexten statt selbstkodierter Diktionäre
 - zunächst Auswahl von Referenztexten die einem gewissen Extrem im pol Raum zugeordnet werden können (evtl auch für Mitte)
 - begleitet Zuweisung von Referenzwerten an die Referenztexte (idealerweise im Kontext bereit schon existierender Expertsurveys)
 - mittels Computerprogramm *Wordscores* zählen wie oft Worte aus Referenztexten in Quelltexten auftauchen und Vergleich (evtl falsch?)
// "computer program Wordscores then counts the number of times each word occurs in the reference texts and compares these counts to word counts from the texts being analyzed"
 - Parteiprogramme werden auf ein Kontinuum zwischen den Referenztexten platziert, basierend darauf wie nah die Word Counts an Referenztexten sind

Wordscores Probleme:

- Nützlichkeit kann durch das Finden von passenden Referenztexten und Referenzwerten gehindert werden
 - keine Einigkeit darüber was politische Extreme betrifft
- Zuweisung selber Gewichtung der Worte im Estimation process
 - somit verzerren häufig auftretende Worte ohne pol Gewicht (neutral) die Einschätzung des Dokuments in Richtung der Mitte
 - Probleme bei Analyse über längeren Zeitraum
 - * pol Lexikon unter dauerhaftem Wandel
 - * schwierig bei Auswahl von Referenztexten

1.3 A Scaling Approach to Party Positions

- Annahme dass Vokabular in Parteiprogrammen Informationen über Position(en) im politischen Raum enthält
- Vorteile:
 - bessere Einschätzungen über Zeiträume
 - benötigt keine Referenztexte, da Annahme einer statistischen Verteilung von Worthäufigkeiten
 - Benutzbarkeit und Evaluierung aller Worte

Naive Bayes:

- zur Analyse von Worten/Texten wird häufig die *naive Bayes* assumption benutzt:
 - ein Text wird als Vektor von Worten repräsentiert
 - Annahme, dass individuelle Worte zufallsverteilt sind
 - * bzw: Wahrscheinlichkeit des Auftreten von Worten ist unabhängig von Positionen/Stellen anderer Worte in dem Text
 - * zwar mehr oder weniger bewiesen, dass diese Annahme falsch ist; naive Bayes dennoch realistische Klassifikation

Poisson:

- Gelehrte haben außerdem versucht eine stat. Verteilung zu finden, welche die Benutzung von Worten am akkuratesten beschreibt

- übliche Verteilungen sind bspw Poisson, negative binominal, zero-inflated(binomial) Verteilungen und andere Poisson Mixtures
- all diese Verteilungen seien stark verzerrt (ebenso ist die Nutzung von Worten)

Analyse von Worthäufigkeiten in Parteiprogrammen und Generierung der Häufigkeiten durch einen Poisson Prozess

- Wahl dieser Verteilung da sie vergleichsweise simpel ist
- nur ein Parameter λ
 - λ ist hier beides, Mittelwert (arithm. Mittel) und Varianz
 - Häufigkeit der Verwendung des Wortes j von Partei i in Wahljahr t wird einer Poisson Verteilung entnommen
 - dieses spezifische Modell ist im Grunde genommen ein *Poisson naive Bayes model*

$$y_{ijt} \text{Poisson}(\lambda_{ijt})$$

$$\lambda_{ijt} = \exp(\alpha_{it} + \Psi_j + \beta_j * \omega_{it})$$

- y_{ijt} ist Anzahl des Wortes j in Partei i 's Programm zum Zeitpunkt t
- α is a set of party-election year fixed effects
 - um Möglichkeit das Parteien vllt in manchen Jahren viel mehr/viel längere Parteiprogramme verfasst haben als in anderen
- Ψ is a set of word fixed effects
 - um den Umstand das einige Worte von allen Parteien häufiger genutzt werden als andere Worte
- β is an estimate of a word specific weight capturing the importance of word j in discriminating between party positions
 - erlaubt Rückschluss darauf, welche Worte sich unterscheiden zwischen Parteipositionen
- ω is the estimate of party i 's position in election year t (therefore it is indexing one specific manifesto)

- Parteiposition in jedem Wahljahr
- Modell behandelt jedes Parteiprogramm als separate Parteiposition und alle Positionen werden zeitgleich/gleichzeitig geschätzt F5 mitte absatz