

Predicting Political Ideology Using Campaign Finance Data

Keri A. McKiernan and Joe A. Napoli

December 11, 2015

Introduction

Public interest has become increasingly focused on the role 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 continues to increase, this remains a critical question. Our project builds largely on previous research by Poole and Rosenthal and makes use of freely available databases of campaign finance 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, the learning algorithm used is Support Vector Regression (SVR), and the output is a score for each candidate in 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 [6]. In particular, they actively develop methods for calculating ‘ideal points’ for candidates. For example, the DW-NOMINATE method calculates a legislator’s overall probability of voting ‘yea’ on a piece of legislation as the sum of a deterministic utility value and a normally distributed random error [2]. 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 [2]. 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 and to draw quantitative comparisons between them.

Much previous work has focused on connecting political contributions to specific votes cast in congress [9, 8]. Furthermore, Adam Bonica of Stanford has developed a correspondence analysis model for predicting DW-NOMINATE scores using data at the resolution of individual contributions to PACs [1]. We think this approach has the deficiency of being computationally expensive compared to 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 [5] and DW-NOMINATE

ideal point [4] datasets. The original scope of our project has been narrowed due to a reduction in the size of our group. We originally planned to investigate anomalous voting behavior and the ability to predict it solely by examining campaign finance data, but shifted our focus to training a model to predict DW-NOMINATE ideal point coordinates.

Dataset and Features

Campaign finance data for 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 [5]. 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. This amounted to 12 features in total. A heatmap of the total feature set is illustrated below.

We also include a plot of the distribution of DW-NOMINATE scores, labeling Democrats and Republicans using their traditional party colors. We note that the dataset appears linearly separable in the first dimension of the ideal point coordinates. This suggests that a classification algorithm may effectively distinguish between the parties in case regression algorithm fails to quantitatively capture the distribution of scores. Finally, we note that for each party the second dimension of the DW-NOMINATE scores is not as easily separable and instead spans the interval $[-1, 1]$. This would be a natural extension of the present work.

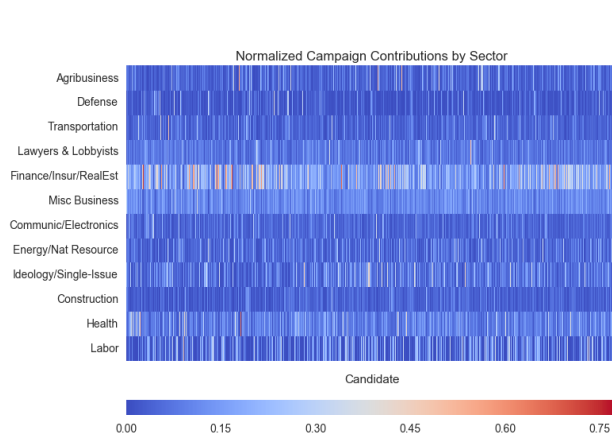


Figure 1: Heatmap of complete feature set

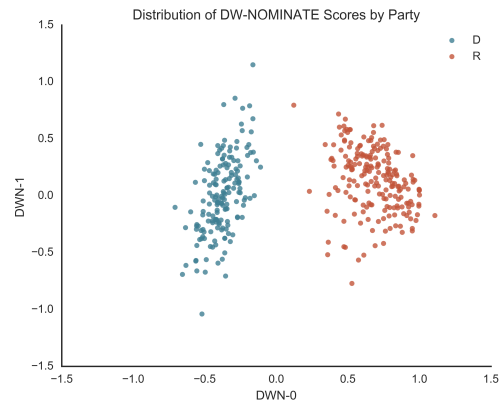


Figure 2: DW-NOMINATE score distribution

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 considered the best “off-the-shelf” supervised learning algorithm [3].

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

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

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

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):

$$\begin{aligned}
\min_{\alpha, \alpha^*} & \frac{1}{2}(\alpha - \alpha^*)^T Q(\alpha - \alpha^*) + \epsilon e^T(\alpha + \alpha^*) - y^T(\alpha - \alpha^*) \\
\text{s.t. } & (\alpha - \alpha^*) = 0 \\
& 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, n
\end{aligned}$$

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:

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

and where γ describes the influence of each individual training example [7].

Experiments/Results/Discussion

We chose a final model after performing a randomized optimization in the space of hyperparameters for ϵ -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 kernel
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 3: 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 scores. We omit results for the polynomial kernel because the calculation did not converge and thus results could only be obtained via a cruder randomized search. 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 hyperparameters we selected represent a compromise between the values of C and

γ . While relatively high values of C 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 is critical in evaluating our final selection.

We find that 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. Deficiencies of the linear kernel are readily apparent; its score saturates at about 0.65 to 0.7, within error, while the RBF kernel achieves 0.8 to 0.85, within error.

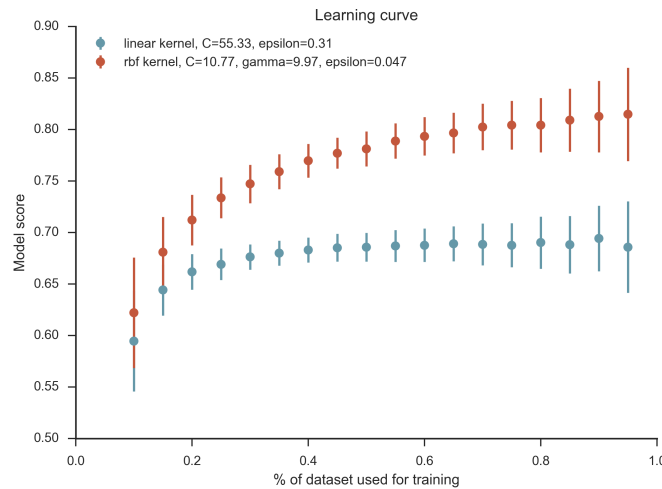


Figure 4: Learning curve comparison

In order to evaluate the structure of the feature data set, and gain qualitative insight into our model, principal component analysis (PCA) was performed on the full training set. The top 3 principal components are depicted below, along with supplementary data regarding the per party distributions of campaign donations by sector.

