A Structural Approach to Legislative Roll Call Vote Prediction

Steve Hanneke HANNEKE@UIUC.EDU

Abstract

I present a structural approach to the task of legislative roll call vote prediction. Given a voting history, consisting of co-sponsorship information and voting records for past legislation, along with a new piece of legislation with sponsor and co-sponsor information, the task is to predict whether the new legislation will pass or be rejected. I propose an approach that constructs a network representing influence relations based on co-sponsorship information in the voting history. This network is then used to predict the votes of the individual legislators, given sponsorship information for the new legislation.

1. The Task: Legislative Vote Prediction

When a new piece of legislation is introduced in the U.S. Senate, it is necessarily sponsored by a particular senator. The legislation may also be co-sponsored by several other senators. After discussion, revision, and other processing, the legislation is voted on and is either passed or rejected based on the number of *Yea* votes. One key part of political analysis in such legislative bodies is determining whether or not a piece of legislation will pass, given the group of legislators who endorse it. This allows politicians and lobbyists to better allocate their efforts to back legislation that has a greater chance of success, or seek more support when the prospects for their proposed legislation are bleak. Unfortunately, this task of vote prediction is currently based on vague intuitions and rules of thumb of experienced political analysts. Even then, the analyst must spend long hours pouring over the details of voting records and positions every time a new legislator is elected, in order to come up with accurate predictions of that legislator's voting behavior. It is therefore desirable to automate this process of vote prediction, to take advantage of the speed and flexibility of computer systems, while maintaining nearly equivalent reliability and accuracy to that of the human political analyst.

1.1 Task Formulation

In the task of interest here, there is a legislative body, consisting of a set of legislators X_i. The set is assumed to be unchanging, which is the case for example in a typical session of Congress. We are given a vote history for this legislative body, consisting of several pieces of legislation, each accompanied by the name of the legislator who sponsored the legislation, the names of any co-sponsors, and the voting behaviors of all legislators for that legislation. Each legislator may be recorded as voting *Yea*, *Nay,* or *Not Voting*. We are then given a new piece of legislation, with information about who the sponsor and co-sponsors are, but no information about the voting behavior. Based on information gathered in the vote history, and the list of sponsor and co-sponsors for the new piece of legislation, we are asked to make a prediction on whether or not the new legislation will pass.

1.2 Outline of the Approach

Thus, this task may be formulated as one of classification, where the instances are drawn from a distribution over the space of possible co-sponsorship combinations, and labeled as *Pass* or *Reject* based on some unknown target function, employing hidden variables. However, the task can also be broken down further into the subtasks of predicting the voting behavior of each individual senator and then applying a majority function to determine the overall vote. This approach has the advantage that it allows one to introduce additional structure in the problem based on knowledge of the task, which restricts the space of possible classifiers to be much smaller, without losing the ability to fit the data reasonably well, thus reducing the amount of vote history necessary to learn an accurate classifier (Vapnik, 1998). This type of restriction becomes necessary here, since the vote history would never realistically be larger than 1,000 votes.

Specifically, in the approach taken here, I define a notion of influence, a relation between pairs of legislators. I assume that the sponsor and co-sponsors vote *Yea*, and that they then try to use their influence to convince others to vote *Yea*. Those who are convinced will vote *Yea*, and also try to use their influence to convince others to vote *Yea*. Thus, the vote is propagated through the network of influence relations. The details of the classifier lie in the function that determines whether or not a senator is convinced to vote *Yea*, based on the votes of those who influence him or her, and also the algorithm used to propagate the vote. As a preliminary approach, I have chosen to use a sigmoid function for the first detail, and a sampling algorithm for the second. The network structure and the parameters for the sigmoid function are learned based on the vote history.

2. Existing Models of Roll Call Voting

There has been much work in the political science literature on analyzing roll call voting behavior in legislative bodies. These theories are centered around a *Spatial Voting Model* (Poole; Ladha, 1994; Clinton, et al, 2004). This framework maps each legislator to an *ideal point* in a low-dimensional space based on voting history, and likewise maps each piece of legislation to two *position points* in the space based on how the legislators voted: one point for *Yea* and one for *Nay*. Each legislator has some utility function, defined for each point in the space, and chooses to vote according to the position point that maximizes that utility.

