Project Proposal

Predicting collective sentiments towards a candidate after a political event

Audrey Glaser, Grace Wright, Minyoung Do, Soowon Jo, Steven Cao

January 24, 2020

Introduction

Today, many Americans take to Twitter to share their thoughts and feelings on the political process. As a result, Twitter data is a promising source of social sentiments towards presidential candidates, especially as those sentiments evolve in reaction to major events in the primary. Our research project proposes to use Twitter data from the days following a major event in the 2020 Democratic presidential primary — likely the results of the Iowa caucus — to train a model which can forecast aggregate sentiment toward the candidates. Our proposed methodology has two main elements. The first element is a classification algorithm, which we will use to label individual tweets containing mention of specific candidates as containing positive or negative sentiment toward that candidate. The second element is a regression model, which we will use to forecast an index of net sentiment toward each candidate over a certain time period.

Though sentiment analysis of social media discussions about candidates have not yet been shown to be a definitive predictor of candidate support or election results, our model may still provide useful information about the shifts in social sentiment occurring around the candidates that ultimately help shape their performance at the polls.

Literature Review

Early work in data science tended to take a critical approach to monitor people's mood signals from social media such as Twitter. Twitter proved to be the most popular microblog platforms on which individual user can publish their reviews, comments, emotions, and opinions, particularly in relation to important political topics (Shamma et al., 2009). To analyze political sentiment using Twitter data, an extensive literature has implemented a sentiment analysis and classifier such as naïve Bayes and Support Vector machines (Raghuwanshi et al., 2017).

In response to such trends in machine learning applications, existing literature has built a model that monitors political sentiment among Twitter users and predicts their voting intentions during the Irish General Election (Bermingham et al., 2011). The primary methodological approach for this study was to combine sentiment analysis using supervised learning and volume-based measures. The novel predictive measures introduced in the study was to avoid any disproportionate influence of a few notable events unrelated to politics on the popularity of candidates and parties. This thus enabled them to measure the proportional share of party mentions in a set of tweets for a given time period. In terms of sentiment analysis, they used classifier trained on annotated data in which they consisted of multiple sentiment classes (positive, negative, mixed, neutral)

(Wilson et al., 2005). Moreover, in order to add tone to texts in tweets that include emoticons and unconventional punctuation, they used tokenizer (Laboreiro et al., 2010) to tag sociolinguistic features to contain all possible sentiment information. As a result, they found that volume played the single most significant indicator for predicting the election outcome and identified a dramatic shift in sentiment towards the parties in the two days before polling day which possibly hints the election outcome.

Another study that focused on sentiment analysis on tweets used a linear regression model to predict the probability of security attack based on the overall sentiment of Twitter users (Hernández et al., 2016). The method for this study employs natural language processing analysis on a collected corpus and analyzes the collective negative sentiment within a context where users are objecting to a particular political, religious or social context. After extracting sentiments using SentiWordnet compendium and the findOrientation algorithm that only identifies negative orientation in Tweets (Feng et al., 2009), they implemented Lasso regression analysis to predict future attacks. The result indicates a high correlation between the sentiment of Twitter users and actual security attacks on the web, meaning that if there is a significant increase in the percentage of negative comments toward a particular event, the probability of cyber attack will be more likely to rise. Through this research project, we hope to monitor collective mood tendencies on political events based on smaller individual samples observed in twitter data.

Data & Method

Before building algorithms, we plan to conduct an exploratory data analysis with `tidytext` in order to estimate the initial model. In this section, we introduce our data collection and preparation process and cast an overview of the algorithms and regression models in the analysis.

Data collection

The primary purpose of this research is to develop a learning algorithm under a time series analysis framework that predicts changes in collective sentiments towards political candidates after a political event by examining the wave of attention observed in tweets. The text data will be extracted directly from Twitter by requesting multiple queries with `rtweet` and `twitteR`. Each query consists of a sequence of topic words: for example, q_1={candidates, debate, policy} and q_2 =: {#election2020, #voteA, #voteB}. The scope of data collection is limited to post-event tweets in a 3-day period for each topic/candidate, as it would be difficult to validate the connection between the event and users' sentiments towards the candidate over time.

