CS229 - Final Report

Predicting Political Ideology Using Campaign Finance Data

Keri A. McKiernan and Joe A. Napoli

December 12, 2015

Introduction

The public interest has increasingly focused on the effect of money in politics. Specifically, the Citizens United v. FEC Supreme Court ruling maintained the legality of unrestricted political expenditures by corporate and union entities. As a result, there has been a proliferation of super PAC, or 'political action committee', organizations. While these organizations are not permitted to make contributions directly to campaigns, they may engage in unlimited independent spending and there is no restriction on the amount of funds they can accept from donors. Such legal developments beg the question: Can one predict generally the influence of donors over politicians to whom they give? Our project largely builds on previous research conducted by Poole and Rosenthal (summarized below), and makes use of freely available databases of financial information. We aim to predict quantitative scores of political ideology using campaign contribution profiles; specifically, the input to our algorithm is a matrix of normalized vectors of financial data, partitioned by industry, and the output is a score for each candidate on the interval $[-1, 1]$ indicating a position on the political ideological spectrum.

Related Work

The research of Poole and Rosenthal has focused on quantifying the political ideology of politicians [3]. In particular, they actively develop methods for calculating 'ideal points' of candidates. The DW-NOMINATE method calculates a legislators overall probability of voting 'yea' on a piece of legislation as the sum of a deterministic utility value and a random error [1]. 'Ideal point' coordinates were obtained for legislators by maximizing the log likelihood function

$$\mathcal{L} = \sum_{t=1}^{T} \sum_{i=1}^{p_t} \sum_{j=1}^{q_t} \sum_{\tau=1}^{2} C_{ij\tau t} \ln P_{ij\tau t}$$

where $P_{ij\tau t}$ is the probability of voting for choice τ (yes or no) and $C_{ij\tau t} = 1$ if that probability accurately predicts the vote [1]. Indexes j, i, and t sum over roll call votes, legislators, and legislative sessions, respectively. Ideal points are constrained to lie within the interval $[-1, 1]$ and are two-dimensional quantities. A common interpretation of the first coordinate is that it reflects the divide between the Republican and Democratic parties, whereas the second coordinate is more highly correlated with intra-party division. While a full congressional voting record would be unavailable for candidates who are new to office, campaign finance data *are* often readily available. It would be useful to be able to predict the ideal point of a candidate even before they have established a congressional voting record. Furthermore, the ability to do so would help elucidate a relationship between monetary contributions to candidates and the voting patterns those contributions may effect.

We have collated freely available campaign finance and DW-NOMINATE ideal point [2] datasets and performed a principal component analysis on the feature set extracted therefrom. The original scope of our project has been narrowed due to a reduction in the size of our group. While originally we had planned to investigate anomalous voting behavior and the ability to predict it solely by examining campaign finance data, we have shifted our focus on training a model to predict DW-NOMINATE ideal points, which directly reflect ideological stances and inter-/intra-party divisions. This shift has been prompted by datasets that were relatively more sparse than expected as well as the challenge of collating them.