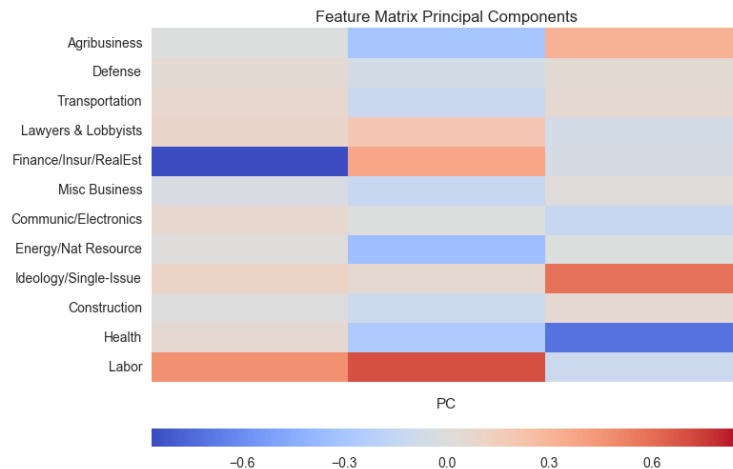


Figure 5: Top 3 principal components

Rank	Sector	Amount	To Cands/Parties	Dems	Repubs	To DEMS To REPUBS
1	Finance/Insur/RealEst	\$508,952,057	\$345,998,502	37.3%	62.5%	
2	Ideology/Single-Issue	\$352,468,939	\$233,985,179	50.2%	49.6%	
3	Other	\$269,909,263	\$233,249,839	53.5%	46.1%	
4	Misc Business	\$229,692,284	\$182,826,316	38.2%	61.6%	
5	Lawyers & Lobbyists	\$151,549,346	\$140,531,911	65.2%	34.6%	
6	Health	\$142,131,127	\$132,215,578	43.0%	56.9%	
7	Labor	\$140,687,552	\$60,582,581	88.9%	10.9%	
8	Communic/Electronics	\$116,177,491	\$90,676,015	59.9%	39.8%	
9	Energy/Nat Resource	\$115,343,359	\$89,158,053	20.9%	79.0%	
10	Agribusiness	\$77,182,810	\$64,541,559	24.9%	74.8%	
11	Construction	\$67,856,115	\$60,053,074	28.3%	71.5%	
12	Transportation	\$60,975,781	\$56,728,333	28.4%	71.5%	
13	Defense	\$25,432,475	\$25,348,450	40.1%	59.8%	

Figure 6: Per party campaign contributions [5]

It appears the financial and labor sectors specify the dimensions of largest variance. It also appears that donations from the finance sector are heavily skewed towards Republican candidates, while donations from the labor sector are heavily skewed towards Democrats. Although SVR models are abstract to interpret, this analysis suggests that these two categories had the largest influence in the prediction of political ideology.

Conclusion/Future Work

We found that the Support Vector regression algorithm with the radial basis function kernel outperformed all other models, achieving a score of 0.8 to 0.85, within error. We think the model would have been even better if we had access to additional years of data. 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 model. More time could be invested in exploring alternative and more representative normalization methods. Additionally, a way to circumvent the feature set normalization issue would be to use an unsupervised learning algorithm such as a neural network.

Had we had an additional group member, we would have liked to explore the correlation of ‘anomalous’ voting behavior, i.e. votes that are not well predicted by DW-NOMINATE scores, and a higher-resolution feature set of individual contributions. Such an analysis would probably also benefit from further degmentation of donations into personal and PAC contributions.

References

- [1] Adam Bonica. “Ideology and interests in the political marketplace”. In: *American Journal of Political Science* 57.2 (2013), pp. 294–311.
- [2] Royce Carroll et al. “Measuring bias and uncertainty in DW-NOMINATE ideal point estimates via the parametric bootstrap”. In: *Political Analysis* 17.3 (2009), pp. 261–275.
- [3] *CS229 class notes on Support Vector Machines*. cs229.stanford.edu/materials/notes/cs229-notes3.pdf. Accessed: 2015-12-11.

- [4] *DW-NOMINATE Scores With Bootstrapped Standard Errors*. <http://voteview.com/dwnomin.htm>. Accessed: 2015.10.01.
- [5] *OpenSecrets.org*. <http://opensecrets.org>. Accessed: 2015-12-11.
- [6] K.T. Poole and H. Rosenthal. *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press, 2000.
- [7] *Scikit-learn SVR implementation documentation*. <http://scikit-learn.org/stable/modules/svm.htmlsvm-implementation-details>. Accessed: 2015-12-11.
- [8] Thomas Stratmann. “Campaign contributions and congressional voting: does the timing of contributions matter?” In: *The Review of Economics and Statistics* (1995), pp. 127–136.
- [9] Thomas Stratmann. “What do campaign contributions buy? Deciphering causal effects of money and votes”. In: *Southern Economic Journal* (1991), pp. 606–620.