Generally, an assumption of *sincere voting* is made, under which the legislators' utility functions remain static from one vote to the next. This assumption significantly simplifies the analysis, but fails to capture the negotiation and other interactions among legislators. Attempts to relax this assumption, while maintaining the spatial model include (Clinton, et al, 2004; Clinton & Meirowitz, 2003), which alter the utility function to account for these interaction factors.

Unfortunately, the spatial voting model is not immediately applicable to the task of vote prediction, since the position points are calculated using knowledge of the voting behavior. Although the ideal points and position points can all be estimated from the historical voting data, the position points of any new pieces of legislation, for which the votes are not yet known, cannot be calculated using existing methods.

3. A Structural Approach

As mentioned above, the *sincere voting* model fails to capture the interactions, negotiations, and influences that are so crucial to a legislator's voting decision. A more realistic approach should make these interactions a central part of the model. With this motivation in mind, I propose a social network model based on the influence structure of a legislative body.

3.1 Influence Networks

In the social networks literature, there has been much work on constructing influence networks for the propagation of ideas and innovations. For example, (Coleman, et al, 1966) provides a classic study. The key to modeling these networks is the temporal order of adoption of an idea, along with the opportunity for an idea to be propagated from one individual to another (i.e., a pre-existing relationship), and some notion of the strength of the influence. The influence relation is then modeled as a directed tie between two actors in the network.

3.2 Co-Sponsorship as Indicative of Influence

Discovering the structure of the influence network in a legislative body, given only the vote history, poses some interesting challenges. However, the nature of the task allows one to make the simplifying assumption that all legislators are in contact with all other legislators in the same legislative body. This assumption is justified given the relatively small size of most legislative bodies.

This still leaves open the question of how to determine which legislators are related by an influence relationship. The approach I propose is to utilize the sponsorship and co-sponsorship data to create directed edges, which seem to capture some aspect of the influence one legislator has on another. Specifically, legislator A is connected to legislator B with a directed edge A→B if and only if there is some legislation on which A is the sponsor and B is a co-sponsor. The value of the edge is determined by an increasing function of the number of such pieces of legislation.

This model of influence, though admittedly not perfect, seems to capture a notion of influence, since legislator A was presumably first to take interest in the issue, and through some method of communication with the others, was able to convince B to be a co-sponsor. It seems natural to conclude that in the future, even when A is unable to convince B to go so far as becoming a co-sponsor, he or she would still have a good chance of convincing B to vote a specific way if the A→B link has a high value.

4. Inference: Predicting the Votes

Once one has the structure of the influence network, the remaining step is to predict the voting behavior of each legislator, given only information about the sponsor and co-sponsor identities. We can comfortably assume that the sponsor and co-sponsors will vote *Yea*. Thus, the task can be viewed as one of classifying the remaining legislators as *Yea* or *Nay*, given some small number of *Yea* votes at known positions in the network. This problem can be formulated in a variety of ways, such as information diffusion, transductive inference, or belief propagation.

4.1 Related Work

Within the social network literature, there is much work on modeling phenomena analogous to the diffusion of information through a communication network. This model has been used to analyze the diffusion of innovative ideas, riot behavior, and viral diseases, each within a network structure. For example, (Granovetter, 1978) discusses using a model in which each actor's decision is based on whether the number of actors connected to him or her exhibiting some behavior exceeds some threshold. Similarly, (Kempe, et al, 2003) discusses the problem of maximizing the diffusion in such a thresholded network by appropriate selection of initiating actors. One could formulate the vote prediction task in a similar way, by assuming that a legislator will vote *Yea* if and only if the total influence on him or her from other *Yea*-voting legislators exceeds some threshold.

In the more general setting of propagating values through a graph structure, recent work has been performed in the field of transductive inference (Blum, et al, 2004; Zhu & Ghahramani, 2002). In the transductive setting, one is given the target values (votes) of some subset of the data points, and asked to infer the target values of the remaining points. Transduction is distinguished from inductive learning in that we are supplied with the "test" data of interest prior to forming any hypothesis. Using the known relations between the initial labeled (sponsoring) legislators, and the remaining (unlabeled) legislators, one can attempt to find a good partition of the legislators into a *Yea* group and a *Nay* group.