Data Preparation and Pre-processing

Considering text data scraped from the web is often not ready for use immediately, we are to perform several different data cleaning operations to remove the noise in the data. First, we will build functions that 1) delete the re-tweets to avoid redundancy, 2) tokenize collected corpus into individual arrays of words, and 3) remove duplicaetd tweets. The sample codes are as follows:

```r
# list of packages for potential use
library(tidyverse) # the mighty tidyverse for data visualization and... everything
library(tidytext) # convertting texts tidy formats
library(topicmodels) # tidying functions for LDA objects
library(SnowballC) # stopwords & tokenization
library(reticulate) # just in case we need python modules for text analysis
library(tm) # for td-idf weighting process
library(rtweet) # twitter data collection
library(twitteR) # twitter data collection

# requesting queries
pol <- get_timelines(
  user = c("BernieSanders", "realDonaldTrump"),
  n = 1000
)

# sample codes for tidying text data
politicians <- pol %>%
  # removing retweets
  filter(is_retweet == FALSE) %>%
  select(text) %>%
  # tokenizing the tweets into individual words
  unnest_tokens(input = text, output = word) %>%
  select(word) %>%
  # filtering out unnecessary information, including numbers and special characters
  filter(!word %in% c("https", "t.co", "amp"),
         !word %in% tolower(pol$screen_name),
         !grepl("^\\d+$", word),
         # back-up function to make sure usernames starting with @ are also removed
         !str_detect(word, "@\\w*"),
         # also a back-up function to remove numbers and special characters
         !str_detect(word, "^[0-9]*$")) %>%
  # removing stop words that do not represent the context of the tweets
         anti_join(stop_words)
```

Algorithm

In the early process of classification, this algorithm detects **words and phrases**(input) based on the sentiment polarity; that is, it assigns positive or negative values to the corpus on the scale of -1 to 1 depending on **the sentiment of each tweet** (output). For the second stage of classificaiton, this project adopts the mood classifier established by Hernandez-Suarez et al. with some degree of modifications. This classifier aims to select terms that best represent user's mood and discard the ones that are not related to the context by weighting relevant features. For this project, instead of SentiWordNet in the original model, sentiment polarity scores built in `qdap` will serve as a standard filter for the primary text data used in this research. Tokenized terms will be categorized into each sentiment group after the tf-idf [1] weighting procedure from which we would be able to obtain a statistical measure that evaluates how important a word is to a document in a corpus.

This analysis also adopts the probabilistic classifier based on Bayes's Theorem from Hernandez-Suarez et al. to learn patterns from tweets [2]. The function for likelihood estiamtes for c and $f(\tau)$ is:

$$P_{NB}(c|t) = \frac{(P(c)) \sum_{i=1}^{m} p(f(\tau)|c)^{n_{i(t)}}}{P(t)} \tag{1}$$

Parameters in the classifier are as follows:
1. A tweet t
2. The class c (negative/positive)
3. The feature $f(\tau)$
4. Count of features $n_i(t)$ present in the tweet t
5. Number of features m

Regression

We will be utilizing different regression models to identify and analyze the causal relationships between a sentiment score of individual Twitter users and collective mood polarity score. This is to test if a certain collective sentiment is observed among users after political events. It is not yet clear which type of regression model would be appropriate for the purpose of this study at this time; we will be reviewing different types from a simple linear regression model to LASSO after the exploratory data analysis.

