A Longitudinal Study of Language and Ideology in Congress

Bei Yu[1]
Daniel Diermeier[2]

Abstract:

This paper presents an analysis of the legislative speech records from the 101st-108th U.S. Congresses using machine learning and natural language processing methods. We use word vectors to represent the speeches in both the Senate and the House, and then use text categorization methods to classify the speakers by their ideological positions. The classification accuracy indicates the level of distinction between the liberal and the conservative ideologies. Our experiment results demonstrate an increasing partisanship in the Congress between 1989 and 2006. Ideology classifiers trained on the House speeches can predict the Senators' ideological positions well (House-to-Senate prediction), however the Senate-to-House prediction is less successful. Our results provide evidence for a long-term increase in partisanship in both chambers with the House consistently more ideologically divided than the Senate.

[1] School of Information Studies, Syracuse University.
[2] Department of Managerial Economics and Decision Sciences (MEDS) and Ford Motor Company Center for Global Citizenship, Kellogg School of Management and Northwestern Institute on Complex Systems (NICO), Northwestern University.

Introduction

Over at least 50 years of research, *ideology* has been used to explain the political behavior of voters, legislators and other elites. In the context of mass political behavior, Converse (1964) has conceptualized ideology as a "belief system." Belief systems give structure to an individual's view on various issues. Intuitively, a political belief system expresses a view of which issue positions go together, the "knowledge of what-goes-with-what" (Poole 2003).

Empirically, belief systems allow us to predict an individual's position on an issue if we know his or her position on another issue. Speaking about belief "systems," however does not necessarily imply a logically consistent political, economic, or social world-view. Indeed, as Converse (1964) argued, the association between issues may just have been contingent and reflect a particular, perhaps cultural, or historical experience. Nevertheless, belief systems do constrain. It is quite unlikely, though not impossible, that a randomly selected U.S. voter who opposes universal health insurance, gun control, affirmative action, environmental regulation, abortion, and higher taxes also supports gay marriage. Converse expresses this idea as follows:

> "Constraint may be taken to mean the success we would have in predicting, given an initial knowledge that an individual holds a special attitude, that he holds certain further ideas and attitudes (Converse 1964, p.207)."

Measuring ideological orientations and belief systems, however, has always been a very difficult task. Unlike party affiliation, for example, ideology is not directly observable. Consequently, scholars have employed different strategies, ranging from survey responses to statistical estimates based on voting records. In legislative politics, and especially the U.S. Congress, the most widely used measure of ideology remains the vote-based score developed and refined by

Poole and Rosenthal (1991; 1997; 2007). The authors estimate ideology in Congress by applying a spatial voting model to Congressional roll call data. Legislators' ideal points are then estimated in choice spaces of various dimensions.

In recent years there has been emerging interest in using automated text analysis methods to measure ideological orientations (Laver, Benoit, and Garry, 2003, Monroe and Maeda 2004, Diermeier et al., 2007). In contrast to traditional statistic analysis based on voting behavior, these new approaches aim to infer legislators' ideological positions based on their speeches. The basic idea is that, just as voting behavior, political speech is an expression of an underlying ideology. Hence, we can use political speech to infer and understand ideological orientations. In Diermeier et al. (2007) we developed an approach based on text-classification for this purpose. Based on automatic text classification methods[3], we conducted an experiment to train "ideology classifiers"[4] using sample Senatorial speeches (labeled by the signs of D-NOMINATE scores as well as the speakers' party affiliations) from 101st-107th Congress, and then use the "ideology classifiers" to predict the ideologies of Senators in the 108th Congress. High prediction accuracy would demonstrate that the ideology concept automatically inferred from the speech data are indeed shared by a group of people (for example, members of the same political party), and are able to predict political positions in various issues and in future periods.

The results, reported in Diermeier et al. (2007), showed that the "ideology classifiers," trained on the 101st-107th Senatorial speech, can reach up to 92% accuracy in predicting the ideology in the 108th Senate. However, because of Congress member turnover, a majority of the

[3] Text classification includes two steps: training and prediction/inference. A collection of text documents with pre-assigned category labels (e.g. "liberal" and "conservative") is needed as training data to build a text classifier. The text classifier can then be used to assign category labels to unlabeled documents.

