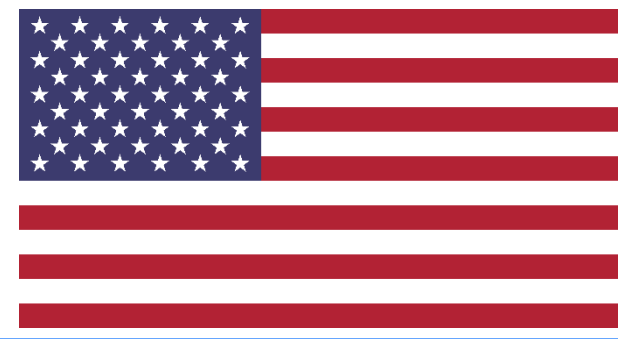
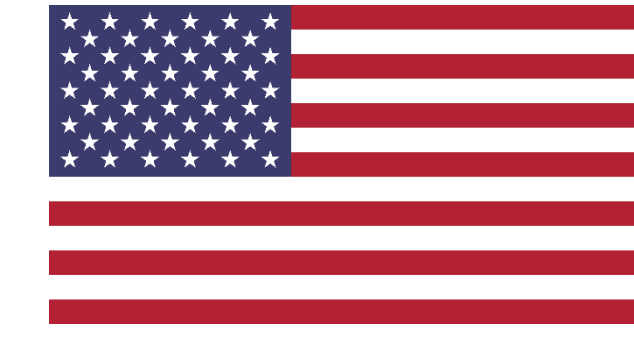


