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? As the flux of money into the political system only continues to increase, this remains a critical question. 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 campaign contribution 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 [**PR**]. In particular, they actively develop methods for calculating 'ideal points' of candidates. For example, 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 normally distributed random error [**NOMBOOT**]. It calculates 'ideal point' coordinates 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 [**NOMBOOT**]. Indexes j, i, and t sum over roll call votes, legislators, and legislative sessions, respectively. These 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. It is important to emphasize that these methods seek to estimate ideal points using roll call voting records. While a full congressional voting record would be unavailable for candidates who are new to office, campaign finance data *are*, ideally, readily available. Thus, we think it would be helpful to be able to predict ideal points for candidates even before they have established a congressional voting record. One could then use these scores to predict how legislators will vote on particular pieces of legislation.

Much previous work has focused on connecting political contributions to specific votes cast in congress [**stratmann1991campaign**, **stratmann1995campaign**]. Furthermore, Adam Bonica's has developed a model for predicting DW-NOMINATE scores using data at the resolution of individual contributions to PACs [**bonica2013ideology**]. This approach is bound to be more computationally expensive than a coarser graining of the space of contributions, and we posit that we can recover the ideal point of a candidate just by using a contribution profile partitioned by sector. We have collated freely available campaign finance [CITE] and DW-NOMINATE ideal point [**DWWEB**] 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. 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.

Dataset and Features

Campaign finance data for a set of 375 candidates (as determined by sampling from the DW-NOMINATE scores) over years 2010, 2012, and 2014 were obtained using the Open Secrets python API [**open˙secrets**]. Of the available data, we chose to featurize by sector. For each sector, individual and PAC contributions were combined and normalized on a per candidate basis. A heatmap of the total feature set is illustrated below.

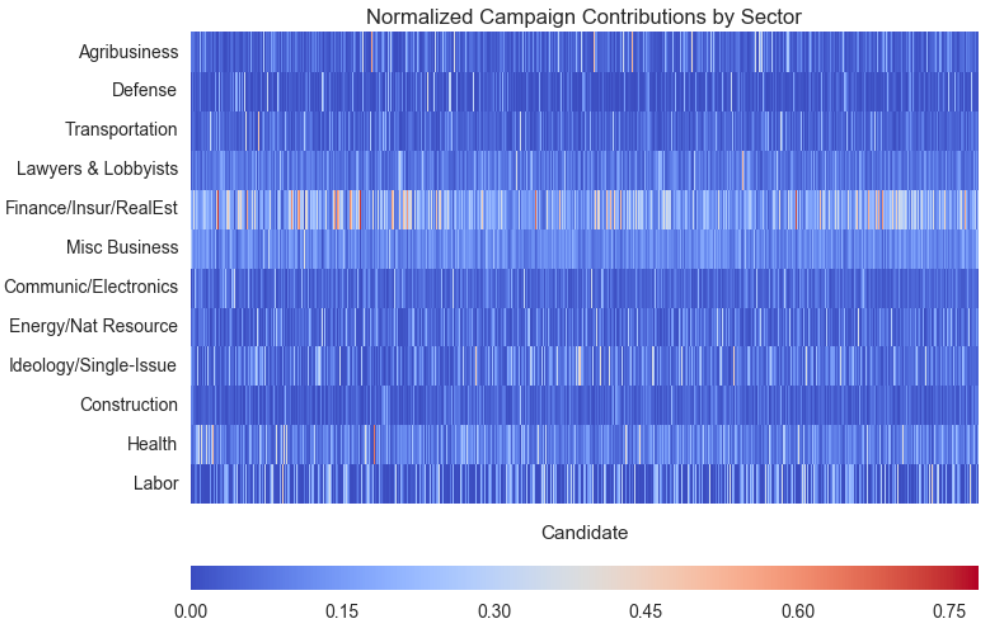

Figure 1: Heatmap of complete feature set

Methods

The learning algorithm used in this work was epsilon-Support Vector Regression (ϵ-SVR), using a radial basis function (RBF) kernel. In ϵ-SVR, a nonlinear function is constructed through the training of a linear function, $f(x)$, in a higher dimensional inner product space defined by a kernel function, $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$. The objective of the function training is to ensure all training data lies within ϵ of all target data. Without supplementary information about a particular data set, SV learning algorithms are consider the best "off-the-shelf" supervised learning algorithm [**class˙svm**].

Given a training set, $S = \{(x_0, y_0), ..., (x_n, y_n) | x_i \in \mathbb{R}\}$, ϵ-SVR can be formulated into the following convex optimization problem, given here in its primal form:

$$\min_{w,b,\zeta,\zeta^*} \frac{1}{2} w^T w + C \sum_{i=1}^{n} (\zeta_i + \zeta_i^*)$$

$$\text{s.t. } y_i - w^T\phi(x_i) - b \leq \epsilon + \zeta_i,$$
$$w^T\phi(x_i) + b - y_i \leq \epsilon + \zeta_i^*,$$
$$\zeta_i, \zeta_i^* \geq 0, i = 1, ..., n$$

where $\phi(x)$ is the mapping to the higher dimension space, C determines the degree of regularization (lower C leads to a smoother solution), and ζ, ζ^* are slack variables which allow for constraint relaxation in the case it is required. This problem is often simpler to solve in its dual form (derived through the method of Lagrange multipliers):

$$\min_{\alpha,\alpha^*} \frac{1}{2}(\alpha - \alpha^*)^T Q(\alpha - \alpha^*) + \varepsilon e^T(\alpha + \alpha^*) - y^T(\alpha - \alpha^*)$$

$$\text{s.t. } (\alpha - \alpha^*) = 0$$
$$0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, ..., n$$

where $Q_{ij} \equiv K(x_i, x_j)$, and Q is the matrix of these values. This problem can be solved to yield the following:

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b \tag{1}$$

Where the RBF kernel is given by the following:

$$K(x_i, x_j) = \exp\left[-\gamma|x_i - x_j|^2\right] \tag{2}$$

where γ describes the influence of each individual training example [**scikit˙learn˙svr**].