[4] Different classifiers can be trained based on different approaches to quantifying the same training speeches.

Senators in the 108th Congress are also members in previous Congresses, leaving a small number of out-of-sample predictions. Such overlap between training and testing Senators can cause problems interpreting the results. On the one hand, the ideology classifiers seem to have captured general ideology characteristics in that the most indicative words ranked by the classifiers do depict the liberal and conservative issues as we expected. On the other hand, the high predication accuracy could also partially result from characteristics of specific speakers or states which could also be captured by the classifiers, e.g. some characteristics of traditional blue and red states, or some characteristics of long-standing Senators. These factors can be controlled to some extent by removing Senators' names and the state names from the feature list, and restricting that a word has to occur in at least five Senators' speech to be an eligible feature. But these solutions are only partial and somewhat *ad hoc* as long as there is substantial overlap between the Senators in the training and testing data. Another option is to train the ideology classifiers on earlier Congresses (e.g. 101st-104th) and evaluate the prediction on recent Congresses (e.g. 108th) in order to reduce the overlap, however the prediction result is less desirable because of the vocabulary drift over time – the words used in the 101st Congress are quite different from the ones used in the 108th Congress. In other words, we may conflate an underlying shift in what it means to be a Conservative or Liberal with a classification task.

Cross-Chamber Design

To improve on this approach it is worthwhile to revisit the concept of a political ideology. Here we focus on three important characteristics. First, ideologies correlate with behavior whether realized as voting or speech. That is, a conservative will vote and talk differently from a liberal.

Second, ideologies are constant across different issues. They constrain which issues positions go together. Third, ideologies are shared among agents. We can group individuals into ideological factions.

In subsequent work we explored the third characteristic of political ideologies. The bicameral structure of the U.S. Congress provides an ideal setting for this approach. To examine whether the ideology classifiers only work well when the test and sample data are drawn from the same population of legislators, we designed a second round of experiments (Yu et al, 2008a) to control the person factor by using speeches from both the House and the Senate. Since there is no overlap between members of the two chambers, we can prove the ideology classifier's person-independence if an ideology classifier trained on one chamber can predict well on the other chamber.

We used the 2005 House speech data collected by Thomas et al. (2006), and our own 2005 Senate speech data to carry out this round of experiments. Since D-NOMINATE scores are not directly comparable across chambers, we used legislators' party affiliations as the signs of their ideological positions. That is, the task now was to classify party affiliation correctly. Specifically, we designed two experiments: in the first one we trained ideology classifiers on the 2005 House speech and tested them on the 2005 Senate speech; in the second experiment we switched the training and testing data.

The experiment result demonstrated that the overall "2005 House to Senate" prediction results (up to 88%) are much better than the "2005 Senate to House" prediction results (up to 68%). This finding supports a common notion that the House is more partisan than the Senate. This finding also suggests that the House speeches might be better suited than the Senatorial speeches to the task of training person-independent ideology classifiers. Further experiments showed that the ideology classifiers trained on 2005 House speech also predicted well on the Senate speech of recent years (2003, 2004, 2006). However, the performance of the ideology classifiers trained on House data decreased when making predictions based on older Senate data, confirming again the classifiers' time dependence due to vocabulary drift.

A longitudinal study

The above result provides evidence that person-independent ideology classifiers can be trained on House speech. The method can capture the idea that ideologies are shared among different individuals. The high classification accuracy of both studies also suggests that respective speech characteristics are highly correlated with party affiliation. What is missing is a longitudinal approach. Our previous cross-chamber results were based on only one year of House data. To obtain more evidence we collected the House speech data from 1989 to 2006 from the Thomas government database, and repeated the "House-to-Senate" and "Senate-to-House" prediction experiments on each year's Congressional speeches. We want to examine (1) whether high-performing ideology classifiers can be inferred from the House speech of every year, and (2) whether House-trained ideology classifiers always outperform the Senate-trained classifiers. If both answers are positive, we can prove that House is more partisan than the Senate and that House speech is suitable data to train ideology classifiers regardless of time.

