Introduction and Motivation

Bills passed by Congress may have a significant impact on corporations and citizens. If we could predict bill passage we could give advance warning to citizens and corporations. Further being able to predict a representative’s vote on a bill could give insight into a representatives actual beliefs and platforms. Also we think its a pretty cool thing to do. Finally previous work in this area of prediction is limited. The only other foray was done by Gerrish and Blei; however, they extended the ideal point model from political science to predict votes and primarily used bill text.

Objectives

We are interested in a few questions in this research space:

Can we predict with high accuracy (above a baseline prediction) how a representative will vote on a given bill in Congress?

What feature of a bill is most important to a representative when deciding how to vote on a bill?

Can we predict whether a bill will pass using individual representatives’ predictions?

Methodology

We used [www.govtrack.us](http://www.govtrack.us) api for accessing their database to pull voting data and bill information for all current representatives of the House of Representatives. We focused upon the House as it tends to vote on bills more often and thus provide us with more data. Each representative had between 150 and 3,500 or so votes on bill passage.

From each bill we extracted certain features, for example the party of the sponsor, district of sponsor, name of sponsor, gender of sponsor, length bill was alive, and year bill was introduced.

We treated each representative as an individual prediction task. Thus we pulled all bills for an individual representative, each bill acting as an example point. We separated the bills into train, validation and test sets and then optimized an SVM. We did this for each representative, thus creating 430 optimized SVMs. We optimized the SVMs from 5 C values.

We compared each representative’s SVM’s performance against the performance of our baseline hypothesis. Our baseline hypothesis was that a representative votes along party lines. Meaning if the bill sponsor is of the same party as the voting representative he or she will vote yes, and no otherwise.

To test our ability to predict bill passage, we took a bill (\*\*\*\*FROM WHERE\*\*\*\*) and predicted representatives’ votes on that bill and then took a weighted sum of votes, where each vote was weighted by our accuracy on the representative’s test set.

Results

(INSERT HERE)

Conclusions

Representatives for with lower baseline accuracy tended to have much higher accuracies with the optimized SVM. When party sponsor is not a predominant factor for a representative our optimized SVM captures other factors and performs well.

Features

* Run McNemar’s test on each representative’s optimized SVM against the baseline for each representative.
* Consolidate McNemar’s tests to determine if optimized SVMs are likely better than baseline hypothesis.
* Separate bills into categories using clustering on bill summaries.
* Consider bill summary and bill categories as further features.
* Consider bill summary and bill categories as isolated feature sets.