Wealth Inequality in the US

Predicting Respondent Ideology from Survey Text Data
Erdos Bootcamp 2020

Matt Spearly, Linnea Turco, and Alec Clott

Department of Political Science, The Ohio State University

https://github.com/spearlymatt/erdosproject

The Problem and Goals

Facts: Of all G7 Countries (U.S., UK, Italy, Japan, Canada, Germany, & France), the U.S. has the highest level of income inequality measured by the GINI Coefficient. (Pew Research Center)

Perception: The majority of the American public recognizes wealth inequality is a problem and supports government policies to reduce it; however, meaningful policy has yet to be implemented.

Question: What psychological factors/biases influence individuals' reported attitudes about wealth inequality?

Goal: Using survey respondent data from an open-ended text question, predict whether individuals are left- or right-leaning.

Data Sources

Data Source: MTurk Survey and Lab Experiment conducted by [Matt Spearly and Shelby Boggs]. Respondents were asked:

Economic inequality refers to the size of the gap between the rich and the poor in terms of income and/or wealth. Do you think that economic inequality is a problem in the United States? Do you think that the government of the United States should pursue policies to reduce economic inequality in this country? What policies, if any, would you like to see implemented in the United States regarding economic inequality? Please elaborate on why you support each of these positions.

Data Information: Survey conducted in early 2019, 370 respondents, average response is 88 words.

Data Cleaning & Approach

- **Hand-Coding:** We [Matt, Linnea, and Alec] hand-coded all responses as:
  - Left (1): Respondent recognizes income inequality is a problem **and** supports government- or state-welfare related policies to alleviate inequality.
  - Not-Left (0): Otherwise.
- **Test/Train Split:** We [Matt, Linnea, and Alec] settled on a 70/30 split that placed lab respondents in the test dataset in order to maintain largest possible group of participants from the experiment in the test set to maximize statistical power. Training set included 231 observations, and the test set 98.
- **Classification Models**: We then applied a series of classification models [Matt: voting ensemble, Linnea: penalized regression, Alec: random forest] to predict whether responses were left- or right-leaning on our test data.

Sample Data

Sure, income inequlaity exists. Not everyone can make the same amount of money as everyone else, and certainly not everyone should. Compare a CEO of a banking corporation to the janitor in the same company. The CEO does much more difficult work and they do not deserve to be paid the same. The governement should not interviene in economic policies; it has been proven that free markets work. If the aforementioned janior wanted to make more money, he can work his way up in the company, starting by applying to manage all the janitors and so on.