Predicting Bill Votes in the House of Representatives



Tom Henighan and Scott Kravitz

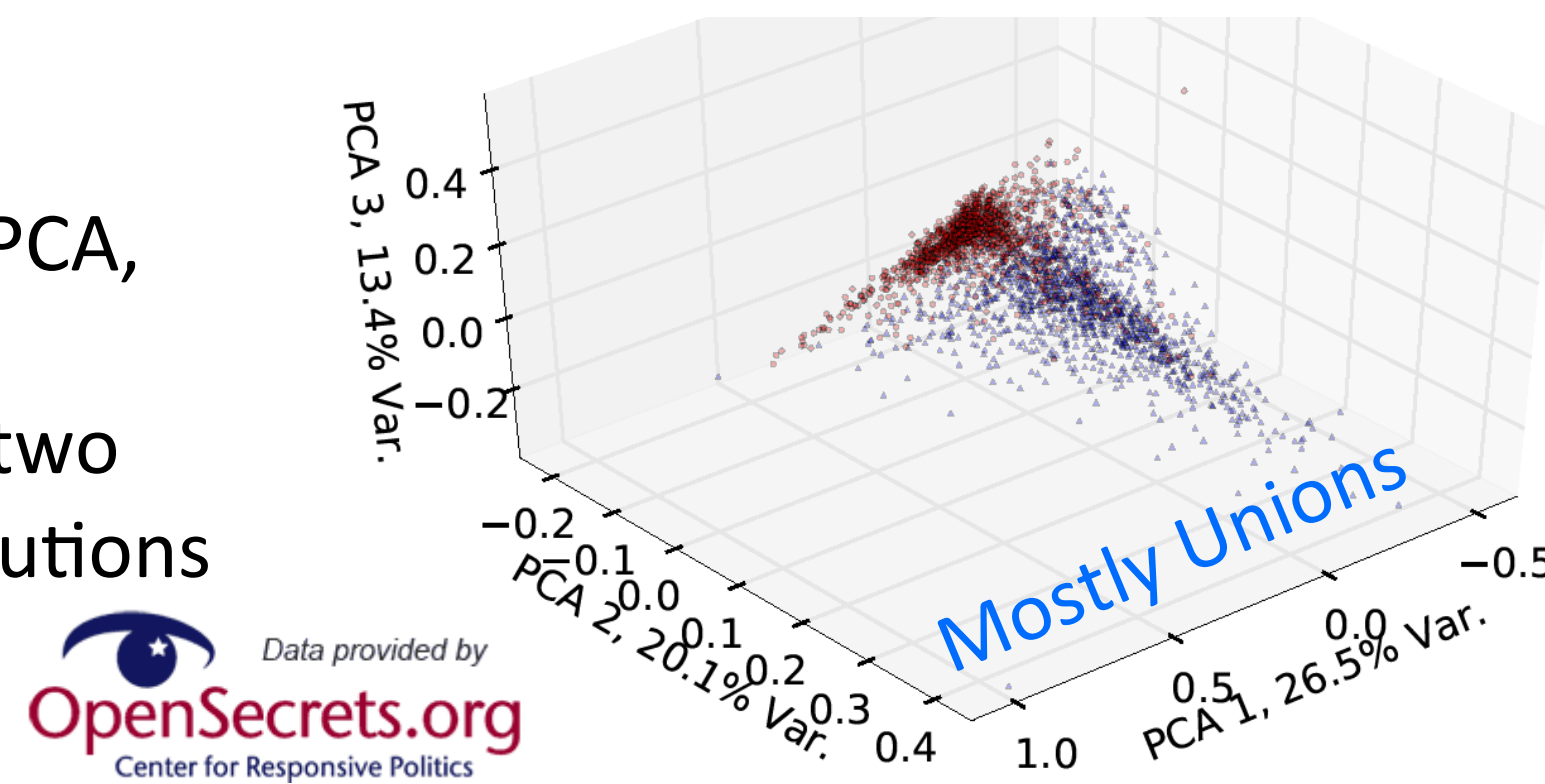


Intro

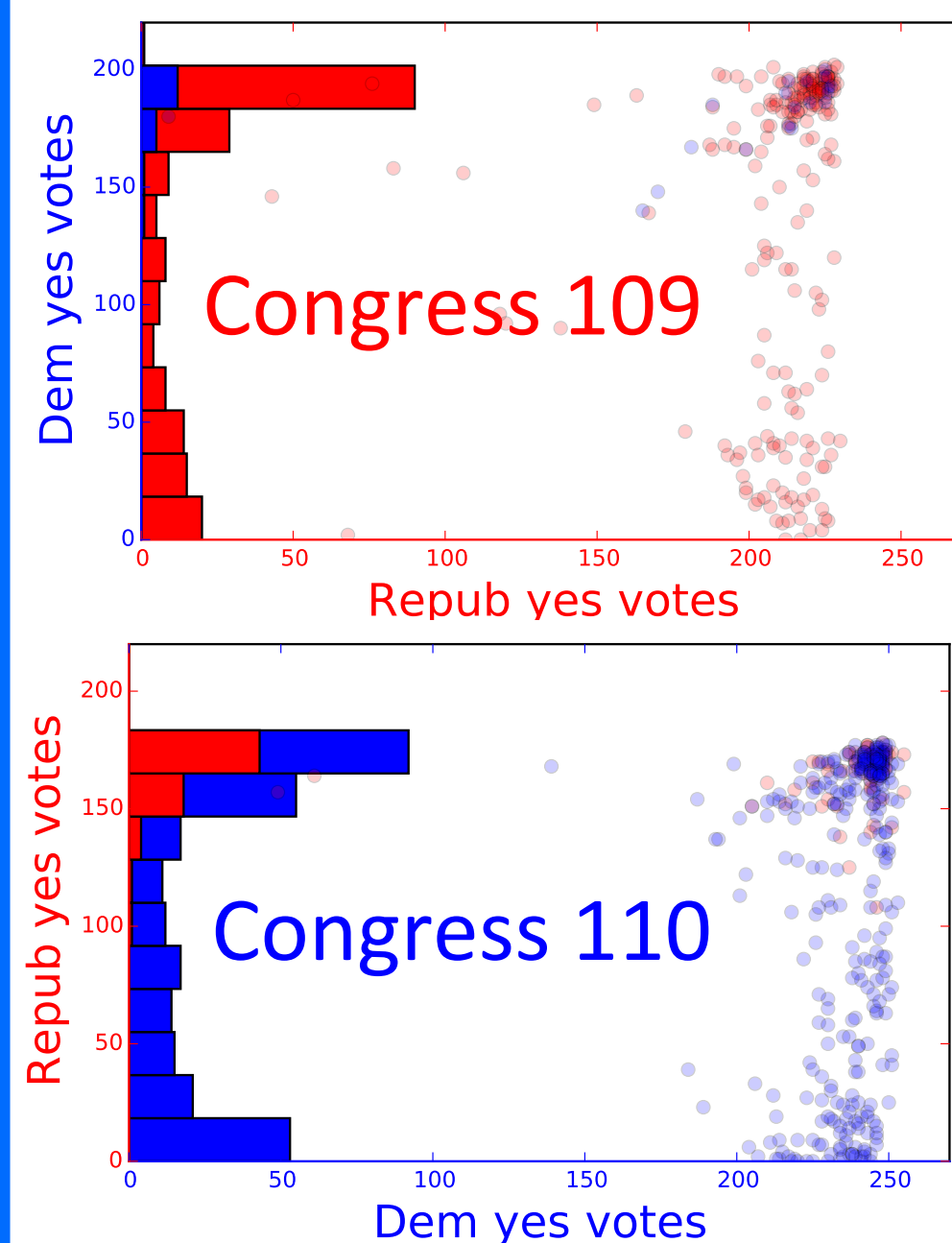
Goal: A Representative voting model generalizable to future Congresses
 We considered Congresses 108-113 (2003-2015)
Representative Features: Political party
 Campaign contributions by sector (13 sectors: Agriculture, Health, etc.)
Bill Features: “Tags”, keywords associated with each bill (~3000 unique)
 Bill sponsor (party and campaign contributions) and cosponsors (party only)

