Replication for: ***The Gender Gap in Elite-Voter Responsiveness Online***

Journal: *Perspectives on Politics*

Author: Zachary P Dickson

Instructions

This repository contains the replication files for the the paper "The Gender Gap in Elite-Voter Responsiveness Online" by Zachary P Dickson.

To replicate the analysis, users will need to have R, Python and Jupyter installed on their machine. Anaconda provides a convenient way to install all three.

There are two folders in this repository: `data` and `code`.

Data

The `data` folder contains the following files:

1. `uk_MIP.xlsx`: Data on public salience of different issues in the UK from YouGov
2. `us_MIP.xlsx`: Data on public salience of different issues in the US from YouGov
3. `individual_data.csv`: Data on the individual behavior of US and UK MPs. Data are indexed by name (legislator), time (survey period), country (US or UK) and issue (issues explained in the article). (**Note:** Twitter prevents the sharing of individual tweets so this data is not included but instructions on how to collect the data are provided below.)
4. `pooled_data.csv`: Time series data that is arranged in a pooled format with one row for men legislators' attention, one row for women legislators' attention, one row for women's issue salience and one row for men's issue salience. This data is simply a transformation of the `individual_data.csv` file. (**Note:** Twitter prevents the sharing of individual tweets so this data is not included but instructions on how to collect the data are provided below.)
5. `confusion_matrix_data.csv`: Data on the confusion matrix for the large language model.

Code

All code files are available in the `code` folder. The `code` folder contains the following files:

1. `main.ipynb`: Jupyter notebook with the analysis code - (**Note:** this is the main file to run to replicate the analysis.)
2. `poisson_estimation.R`: R script to estimate the Poisson regression model.
3. `requirements.txt`: Python package requirements for the analysis.

4. `functions.py`: A python file of all functions for the analysis/figures

Data Collection

Twitter data: Although Twitter (now X) prevents the sharing of data, there are several sources from which the data can be collected that appear to disregard guidelines. For example, all legislators' tweets for the past 8 years are available on Github.

Issue salience data: The issue salience data is collected from YouGov. The data is available for download from the YouGov website. For the UK, data are available here. For the US, data are available here. You can navigate to the bottom of the page and download the data in Excel format.

Election data: The fixed effects regressions condition on vote share at the previous elections. These data are available from the House of Commons Library and from the MIT Election Lab for the US.

Language Model

The language model used for classification has been made public on Huggingface's model hub. The model can be accessed here. The model is a fine-tuned version of the `bert-base-uncased` model.

The language model can be used in the following way:

```
from transformers import AutoModelForSequenceClassification
from transformers import TextClassificationPipeline, AutoTokenizer

mp = 'z-dickson/issue_classification_tweets'
model = AutoModelForSequenceClassification.from_pretrained(mp)
tokenizer =  AutoTokenizer.from_pretrained('bert-base-uncased')

classifier = TextClassificationPipeline(tokenizer=tokenizer,
                    model=model,
                    device=0)

classifier("""
We can't count on SCOTUS to protect our reproductive freedom. \\
The Senate must pass the Women's Health Protection Act now. \\
""")
```

In the case that you're unable to find any of the required data, or have any issues with the replication, please don't hesitate to contact me at zachdickson94@gmail.com.