For every year's House and Senate speech data, we train ideology classifiers using three classification methods (SVM-bool, SVM-ntf, and SVM-tfidf). The three SVM classification methods are based on different text representation models and have been proven effective in our previous experiments. SVM-bool represents every legislator’s speech as a vector of word presence or absence. SVM-ntf represents every legislator’s speech as a vector of word frequency normalized by the length of the speech. SVM-tfidf represents every legislator’s speech as a vector of word frequency normalized by word’s document frequency. We used the SVM-light package with its default parameter settings for the implementation of this study.

For each method we used two measures to test the trained classifier’s performance. The first is the classifier’s Leave-One-Out Cross Validation (LOOCV) accuracy on the training chamber. Leave-One-Out Cross Validation is a special case of N-fold cross validation, which partitions a dataset into N folds and runs the classification experiment N times, each time using one fold of data as the test set and training the classifier on the remaining N-1 folds, and at the end averages the classification accuracy of N times. The second measure is the classifier’s cross-chamber prediction accuracy. A classification method’s LOOCV performance indicates the level of distinction between the liberal and conservative ideologies in the training set. A high LOOCV accuracy suggests that the ideology concepts are highly separable using the current method, and that the trained classifier is expected to perform well on out-of-sample test data if the training and test data are homogeneous. In contrast, a low LOOCV accuracy suggests that the current method was not able to separate the liberal and conservative ideology concepts from the training set and thus will not be able to predict on out-of-sample data either.

The 1989-2005 Senate LOOCV test and the “House to Senate” prediction results are reported in Table 1 and visualized in Figure 1. In Figure 1, the solid line is the Senate majority baseline indicatingthe accuracy that a trivial majority vote classifier can reach. For the Senate data, the majority vote accuracy ranges from 50% to 57% over the years. A non-trivial ideology classifier should at least outperform the majority baseline to be deemed as useful. In Figure 1, the three thin dotted lines indicate the Senate LOOCV test results (corresponding to columns 4, 6, 8 in Table 1). The three thick dotted lines indicate the House to Senate prediction results (corresponding to columns 3, 5, 7 in Table 1). Over the 18 years, the average majority baseline is 55% (see the last row in Table 1). The average Senate LOOCV accuracies are at the 60% level (63%, 62%, and 65% for SVM-bool, SVM-ntf, SVM-tfidf respectively). The average House to Senate prediction accuracies are at the 70% level (79%, 72%, 76% for SVM-bool, SVM-ntf, and SVM-tfidf respectively). The trend in Figure 1 also shows that for every year the “House to Senate” prediction accuracy is almost always higher than the Senate LOOCV accuracy, and of course is higher than the majority baseline also. These patterns indicate that the Senate-trained ideology classifiers generally perform better than the majority baseline, but the House-trained ideology classifiers can predict even better on the Senate data.

The 1989-2005 House LOOCV test and the “Senate to House” prediction results are reported in Table 2 and visualized in Figure 2. For the House data, the majority vote accuracy ranges from 51% to 62% over the 18 years. In Figure 2, the three thin dotted lines indicate the House LOOCV test results (corresponding to columns 4, 6, 8 in Table 2). The three thick dotted lines indicate the Senate to House prediction results (corresponding to columns 3, 5, 7 in Table 2). Over the 18 years, the average majority baseline is 55%. The average House LOOCV accuracies are at the 70-80% level (79%, 80%, and 75% for SVM-bool, SVM-ntf, SVM-tfidf respectively). The average Senate to House prediction accuracies are at the 50-60% level (55%, 63%, 68% for SVM-bool, SVM-ntf, and SVM-tfidf respectively). The trend on Figure 2 also shows that for every year the “Senate to House” prediction accuracy is always lower than the House LOOCV accuracy.

Combining the results in both experiments we summarize the classification accuracies (averaged over years) in Table 3. The summary clearly shows that the House-trained ideology classifiers outperform the Senate-trained classifiers in both cross-validation tests and the cross-chamber prediction tests. This finding once again indicates that the ideology concepts in the House speeches are consistently more separable from 1989 to 2006, suggesting that the House is consistently more partisan than the Senate.

