The Voice of the People

A Sentiment Analysis of Tweets Containing "Trump" and "Clinton"

Christopher Maddox and Kaija Petrone • Prof. Zhi Li • MIS3640-01: Problem Solving and Software Design • Assignment 3: Text Mining and Analysis • 01 November 2016

Project Overview:

The code submitted for this assignment performs sentiment analysis on one-hundred Twitter entries about Donald Trump and one-hundred Twitter entries about Hillary Clinton. The reason "Twitter entries" as opposed to "tweets" is used to describe the content is that the code analyzes both tweets *and* retweets, not exclusively tweets. (A comment in the code indicates which line to comment or uncomment to analyze only tweets and thus ignore retweets.) It does so by searching Twitter for tweets mentioning the names of these candidates and determining the sentiment of those tweets. The code then determines if the overall tone of the tweets is positive, negative or neutral.

Implementation:

To obtain the data used for analysis, the search function was used with "q," or the phrase, set equal to "Trump," and the count set equal to one-hundred, the maximum amount of tweets that can be returned. Once the tweets were found, they were imported into a new text file so that the analysis could be performed on that file as opposed to on new data that was collected each time the code was run. The "sentiment" function was then used to determine whether the words in each tweet were predominantly positive, negative or neutral. After the sentiment was returned for each tweet, the code aggregated the sentiments into three separate lists: a "positive" list, a "negative" list, and a "neutral" list, and the length of each list was returned to reveal how many tweets were positive, how many were negative, and how many were neutral. These lengths were then compared to determine whether "positive," "negative," or "neutral" was the most common sentiment.

When we ran the code, we noticed that it would print a statement indicating that the overall consensus about a candidate was neutral. However, for the sake of providing more valuable information, we decided to have the code go one step further if the general consensus was neutral. Thus, we programed the code so that, if the general consensus was neutral, it would return an additional output that ignored the neutral tweets and determined whether the remaining tweets had an overall positive or negative connotation. We felt it was necessary to incorporate this step because we do not believe that having a "neutral" sentiment is the most informative.
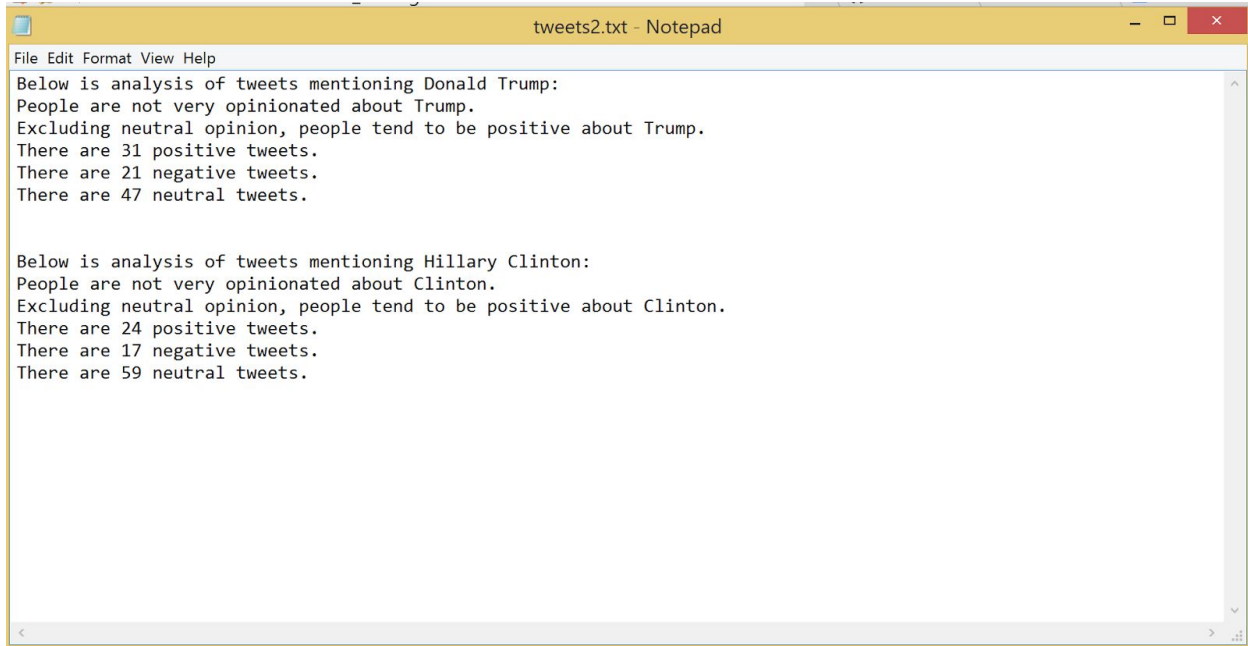
An additional design consideration we encountered was whether to have the text of the tweets returned in the output or not. We felt that an analyst would be curious about the exact words used to describe each candidate, however we also felt that an analyst would primarily be concerned about the overall tone than about the