An intimately related perspective is that of belief propagation in a relational model (Pearl, 1988; Lu & Getoor, 2003; Richardson, 2004). Here, we are given a model representing the statistical dependencies within a data set. We are then given the values of some of the data points and asked to infer the values of the others. A typical approach to this task employs a Gibbs sampling technique, which proceeds in iterations, each time fixing the values of all data points except one, and sampling the value of that point from its distribution conditioned on the values of the other points. After all points have been sampled in this manner, the old values are replaced with the new sampled values and the process is repeated. After many iterations, one can estimate the probability that a given point has a specific value using the frequency with which it held that value during the sampling process. Given a way to convert the influence information in the network of legislators to probabilities, this technique could potentially provide a way to estimate the probabilities of a legislator voting *Yea* or *Nay.* After applying a threshold of ½, one can predict the voting behavior of the legislators.

4.2 A Sampling Approach

By combining elements of each of the above perspectives, one can construct an algorithm that estimates the voting behavior of the legislators in an influence network. Drawing inspiration from the threshold function used to model diffusion processes, define a probability distribution for legislator X_i voting *Yea*, given the votes of all other legislators as follows.

$$P\left\{X_i = Yea \middle| X_{-i}\right\} = S\left(\sum_{j \neq i} Influence(X_j, X_i)\right)$$

Here, $S(x) = \frac{1}{1 + e^{-\sigma(x-\theta)}}$ is the sigmoid function with threshold parameter θ and slope parameter σ. This function becomes the simple step function used in the information diffusion model as σ approaches ∞. Also, $Influence(X_j,X_i)$ is the influence the j^{th} legislator has on the i^{th} legislator, as defined by the number of times X_i co-sponsored legislation that X_j sponsored. Note the resulting conditional distribution does not necessarily reflect the actual frequency in the data, but will be a convenient quantity in the inference procedure.

Now, given this notion of probability, we can perform a sampling algorithm over this network. The complete inference algorithm is given below.

1. Fix the sponsor and co-sponsors as voting *Yea*.
2. Initialize the remaining legislators randomly.
3. Perform Gibbs sampling to recover the *Yea* frequency estimates for each unknown legislator.
4. Apply a threshold of ½ to obtain an actual vote for each legislator.
5. Predict *Pass* if and only if the number of *Yea* votes exceeds the number necessary for the legislation to pass.

5. Experimental Results

I evaluate the proposed approach on data consisting of all roll call votes taken in the 108^{th} U.S. Senate through October 2004. The first 90% of the roll call votes are used as a vote history, while the last 10% are used to evaluate the prediction accuracy.

As mentioned above, previous approaches to modeling roll call voting behavior are not applicable to the task of vote prediction. As such, the comparison is between the proposed algorithm and several baseline algorithms. The baseline algorithms were selected to represent approaches to the problem that do not employ a network structure.

5.1 Learning the Parameters

In the proposed algorithm, the network is constructed from co-sponsorship information as described above. The sigmoid function is used, and the inference procedure employs the sampling approach discussed in Section 4.2. The sigmoid function has two parameters: the threshold θ and slope σ. In the experiments that follow, these parameters are optimized by a simulated annealing procedure, using only the voting history for fitness evaluation (Rose, et al, 1990). The final values for this particular data set are $\theta = 8.5$ and $\sigma = 0.5$.

5.2 Comparison with Non-Structured Approaches

To provide a relative comparison with other methods, I have also evaluated several algorithms that perform the classification directly, without taking advantage of a network structure. The first such algorithm is a standard logistic regression algorithm, taken from the weka data mining library (Hastie, 2001; Witten & Frank, 2000). The second algorithm is a support vector machine, using a quadratic kernel; for these experiments, I use the popular SVM^{light} implementation (Joachims, 1999). As a baseline accuracy value, I provide results for a simple function that predicts *Pass* if the number of cosponsors is over a given threshold. This threshold is optimized on the voting history, but it turns out that for this data set the optimized value is 0 (i.e., always predict *Pass*). The results are displayed in Table 1.

Algorithm	Sampling	Logistic	SVM	Baseline
Prediction Accuracy	**81.3%**	62.5%	52.1%	58.5%

Table 1 displays results for the roll call vote prediction task using several algorithms. The approach utilizing the influence network provides a clear advantage.

