Learning Alignments from Legislative Discourse

Daniel Kauffman[1,2], Foaad Khosmood[1,2], Toshihiro Kuboi[1], Alex Dekhtyar[1,2]

1. Institute for Advanced Technology and Public Policy
2. California Polytechnic State University, San Luis Obispo, Department of Computer Science and Software Engineering

Problem Statement

In this work, we seek to quantify the extent to which a legislator's spoken language indicates their degree of alignment toward an organization that has a taken a documented position on some legislation. To perform this study, we use a corpus of bill discussion transcripts provided by Digital Democracy, an online platform that promotes transparency in American state governments. We then apply proven learning methods in the field of natural language processing to predict alignment scores between each member of the California state legislature and a select set of state-recognized organizations. Our methods surpass established baselines, achieving up to 78% predictive accuracy using a combination of discourse and legislator-related features.

Digital Democracy

In 2012, former California State Senator Sam Blakeslee founded the Institute for Advanced Technology and Public Policy (IATPP), a non-profit organization housed at California Polytechnic State University. Through private donations and student development, the IATPP launched Digital Democracy, a web service for increasing government transparency and accountability.

One of the primary goals at Digital Democracy is to provide searchable transcripts of video-recorded bill discussions from state legislatures. As of 2018, the Digital Democracy database contains over one million transcribed utterances from the California state legislature.

Alignment Scores

Project staff have manually labeled the bill positions of several organizations, indicating whether each organization is for or against a particular bill. Used with legislator voting records, we can determine how often a legislator's vote agrees with an organization's position.

Aggregated across all bills for which tagged data exist, we calculate scores that indicate the degree of each legislator-organization alignment. These alignment scores are used as the ground truth when measuring the accuracy of our experiments.

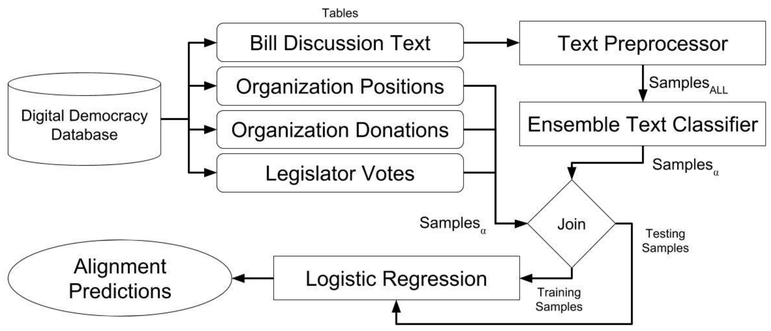
Data Set

Category	Description
Organization Positions	Stances (For or Against) over a set of bills
Organization Donations	Amount given to legislators in a session
Legislator Votes	Votes (Aye or Nay) on bills
Bill Discussion Text	Transcribed discussions during a hearing

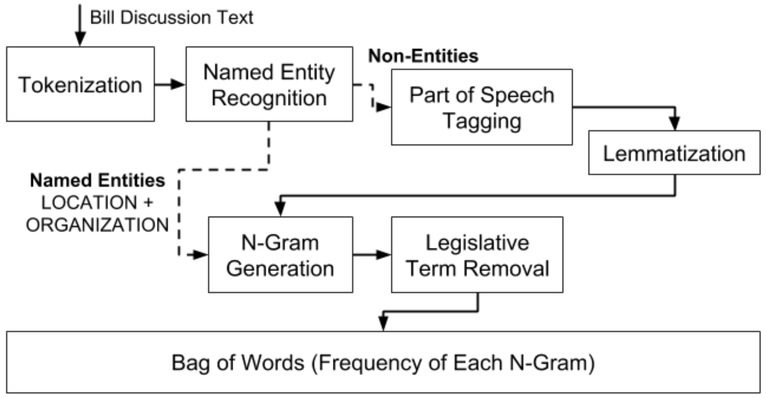
Data Sources

California LegInfo	Cal-Access	MapLight
Legislator Voting Records Bill and Text Analysis	Lobbying Activity	Donations
http://leginfo.legislature.ca.gov/	http://cal-access.sos.ca.gov/	https://maplight.org/

