[entry]nyt/global/

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 Twitter messages by German Members of Parliament (MPs)
- 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 **blei2003latent** as first probabilistic topic model
- Based on LDA and other topic models: *Structural Topic Model* (STM), by **roberts2016model**

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 available number of 3200 Tweets per MP downloaded using the *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