Dataset and Features

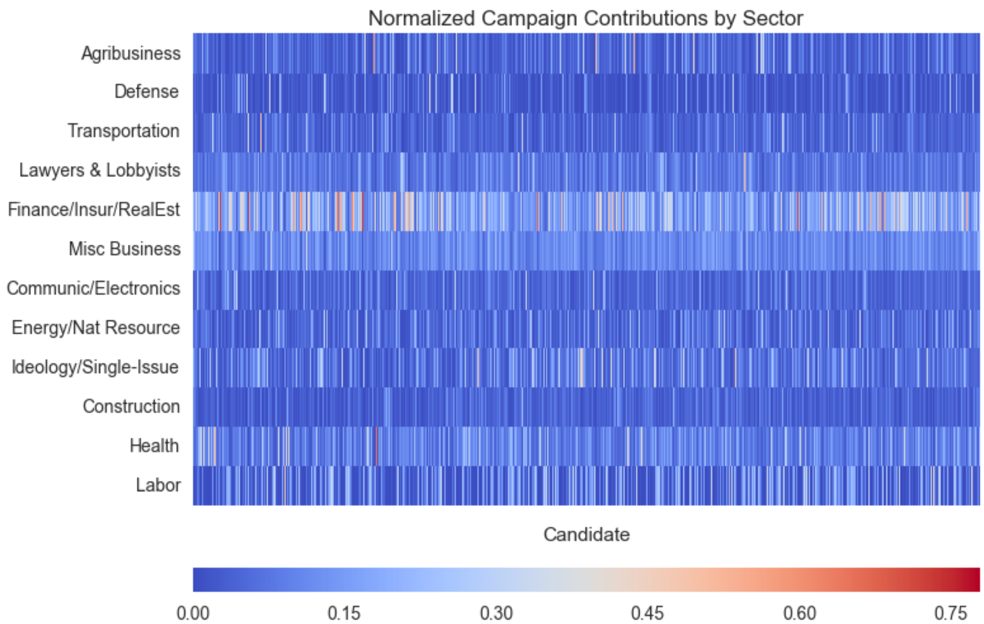

Figure 1: Heatmap of complete feature set

Methods

Experiments/Results/Discussion

We chose a final model after performing a randomized optimization in the space of hyperparameters for SVR, for multiple kernel types. This required specifying distributions over which to sample the hyperparameters. We specified a separate exponential distribution for each hyperparameter, and estimated the decay scale of each distribution by first performing an exhaustive grid search over values of the hyperparameters spanning several orders of magnitude. In both the grid and randomized search schemes, the coefficient of determination (R^2) was used as the scoring metric for ranking all sampled models and was estimated for each candidate model using 5-fold cross-validation on a randomly sampled set of training data that comprised 80% of the data set. Each randomized hyperparameter search proceeded for $30,000$ iterations.

	RBF kernel	**Linear kerel**
C	10.77	55.33
ϵ	0.047	0.31
γ	9.97	n/a
Training score	0.82	0.70
Test score	0.83	0.64

The final hyperparameters for the linear and radial basis function (RBF) kernels are summarized in the table below alongside their corresponding training and test errors. We note that results for the polynomial kernel were not as trustworthy because it proved to be much more computationally intensive and thus was unable to be optimized as exhaustively as the other two kernel types. Therefore, we omit those results. We also note that of the best models identified by the randomized search, for each kernel type there were several distinct sets of hyperparameters that yielded very similar training scores. For the RBF kernel specifically, the hypermarameters we selected represent a compromise between the values of C and γ. While relatively high values of C tend to correspond to overfitting by overpenalizing large deviations from the target, constraining C too aggressively would encourage the bandwidth of the kernel function, γ, to expand and potentially overfit the data. Therefore, cross-validation will be critical in evaluating the final selection.

As expected, the RBF kernel is better able to capture the nonlinearities and regional variation of the data by virtue its implicit infinite-dimensional feature mapping. In order to characterize the performance of the final model further as well as to illustrate the deficiencies of the linear kernel, we compute a learning curve of the test score as a function of fraction of data used for training.

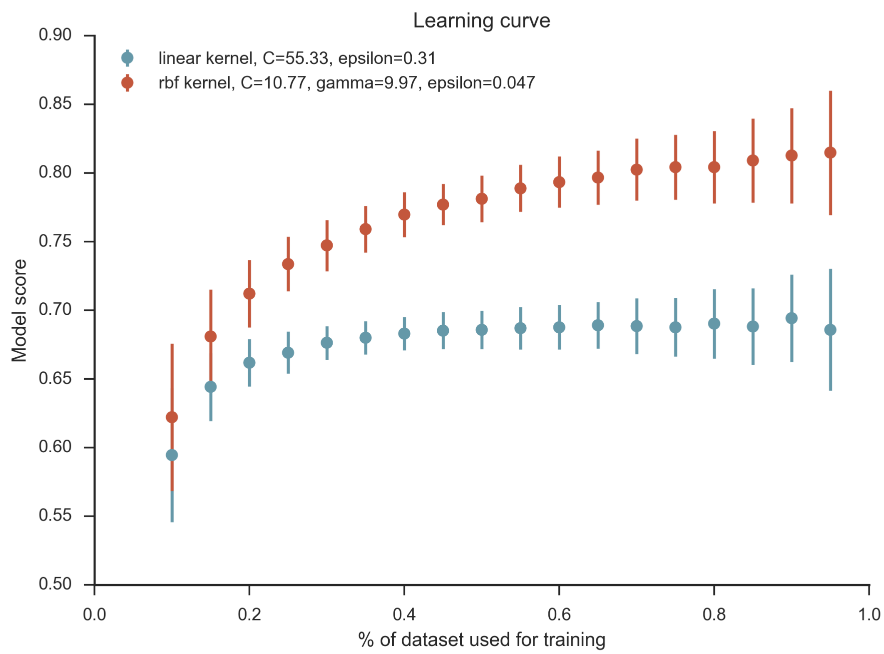

Figure 2: Learning curve comparison

Conclusion/Future Work