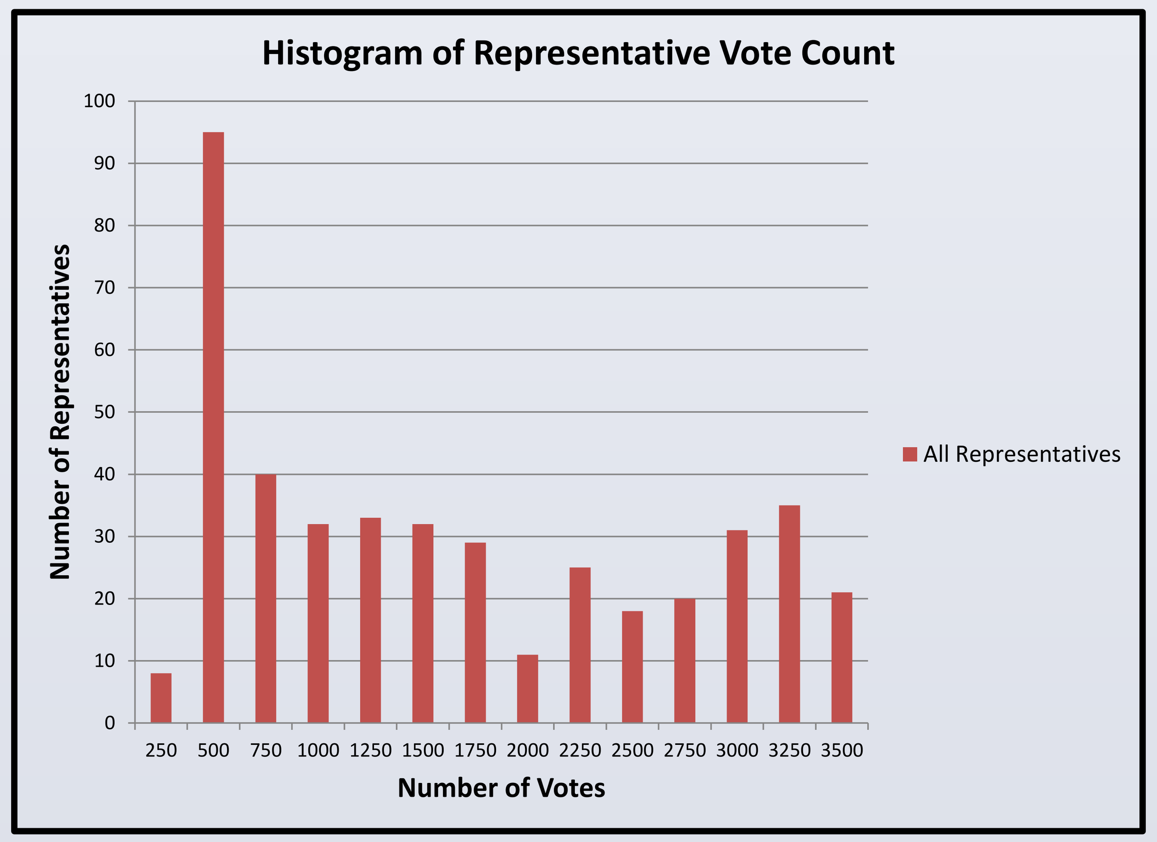
Vote Prediction in the House of Representatives

Jisha Kambo, John Oliver, Benjamin Shulman

Department of Computer Science, Cornell University

Introduction

- Bills passed by Congress may have huge effects on corporations and citizens.
- Being able to predict outcome of votes could give corporations and citizens advanced warning.
- Representative vote prediction could give insight into representatives' actual beliefs and platforms.
- It's cool to predict political actions.
- Previous work by Gerrish and Blei focused upon bill text.
- Previous work extended the ideal point model from political science.



A histogram of our data sets, showing distribution of size of sets.

Objectives

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?