The trends in Figures 1 and 2 also show that the House-trained ideology classifiers do not maintain constant performance over the years. Both the LOOCV curves and the House to Senate prediction curves demonstrate a fluctuating yet increasing trend. This trend also indicates that the Congress is increasingly partisan in recent years.

Why is the House data better for training ideology classifier?

Our experiment results indicate that the House speech data are better suited than the Senatorial speeches to the task of training person-independent ideology classifiers. We have interpreted this as evidence that the House is more partisan than the Senate. From a classification theory perspective, however, some characteristics of the House speech, unrelated to higher partisanship, could possibly make it more suitable for training ideology classifiers from a classification theory perspective. If this is true our findings that the House is more partisan than the Senate would need to be qualified. Here we analyze two major differences between the Senate and the House speeches, which might affect classifiers' performance. The first is the number of training examples, and the second is the organizational difference between the House speech and the Senate speech.

First of all, the number of House representatives is more than four times the number of Senators. In other words, the House data provide more training examples than the Senate data in that each legislator forms one training or testing example. According to machine learning theory, larger numbers of training examples help prevent the trained classifier from over-fitting the training data. We have seen that the House-trained ideology classifier achieves comparable LOOCV accuracy (row 1 in Table 3) and cross-chamber prediction accuracy (row 2 in Table 3), except that for the SVM-ntf classifier the cross-chamber prediction accuracy is lower than the House LOOCV accuracy. This indicates that the House-trained ideology classifier generalizes well on the Senate data, and thus does not over-fit the House data. Likewise, if the Senate-trained ideology classifier over-fits the training data, its LOOCV accuracy (row 3 in Table 3) should be higher than its cross-chamber prediction accuracy (row 4 in Table 3). The results in rows 3 and 4 in Table 3 show that it is true for SVM-bool, whose average Senate LOOCV accuracy is 63% and the average Senate to House prediction accuracy is only 55%, the same as the majority vote. This means this Senate-trained SVM-bool classifier is over-fitting the Senate data. However, we did not observe over-fitting from the Senate-trained SVM-ntf and SVM-tfidf classifiers, both of which achieved comparable accuracies on LOOCV tests and cross-chamber prediction tests. Therefore, over-fitting is not a convincing explanation for the lower performance of Senate-trained ideology classifier.

Second, Senate speeches are considerably longer and unconstrained by the germaneness rule, which permits a less restricted expression of political ideologies (Table 4). In contrast, the House speeches are more organized and thus reflect more consistent positions. Yu et al. (2008b) have found that 62% of the beginning sentences and 72% of the ending sentences in House speeches clearly indicate the overall position (support or opposition) of each speech. In contrast the percentages drop to 49% and 35% in the Senate speeches. From a text classification perspective, speeches with clear and consistent ideology or opinion indicators would be easier to classify. As a result, the above differences between Senate speeches and House speeches could make the ideology classification of Senate speeches more difficult. However, if this is the sole reason for the lower performance of Senate-trained ideology classifier, we should expect a constantly low performance over time. However, the Senate LOOCV curves in Figure 1 have demonstrated an increasing trend, which means the ideology concepts are more separable in recent-year Senatorial speeches. This suggests that both House and Senate are becoming more partisan.

In summary although the characteristics of the House speech seem to make it more suitable to train ideology classifiers, our findings still hold that the Congress is increasingly partisan in recent years and that the House is more partisan than the Senate.

Conclusion

In this article we used automatic text classification methods to study the ideology in Congress. Our new longitudinal study results provide strong evidence that both the House and the Senate are becoming more partisan in recent years, and the House is more partisan than the Senate. As a next step, our goal is to understand the content and changing nature of partisanship in more detail.

