



Predicting U.S. Political Party Affiliation on Twitter

Catherine Lee, Jacob Shiff, Sridatta Thatipamala
{catfranl, jshiff, sthatip}@stanford.edu



Introduction

Political discourse on Twitter in the US is highly partisan, with little cross-party dialogue.

In this project, we develop a classification algorithm to identify a user's political affiliation (not explicitly captured by Twitter) based on a single tweet. This classification can ultimately be used to facilitate cross-party dialogue.

Data

We used the Congressional Tweets Database, a database of Twitter tweet IDs of the 114th U.S. Congress members from 2014 through 2017.

Raw data was enriched with tweet text and preprocessed, including tokenization, padding, and placeholder substitution (URLs). A custom word embedding was also trained.



Democrat

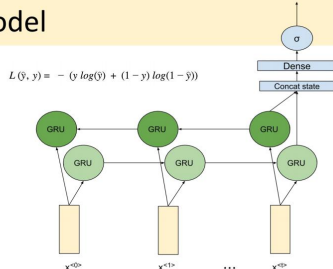


Republican

Figure 1: Sample entries

n = 86,472 tweets | Train/dev/test split of 70/15/15

Model



We projected the text using GloVe 300d embeddings and ran it through a bidirectional GRU model with a cross entropy loss. To classify, we passed the GRU hidden state to a fully-connected layer, followed by a sigmoid activation.

Results

Model	Train Acc	Test Acc
1. GRU	91%	86.4%
2. GRU (trainable embeddings)	94.4%	87.9%
3. Bidirectional GRU	95.9%	88.7%
4. Bidirectional GRU (trainable embeddings)	98.3%	91.6%

Model 4	Predicted Dem	Predicted Rep
Actual Dem	5632 (43.3%)	585 (4.5%)
Actual Rep	512 (3.9%)	6267 (48.2%)

Accuracy = 91.6%; Precision = 91.7%; Recall = 90.6%;

Discussion

Our model is able to classify party affiliation with high accuracy. We implemented several regularization techniques (dropout, L2) to reduce variance, but variance could be potentially be reduced further with more training data.

Our model struggled with classifying:

- *Mentions of specific people*: "i call on speaker boehner to convene the house to debate and vote on syria by wednesday of next week" (predicted: R; actual: D)
- *'This' contextual references*: "this SCOTUS ruling creates even more uncertainty for americans trying to comply with the ACA." (predicted: R; actual: D)

Future work

To evaluate and tune our model on non-politician users, we could survey Twitter users to explicitly collect party affiliation.

A related, but distinct problem, would be to evaluate political bias amongst various parties. Building such a model would require a database of tweets labeled by degree of political bias.

References

- [1] Congressional Tweets Database: https://github.com/kmjohanson/twitter-framing/tree/master/congressional_tweets_dataset_2017
- [2] Tweet Downloader: <https://github.com/DocNow/twarc>

Acknowledgements

We would like to thank the teaching staff, especially Abhijeet Shenoj, for their guidance with our project.