Description of Replication Material

The Temporal Focus of Campaign Communication

The Journal of Politics

Stefan Müller

School of Politics and International Relations
University College Dublin
stefan.mueller@ucd.ie

Introduction

This document describes the datasets and R scripts to reproduce all plots and tables included in the paper and the Supporting Information.

The header of each R script lists the package versions that were used to run the code.

If you are using RStudio, I recommend to create a new R Project[1] in the folder that contains the scripts and data. Working with an RProject will not require that you manually change the working directory in the scripts to load the files. Alternatively, you need to specify the path to the folder that contains the data at the beginning of each script using `setwd()`.

Summary and Descriptions of all R Scripts and Required Data

- `00_example_classification.Rmd` and `00_example_classification.html`: An R Markdown script that shows how to apply the classifier of the temporal focus to a *quanteda* text corpus and how to reshape a text corpus to the level of sentences. Users who prefer to segment the corpus by specific delimiters should apply `corpus_segment()` instead of `corpus_reshape()`. This file is only intended for demonstration purposes and to summarize the typical workflow. It requires the following file:

[1] https://support.rstudio.com/hc/en-us/articles/200526207-Using-Projects

    - **data_sentences_classified_english.rds**: dataset with English coded sentences (one observation per sentence; sentences were coded by crowd workers and research assistants)
- `01_load_and_merge_data.R` contains the code to merge the Manifesto Corpus with metadata about each manifestos and the government status for each party (at the time when the manifesto was written). Note: to run this script, you need to get access to an API from the Manifesto Project and download the manifesto corpus. This script requires the following datasets:
    - The Manifesto Corpus [**cmp_corpus(year >= "1945")**] from the *manifestoR* package[2]. Set mp_use_corpus_version(versionid = "20191220124521") to get the corpus used in this analysis (version 2019-2). Note that you require a valid API key to access the corpus (see the description of the manifestoR package (Merz et al. 2016))
    - **parlgov_2018_parties.csv**: 2018 version of the party information from ParlGov[3]
    - **parlgov_2018_cabinets.csv**: 2018 version of the cabinet information from ParlGov
    - **parlgov_2018_elections.csv**: 2018 version of the election information from ParlGov
    - **merge_pg_cmp.csv**: A dataset (kindly provided by Martin Mölder) that contains adjusted CMP/ParlGov codes to merge the datasets
    - **cmp_maindata_2019.rds**: 2019 version of the Manifesto Project Main Dataset (one observation per manifesto)[4]
    - **data_economic.csv**: economic indicators, collected for this paper
    - **CPDS-1960-2016-Update-2018.dta**: 2018 version of the Comparative Political Data Set[5]
    - **data_missing_cabinet_status_added.csv**: spreadsheet that includes the government status of parties not included in ParlGov. The details in this spreadsheet were collected manually.
- `02_classify_manifestos.R` contains the code to split all manifestos to the level of (quasi-)sentences, classify each sentence in terms of the temporal focus, and apply the sentiment dictionaries. Note that the LIWC dictionaries cannot be added to the Dataverse because of copyright reasons. If you have access to the dictionaries, copy the .dic files into the working directory.[6] Note that the part-of-speech tagging requires that you have spaCy and Python installed.[7] If you want to reproduce the findings of the paper or appendix and do not have the LIWC dictionaries, you can skip this file and continue with the remaining R scripts. This script loads the following datasets:
    - **data_merged.rds**: Dataset constructed in the previous script (`01_load_and_merge_data.R`)
    - **data_sentences_classified_english.rds**: dataset with English coded sentences (one observation per sentence; sentences were coded by crowd workers and research assistants)
    - **data_sentences_classified_german.rds**: dataset with German coded sentences (one observation per sentence; sentences were coded by the team of the Austrian National Election Study)
    - **data_dictionary_rauh.rds**: Rauh's sentiment dictionary (from the development version of the *quanteda.dictionaries* package, but saved as an .rds file to ensure reproducibility)

