

Project Writeup

Kevin Crispie

26 February 2015

Project Overview

For this project, I wanted to find out how people on twitter viewed certain political figures in America. I created a sentiment analyzer that I used to analyze tweets on Twitter regarding several prominent American political figures. I utilized web-scraping tools to take information off of a website in order to create the sentiment analyzer, and I employed an API in order to obtain tweets from Twitter. I used Python's string and list operations in order to process the tweets, and I used Matplotlib to portray the data I obtained from my sentiment analysis.

Implementation

The project consists of two files, each representing a discrete part of the project. In the first file, I used Pattern's Twitter API to obtain tweets given my search terms from Twitter, which it then wrote to a text file. In the second file, I process the tweets in the text file, eliminating the text labels I added as part of the file writing process and separating each set of tweets into a corresponding list. From the list, I could then run a sentiment analysis on each set of tweets.

The sentiment analysis was rather straightforward. I found a list of positive words and negative words in plaintext on a website. I used these words to form the basis of my sentiment analysis. I also incorporated some Twitter specific items, such as the hashtag symbol. Since many people on Twitter use the hashtag symbol for expressing excitement or anger, as well as many other

feelings associated with both positive and negative sentiment, I decided that the hashtag and other elements of a tweet, such as the use of an exclamation point or prolific use of capital letters, would amplify whatever score the sentiment analyzer would have returned. In order to keep the impact of these other tweet elements in the appropriate proportion to the use of positive or negative words, I scaled each of these impacts by a constant. I also tracked what I perceived to be the amount of emotion or passion within a tweet. From personal experience, many Twitter users may have both positive and negative sentiments within the same tweet. In my implementation, positive and negative word usage would cancel out. In order to get a numerical score for the amount of emotion in a tweet, I used only positive numbers in the analysis. I plotted the results for each term in a scatter plot with Matplotlib, scaling the numerical scores for sentiment and emotion to better fit a reasonable axis scale.

I chose to use pattern's Twitter API as opposed to another API because of its superior documentation and relative ease of use. I investigated multiple other APIs designed to pull tweets from Twitter, but none of them were as well documented as Pattern's and it was much easier to pull only the text from a given tweet with Pattern's API than it was for another organization's. I also chose to have a separate program for importing tweets into a text file because I wanted to avoid having to import tweets each time I was testing my main program. Twitter has very strict rules about how developers can use APIs that pull tweets from its site. I wanted to avoid hitting the limits Twitter imposes, so I got all the tweets I wanted to use into one text file, then pulled the tweets back from that text file. In this way, I was able to test my program very easily.

Results

From looking at the results graphs, it seems that many people on Twitter have vastly different opinions on political leaders in America. Many tweets have generally positive views, but from the perspective of my sentiment analysis, some of the most extreme or emotion tweets are very negative.

I was interested to see how people viewed the President, especially in light of his recent veto of the Keystone XL pipeline. This topic made up a large

portion of the tweets that involved the President.

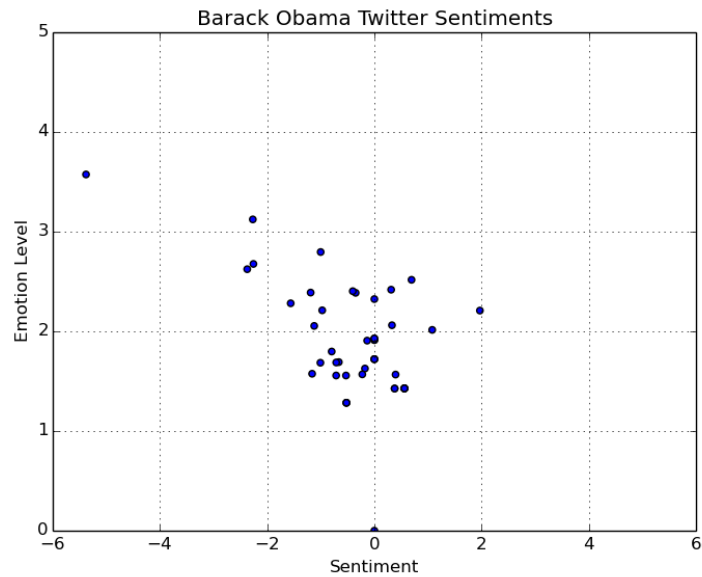


Figure 1: search term “Barack Obama” Sentiment Analysis

In particular, there is one tweet on the overwhelmingly negative side. Considering the location of the other values, this could be considered an outlier, but it is clear there are some tweets out there that express extreme negative feeling when also discussing President Obama. Overall, there is a substantial spread in sentiment for tweets mentioning Barack Obama

