Sam Lagoe and David Dominguez

Prof. Lirong Xia

Computational Social Choice

May 14, 2016

**Overview:**

With the growing presence of social media in today’s society, more and more people are using social media as their primary news source. We have also seen a rise in how involved and influential social media companies are in government and politics.

Recent research done at Dartmouth College has shown that, “modern technology has shortened attention spans and left little time for contemplative thought.”

This recent trend and researched prompted us to come up with the idea for this project. We developed an application to find a correlation, if one exists, between Twitter data and the results of the current presidential election.

**Scraping Twitter:**

Tweepy API

We used the Tweepy API to scrape Twitter for tweets relevant to the presidential candidates. We searched for tweets which contained the full name of one of the candidates. Depending on the privacy settings of each user, we were able to grab the tweet along with the ID number of the user, the ID number of the tweet, and the user’s location.

**Natural Language Processing:**

Sentiment Analysis API by Meaning Cloud

After grabbing the tweets we need to use a sentiment analysis API in order to determine the overall polarity of the tweet. This allows us to infer whether the tweet is in support of a certain candidate or not. While conducting the sentiment analysis, we were only concerned with three of the statistics calculated during the analysis.

* Score (P+, P, NEU, N, N+, NONE): Polarity of the entire text.
* Confidence(0-100): Confidence in the overall sentiment analysis performed.
* Agreement (agree or disagree): Marks the agreement between sentiments in the text.

By using the score and agreement statistics, we are able to infer the general opinion that the Twitter user has about the candidate they are talking about. Here are a couple examples of actual tweets that we performed sentiment analysis on.

Bernie Sanders tax plan might represent the “largest tax increase in peacetime history”

* Score: N+
* Agreement: Agreement
* Confidence: 95

I get email from “democratic” organizations. I always email back “I only donate directly to Bernie Sanders”

* Score: P
* Agreement: Agreement
* Confidence: 100

The first tweet received a score of N+, or strongly negative, and the statement is in agreement with itself. Just based off the analysis without reading the actual tweet, we could assume this Twitter user has a negative opinion of Bernie Sanders and will most likely not vote for him.

The second example has a positive score and is in agreement with itself. Therefore, we can infer this user supports Bernie Sanders and will vote for him.

**Conclusion:**

We have a couple possible sources of error in our simulated vote calculations. First, Some of the tweets we had grabbed were not written in english. We are only able to accurately perform sentiment analysis on English text. A second source of error stemmed from the sentiment analysis API we used. We observed a small amount of incorrect analyses while manually testing the API ourselves. We also may have some error due to Twitter’s privacy settings. In some cases we were not able to grab the user’s ID number. This may have led us to interpret multiple tweets as separate votes while they may have been posted by a single user.

**Future:**

In the future we would like to develop a much larger data set. This would allow our conclusions to be much stronger. We also would like be refine our Twitter scraping methods to allow us to pull more relevant tweets in a more efficient manner. Another feature we would develop in the future is the automation of our Twitter scraping and sentiment analysis. This would aid our goal of developing a larger data set.

Sam Lagoe, David Dominguez

Prof. Lirong Xia

Computational Social Choice

April 07, 2016

**Proposal Topic**

Extraction and Collection of Pertinent Election Twitter Activity Used to Predict and Analyze Future Races and Ongoing Trends Viewed Under A Scope Of Computational Social Choice

**Abstract**

Social media, specifically Twitter, has played a significant impact in the 2016 Presidential Elections similarly to previous innovations in media decades earlier with newspaper, radio and television. Twitter provides an convenient 140-character channel of communication for candidates and their teams to reach a growing base of followers (Donald Trump: 7.49 million,1 Hillary Clinton: 5.89 million,2 Bernie Sanders: 1.82 million3) almost instantaneously with their responses and opinions regarding trending topics/issues throughout the process. In sponsorship with cable news networks, large tech giants, Google, Twitter and Facebook, have recently started providing data virtualization tools to better understand the general bodies responses to candidate decisions/actions during debates.4 This project aims to apply voting rules to data activity across twitter regarding the election. By collecting relevant tweets about candidates and the users who posted these tweets, one could generate a primitive set of preferences for each user and quantify users’ voting decisions effectively predicting the results of a race. Due to current position in the cycle we will be focusing mainly on primaries and caucuses already held to determine our trends.

**Schedule**

|  |  |  |
| --- | --- | --- |
| Sprint | Task(s) | Deadline(s) |
| 01 | Wireframes/MockUp/Set Up Git | 04/15 |
| 02 | Collect Data & Create Visuals for Previously Held Races (Primaries/Caucuses) | 04/22 |
| 03 | “Predict” Future Results based off of Trends from Collected Data | 04/29 |
| 04 | Alpha & Beta Testing | 05/06 |
| XX | Project Presentation | 05/10 |

Sprint 03: Potentially postponed if Sprint 02 deadline is not met.