[2] https://manifesto-project.wzb.eu/information/documents/manifestoR
[3] http://www.parlgov.org
[4] https://manifesto-project.wzb.eu/datasets
[5] https://www.cpds-data.org
[6] You can buy the LIWC software and the accompanying dictionaries at https://liwcsoftware.onfastspring.com
[7] You can find information on how to install space at: http://spacyr.quanteda.io/index.html

    - **data_dictionary_lsdgerman.rds**: German translation of Lexicoder Sentiment Dictionary (from Proksch et al. (2019) replication material)[8]
    - the German (2001) and English (2007/2015) **LIWC sentiment dictionaries**, which cannot be added to the Dataverse because of copyright restrictions. You will need to purchase these dictionaries and add copy them into the folder of your working directory. Alternatively, you can skip this script and continue the reproduction of the results in `03_analysis_paper.R`.
- `03_analysis_paper.R` contains the code to create all figures from the main paper. This script loads only one dataset:
    - **data_manifestos_classified.rds**: Dataset constructed in `02_classify_manifestos.R`. It includes the classified manifesto sentences which are also scored in terms of sentiment
- `04_analysis_appendix_si_a.R`: contains the code to create all figures from SI Section A. The script loads the following datasets:
    - **data_manifestos_classified.rds**. Dataset constructed in `02_classify_manifestos.R`. It includes the classified manifesto sentences which are also scored in terms of sentiment
- `04_analysis_appendix_si_c.R`: contains the code to create all figures and tables from SI Section C.[9] This script uses the *quanteda.classifiers* package which was under development when the paper was published. It may be the case that the MLP classifier changes in the future. You can reproduce the results, however, when you install v0.3 of the *quanteda.classifiers* package
    - **data_sentences_classified_english.rds**: dataset with English coded sentences (one observation per sentence)
    - **data_sentences_classified_german.rds**: dataset with German coded sentences (one observation per sentence)
    - the **German (2001) LIWC dictionary** which cannot be included to the Dataverse because of copyright restrictions. You will need to purchase the dictionary and copy it into the folder of your working directory
    - **data_manifestos_classified.rds**: Dataset constructed in `02_classify_manifestos.R`. It includes the classified manifesto sentences which are also scored in terms of sentiment
- `04_analysis_appendix_si_d.R`: contains the code to create all figures and tables from SI Section D
    - **data_manifestos_classified.rds**: Dataset constructed in `02_classify_manifestos.R`. It includes the classified manifesto sentences which are also scored in terms of sentiment
- `04_analysis_appendix_si_e.R`: contains the code to create all figures and tables from SI Section E
    - **data_manifestos_classified.rds**: Dataset constructed in `02_classify_manifestos.R`. It includes the classified manifesto sentences which are also scored in terms of sentiment
    - **data_dictionary_lsdgerman.rds**: German translation of Lexicoder Sentiment Dictionary (from the Proksch et al. (2019) replication material)
- `04_analysis_appendix_si_f.R`: contains the code to create all figures and tables from SI Section F

[8] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ALFLK6
[9] Note that SI Section B does not consists of any quantitative analysis – therefore 04_analysis_appendix_si_b.R does not exist.

- **prokschetal_3_master_data.Rdata** and **proksch_etal_3_positions_before_senti.Rdata**: both files are part of the replication material provided by Proksch et al. (2019). These files contain the Irish budget speeches used for the analysis of sentiment and the classification of the temporal focus of sentences
- **ZA5710_v2-2-0.dta**: dataset containing the content analysis of the 2013 German TV debate. Since a registration is required to access this dataset, the file cannot be uploaded to the Dataverse. You can download the dataset on the GESIS website for free at https://dbk.gesis.org/dbksearch/sdesc2.asp?no=5710&db=e&doi=10.4232/1.13203.

```
> sessionInfo()
R version 3.6.0 (2019-04-26)
Platform: x86_64-apple-darwin15.6.0 (64-bit)
Running under: macOS  10.15.5

Matrix products: default
BLAS:
/System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions/A/libBLAS
.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib

locale:
[1] en_IE.UTF-8/en_IE.UTF-8/en_IE.UTF-8/C/en_IE.UTF-8/en_IE.UTF-8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

loaded via a namespace (and not attached):
[1] compiler_3.6.0 tools_3.6.0
```