Group ID: 12

Data Collection

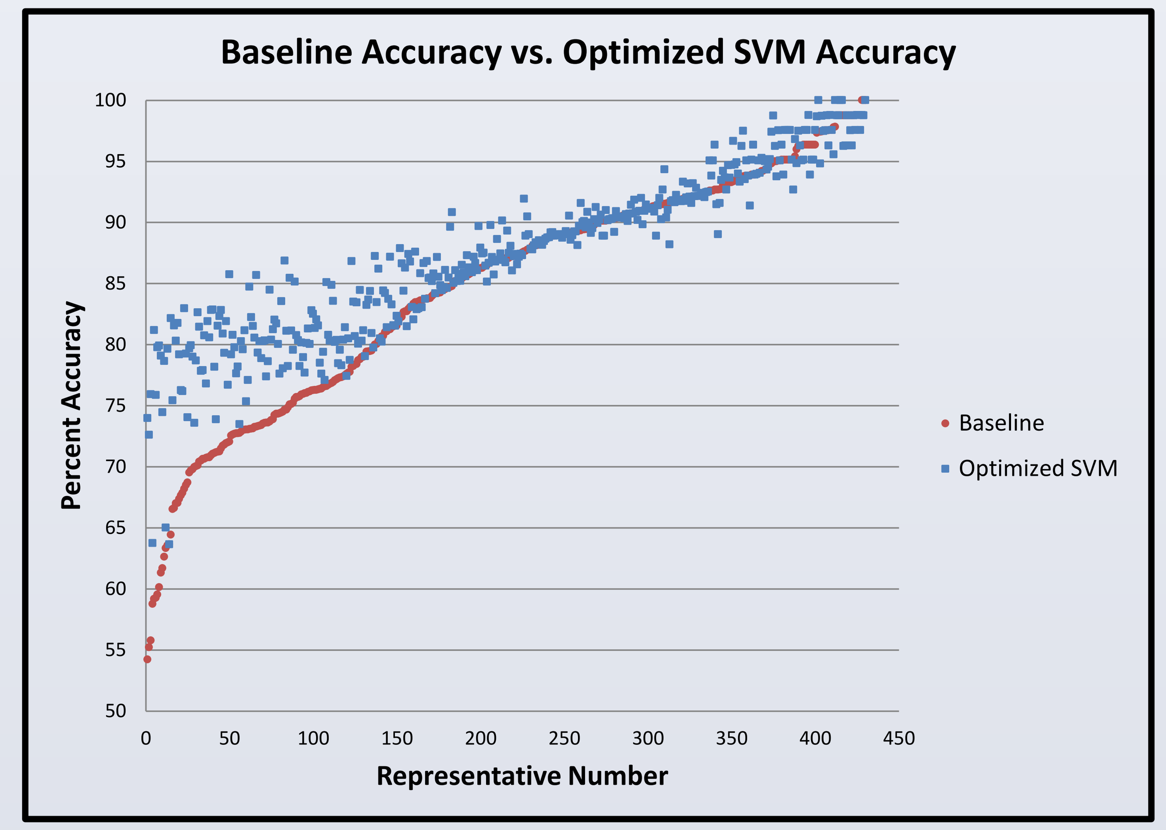
- Limited scope to House of Representatives
- Data was taken from www.govtrack.us
- Pulled all current members of the House of Representatives
- Pulled all votes made by each current member (200-3,500)
- Pulled information for each bill
- Features of SVM:
  - Party of bill sponsor
  - Name of bill sponsor
  - District of bill sponsor
  - Gender of bill sponsor
  - Date bill sponsor joined House of Representatives
  - Whether bill sponsor has Twitter
  - Whether bill sponsor has a website
  - Whether sponsor has a nickname
  - Length of time bill has been alive
  - Congressional Session
  - Year the bill was introduced
  - Day the bill was voted on
  - Month the bill was voted on
  - Year the bill was voted on
  - Year the bill was voted on mod 2
  - Year the bill was voted on mod 4
  - Year the bill was voted on mod 6
- Name of sponsor feature was numerous features, each a binary feature corresponding to one bill sponsor in the task.
- District of sponsor feature was numerous features, each a binary feature corresponding to one bill sponsor in the task.

