Twitter in the Parliament - A Text-based Analysis of German Political Entities

Patrick Schulze, Simon Wiegrebe

Supervisors:
Prof. Dr. Christian Heumann, Prof. Dr. Paul W. Thurner

7. Juli 2020

Main Contributions

- Construction of data set containing more than 500k Tweets and more than 90 variables using web scraping and twitter scraping
- Topic modeling of German MP Tweets
- Extension of analytical tools available for examination of topic-metadata relationships
- Discussion of causal Inference framework within a topic modeling context

Topic Modeling

- *Latent Dirichlet Allocation* (LDA) by Blei, Ng, and Jordan (2003) as first probabilistic topic model
- Based on LDA and other topic models: *Structural Topic Model* (STM), by Roberts, Stewart, and Airoldi (2016)

Data

- MP-level data scraped from *www.bundestag.de/abgeordnete* using BeautifulSoup and Selenium Web Driver
- Electoral-district-level social-economic data extracted from *www.bundeswahlleiter.de*
- German federal election 2017 results retrieved from *www.bundeswahlleiter.de*
- Maximum number of 3200 Tweets per MP downloaded using Tweepy API

Results

- Hyperparameter search yields 15 distinct topics
- Topic labeling conducted manually (human judgment)
- Descriptive discussion of relationship between metadata and topics
- Causal Inference: estimation of cause-effect relationships between document-sepcific features (e.g. political party) and topics

Bibliography

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003). "Latent dirichlet allocation". In: *Journal of machine Learning research* 3.Jan, pp. 993–1022.

Roberts, Margaret E., Brandon M. Stewart, and Edoardo M. Airoldi (2016). "A model of text for experimentation in the social sciences". In: *Journal of the American Statistical Association* 111.515, pp. 988–1003.