Predicting Opinion Change in Deliberative Groups (Natural Language Processing) *

Ross Dahlke (rdahlke@stanford.edu)

1. Introduction

Deliberative Democracy is when a representative sample of people come together for a short period of time to discuss civic issues. The idea is that through a deliberative process, where people are access to objective information and good conditions to discuss, people can come to conclusions on civic issues that are most representative of what people actually want [cite Fishkin book]. In a time where so much discussion about political and civic issues happens in online information bubbles, Deliberative Democracy offers a chance for people to discuss issues on their merit and the freedom to change their opinions in order to reveal a better representation of public opinion.

Deliberative Democracy Polling is a technique pioneered by Stanford University Professor Jim Fishkin and the Center for Deliberative Democracy (CDD). A Deliberative Poll is where individuals are polled on their opinions before and after a weekend of small-group deliberation. This method allows researchers to measure the change in people's opinions as a result of the deliberative sessions.

This project uses the results of surveys and transcripts from the CDD's Deliberative Polls to predict group opinion change based on the transcripts from the groups' deliberative sessions. In talking with members of the CDD, they theorized that computational approaches to predicting opinion change might be difficult because of the narrow focus of each of the deliberative sessions. For example, one whole session might be devoted to discussing water policy. And so, water is the main topic of discussion the entire time. Folks from the CDD have also concluded that participants often don't explicitly say that they have changed their opinion on the topic of the deliberative session, even if they reveal a change in opinion in the survey. Given this prior information, I will work on building a model to predict opinion change in two steps:

1. (For this milestone) Experimenting with predicting opinion change based on the words and text of the deliberative sessions.
2. (For the final submission) Experimenting with predicting opinion change by incorporating linguistic features

Based on previous discussions with members of the CDC, I will first try to find a baseline model based on the text found in the transcripts. For the final project I will try to incorporate linguistic features into the models. Incorporating linguistic features will be a key component of this project because I hypothesize that it is not *what* people talk about in the deliberative sessions, it is *how* they talk about the issues that indicates opinion change. For example, how subjective or objective are the statements people are making in the deliberative setting, or how are sentences structured in groups that show opinion change versus don't show opinion change?

---

*Code and data available at: github.com/rossdahlke/cs_230_project

2. Related work

This project derives its inspiration from two main previous papers: "Attitude Change on Reddit's Change My View" (Priniski and Horne 2018) and "A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media" (Zhang and Pan 2019). The former paper found that the "the amount of evidence provided in a discussion predicts attitude change" on the Change My View Reddit forum. The latter paper was able to identify and collective action events in China based on image and text data and provides inspiration on using text data to identify collective behavior that has real-world application and can be a component of future computational social science research.

3. Dataset and Features

This project thanks the CDD for sharing their data. For this project, I have access to surveys and transcripts from three Deliberative Polls, each consisting of about 14 groups, with each group discussing in two or three different sessions. The surveys allow me to calculate opinion change on the issues discussed in each session before and after the deliberations. I use these pre- and post-surveys (responses on a 1-10 scale) to calculate an average delta in opinion change for each deliberative group for each topic session which I use as my dependent variable. The transcripts are broken down by group by session, but they are not attributed to specific individuals. While the additional granularity of individually-attributed transcripts would have great and is something to consider devoting resources to in the future, these group-level opinion deltas and transcripts provide enough data to complete the task. In total, there are 101 examples. examples.

4. Methods

To capture changes in the differences in the words that discussants use through their deliberations, I break each transcript down into 1/10th chunks. I then calculated the similarity of each 1/10th chunk to each other by comparing the word embeddings of each chunk (Honnibal and Montani 2017). For example, of a given transcript, there might be a similarity of .9 between the first and second 1/10th chunks and a similarity of .7 between the third and fourth 1/10th chunks. I use these similarities from each transcript to build two simple OLS models (Pedregosa et al. 2011). The first baseline model just uses the similarity of the second and tenth chunks (the first chunk is generally devoted to introductions, welcomes, etc.) to predict opinion deltas. The second model uses all of the similarities to try predict opinion deltas. I also train a basic neural network in Pytorch (Paszke et al. 2019) to try to predict the opinion deltas from all of the similarities.

In addition to this approach of using word embeddings, I also build a baseline model using Long Short-Term Memory in Pytorch on the sequences of words in the transcripts.

All code examples that I used to build these models can be found commented in my code.

5. Experiments/ Results/ Discussion

The main loss function that I use is Mean Squared Error since I am trying to predict a specific value (as opposed to a discrete prediction of whether a group of people changed their opinions at all). Below is a table of the MSE of the three methods mentioned in section 4.

method	MSE
One word embedding	0.149
All word embeddings	0.289
Basic NN	0.064
LSTM	0.130

Using the one similarity of the word embeddings has a lower loss than using all of the similarities of the word embeddings. But both of these perform worse than a basic neural network. The neural network that I used was very simple with only two linear activation functions and one ReLU that took in all 45 of the similarities from the combination of word embeddings.

Despite the varying levels of loss, all of the models seem to be performing poorly upon closer inspection. If you actually examine the predicted values, you see that they are all basically the same. Particularly the basic NN, which has the lowest loss, and the LSTM are simply finding the best average value and applying that "prediction" to all test examples. To try to circumvent this problem, I am going to try to incorporate data from the Priniski paper.

6. Next Steps

These basic models provide a good baseline estimate of how well one can predict opinion change based on the words that people use in deliberations. My next steps are to incorporate linguistic features. I plan to incorporate the findings of the Priniski and Horne paper that the amount of evidence provided in discussions predicts opinion change. I also examine the preditive ability of polarity and subjectivity of statements in the transcripts. One limitation of this project is the relatively small number of training examples available. And so, I will also leverage transformers to see if they provide the ability to better predict opinion change.

References

Honnibal, Matthew, and Ines Montani. 2017. "spaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing."

Paszke, Adam, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, et al. 2019. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." In *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. Alche-Buc, E. Fox, and R. Garnett, 8024–35. Curran Associates, Inc. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30.

Priniski, John Hunter, and Zach Horne. 2018. "Attitude Change on Reddit's Change My View." *Conference: Proceedings of the 40th Annual Meeting of the Cognitiee Science Society, Madison, Wisconsin*. https://www.researchgate.net/publication/333175400_Attitude_Change_on_Reddit's_Change_My_View.

Zhang, Han, and Jennifer Pan. 2019. "CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media." *Sociological Methodology* 49 (1): 1–57.