Methodology

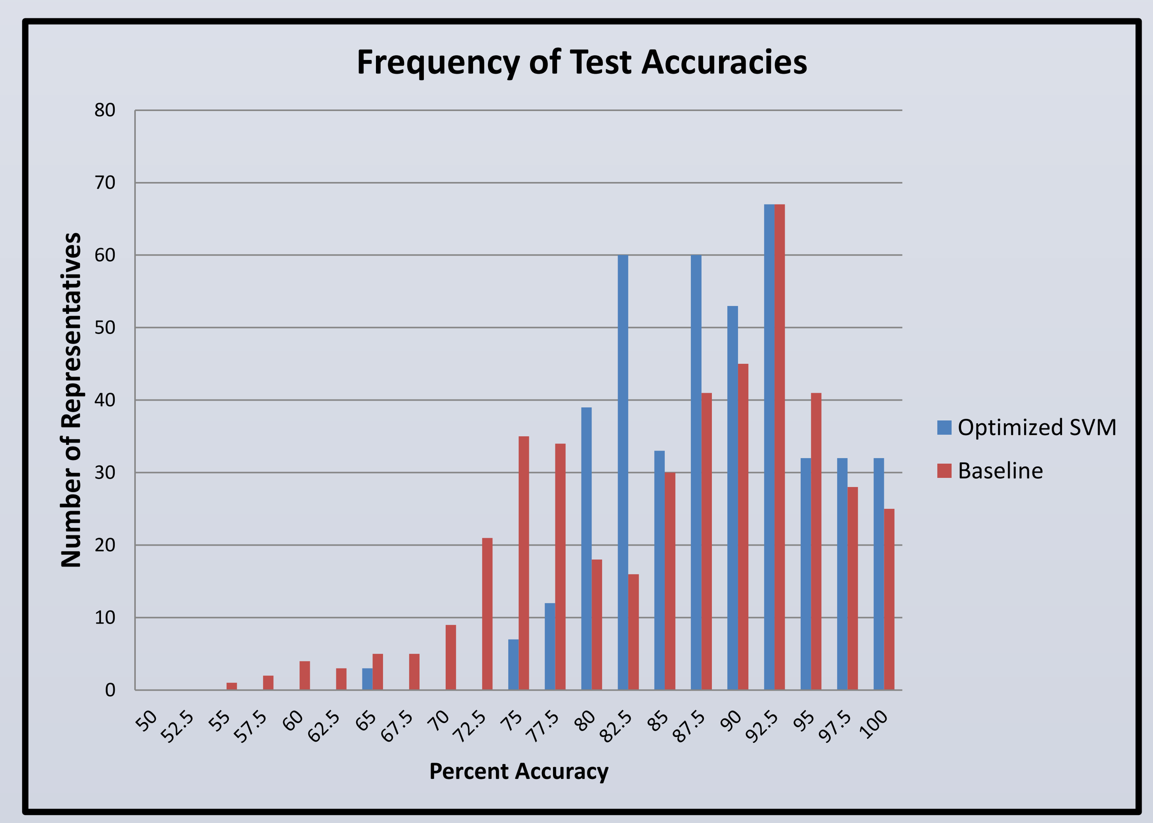
- Individual prediction task for each representative
- One SVM per representative
- Each SVM trained and tested on its representative's votes
- SVM Optimization
  - Split train set into train and validate
  - Trained 5 different SVMs, C values of [0.0001, 0.001, 0.01, 0.05, 0.1]
  - Picked SVM with best validation accuracy
- Compared each SVM against the baseline hypothesis.
- Baseline Hypothesis: representative will vote along party lines
  - Vote yes (1) if party of rep = party of sponsor
  - Vote no (0) otherwise

Results

- Average test accuracy over all representatives for baseline: 80.80%
- Average test accuracy over all representatives for optimized SVM: 84.62%
- Range of test accuracies for baseline: [54.22%,100%]
- Range of test accuracies for optimized SVM: [63.64%,100%]
- Generally Optimized SVM performed better than baseline on a per representative basis.
- Optimized SVM C values were predominantly C=0.1. A few instances of C=0.01 and C=0.05.