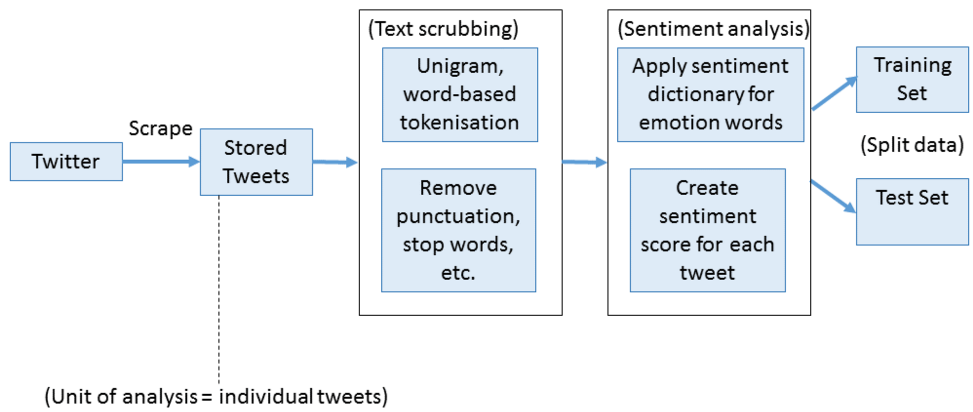
```
# classification algorithm: average of polarity scores of tweets
sentiment_polarity <- function(x){
  polarity(
    str_replace_all(
    unique(
      sapply(
        strip_retweets(searchTwitter(x, n=1500, lang = "en", resultType = "recent")),
        function(x) x$getText())), "[^[:alnum:]]", " "))$group[4]
  }

# regression function to identify and evaluate potential causal relationships (TBD)
```

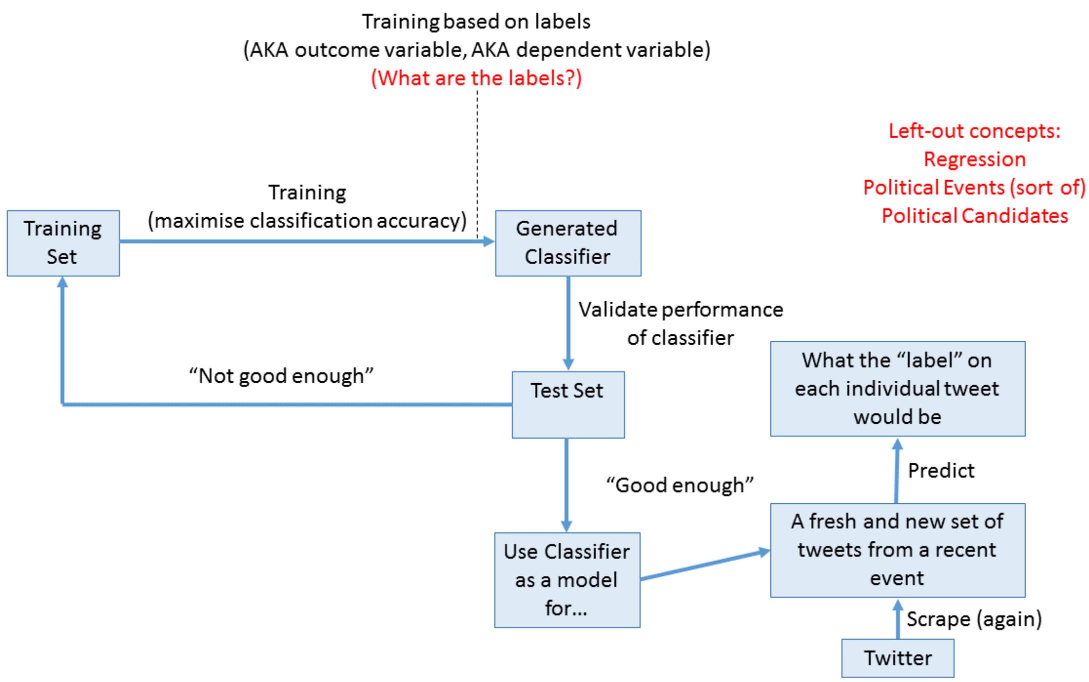
Current Caveats

There are a number of conceptual caveats and considerations that we would like to receive comments on. The flowcharts we have attached to this proposal are included not to definitively outline what our workflow is, but to express points of confusion which we are working to clarify. They should hopefully illustrate some of the conceptual roadblocks we have run into; the rest of

[1] Term Frequency Inverse Document Frequency
[2] See Hernandez-Suarez et al. text, *Predicting political mood tendencies based on Twitter data* for more details.

this section is an elaboration of those roadblocks.





The essence of the project is to be able to make fruitful characterizations of the political landscape, which we can then use to predict other interesting features and outcomes of that landscape. That being said, we are looking to establish some kind of relationship between important political events, and the effects they have on the ensuing political atmosphere. We explore one facet of this "ensuing political atmosphere" by gauging the populace's general sentiments on Twitter.