Table 1: 1989-2005 House to Senate (H2S) Prediction Accuracies (percent)

Year	Majority	H2S prediction SVM-bool	Senate LOOCV SVM-bool	H2S prediction SVM-ntf	Senate LOOCV SVM-ntf	H2S prediction SVM-tfidf	Senate LOOCV SVM-tfidf
1989	55.0	65.0	56.0	59.0	55.0	69.0	59.0
1990	55.0	63.0	62.0	62.0	55.0	65.0	60.0
1991	56.6	63.6	55.6	63.6	56.6	74.8	63.6
1992	56.6	82.8	56.6	66.7	56.6	77.8	57.6
1993	57.0	73.0	58.0	78.0	57.0	72.0	62.0
1994	56.6	70.7	58.6	83.8	56.7	83.8	67.7
1995	54.1	93.9	60.2	79.6	77.6	85.7	59.2
1996	53.5	82.8	51.5	61.6	53.5	77.8	52.5
1997	55.6	73.7	55.6	72.7	55.6	83.8	56.6
1998	55.0	65.0	63.0	68.0	55.0	76.0	65.0
1999	54.6	85.9	65.6	72.7	60.6	82.8	70.7
2000	54.0	82.0	71.0	74.0	68.0	82.0	63.0
2001	50.0	85.0	55.0	85.0	64.0	66.0	70.0
2002	50.0	80.0	59.0	70.0	63.0	72.0	78.0
2003	51.0	85.4	76.0	79.2	73.0	69.8	76.0
2004	51.5	86.9	76.8	79.8	60.6	71.7	72.7
2005	55.0	85.0	75.0	73.0	75.0	78.0	75.0
2006	55.0	94.0	76.0	75.0	76.0	76.0	66.0
Avg	54	79	63	72	62	76	65
St. dev.	2	10	8	8	8	6	7

Table 2: 1989-2005 Senate to House (S2H) Prediction Accuracies (percent)

Year	Majority	S2H prediction SVM-bool	House LOOCV SVM-bool	S2H prediction SVM-ntf	House LOOCV SVM-ntf	S2H prediction SVM-tfidf	House LOOCV SVM-tfidf
1989	59.7	59.7	70.2	61.3	70.2	62.8	63.5
1990	59.7	59.4	70.7	63.5	72.6	62.5	61.3
1991	61.6	61.6	71.2	62.5	74.5	63.4	65.3
1992	61.6	61.9	73.4	63.1	75.5	71.0	67.4
1993	58.5	58.5	79.9	63.1	81.8	68.3	72.9
1994	58.5	58.7	76.4	60.3	79.4	70.1	74.1
1995	53.8	53.8	81.1	61.5	84.7	79.9	79.9
1996	53.8	54.5	83.8	55.0	82.1	70.7	72.4
1997	52.9	52.9	75.5	52.9	78.1	67.8	73.2
1998	52.9	52.6	79.3	55.9	74.8	66.3	73.6
1999	51.5	51.6	81.7	66.6	82.4	71.1	78.4
2000	51.5	51.3	85.1	62.1	80.8	67.8	79.9
2001	51.4	51.9	77.9	61.9	80.9	63.3	77.7
2002	51.4	52.3	80.1	61.1	82.7	60.3	82.0
2003	52.6	52.9	84.8	74.6	86.1	70.7	84.3
2004	52.6	53.4	81.6	67.4	84.0	71.6	83.7
2005	53.8	54.9	80.3	68.5	85.6	73.8	84.0
2006	53.8	54.4	80.5	71.5	85.8	71.2	82.3
average	55	55	79	63	80	68	75
St. dev.	4	4	5	5	5	5	7

Table 3: Ideology Classification Accuracies averaged over years (percent)

Evaluation measure	SVM-bool	SVM-ntf	SVM-tfidf
House LOOCV	79	80	75
House to Senate	79	72	76
Senate LOOCV	63	62	65
Senate to House	55	63	68

Table 4: statistics of 101-108th Senate and House speeches

Congress	House members	Avg words	Total words	Senators	Avg words	Total words
101	431	25,648	11,054,179	101	170,170	17,187,122
102	433	35,077	15,188,322	100	267,797	26,779,721
103	429	28,688	12,307,100	102	210,302	21,450,821
104	425	38,649	16,426,028	101	224,757	22,700,438
105	433	31,623	13,692,820	100	186,890	18,689,025
106	431	40,691	17,537,719	100	194,398	19,439,828
107	437	30,908	13,506,791	103	172,019	17,717,979
108	432	35,291	15,245,857	100	180,683	18,068,293
109	428	36,483	15,614,603	101	154,594	15,459,406

