Party Predictor: Predicting Political Affiliation

Nathaniel Roth, Brandon Ewonus, Bryan McCann

Stanford University

1 Overview

Division among the political parties in the United States has become an increasingly large problem.

This paper outlines a variety of supervised and unsupervised techniques employed in an effort to flesh out these divisions.

2 Data

- ❖ 348 (175 Democrat/173 Republican) speeches
- ❖ A majority of the speeches are speeches delivered by presidents dating back to, but not including, the presidency of Franklin Roosevelt.
- ❖ Political lines prior to the presidency of FDR become increasingly difficult to relate in a one-to-one fashion to the political parties today.
- ❖ All of the data was collected by scraping online sources for text

3 Starting Naïve

As a first step, we implemented a Naïve Bayes Model with Laplacian smoothing, which achieved a very respectable error of 22.7%, but before cross-validation.

Below are the words most indicative of political affiliation according to the Naïve Bayes Model:

Key Democrat Features
internet, algeria, bosnia, gay, assad, tunisia, negro, online, algerian, lgbt, barack, conversation, newtown, womens, ghana, secondly, cyber, digital, kosovo', rwanda

Key Republican Features
russias, iraqi, conservatives, narcotics, abortion, iraqis, sdi, tea, heroin, unborn, whittier, liberals, rehabilitation, palin, 1974, 1982, duke, eisler, gorbachev, inflationary

4 Switching to SVM

Approach

In order to fully explore what SVMs had to offer we employed a grid search cross validation technique.

We ran grid search over linear, rbf, poly, and sigmoid kernels, with penalties ranging from 1-10.

Optimal SVM Model

The optimal model for the SVM: used a linear kernel $K(x,z) = \langle x, z \rangle$ with a regularization penalty of 1 and a γ = 1;

It achieved a LOOCV error of 24.7%

5 Logistic Regression

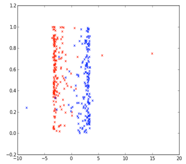
More Search

We continued our search for the best overall model using a grid search of l1/l2 regularizations with varying penalties.

Optimal LR Model

The optimal LOOCV error of .2355 was achieved with l1 with an inverse penalty of .5. 10-fold CV achieved .25

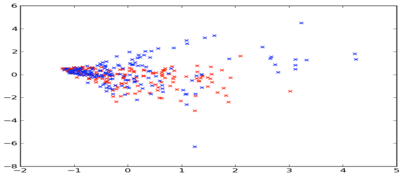
6 LDA



With Linear Discriminant Analysis we achieved a LOOCV test error of 22.5%, by finding the linear combination of features which best explained the variance between parties.

7 Logistic Regression with PCA

We used Principal Component Analysis for dimensionality reduction, in conjunction with logistic regression to evaluate the performance of the resulting components. We obtained a nearly identical LOOCV error of 22.4% using only 25 features. A PCA plot of the speeches using the first 2 components is shown below (blue = Democrat, red = Republican).



8 Ranking Presidents

Ranking System and H2H

In order to rank our presidents by political affiliation, we, for each pair of (Democrat, Republican) presidents, held out the training examples for those presidents, and then obtained the probability of each president being Democrat. We used a president's average probability of being Democrat to sort our rankings.

We also calculated the proportion of times that we correctly identified the Democrat as more of a Democrat than the Republican. We called this the H2H (Head-to-Head) score

Final Rankings

1. Nixon
2. Bush Sr
3. Truman
4. Ford
5. Carter
6. Clinton
7. JFK
8. LBJ
9. Reagan
10. Obama
11. Bush

What happened?

Looking at the rankings to the right, one can clearly see that this ordering is simply not accurate. Our final H2H score was 53.3%, indicating that we did little better than chance.

Combining this with the PCA plot in the previous section, we suspected that our high accuracies in supervised learning were not a result of political affiliation. Rather, it resulted from learning enough about an individual's speech style to associate that back to political party.

9 K-Mean Clustering

Raw Results

Initially we found that for k > 1, two speeches consistently each appeared in their own cluster (Cuomo's DNC keynote and Carter's 1981 SOU). After removing them from our data set, we ran K-means with 8 clusters, and found the following:

- ❖ A cluster of 12 Obama speeches, a cluster of 8 Clinton speeches, and a cluster of 6 Nixon speeches
- ❖ Two large clusters: one of size 80 with 69% Rep. speeches, and one of size 223 with 53% Dem. speeches
- ❖ Three small clusters with one speech each: two JFK speeches and one Clinton speech

10 Unsupervised Insights

- ❖ Most of the speeches we analyzed are similar to each other
- ❖ Many speeches by Obama and Clinton (and Nixon to some extent) stand out from the rest, and from each other
- ❖ There tends to be greater variability among Democrat speeches than there is among Republican speeches

11 Leave One Out Ranking Presidents

In addition to the H2H ranking detailed above, for every speech, we trained a model holding speech out and including the rest of the data. We then predicted the probability the held out speech belonged to a Democrat. Finally, for each president, we averaged the probabilities of all their speeches to rank them as Republican or Democratic; a high number means the president is very likely a Democrat, while a low number means they are probably a Republican. As the table to Right shows, this model fits our intuition much better than the H2H model classifying all the presidents correctly. We believe the better performance may have to do with the fact that we are training on speeches from every president.

1.	Reagan	0.21
2.	Bush	0.23
3.	Nixon	0.26
4.	Ford	0.30
5.	Bush Sr	0.37
6.	Carter	0.62
7.	LBJ	0.62
8.	JFK	0.67
9.	Clinton	0.74
10.	Truman	0.80
11.	Obama	0.86

12 Conclusions

It appears that the predictive power of our learning algorithms may have less to do with political party affiliation then we had originally thought, and more to do with individual differences between politicians. Republicans appear to be more similar to each other in terms of rhetoric, whereas Democrats appear to speak different not only from Republicans, but also from each other; they are different in different ways.