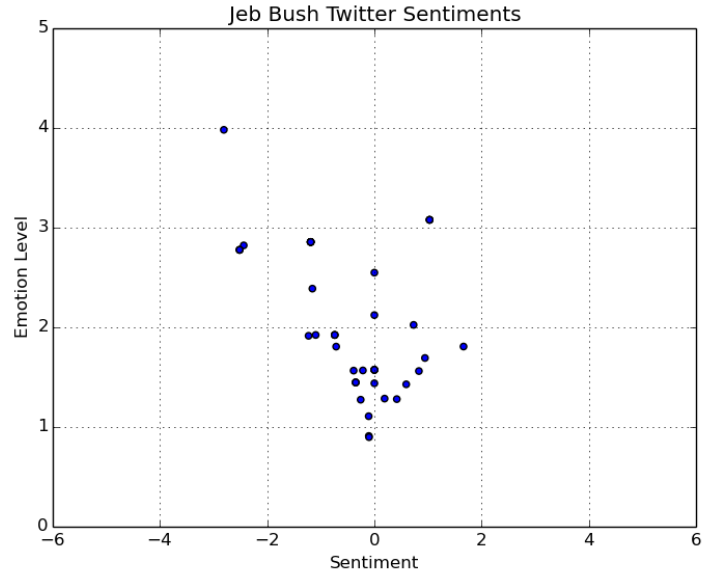


Figure 2: search term “Jeb Bush” Sentiment Analysis

There is also a substantial spread in sentiment for tweets mentioning Jeb Bush, a possible Republican Presidential candidate and brother of former president George W. Bush. The spread isn’t as large as Obama’s, but there is still a lot of division on Twitter. It’s possible that Bush hasn’t been in the limelight long enough. President Obama is the most high profile of any of the people I analyzed, so people may have stronger opinions on him because they have had more time to form strong opinions on him.

Compared to the other political figures in my analysis, the Speaker of the House of Representatives, John Boehner, is viewed in a more positive light, at least within the tweets that I analyzed.

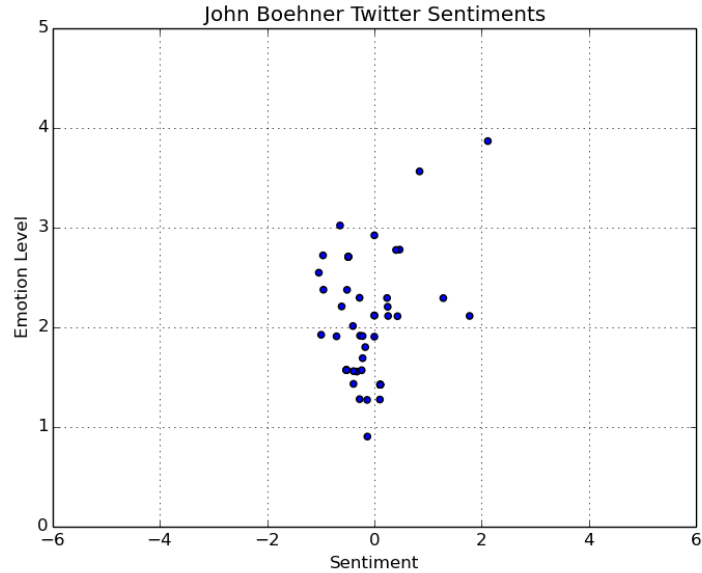


Figure 3: search term “John Boehner” Sentiment Analysis

Another item of note is that his range of sentiments, especially on the negative sentiment side, is quite narrow. In the tweets analyzed, there is a much more consistent sentiment value than for other political figures.

