Experiment Report

Hao Yan

April 18, 2017

This document will explain the result files on GitHub in detail.

1 Bootstrap experiments

In following experiments, we run bootstrap experiments based on Congressional Record and Press Releases.

1.1 Pooled bootstrap

We first run "Polled Bootstrap" experiment. For each Congress people in Press Releases and Congressional Record, we aggregate his/her data into one single "article". Then, we run 1000 times of bootstrap experiments. Each time, we randomly subsample 80% of Congress people as training data. Then train a support vector regression model and test on the rest 20% of Congress members. And therefore each of them will get a political ideology score.

After we run all 1000 times of experiments, each Congress people actually will get a score vector. In the end, we calculate the mean and standard deviation.

1.2 Together bootstrap

"Together bootstrap" experiments is slightly different from "Pooled bootstrap" experiment. We do **not** aggregate articles from two datasets. For each bootstrap round, we again randomly subsample 80% of **Congress people** as training data and train a support vector regression model. Finally, we test on the rest 20% of Congress members' Congressional Record and Press Releases.

1.3 Separate bootstrap

In "Separate bootstrap" experiments, we actually run two groups experiments: one for Congressional Record and one for Press Releases. This is different from "Together bootstrap" experiments. In "Together bootstrap" experiments, we just run one group experiments and each Congress people's Congressional Record and Press Releases are treated as two "hypothetical" people.

1.4 Results

All the results have been uploaded on GitHub. "separate" folder contains the results from "Separate bootstrap" and "Pooled bootstrap" experiments. "together" folder contains the results from "Together bootstrap" and "Pooled bootstrap" experiments. I will show and explain all the files in two folders.

First, in each folder, there is a "figures" folder and two .csv files: congress_people_info.csv and non-overlap.csv. In "figures" folder, we have Democratic_*.png, Republican_*.png and histogram_*.png.

Democratic_*.png, Republican_*.png are the figures that show the mean and standard deviation of the regression results for each House representative, including Congressional Record, Press Releases and Pooled text. X-axis is the names of Congress members. Y-axis is the normalized regression scores.

histogram_*.png show the mean and median scores distribution of Republican and Democrats, including Congressional Record, Press Releases and Pooled text.

congress_people_info.csv contains following information: name of Congress members, state, party affiliations, DW-Nominate scores, Legislative Efficacy Scores, 2012 Election Misalignment results, and regression score statistical results for three text data, including mean, median, standard deviation, 97.5% and 2.5% interval.

non-overlap.csv contains the Congress members whose CR [2.5%, 97.5%] interval and PR [2.5%, 97.5%] interval are disjoint.

1.5 Conclusion

1.5.1 together experiments

- First, we focus on the House representatives who have big deviation between Congressional Record and Press Releases. According to Figure 1, 19 out of 20 these Republican representatives show that their Congressional Record data is more neutral. However, as we can see in Figure 2, only 11 of Democratics representatives show this property.

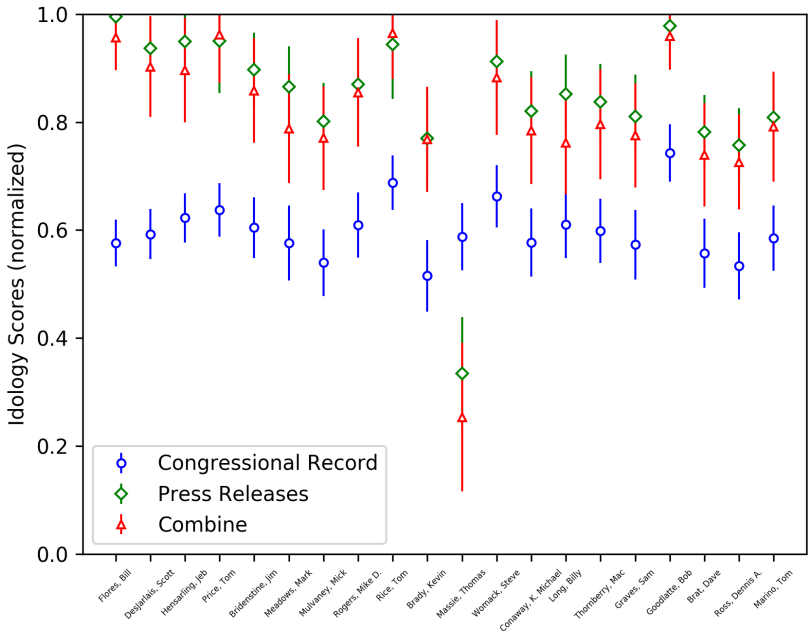

Figure 1: Republican House representatives that show big deviation between Congressional Record and Press Releases in "together bootstrap" experiments (top 20).

