README

Overview

The files contained in this repository can be used to replicate the results reported in *"RELATIO: Text Semantics Capture Political and Economic Narratives", Elliott Ash, Germain Gauthier, and Philine Widmer, Forthcoming, Political Analysis, 2022.*

The results were computed using Python 3.7.4. We recommend using this Python version for replication.

Processing the U.S. Congressional Record

This part cannot be run on a personal computer.

We relied on a High-Performance Computing (HPC) cluster to process the U.S. Congressional Record and estimate different narrative models. Our implementation requires approximately 130 days of computing time (when not parallelized). Each job was run with 35GB RAM and one core, except for *summary_statistics.py*, which required 80GB RAM.

We detail the implementation in *process_data.sh* and provide all our codes. The codes cover the processing of the data, the estimation of the models, a human validation procedure to choose the optimal model, and some manual relabeling of clustered entities.

The full-fledged replication folder (with the necessary seed data files) may be found here.

Replicating Figures and Tables

To save some computing time, we provide the final analysis data frames in */data/gpo_final_data.* From these data frames, one can replicate the tables and figures in the paper and supplementary materials on a personal computer.

We recommend using a 64-bit operating system with a minimum of 16GB of RAM.

Below is a summary of the main steps to follow. These commands are designed to be run from the operating system's command-line interface.

1. Download and extract the Dataverse repository files, then enter the directory:

```
unzip dataverse_files.zip
cd dataverse_files
```

2. Install the Python *virtualenv* package (if needed):

```
pip install virtualenv
```

3. Run the *make_virtualenv.sh* file:

```
bash make_virtualenv.sh
```

NB: This file creates the virtual environment and installs dependencies. Interested readers may find the list of dependencies at the end of this README.

4. Run the *main_results.sh* file:

```
bash main_results.sh
```

The code takes approximately 15 minutes to run.

Fitting New Models on Different Corpora

The Python package RELATIO is also made available on PyPi to run similar types of analyses on other corpora. Consult the GitHub repository for additional information and ongoing developments.

Disclaimer: This replication repository relies on an early-development version of RELATIO. A toy example for this version is provided as part of this replication package (see *toy_example.ipynb*). For new applications, however, we strongly recommend using the latest version of the package on PyPi.

License

These materials are released under a Creative Commons Zero (“CC0”) license. Third-party software libraries are subjected to their respective licenses.

Table: Main Results and Associated Codes

Table / Figure	Associated Code
MAIN TEXT	
Table 1	print_descriptive_tables.py
Figure 3 Panel (a)	plot_war_on_terror_narratives.py
Figure 3 Panel (b)	get_sentiment_narratives_horizontal.py plot_sentiment_narratives_horizontal.py
Figure 3 Panel (c)	get_partisan_narratives_horizontal.py plot_partisan_narratives_horizontal.py
Table 2	divisive_narratives_entities.py
Figure 4	plot_worldviews.py
Partisanship of entities vs. narratives as measured by a penalized logistic regression	prepare_document_term_matrices.py predict_partisanship_narratives.py predict_partisanship_entities.py
Footnote on the average number of narratives per character/words for Republicans and Democrats	get_narratives_per_words_or_chars.py
APPENDIX	
Figure A.1	clusters_pca_visualization.py
Tables C.1-2	inspect_entities.py
Tables C.3-4	inspect_entities.py
Table D.1	summary_statistics.py
Figure D.1	inspect_entities.py
Sampled sentences and their associated narratives	print_descriptive_tables.py
Most frequent narratives and examples of sentences	print_descriptive_tables.py
Tables F.1-3	print_descriptive_tables.py

Figure G.1	plot_individual_narratives.py
Table G.1	print_descriptive_tables.py
Figure H.1	plot_partisan_narratives_vertical.py
Table H.1	divisive_narratives_entities.py
Table H.2	divisive_narratives_entities.py
Figure H.2	plot_worldviews.py
Figure H.3	plot_worldviews.py
Figure I.1	narrative_embeddings.py
Table I.1	narrative_embeddings.py
Figure J.1	See the GitHub Repository (d5ee46e)

List of Python Dependencies

pip==22.3.1
pandas==1.2.4
nltk==3.6.2
spacy==3.0.5
gensim==3.8.3
scikit-learn==0.24.1
allennlp==2.1.0
allennlp-models==2.1.0
networkx==2.5.1
pyvis==0.1.9
tensorflow==2.4.1
tensorflow-hub==0.12.0
matplotlib==3.4.1
scipy==1.6.2
numpy==1.19.5
joblib==1.0.1
threadpoolctl==2.1.0
seaborn==0.11.2
plotly==5.9.0
imbalanced-learn==0.8.0