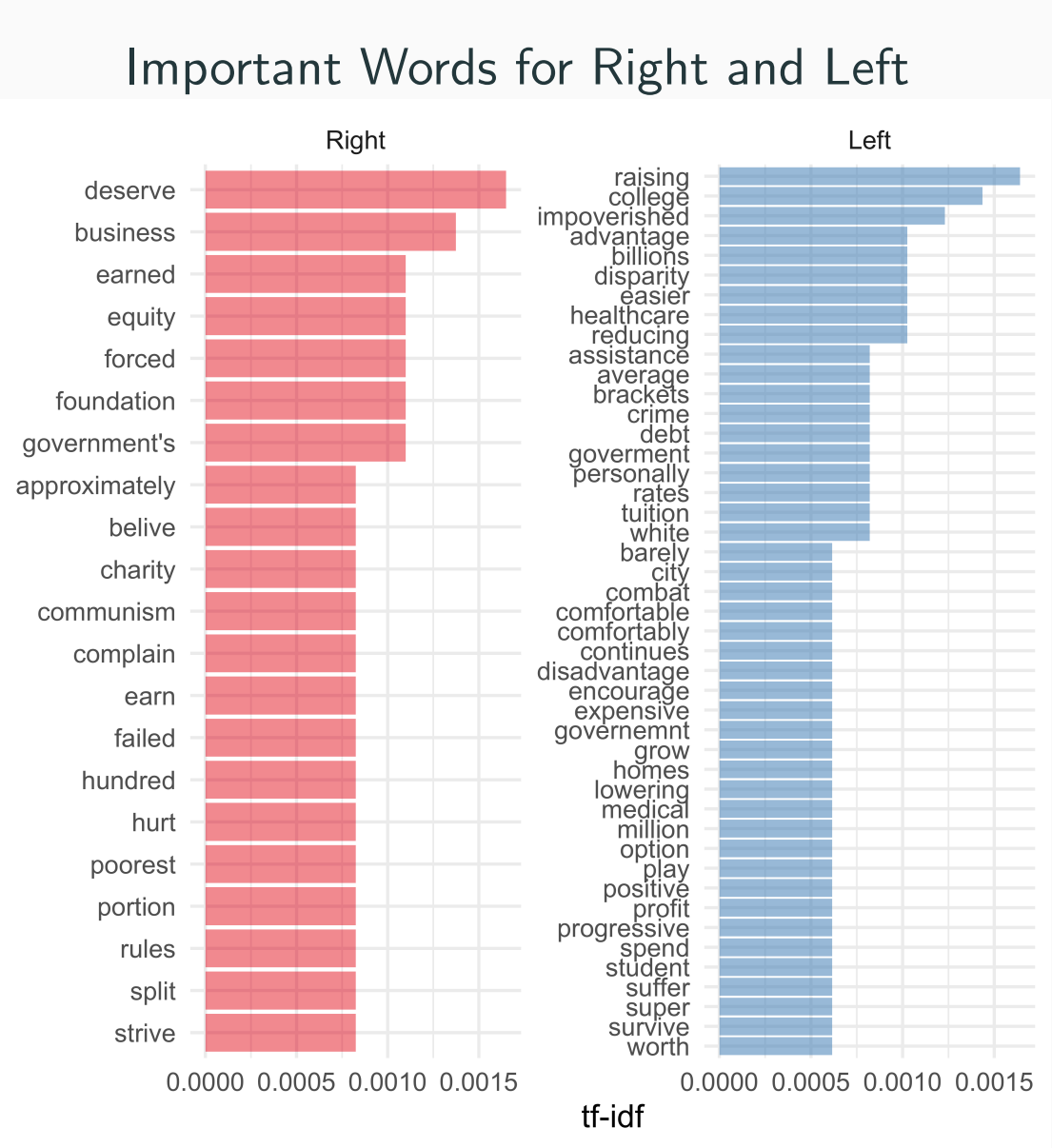
Coded as Left=0 (Right-Leaning)

Sample Data

I think it's a massive problem. Apparently, six members of the Walton family own more than around 150 million Americans (about 50% of the US population). Jeff Bezos also got a subsidized helipad and paid zero dollars in federal income tax in 2017. Why should those that own everything be given free handouts on top of what they already have? I think our government should tax those that are at the very least in the top 1% a lot more. All of their wealth was made from our society, so it's time for them to give back to it. Some policies I saw proposed were rolling back the Bush era tax cuts, using the recent military budget increase to pay for free college tuition, and also rolling back the Trump tax cuts as well. These and many other policies would make it so poorer and middle class Americans can get a good education, not be burdened by crippling student loan debt for decades, and also not have to worry about how they will pay their medical bills or of going bankrupt because of them. I think in this sense, the country would go back to protecting the welfare of its citizens and become more of a democracy.

Coded as Left=1 (Left-Leaning)

Descriptive Statistics



Methods and Packages Used

Data Cleaning: RStudio & Excel

Python Packages: `matplotlib`, `seaborn`, `pandas`, `numpy`, `sklearn`

Models (in Python): Penalized classification models (ridge, elastic net), random forest, naive bayes, SVM, ensemble, bagging and boosting

Results and Challenges

In terms of predictive accuracy...

- Overall: Every model hit roughly 70% predictive accuracy on the test set.
- Ensemble/RF: The voting ensemble hit 75%, but since we were also able to get a RF close to 75%, we opt for the latter as our "final" model since it wins the interpretability/performance tradeoff between these two methods.
- Challenges: Nature of the data (different connotation of same words), MTurk vs. Lab responses, limited sample size, limited response length, nuanced targets

Next Steps

1. Model validation by seeing which responses are incorrectly classified
2. Building a classifier that codes for extremity of ideology
3. Analyze results of pilot experiment
4. Collect more data to improve classifier
5. Explore out-of-sample data

Thanks!

Contact Information (Team of PhD Candidates in Department of Political Science at Ohio State):

- Matt Spearly (spearly.1@osu.edu)
- Linnea Turco (turco.7@osu.edu)
- Alec Clott (clott.1@osu.edu)