- The "Combined bootstrap" experiments' results are close to Press Releases results. It is because the length of Press Releases text data is longer than Congressional Record for most of the Congress members.
- In general, Republican are more consistent than Democratic. The distribution of mean score for Democratic is more diverse than Republican.
- According to Figure 3, House representatives' regression (median) score distribution are consistent.
- There are 26 representatives whose CR [2.5%, 97.5%] interval and PR [2.5%, 97.5%] interval are disjoint. 12 of them are Republican. We also locate their DW-nominate scores to see which interval, Congressional Record or Press Releases, is closer. 17 of them show that their DW-nominate scores are closer to Press Releases' interval.

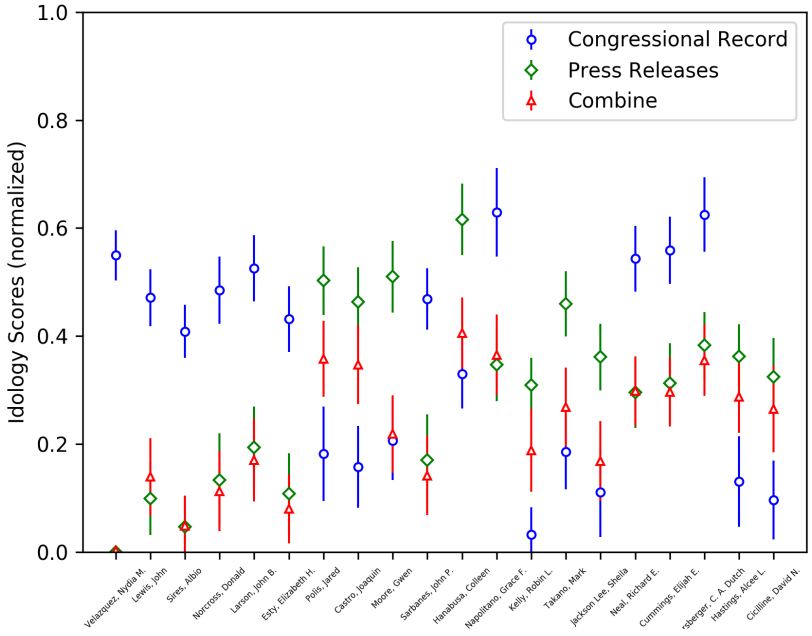

Figure 2: Democtratic House representatives that show big deviation between Congressional Record and Press Releases in "together bootstrap" experiments (top 20).

1.5.2 separate experiments

- We first take a look at the histogram figures. According Figure 6, the median score distributions based on Congressional Record and Press Releases are more different than Figure 3. Democtratic party are much more diversity in "separate bootstrap" experiments and the split point between two parties moves to Republican part.

- Since in "separate bootstrap" experiments, Congressional Record and Press Releases are processd separately, we can't compare each Congress member's Congressional Record and Press Releases directly.

- There are 48 representatives whose CR [2.5%, 97.5%] interval and PR [2.5%, 97.5%] interval are disjoint. 12 of them are Republican. We also locate their DW-nominate scores to see which interval is closer. It turns out that 10 of them show that their DW-nominate scores are closer to Congressional Record's interval.

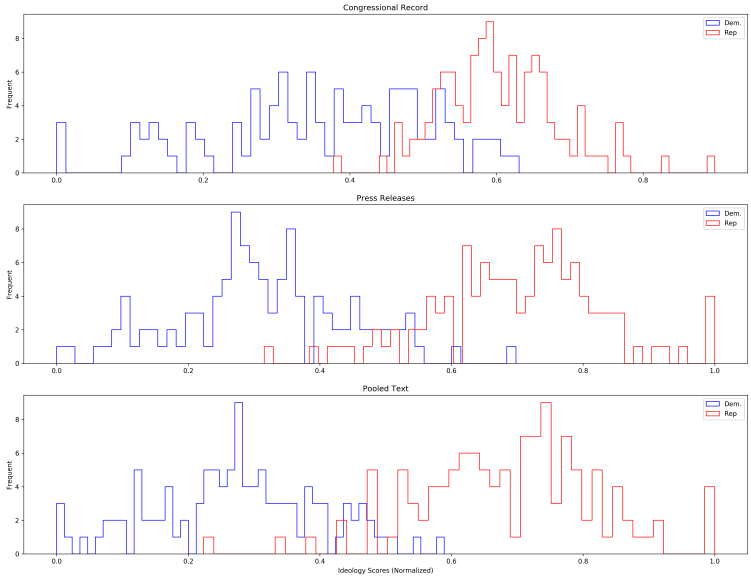

Figure 3: House representatives median scores distribution in "together bootstrap" experiments.

References

[1] Pablo BarberÃ¡. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, 23(1):76, 2015.

[2] Adam Bonica. Mapping the ideological marketplace. *American Journal of Political Science*, 58(2):367–386, 2014.

[3] Keith T Poole and Howard Rosenthal. *Congress: A political-economic history of roll call voting*. Oxford University Press, 1997.