System Design



Text Preprocessor



Ensemble Text Classifier

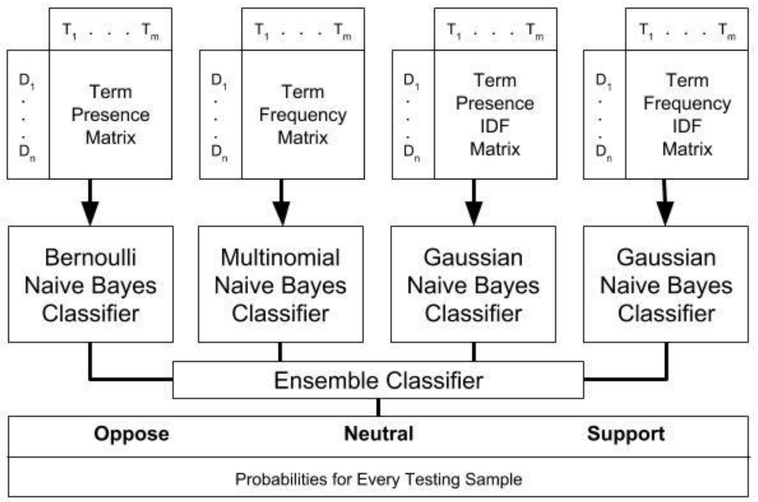
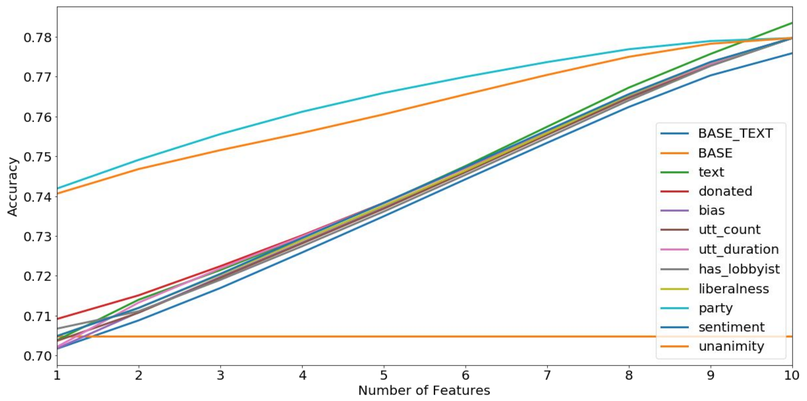


Table of Features

Name	Description
Utterance Text	Transcribed legislator speech
Utterance Count	Average number of utterances spoken by legislator
Utterance Duration	Average number of seconds spoken by legislator
Bias Corpus Hit Rate	Count of words from bias lexicon in each utterance
Sentiment Score	Average scores from VADER sentence analysis
Donations	Cumulative sum of donations, gifts, and behests
Political Party	Partisan affiliation of each legislator
Liberalness	Proportion of Democratic members on committee
Bill Vote Unanimity	Degree of vote agreement of committee members
Lobbyist Present	Whether organization lobbyist was at discussion

Results

The line plot below visually represents the impact that testing with every combination of features has on accuracy. The x-axis indicates how many features were used for a test while the y-axis displays the average accuracy for tests with that number of features. Each line thus represents all tests for which a specific feature was present, with the average accuracy of a feature changing from left to right as the number of other features tested with it increases.



While the party feature is expectedly dominant, the unanimity feature performs comparably well. We hypothesize that, since the Democratic party held a majority during this session, agreements with unanimous votes are easier to predict. At 78% accuracy, we find that combining all features produces the best results, increasing past the baselines and the dominant features individually.

Summary

In this work, we sought to predict the degree of alignment between California state legislators and a select group of organizations. Using the Digital Democracy database, we developed a system that combines this copious legislative information for use with a learning algorithm. We combined text features from transcriptions with other discourse-related features into a sample set that we trained on a Logistic Regression classifier, achieving 78% accuracy when aggregating data by discussion.