Figure 1

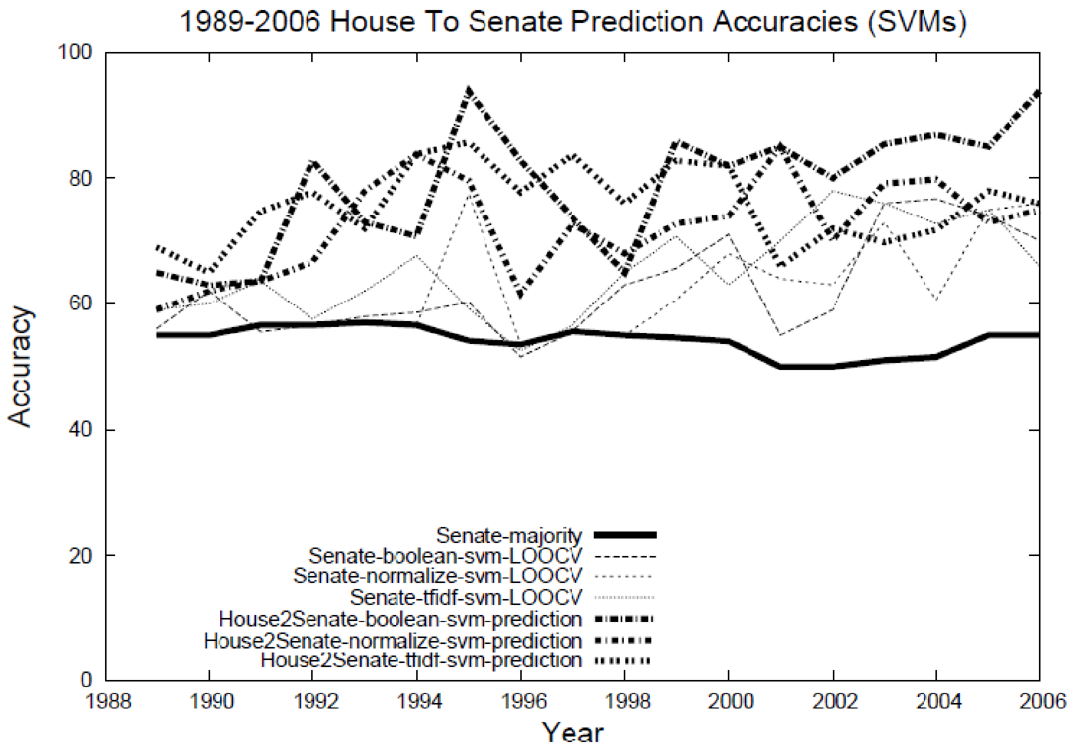
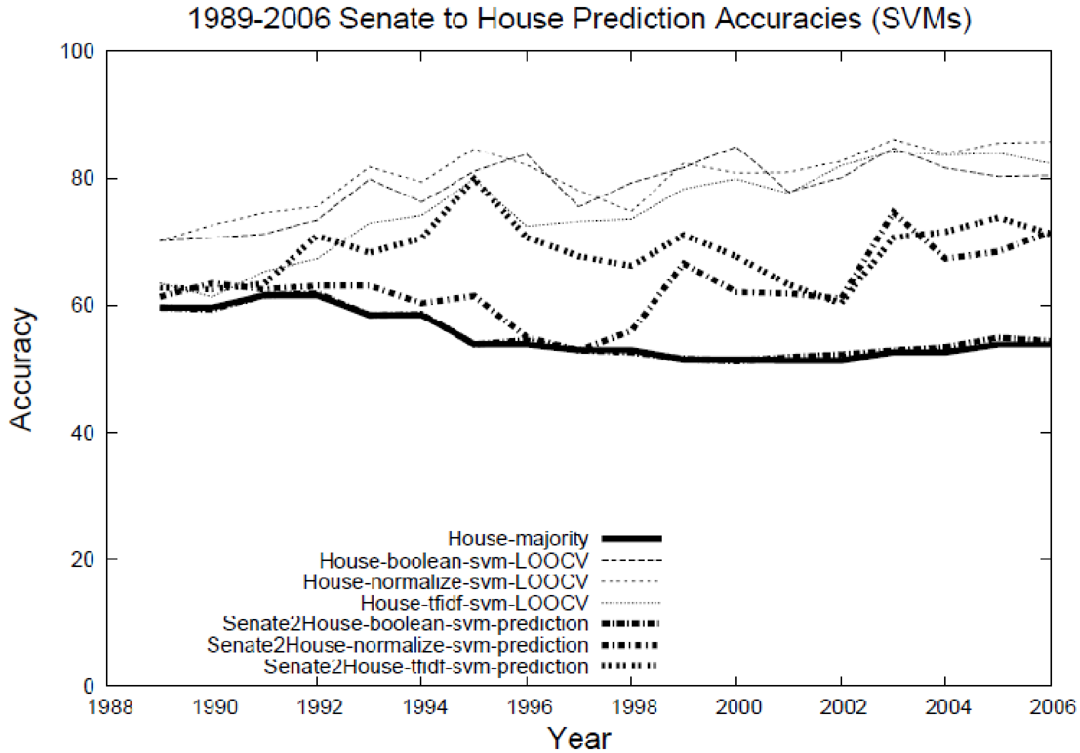