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

Figure 2: Summary of hyperparameters and model scores

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 of its implicit infinite-dimensional feature mapping. In order to characterize the performance of the final model further and to compare it to the linear kernel model, we compute learning curves for both showing the test score as a function of the fraction of data used for training. The deficiencies of the linear kernel are readily apparent; its score saturates at about 0.65 to 0.7, within error, while the RBF kernel acheives 0.8 to 0.85, within error.

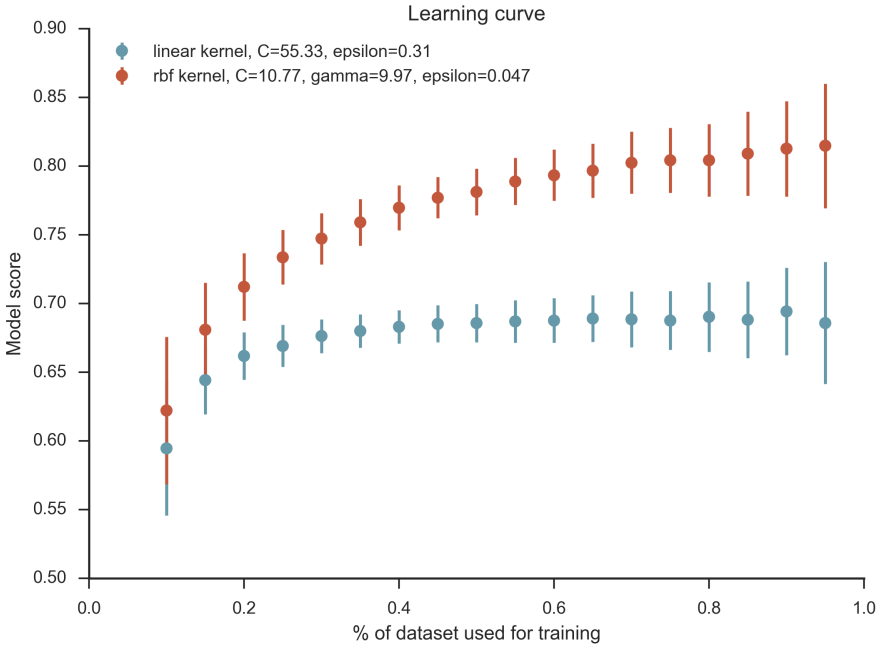

Figure 3: Learning curve comparison

In order to evaluate the structure of the feature data set, and gain insight into our model, PCA was performed on the full training set. The top 3 principal components are depicted below.

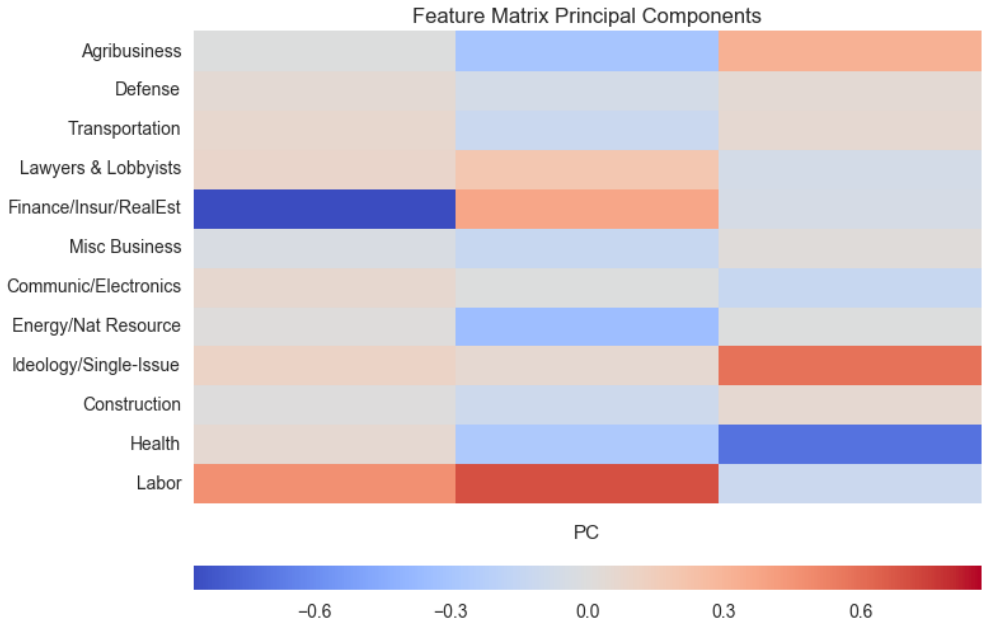

Figure 4: Heatmap of top 3 principal components

It appears the financial, and labor sectors specify the dimensions of largest variance. Where the finance sector is most characteristic of a republican candidate, and the labor sector is most characteristic of democratic candidates.

Conclusion/Future Work

We found that the Support Vector regression algorithm with the radial basis function kernel outperformed all other models. The nonlinearity of this problem is not especially surprising given the regional variability of both political ideology and of the contributing industries. We note that the normalization procedure employed here likely washed out some of the variability, as some candidates surely received greater amounts of contributions overall compared to others; however, the unnormalized feature matrix was prohibitively costly to fit to. One future direction may be to look at incorporating this variability in the mode. Moreover...

BLAH BLAH BLAH.