Traditionally, previous research has used sentiment scores of tweets as a predictor for other outcomes, such as the number of cyber attacks (Hernandez-Suarez et al, 2016). Our aims take this in the reverse direction: take some feature(s) of an important political event and use that to predict general sentiments on Twitter. While this would require a regression model, we are currently unsure how exactly to fit it in, partly for the reason that we have not currently decided on a clear definition of what these "features of an important political event" would be.

We are also interested in something more specific than just predicting "the general sentiment of Twitter following some event". If a disaster happens, then sentiments will be negative, but that is not necessarily very interesting (in terms of prediction). We want to predict the populace's sentiments *regarding some political candidate*, but we have not currently conceived of a clear way to approach that. (We do suspect that it will involve topic-modeling, in which each candidate is a "topic", in the sense that a particular tweet may be talking about a certain candidate.)

Last but not least, we are not immediately clear on what exactly our classifier would be classifying. Our observations (individual tweets) have one feature associated with it: sentiment score. But we aren't sure what kind of feature the labels would represent, hence we aren't sure how we would classify the observations, i.e. what the outcome feature would be. We were considering "polarity of a tweet" as a candidate, but it's not clear that "the polarity of a tweet" would say something particularly different from "the sentiment score of a tweet".

Implications

There are real world implications for building a model that can predict sentiment trends after a political event. For example, if campaigns were able to predict the general sentiment of the people after a significant political event, the candidate could potentially voice his or her stance on the topic well before her or his peers even recognized the significance of said event. More concretely, this model could provide support for a new methodology to understanding the sentiments of the people. Namely, a shift from traditional campaign methodology such as polling, to a new, yet quickly evolving, methodology utilizing social-media acquired data.

Since social media is the individual's personal platform, the data gathered on a social media site, such as Twitter, could more accurately represent the individual's views than the data gathered from survey polling. On the flip side, there is also an evident bias in the data acquired from social media. Not only does the individual need to have a Twitter, or said social media, the individual also has to be active enough and politically engaged enough to post about a political event. This narrows our sample significantly. At this stage of history, our pool would largely consist of younger individuals with strong political convictions, inherently more likely to be on the ends of the political spectrum. Nonetheless, if this model did accurately gauge the sentiment of the people, even if from a biased sample, it could have broad implications for the use and development of further machine-learning models to predict political sentiments from twitter-based data.

References

Bakliwal, A., Foster, J., Puil, J.V., O'brien, R., Tounsi, L., & Hughes, M. (2013). Sentiment Analysis of Political Tweets: Towards an Accurate Classifier.

Bermingham, A., & Smeaton, A.F. (2011). On Using Twitter to Monitor Political Sentiment and Predict Election Results.

Conover, M.D., Gonçalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011). Predicting the Political Alignment of Twitter Users. 2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing, 192-199.

Feng, S., Wang, D., Yu, G., Yang, C., & Yang, N. (2009). Sentiment Clustering: A Novel Method to Explore in the Blogosphere. APWeb/WAIM.

Hernandez-Suarez, A., Sanchez-Perez, G., Martinez-Hernandez, V., Meana, H.M., Toscano-Medina, K., Nakano-Miyatake, M., & Sanchez, V. (2017). Predicting political mood tendencies based on Twitter data. 2017 5th International Workshop on Biometrics and Forensics (IWBF), 1-6.

Hernandez-Suarez, A., Sanchez, V., Sanchez-Perez, G., Pérez, H., Olivares, J., Toscano-Medina, K., Nakano-Miyatake, M., & Martinez-Hernandez, V. (2016). Security attack prediction based on user sentiment analysis of Twitter data. 2016 IEEE International Conference on Industrial Technology (ICIT), 610-617.

Laboreiro, G., Sarmento, L., Teixeira, J., & Oliveira, E.C. (2010). Tokenizing micro-blogging messages using a text classification approach. AND '10.

Raghuwanshi, A.S., & Pawar, S. (2017). Polarity Classification of Twitter Data using Sentiment Analysis.

Shamma, D.A., Kennedy, L., & Churchill, E.F. (2009). Tweet the debates: understanding community annotation of uncollected sources. WSM '09.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. HLT/EMNLP.