

Politweet

Aaron Chan, Jonathan Dutson, and Austin Kolander

Level of Interest: 10/10

Project Description

We would like to investigate the feasibility and extent to which we can label Tweets by their political affiliation. We are interested in being able to use tweets from politicians to predict the extent to which they lean toward conservative and liberal ideologies. We will use a deep neural network to learn which features are important in determining political affiliation. In preprocessing features, we consider it likely that we will transform text data into sentiment vectors using a Doc2Vec model.

Some politicians use Twitter to communicate with their constituents. It is beneficial to those who elect politicians to know where their candidates stand ideologically. If our neural network is capable of correctly classifying politicians, perhaps it will also be able to classify others, such as students, celebrities, and executives.

Data Set Features

The Twitter API provides an abundance of tweet data organized by user. We will select features which we consider potentially important from this dataset, which could include the following:

Tweeter	Text Sentiment Vector	Hash tags	Age	Gender	Target: Political ideology rating (1 is conservative, 0 is liberal)
Sen. Elizabeth Warren	vector representation of tweet text	#EndCorruptionNow	69	F	0.23
Sen. Mike Lee	vector representation of tweet text	#utpol	47	M	0.88

Gathering and Labeling Data

We will gather the data by obtaining a Twitter API key (documentation: <https://developer.twitter.com/en/docs.html>) and requesting tweet data from politicians of the United States Senate and House of Representatives. We will use the ideology report cards provided by govtrack.us (<https://goo.gl/fYnzdq>) to assign each senator and representative a political affiliation rating on the liberal/conservative scale. These ratings will be our labels for training.

We will use Twitter's Python toolkit to retrieve the data, and our own script to extract features, as well as gensim for Doc2Vec conversion of text bodies to sentiment vectors. We will use either the Keras or Theano implementation of a deep neural network or alternatively our own homegrown deep learning model for training and prediction.