Communicating in an eventful campaign: A case study of party press releases during the German federal election campaign 2021

Description of replication material

Christoph Ivanusch, Lisa Zehnter and Tobias Burst

Introduction

This file describes the code and data sets required to rerun and reproduce the analyses reported in the paper "Communicating in an eventful campaign: A case study of party press releases during the German federal election campaign 2021". To run the code successfully, all files need to be stored in a single folder.

Scripts

The following scripts are necessary for replicating the full empirical analyses. The scripts (R, Python and RMarkdown) are named numerically (01-07) and listed in the correct order to perform the analysis.

- *01_Classification.py*: This script performs the classification of press releases (*Corpus_Pressreleases_20210401_20210926.csv*). The script downloads manifesto data from the Manifesto Project corpus (Lehmann, Burst, Lewandowski, et al. 2022) by using the Manifesto API, trains ("finetunes") a multilingual BERT model on them and creates predictions for the press release data set.
  Attention: To run the script users need to provide a Manifesto Project API key for downloading the manifestos!
  Note: We recommend to load and run the script as a Google Colab notebook.
  Note: The output from this script is included in the replication material (*Classified_Pressreleases_20210401_20210926.csv*) to speed up the replication process or in case of unavailability of the raw text data.

- *02_Validation.R*: Calculation of Validation statistics. The script loads a human coded gold-standard and compares it with the output generated by the file *01_Classification.py*. The script reports accuracy, f1-score and cohen´s kappa for the classification of press releases into type (policy-related/not policy-related) and issue categories.

- *03_Preparation_Manifestos.R*: This script downloads manifesto data from the Manifesto Project corpus (Lehmann, Burst, Lewandowski, et al. 2022) for the German federal election campaign in 2021 and prepares the data for further use during the analysis.
  Attention: To run the script users need to provide a Manifesto Project API key for downloading the manifestos!

- *04_Preparation_GLES.R*: This script prepares the responses to the open-ended questions from GLES survey data (GLES Panel, Wave 16-19) for further use during the analysis.
  Attention: To run the script users need to provide the files with the responses to the open-ended questions from the GLES data available via GESIS. The links are provided below:

  - GLES Panel, Wave 16 (GLES 2021):
    https://search.gesis.org/research_data/ZA7722
  - GLES Panel, Wave 17 (GLES 2022a):
    https://search.gesis.org/research_data/ZA7723
  - GLES Panel, Wave 18 (GLES 2022b):
    https://search.gesis.org/research_data/ZA7724
  - GLES Panel, Wave 19 (GLES 2022c):
    https://search.gesis.org/research_data/ZA7725

  Note: The output from this script is included in the replication material (*GLES_Data.RDS*) to speed up the replication process or in case of unavailability of the raw survey data.

- *05_Preparation_Polls.R*: This script prepares poll data for further use during the analysis.

- *06_Analysis_Preparation.R*: This script creates the main data set used for the statistical analysis in the paper.

- *07_Figures_and_Table.Rmd*: This script reproduces the Figures and Tables included in the paper and appendix.

Data sets

The following data sets are necessary for replicating the empirical analyses or are intermediate products that can be used to speed up the replication process.

- *Analysis_Data.RDS*: This file contains the main data set used for the analysis reported in the paper and to create the Figures and Tables included in the paper and appendix.
- *Codebook.xlsx*: This file contains information on our codebook and relates the issue categories from our coding scheme to the coding scheme of the Manifesto Project (Lehmann, Burst, Matthieß, et al. 2022).
- *Corpus_Pressreleases_20210401_20210926.csv*: This file contains the main corpus of press releases used in the analysis. Unfortunately, we are not allowed to include the full texts because of copyright reasons. Therefore, this file only contains meta information on the press releases (e.g. party, date, URL, title). This meta information - in particular the URL - can be used to navigate to and collect the press releases included in the corpus.
- *Classified_Pressreleases_20210401_20210926.csv*: This file contains the classification results for each press release. The file can be used to bypass the script *01_Classification.py* and speed up the replication process. Unfortunately, we are not allowed to include the full texts because of copyright reasons.
- *GLES_Data.RDS*: This file contains the classified responses to the open-ended questions from GLES survey data (GLES Panel, Wave 16-19). The file can be used to bypass the script *04_Preparation_GLES.R* and speed up the replication process.
- *Validation_Data_coded.xlsx*: This file contains human coded press releases for validation (see: *02_Validation.R*).

BERT model

The following files constitute the BERT model that we trained on labeled manifestos from the Manifesto Project corpus (Lehmann, Burst, Lewandowski, et al. 2022) and applied to classify the press releases. The model can be used to bypass the model training procedure in the script *01_Classification.py* and speed up the replication process.

- *config.json*
- *pytorch_model.bin*

References

GLES. 2021. *GLES Panel 2021, Wave 16* [in de]. GESIS, Köln. ZA7722 Datenfile Version 2.0.0. https://doi.org/10.4232/1.13857.

———. 2022a. *GLES Panel 2021, Wave 17*. GESIS, Köln. ZA7723 Datenfile Version 2.0.0. https://doi.org/10.4232/1.13935.

GLES. 2022b. *GLES Panel 2021, Wave 18*. GESIS, Köln. . ZA7724 Datenfile Version 2.0.0. https://doi.org/10.4232/1.13936.

———. 2022c. *GLES Panel 2021, Wave 19*. GESIS, Köln. ZA7725 Datenfile Version 1.0.0. https://doi.org/10.4232/1.13920.

Lehmann, Pola, Tobias Burst, Jirka Lewandowski, Theres Matthieß, Sven Regel, and Lisa Zehnter. 2022. *Manifesto Corpus. Version: 2022-1.*

Lehmann, Pola, Tobias Burst, Theres Matthieß, Sven Regel, Andrea Volkens, Bernhard Weßels, Lisa Zehnter, and Wissenschaftszentrum Berlin Für Sozialforschung (WZB). 2022. *Manifesto Project Dataset.* https://manifesto-project.wzb.eu/doi/manifesto.mpds.2022a.