Campaign contributions

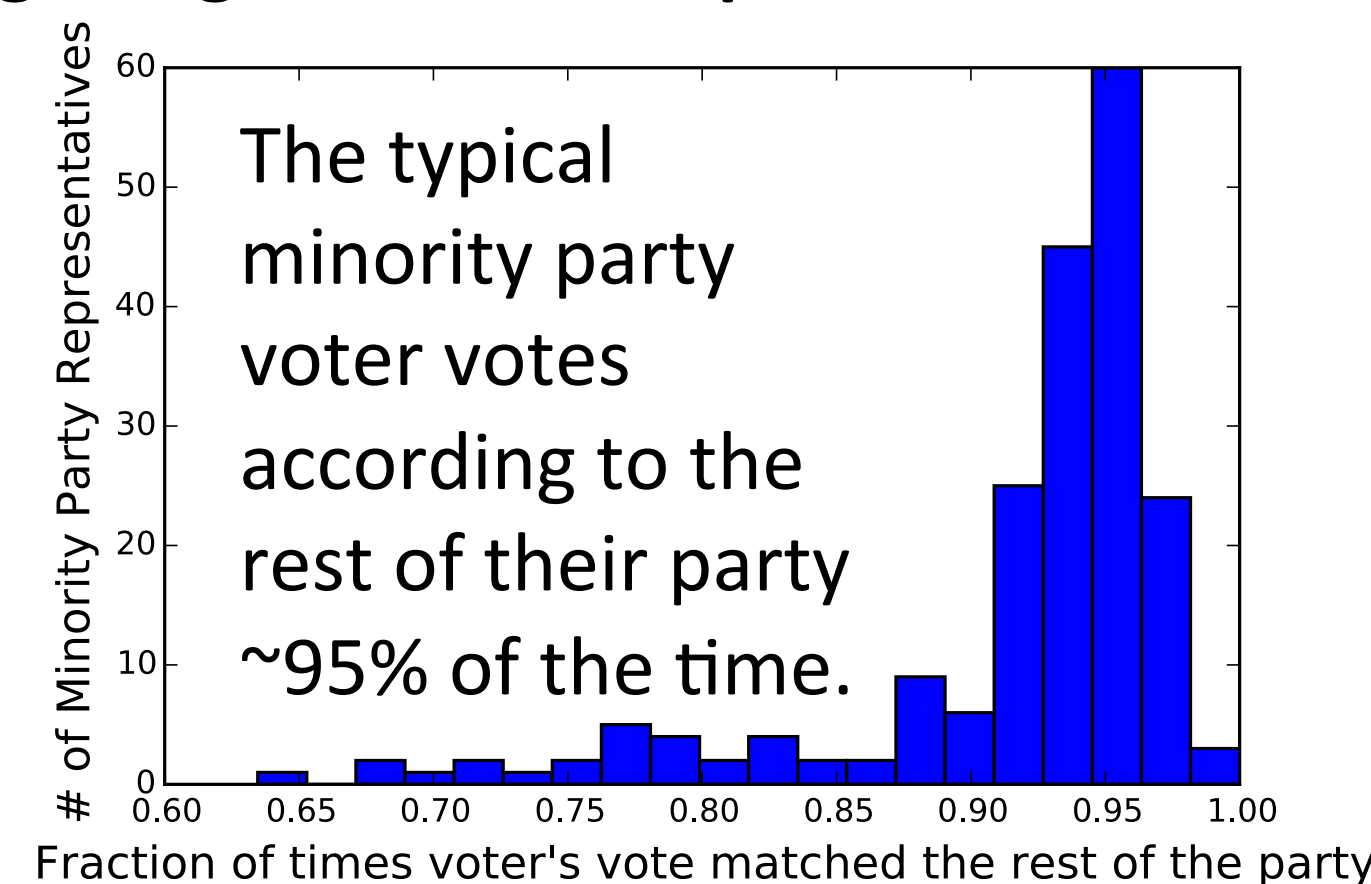
Reduced to 3 dimensions with PCA, capturing 65% of variance.
 Axis which most separates the two parties is dominated by contributions from labor unions.