For Hillary Clinton, who is leading in the polls for the Democratic Party nomination for President, there is also a wide spread. There are tweets mentioning Clinton that are generally positive, but even more that are negative, and some that are very negative. Hillary Clinton has been a polarizing figure for a while, though the negative rating may also be due in part to the tweets mentioning issues that have a negative connotation, as opposed to expressing a negative opinion about Clinton.

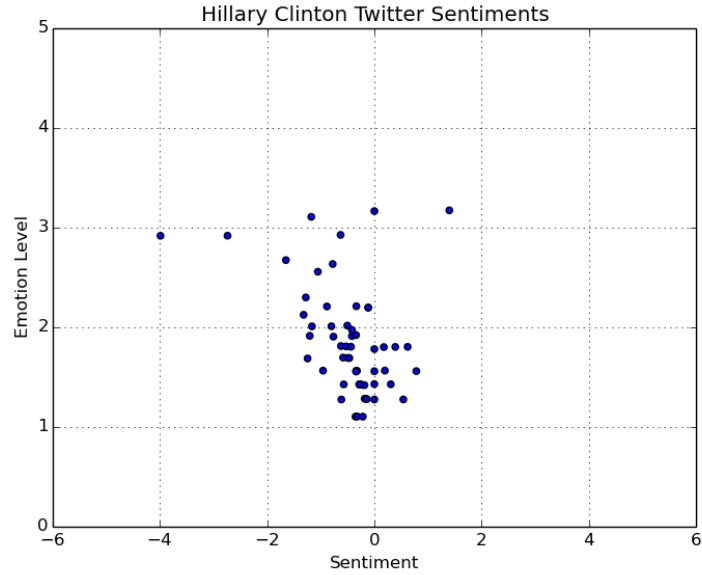


Figure 4: search term “Hillary Clinton” Sentiment Analysis

Combing over some of the tweets myself, I saw that many of them dealt with issues of gender and economic inequality, issues that may have more negative wording assoicated with them. Further analysis would be needed to determine the nature of many of these tweets.

Vice President Joe Biden, another potential candidate for the Democratic Party nomination for President, has a very erratic sentiment spread. There is a central cluster of positive and negative tweets, one very positive tweet, and several extremely negative tweets.

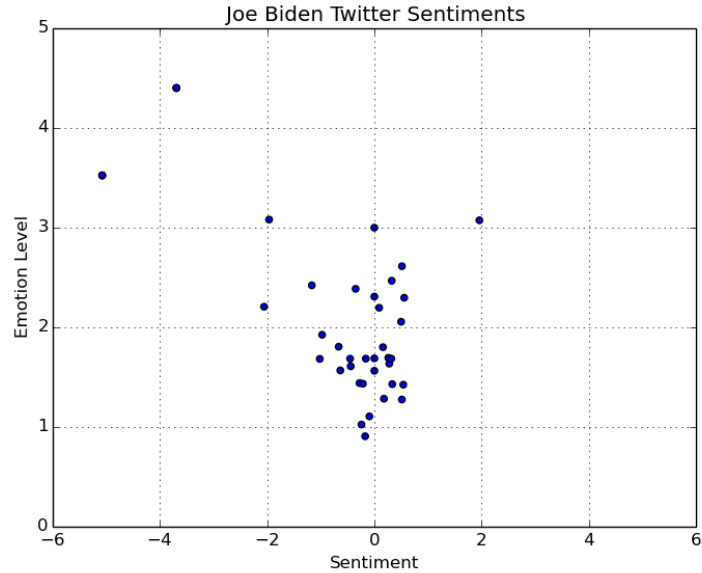


Figure 5: search term “Joe Biden” Sentiment Analysis

Joe Biden has also been a polarizing figure, even among Democrats. He has a tendency to get into trouble with things he says or does, and that may be reflected in the sentiment analysis for tweets mentioning him.

Governor Scott Walker of Wisconsin also has a reasonably wide spread, though not nearly as wide as someone like Joe Biden. Many of his tweets are clustered around the middle, with an ever so slight skew towards the negative side.

