Issue Competition in Parliamentary Speeches? A Computer-based Content Analysis of Legislative Debates in the Austrian Nationalrat

Description of replication material, data sets and R scripts

Introduction

This file describes all data sets and R scripts required to rerun and reproduce the analyses reported in the paper "Issue Competition in Parliamentary Speeches? A Computer-based Content Analysis of Legislative Debates in the Austrian Nationalrat". The analysis was carried out in R 4.1.2 and information on the required R packages is stored in the file *renv.lock*. To run the scripts successfully, all files need to be stored in a single folder.

R scripts and data sets

The following files are necessary for replicating the empirical analyses. The R scripts are named numerically (01-07) and listed in the correct order to perform the analysis. The included data sets (.RDS) files are necessary for the analysis.

- *01_CreateCorpus.R*: Combines data provided by the ParlSpeech V2 data set (Rauh and Schwalbach 2020) with data scraped by the author. Furthermore, it filters the data to create the corpus for analysis. In order to run this script, the data set provided by Rauh and Schwalbach (2020) for Austria needs to be downloaded: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/L4OAKN
- *02_Preprocession.R*: Preprocession of corpus texts to prepare them for classification task.
- *03_Classification.R*: Classification of speeches into issue categories at sentence-level with a semi-supervised topic classification technique following Watanabe and Zhou (2022).

- *04_Validation.R*: Calculation of validation statistics.
- *05_Analysis_Preparation.R*: Preparation of data for statistical analysis.
- *06_Analysis_Preparation_Alternative.R*: Preparation of data for alternative statistical analysis (see Appendix I).
- *07_Analysis.RMD*: Statistical analysis and creation of figures and tables.
- *Analysis_Data.RDS*: Main data set used for producing the statistical analysis, tables and figures. This data set is called from script *07_Analysis.RMD*.
- *Analysis_Data_Alternative.RDS*: Additional data set used for producing alternative statistical analysis (see Appendix I). This data set is called from script *07_Analysis.RMD*.
- *Classification_Data.RDS*: Data set storing the classified speeches at sentence level. It can be used to save time because the classification task (*03_Classification.R*) does not have to be performed then. The text is, however, blinded because most of the original speech data was not collected by the author of this paper but by Rauh and Schwalbach (2020). The complete and unblinded data set can be reproduced by running scripts *01_CreateCorpus.R*, *02_Preprocession.R* and *03_Classification.R*.
- *Corp_Nationalrat_20190101_20191022.RDS*: Corpus of parliamentary speeches from 01.01.2019 to 22.10.2019. This data set only contains speeches scraped by the author. In order to create the whole corpus used in the paper, it is necessary to download the data set provided by Rauh and Schwalbach (2020) for Austria and run the script *01_CreateCorpus.R*.
- *GovOpp_Austria.RDS*: Data set storing information on government and opposition parties. Data set is required for R scripts *05_Analysis_Preparation.R* and *06_Analysis_Preparation_Alternative.R*.
- *IO_Austria.RDS*: Data set storing information on issue ownership of parties. Data set is required for R scripts *05_Analysis_Preparation.R* and *06_Analysis_Preparation_Alternative.R*.
- *ManifestoSalience_Austria.RDS*: Data set storing manifesto data based on Müller et al. (2020). Data set is required for R scripts *05_Analysis_Preparation.R* and *06_Analysis_Preparation_Alternative.R*.
- *PartySize_Austria.RDS*: Data set storing information on party size in parliament. Data set is required for R scripts *05_Analysis_Preparation.R* and *06_Analysis_Preparation_Alternative.R*.
- *renv.lock*: This file stores information on the required R packages.
- *Validation_Data.RDS*: Data set storing manually coded speeches for validation. Data set required for R script *04_Validation.R*.

References

Müller, Wolfgang C., Anita Bodlos, Martin Dolezal, Nikolaus Eder, Laurenz Ennser-Jedenastik, Christina Gahn, Elisabeth Graf, et al. 2020. *AUTNES Content Analysis of Party Manifestos: Cumulative File (SUF edition)*. AUSSDA. https://doi.org/10.11587/DEHBL2.

Rauh, Christian, and Jan Schwalbach. 2020. *The ParlSpeech V2 Data Set: Full-text Corpora of 6.3 Million Parliamentary Speeches in the Key Legislative Chambers of nine Representative Democracies.* Harvard Dataverse. https://doi.org/10.7910/DVN/L4OAKN.

Watanabe, Kohei, and Yuan Zhou. 2022. "Theory-Driven Analysis of Large Corpora: Semisupervised Topic Classification of the UN Speeches." *Social Science Computer Review* 40 (2): 311–331. https://doi.org/https://doi.org/10.1177/0894439320907027.