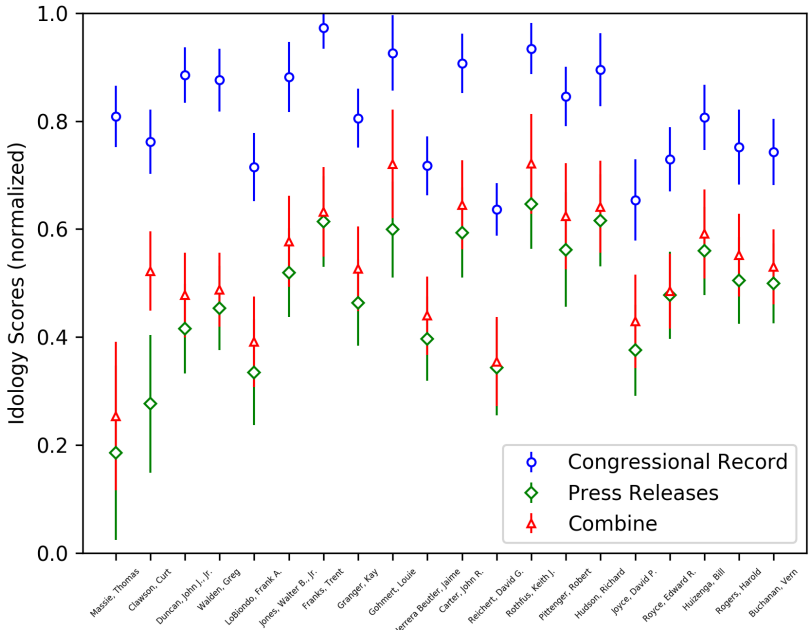

Figure 4: Republican House representatives that show big deviation between Congressional Record and Press Releases (top 20).

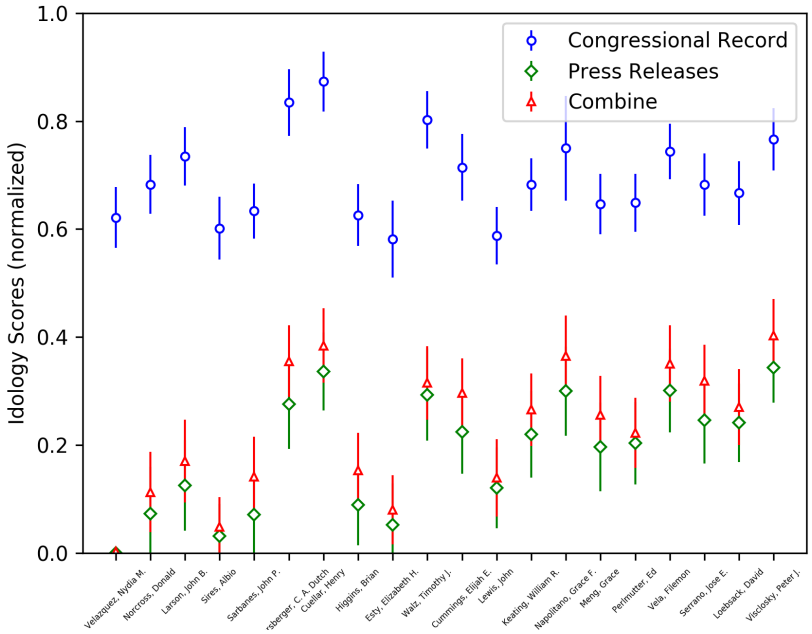

Figure 5: Democtratic House representatives that show big deviation between Congressional Record and Press Releases (top 20).

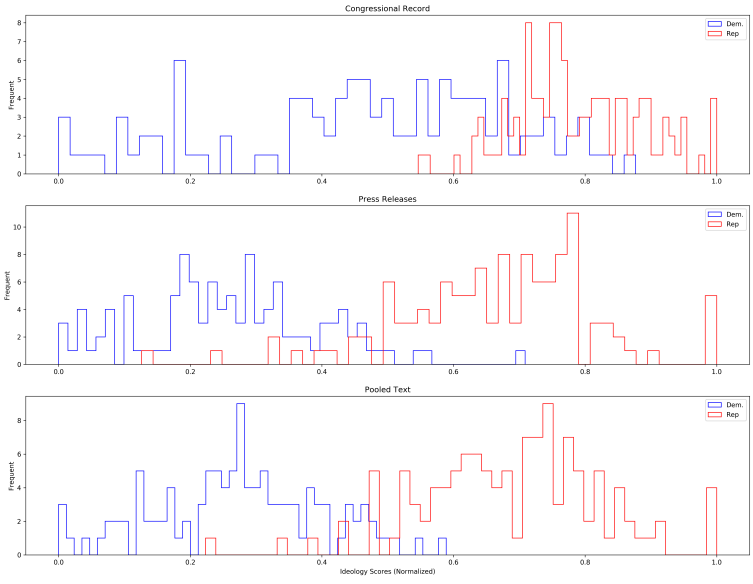

Figure 6: House representatives median scores distribution.