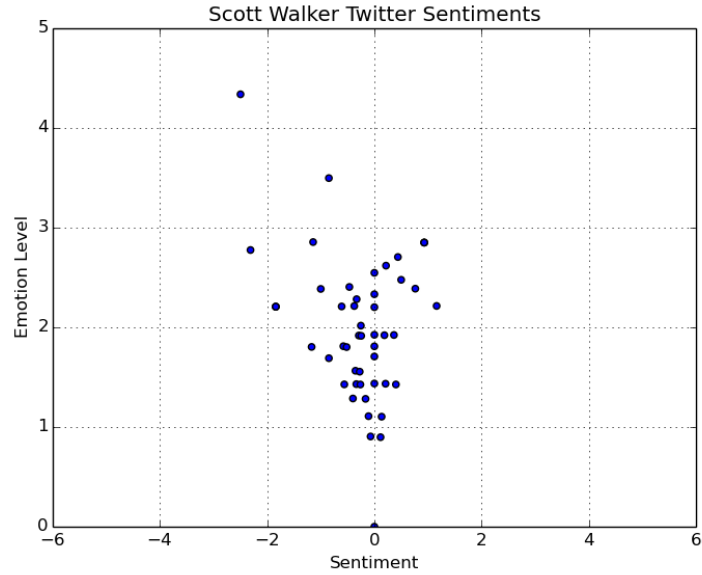


Figure 6: search term “Scott Walker” Sentiment Analysis

The lack of a wide spread of large split in sentiments could be attributed to Walker’s lack of name recognition and his nascent Presidential campaign. It appears as if he will run for the Republican nomination, but he doesn’t have the national name recognition of other potential candidates. Perhaps his spread would change given time and more media exposure.

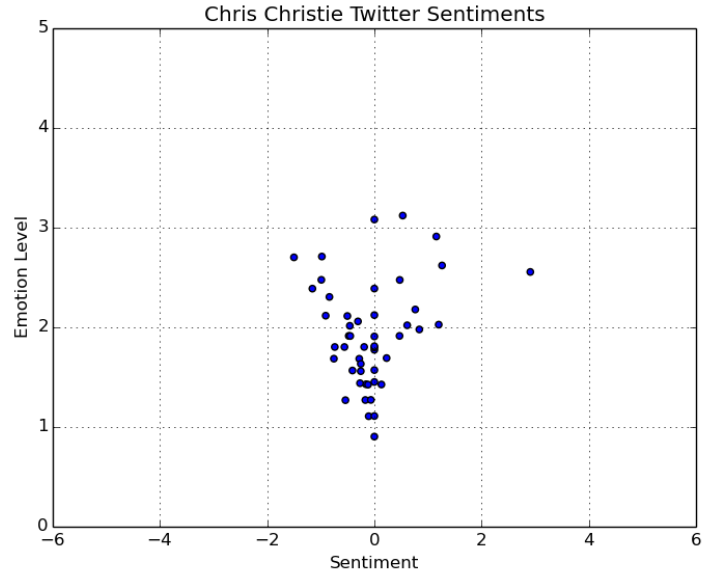


Figure 7: search term “Chris Christie” Sentiment Analysis

Chris Christie has a relatively balanced spread of sentiments for tweets mentioning him, with perhaps more tweets on the negative side, but tweets with a larger magnitude of sentiment on the positive side. It is somewhat surprising that Christie’s sentiment analysis would be relatively unpolarized. Even though Christie himself is a moderate, many people don’t like his style, both Democrats and Republicans. Granted this is a small sample size, but perhaps people on Twitter don’t have the same strong opinions of Christie that many in the media have.

Reflection

I felt that my project was appropriately scoped. I accomplished what I set out to do. Obtaining tweets from Twitter was a bit harder than I anticipated, and I had to use a few workarounds in order to obtain a reasonable amount of tweets for analysis. I ended up having to copy paste some tweets in because of the restrictions of getting tweets from Twitter.

If I were to take this project further, I would like to run the analysis

with more tweets over a longer time. Sometimes, certain particular events contribute a large amount of tweets to a particular political figure's twitter search results. In order to get a more well-rounded analysis, I would want to get tweets from a variety of different time periods. I would also want to improve my sentiment analysis. I felt that it was a good first attempt at a sentiment analyzer, but I didn't really put too much thought into crafting a more serious product. I would want to research more the way others have completed sentiment analysis, and try to make mine more sophisticated. I would also be interested in doing word frequency analysis on some of the tweets, to figure out which issues are associated with which political figures.