5.3 Discussion

The results are encouraging for the structural approach, which displays a clear advantage over the other algorithms. One can interpret these results as being caused by a combination of intelligent modeling choices and the expressiveness of the model. That is, the baseline model is clearly not expressive enough, and as such achieves poor results. The logistic regression model does not capture the interactions between legislators, and may also be too expressive to accurately approximate the target, given the relatively small quantity of data in the voting history. The SVM model is almost certainly too expressive for this task. This hypothesis is confirmed by increasing the degree of the kernel, which yields progressively worse results. It seems that the network approach restricts the model to allow tuning the parameters in a way that can still generalize to new data, but restricts it in an intelligent way so as not to sacrifice the ability to represent a fairly accurate classifier.

6. Promising Future Directions

Legislative roll call vote prediction is a relatively unexplored, but interesting problem. The above results demonstrate the potential benefits of using a structural representation for this task, but they are still quite preliminary and allow much room for possible improvement.

One possible direction for improving upon this approach is to incorporate ideas from the spatial voting model. As mentioned before, the spatial voting model in its current form requires one to know the voting behavior of the legislators in order to determine the position points of the legislation. However, it might be possible to train an estimator that maps new legislation to approximate position points. The features for this estimator could possibly be drawn from the text of the legislation itself, or from expert knowledge of the legislation. It would then be possible to add an additional term to the inference algorithm, which accounts for the probability that the legislator votes sincerely.

Another possible direction is to introduce a notion of reciprocity, a key factor in legislative negotiations which is missing from the current model. The introduction of this notion forces one to view the data as a time series, with each legislator owing favors to or being owed favors by other legislators at any given time. Other additions to the model are also possible, such as vote buying or party influence (Groseclose & Snyder, 1996; McCarty, et al, 2001).

Additionally, it is desirable to design the inference procedure in a more principled way. That is, I chose the sigmoid function above based on an intuition of how influence should work, but there may be a more interesting function that better models the interaction. Such a function might be learned via genetic programming or other techniques (Koza, 1992).

References

Blum, A., Lafferty, J., Rwebangira, M., and Reddy, R. (2004). Semi-supervised learning using randomized mincuts. *ICML 2004*.

Clinton, J., Jackman, S., and Rivers, D. (2004). The statistical analysis of roll call data. In: *American Political Science Review* Vol. 98, No. 2.

Clinton, J., and Meirowitz, A. (2003). Integrating voting theory and roll call analysis: a framework. In: *Political Analysis 11: 381 – 396.*

Coleman, J., Menzel, H., and Katz, E. (1966). *Medical Innovations: A Diffusion Study.* Bobbs Merrill.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*. 83(6): 1420-1443.

Groseclose, T., Snyder, J. (1996). Buying supermajorities. *American Political Science Review* 90:303-315

Hastie, T., Tibshirani, R., Friedman, J. (2001). *The Elements of Statistical Learning.* Springer-Verlag, New York.

Joachims, T. (1999). Making large-scale SVM learning practical. *Advances in Kernel Methods – Support Vector Learning.* Schölkopf, B., Burges, C., and Smola, A. (ed.), MIT Press

Kempe, D., Kleinberg, J., and Tardos, É. (2003). Maximizing the spread of influence through a social network. *SIGKDD 2003*.

Koza, J. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* Bradford Books.

Ladha, K. (1994). A spatial model of legislative voting with perceptual error. *Public Choice. 68:151-17*.

Lu, Q. and Getoor, L. (2003). Link-based classification. *ICML 2003*: 496-503.

McCarty, N., Poole, K., Rosenthal, H. (2001). The hunt for party discipline in congress. *American Political Science Review* 95: 673-688.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann.

Poole, K. *Spatial Models of Parliamentary Voting.* Forthcoming book. Temporarily available at http://voteview.uh.edu/Spatial_Models_of_Parliamentary_Voting.htm

Richardson, M., (2004). *Learning and Inference in Collective Knowledge Bases.* Doctoral Dissertation, University of Washington.

Rose, R., Eitan, G., Fox, G. (1990). Statistical mechanics and phase transitions in clustering. In: *Physical Review Letters* Vol. 5, Number 8, pp. 945-948.

Vapnik, V. (1998). *Statistical learning theory*. John Wiley & Sons, Inc.

Witten, I., and Frank, E. (2000). *Data Mining: Practical Machine Learning Tools with Java Implementations.* Morgan Kaufmann, San Francisco.

Zhu, X. and Ghahramani, Z. (2002). *Learning from labeled and unlabeled data with label propagation.* Technical Report, Carnegie Mellon University.