specific words used. Thus, we kept the line of code (`print(tweet.text.encode("utf-8")))` in the code file **but commented it out so** that a user could easily read the specific tweets used for further analysis, if so desired.

Results:

Based on output generated by the code, it was suggested that the overall sentiments of the tweets about both Trump and Clinton were neutral. However, when our code determined that the overall sentiment was neutral, it would then go deeper to determine whether the overall consensus, when ignoring the neutral tweets, was positive or negative. Running the code a number of times suggests that, on average, the overall sentiment of tweets containing Trump's name was positive and the overall sentiment of tweets containing Clinton's name is positive as well. The other interesting result we discovered was that depending on when we ran the code sometimes tweets were significantly more negative towards a candidate. Overall tweets for both candidates were 75% positive and 25% negative.

Sample results output:

```
tweets2.txt - Notepad
File Edit Format View Help

Below is analysis of tweets mentioning Donald Trump:
People are not very opinionated about Trump.
Excluding neutral opinion, people tend to be positive about Trump.
There are 31 positive tweets.
There are 21 negative tweets.
There are 47 neutral tweets.


Below is analysis of tweets mentioning Hillary Clinton:
People are not very opinionated about Clinton.
Excluding neutral opinion, people tend to be positive about Clinton.
There are 24 positive tweets.
There are 17 negative tweets.
There are 59 neutral tweets.
```

Reflection:

From a process point of view, the collaborative nature of the assignment worked well. We were able to brainstorm ideas together and help each other resolve issues. Moving forward, we will use the knowledge gained from the project to help us in extracting data from Twitter for future analysis. This skill can be extended to learn how to potentially extract information from other sources as well, such as text from a news site such as CNN.com, or locations from which Pinterest posts are posted. Prior to beginning the assignment, it would have been helpful to understand how to the elasticsearch function works so that we could have more easily presented the data in an effective manner, such as a pie chart of the three types of sentiment.

As a partnership, we began by determining the source from which we wanted to extract our data and then worked to determine our objective. We met and had planned to have one team member work to obtain the data from Twitter and the other to input that data into a new textfile. We had significant difficulty importing the data into a new textfile, so we contacted Professor Li for assistance and then met in person as a team to work through a solution. We spent days working to debug our code. There were moments when we considered pivoting to something less challenging, but after many attempts we were finally able to gather usable data. After overcoming this challenge, we had the additional challenge of determining the best way to get usable information from the data. After much dedication to the project, we are happy with our overall result. When we met with each other, we discovered additional problem areas and worked to resolve those problems by utilizing online resources such as docs.python.org and YouTube; in doing so, we were inspired to add additional capabilities to our code and to make slight adjustments to its functionality. If we were to perform this assignment again, we would clearly delineate the plan for our code more thoroughly so that we could determine what exactly is required to run it and ask Professor Li more specific questions. Ultimately, working on this assignment was a great opportunity to improve not only our coding skills, but also our collaborative and communication skills.