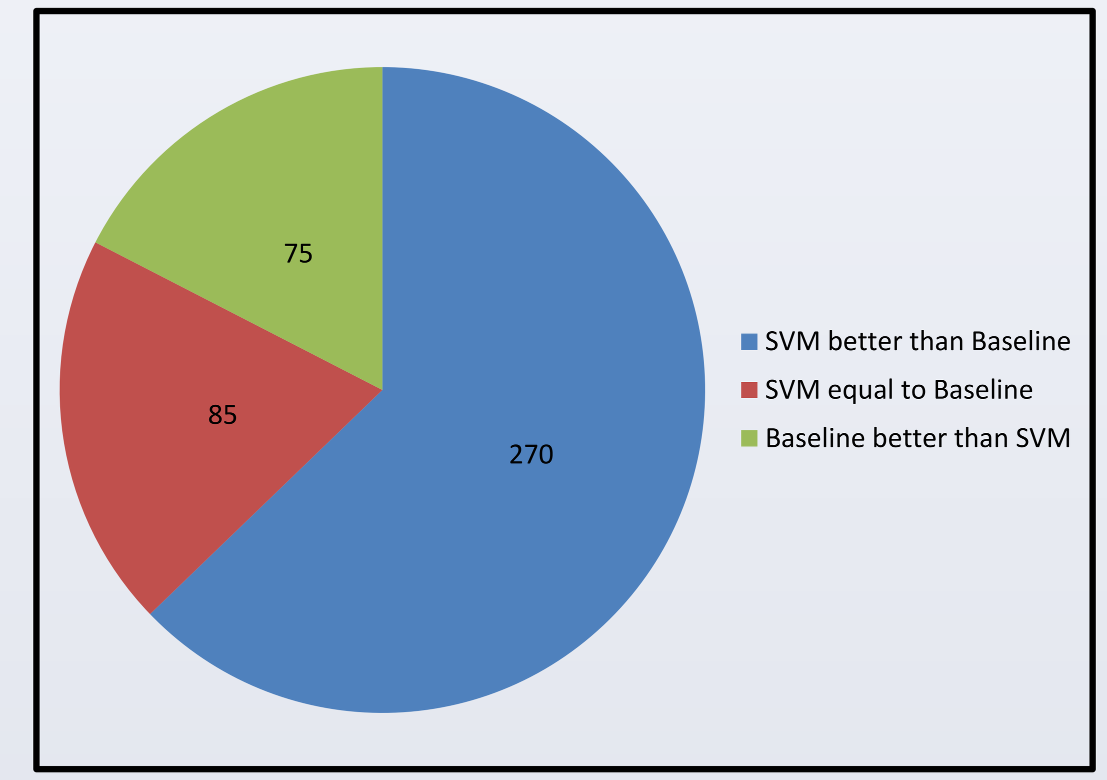
Accuracy of baseline compared to optimized SVM. Reps are sorted by increasing baseline accuracy.



Histogram showing frequency of representative accuracies in 2.5% bins.

Conclusions

- Representatives 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.



Each section represents the number of representatives satisfying the category label on the test set.

Future Work

- Predict bill passage using sum of predicted votes, weighted by test set accuracy.
- 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.

References

Gerrish, S. M., & Blei, D. M. (2011). Predicting legislative roll calls from text. *International Conference on Machine Learning*

Acknowledgements

Cornell University
Thorsten Joachims
Igor Labutov
Govtrack.us
CS4780 Teaching Assistants and Consultants

Contact

Jisha Kambo:	jk2228@cornell.edu
John Oliver:	jro67@cornell.edu
Benjamin Shulman:	bgs53@cornell.edu