

ABSTRACT

Partisan redrawing of voting maps in the United States is a central strategy for collecting and maintaining political power. Advances in algorithmic tools for this purpose foreshadow a deepening and hardening of these ideological bastions. Preparing responses to this potential future requires an understanding of the relationship between voter and representative ideology. Linking precinct demographics to candidate ideologies through vote share provides a proxy for the ideological spectrum at the precinct level. Feeding this information through predictive models predicated on potential redistricted maps expresses the expected ideological score of a potential district, and therefore the anticipated ideological position of the "most appropriate" representative. Analyzing the delta between these predictions and future outcomes provides the public with a window into the extent to which their beliefs and interests are being accurately represented by their Congressperson.

DATASETS

Our work leverages data from the following four datasets. Our models will be tested on Michigan's voting maps and political information. It will be trained on the political information from Pennsylvania, Wisconsin, and Minnesota which share many demographic, regional, and political features with Michigan.

Candidate Ideology Scoring:

Stanford's Database on Ideology, Money, and Elections (DIME) includes Candidate ideology scores. These scores are assigned based on the ideological positions of the Candidate's donors, as well as those causes/candidates to which the Candidate donates. Scores are provided for Candidates who reach the General Election.

Precinct-Level Voting Results:

Harvard's dataverse provides access to the precinct-level vote share for all candidates in the General Elections for Congressional seats.

Voting Maps:

Michigan's public redistricting process provides access to a number of proposed maps. We use the current map as a control, and expand to the proposed maps providing varied precinct-to-district configurations for analysis.

Demographic Data:

We use the American Community Survey's 5-year estimates on demographic data. As redistricting is often focused on voters' demographic qualities, the

intersection of demography and ideology is a window to the strategic decisions being made by candidates and party bodies as they construct political geography.

MODEL DETAILS

At first stage we will be constructing two simple models: a decision tree and a logistic regression.

The decision tree is intended to simulate a simplified strategic decision-making process, wherein the tree acts as a redistricting committee splitting on demographic features. It will label potential districts with ideological leans based on predictions made by categorizing their component precincts.

The logistic regression is intended to provide a deeper understanding of the underlying probabilities that certain features result in particular ideological skews. At the first stage, it serves as a backup to ensure that our probability measurements are working.

THEORETICAL RESULTS AND ASSUMPTIONS

For this initial round we focused only on the results from Michigan. In future iterations, we will use similar states as training data in order to fully predict the state of Michigan through multiple potential redistricting scenarios.

From the logistic regression we receive an 81.81% success rate on predicting an election's ideological outcomes. These predictions represent potential ideological outcomes of voting districts based on their underlying precinct makeups in a variety of possible redistricting situations.

From the decision tree we found the same success rate and output the classifications of the districts split out for testing.

LIMITATIONS

We expect our analysis to be limited by the following factors:

The DIME dataset is missing a small number of candidates

The DIME dataset is back-out of ideology, relying on 3rd party contributors as measurement. It has the advantage of representing more candidates who received votes, thereby increasing our ability to understand the granular aspects of precinct voting preferences. However, it may be a less precise metric than scoring systems that narrow their scope to election winners and label politicians based on their voting records.

The census datasets do not label their geographic units in the same way as our other datasets. For this round we defaulted to county level measurement. We are continuing to develop our matching strategy at the precinct level.

NEGATIVE SOCIETAL IMPACTS

There is the potential that this work is used by political operatives to further understand and abuse differences in ideology and demography. Rather than a tool for changing candidate behavior and encouraging a closer match of representative to represented, it may be leveraged to instead alter the playing field in favor of particular candidates and their ideological beliefs.

ATTRIBUTIONS

Bonica, Adam. 2019. Database on Ideology, Money in Politics, and Elections: Public version 3.0 [Computer file]. Stanford, CA: Stanford University Libraries. <<http://data.stanford.edu/dime>>.

Bonica, Adam. 2018. "Inferring Roll-Call Scores from Campaign Contributions Using Supervised Machine Learning" *American Journal of Political Science*, 62, (4): 830-848. (<https://onlinelibrary.wiley.com/doi/full/10.1111/ajps.12376>).

U.S. Census Bureau. (2012). 2020 American Community Survey. Retrieved from <https://www.census.gov/data/datasets.html>.

Voting and Election Science Team, 2019, "2018 Precinct-Level Election Results", <https://doi.org/10.7910/DVN/UBKYRU>, Harvard Dataverse, V59.