1989-2006 House To Senate Prediction Accuracies (SVMs)
Accuracy
Year
100
80
60
40
20
0
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
Senate-majority
Senate-boolean-svm-LOOCV
Senate-normalize-svm-LOOCV
Senate-tfidf-svm-LOOCV
House2Senate-boolean-svm-prediction
House2Senate-normalize-svm-prediction
House2Senate-tfidf-svm-prediction

Figure 2


1989-2006 Senate to House Prediction Accuracies (SVMs)
Accuracy
Year
0
20
40
60
80
100
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
House-majority
House-boolean-svm-LOOCV
House-normalize-svm-LOOCV
House-tfidf-svm-LOOCV
Senate2House-boolean-svm-prediction
Senate2House-normalize-svm-prediction
Senate2House-tfidf-svm-prediction

References:

Converse, Philip E. 1964. “The Nature of Belief Systems in Mass Publics.” In *Ideology and Discontent*, edited by D.E. Apter. New York: Free Press.

Diermeier, D., Godbout, J-F, Yu, B., & Kaufmann, S. (2007, April). Language and ideology in Congress. Paper presented at the annual meeting of the Midwest Political Science Association (MPSA’07), Chicago.

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, 97, 311-337.

Monroe, B. L. & Maeda, K. (2004). Rhetorical ideal point estimation: Mapping legislative speech. Society for Political Methodology, Stanford University, Palo Alto.

Poole, K. T., & Rosenthal, H. (1991). Patterns of Congressional Voting. *American Journal of Political Science, 35*(1), 228-278.

Poole, Keith T., and Howard Rosenthal. 1997 Congress: A Political-Economic History of Roll Call Voting. New York: Oxford

Poole, Keith T., and Howard Rosenthal. 2007. Ideology and Congress. Transaction Publisher.

Poole, Keith T. 2005. *Spatial Models of Parliamentary Voting*. Cambridge: Cambridge University Press.

Poole, Keith T. 2003. Changing Minds? Not in Congress. Unpublished Manuscript. Thomas, M., Pang, B., & Lee, L. (2006). Get out the vote: Determining support or opposition from

Congressional floor-debate transcripts. *Proceedings of the 2006 Conference on Empirical Methods in Natural LanguageProcessing (EMNLP'06)*, 327–335.

Yu, B., Kaufmann, S. and Diermeier, D. (2008a) Classifying party affiliation from political speech. *Journal of Information Technology and Politics*, 5(1), pp. 33-48.

Yu, B., Kaufmann, S. and Diermeier, D. (2008b) Exploring the characteristics of opinion expressions for political opinion classification. *Proceedings of the 9th Annual International Conference on Digital Government Research* (dg.o 2008), Montreal, Canada, May 2008, pp. 82-91.