Voting Behavior



Each point (left) is a bill, colored according to sponsor party (typically the majority party). Voters in the majority party vote “yes” 95% of the time, while minority party voting is bimodal. **Identifying minority party collective decision is enough to get >95% vote prediction accuracy.**



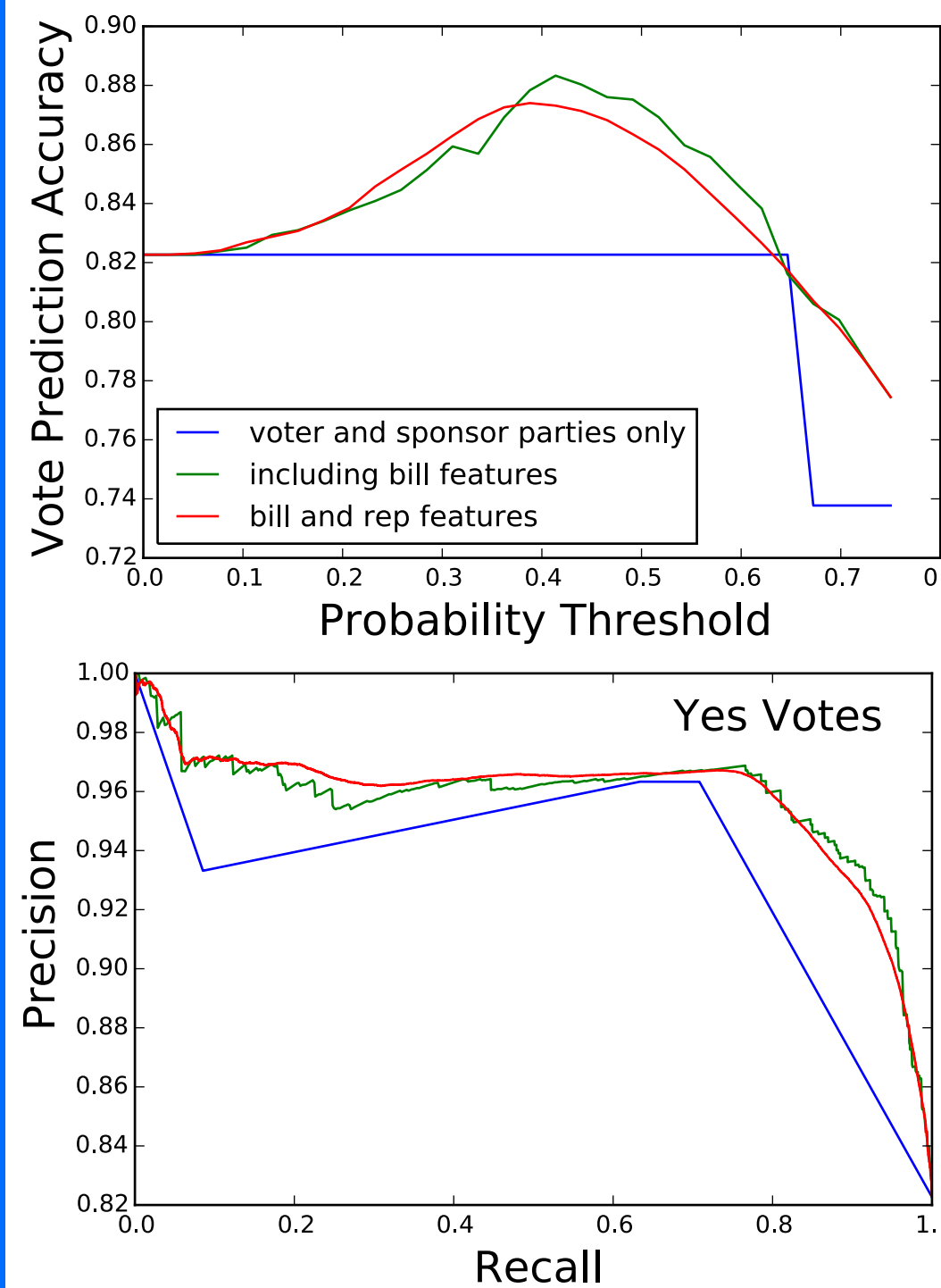
Tags

Found fraction of “yes” votes on past bills for each bill “tag” – which is included as a bill feature

Most Controversial Tags: Least Controversial Tags:

- Health (e.g. diseases, terminally ill)
- Military (e.g. the draft)
- Sports
- Nature (e.g. flowers, aquariums)
- Radioactivity
- Social Studies

Results



Performed logistic regression with different features
 Used congresses 108-112 as training, 113 as test
 Training & test error nearly equal: no overfitting
 Campaign contribution data adds little value

Similar performance from training and testing only on data within a single congress: **the results appear to be generalizable to future congresses.**

Future Work

- Improve identification of controversial bills (currently 75% accuracy)
 - See what information political scientists use to identify controversial bills
 - Use Naïve Bayes on tags
- Identify “maverick” voters who often don’t vote with the rest of the party.

