Alex Grant
Prof. Steele
Software Design
3 October 2016

Project Overview

I used tweets from the First Presidential Debate and ran a sentiment analysis on them individually to track the sentiment over time. I learned how to use Twitter as a data source and how to use API keys to run additional analysis on a body of text. I also learned to plot the data derived from the sentiment analysis on a graph using pyplot.

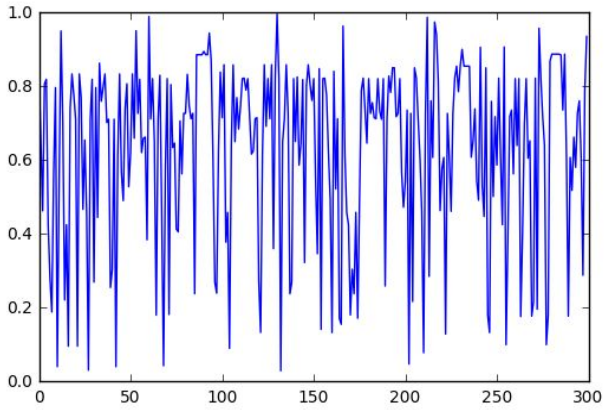
Implementation

In the first cell, I used pattern to import tweets with the hashtag #debates. This was the official hashtag of the First Presidential Debate so I figured it would have some relevant content to analyze. When pulling the Twitter, I recognized that it only pulls the most recent 300 tweets. I made the design decision here because I wanted to measure a general sentiment of the debate at this point in time.

Once the tweets were pulled, I used Indico's sentiment analysis tool to run an individual analysis on each of the tweets. These numbers range between 0 and 1. In order to make sure the tweets should up in chronological order on my graph I had to reverse my list of sentiment values.

Results

Interestingly enough, there was much variance in the sentiment values. I included the final graph that I generating using the matplotlib in pyplot.



To some extent, I am surprised by the results because I was expecting a consistent trend over time to be represented in the overall sentiment. However I recognize the limitations of sentiment analysis in both a programming sense and a realistic sense.

From a programming perspective, sentiment is limited to identifying keywords that could be positive or negative; therefore making an assumption on a body of text. Because of this, there is a chance that some tweets were wrongly categorized positive when they were negative and vice versa.

From a realistic perspective, I understand that there were probably better tests to run in order to truly understand the intentions of these tweets sourced. For example, Indicio offered a political analysis tool. It would have been interesting to see if there was a correlation between sentiment and political alignment.

Reflection

Overall, what went well, - frankly - is the fact that I created a program from scratch and was able to have it run without a bug in the end. The program could really be improved if I could have sourced tweets and overlayed that data on events that were happening in real-time during the debate. This would have shown a cool correlation between who was speaking and the relative reaction of the Twittersphere to what was occurring. This being my first real project, I was not quite sure how to scope, but moving forward I wish I had written out the objectives of the program and mapped out a few logical ways for me to get there instead of simply starting to write code.