Analyzing Twitter Sentiment for Presidential Elections 2016

Onkar Abhishek Tiwari,
MAIT, GGSIP University,
New Delhi, India

Bhaskar Kapoor,
Assistant Professor,
MAIT, GGSIP University, New Delhi, India

Abstract-Analyzing the mood of the public regarding the presidential nominees is the prime objective here. The micro-blogging website twitter is common arena where cyber masses and the politicians express their views. We have used R programming language, and have programmed the analyzer using R Studio. We have pitted superimposed valence plots of the two candidates: Trump and Hillary, to assess to the sentiment comparison as per as the twitter data collected and analyzed by 21st of October,2016.

1. INTRODUCTION

Twitter is a popular micro blogging service where users create status messages (called "tweets"). These tweets sometimes express opinions about different topics, these opinions are sentiment. We propose to build an analyzer which will extract sentiment from a tweet. The tweets are broadly divided into three categories which are as positive, negative and neutral. We have followed a unigram approach i.e. sentiment derivation from an individual word in the tweets.

2. TWITTER API

To import tweets in RStudio we used my existing twitter account to log into twitter developer section. There we created my own application (a unique name of application was required). After creation of the application we generated my access tokens; Access Token and Access Token Secret.

We also had to generate consumer keys; Consumer Key (API Key) and Consumer Secret (API Secret). The Access tokens coupled with Consumer keys paved way for authorization of API in R. After setting up twitter authorization we made a search query with the candidate's name, which was "trump" and "hilary" in this case. The search query enabled the API to download tweets containing this keyword.

3. PREPROCESSING

After the retrieval of Twitter data from the API, We subjected it to certain preprocessing cases such as:

- Removing retweet entities: Because repetitive tweets will affect the meaningful analytics.
- Removing punctuation marks, numbers: Because they convey no sentiment.
- Remove html links: Because we are concerned only with tweets.

- Removing at people: Because their names are irrelevant for sentiment evaluation.

These preprocessing cases were applied in order to improve accuracy of our results and to speed up our calculations. The data after preprocessing was saved to a text file.

4. SENTIMENT EXTRACTION

To derive sentiment from this text file we used the get_sentiment() function method defined in syuzhet package. We used syuzhet method as the argument in this function. This function returns numerical values corresponding to emanating sentiment from each of the sentences. We stored these values in a vector. These values were then plotted on Y-axis in the sentiment plots i.e. each sentiment value corresponded to its equivalent tweet on the X-Axis.
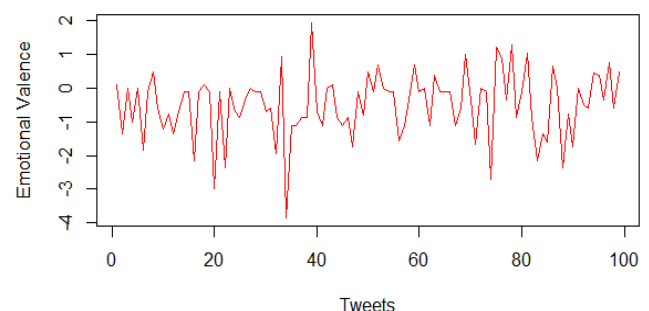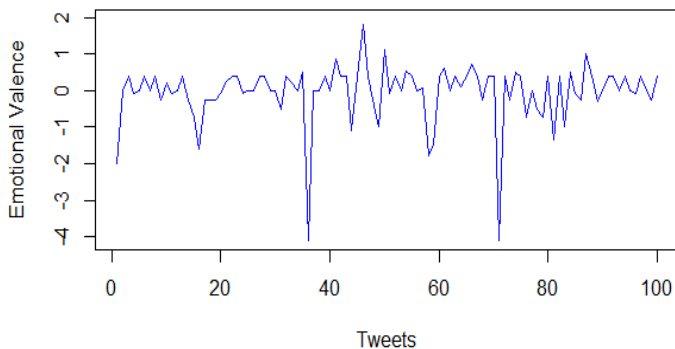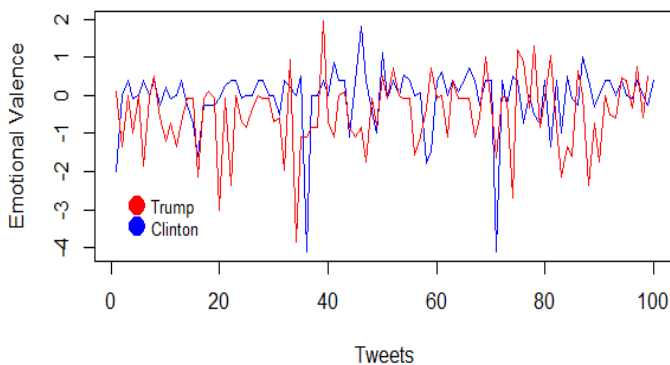
5. SENTIMENT INTERPRETATION

Attributing to the continuous down spikes we interpret that Trump is more criticized by the twitter users. The occasional up-spikes suggests a tweet by trump supporter supporting him.

As for as Hilary, we can say that the she is fairly above the neutral line in most of the occasions, suggesting she is supported by majority of twitter users. However, the lack of more significant positive peaks suggest that she isn't liked very much by the twitter users either, they might possibly be supporting her due to their preference of her over   Trump

Sentiment Tweet Variance Plot Of Trump

Special Issue - 2017

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICCCS - 2017 Conference Proceedings

Sentiment Tweet Variance Plot Of Clinton

Sentiment Comparison Of Trump and Hillary

6. WORD CLOUDS

We also generated Word clouds for both of the candidates in order to get an overview of the words that are most frequently associated with the tweets attributed to the respective candidates. Word clouds provide a pictorial depiction of candidate's image in the public. Also, they can give us an idea as to why the sentiments for the candidates are positive or negative.

7. DISCUSSION

Though Trump has won the elections. Which went against our prediction. It led to some powerful revelations. One of them being: Twitter alone can't be a dependent tool for predicting election outcome. Mining of Twitter Data cannot address "The Silent Majority". The silent majority in sentiment analysis could be related to an analogy of dark matter in the universe.

Secondly, in order to make more accurate predictions, one has to incorporate the global sentiment in regards to the issue. In this experiment, had we incorporated the rising protectionism sentiment of the world, we could have gotten a more accurate result.

Thirdly, ignoring non-English tweets is a bad idea. Reason being, that non-English tweet can influence behavior of another person corresponding to that ethnicity. Setting aside this data can lead into an erroneous evaluation.

8. CONCLUSIONS

The world is getting more digital than ever, it is highly possible in future that even elections will be carried out online. It is safe to suggest that cyber-sentiment can play a pivotal role in its outcome. Therefore, the analysis of sentiment can prove to be of high significance.

We finally reach to a powerful conclusion that, in order to generalize the derivations of sentiments engendering from the aggregate of Twitter data in regards to general public, one must take into account the silent majority, the global chain of factors that are directly or indirectly associated with it in order to ensure more precise predictions.

9. REFERENCES

[1] Delenn Chin, Anna Zappone, Jessica Zhao: Analyzing Twitter Sentiment of the 2016 Presidential Candidates.
[2] Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau: Sentiment Analysis of Twitter Data.
[3] Shruti Wakade, Chandra Shekar, Kathy J. Liszka and Chien-Chung Chan:Text Mining for Sentiment Analysis of Twitter Data.
[4] Efthymios Kouloumpis, TheresaWilson, Johanna Moore: Twitter Sentiment Analysis: The Good the Bad and the OMG!
[5] Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014: The Stanford CoreNLP Natural Language Processing Toolkit