Using Sentiment Analysis to Understand Tweets and their Emotional Value

Final Project Report
"Introduction to Affective Computing" (CMPSCI 590AC)
Summer Session II, 2016

Rohan Tandon
School of Computer Science, UMass Amherst
Junior
rohantandon@umass.edu

Abstract

These days, with the emergence of technology and ease of access to the internet, most people have an online presence and as a result, there exists huge amounts of data about people and their preferences, activities online. This has resulted in corporations trying to study this "big data" to analyze people's views and opinions on microblogging websites such as Twitter, Facebook, Tumblr etc. People often use these websites to express their opinions and feelings about various personal and global issues by posting messages on these websites and friends, family can subscribe to their messages by following (Twitter) or adding them as friends (Facebook).

In this paper, I have used the R programming language to implement sentiment analysis on twitter messages to understand how people talk about the two presidential candidates "Donald Trump" and "Hillary Clinton" in terms of positive or negative sentiment and to verify whether tweets can be used to understand the emotions of the voters. The results have shown that 53 % of tweets pertaining to Donald Trump contain positive sentiment, while 51% of the tweets related to Hillary Clinton contain positive sentiment.

Overall, it was found that sentiment analysis can be used to understand the emotion direction of tweets but it is very difficult to translate the sentiment into a particular emotion as text often fails to express the context of the content and as a result cannot affectively capture intricacies like sarcasm, irony and the tone of mocking someone.

1. Introduction

The main objective of the project was to understand how the general public uses a microblogging site like Twitter to express its emotions. This topic was taken up as Twitter provides an enormous repository of people's uncensored emotions and feelings in the form of tweets, about topics they care about and hence would be their natural reactions. Also, by understanding the emotional significance of text, we can further this implementation to create affective agents which can use text as a model of emotion recognition. The learning motivation was to familiarize myself with sentiment analysis and natural language processing.

1.1 Background Information

Twitter is an excellent platform for companies, politicians and general public to create awareness about various issues and to voice their opinions on current global events. People can use the "#" symbol followed by small text inside their post on twitter to add tags to their tweet and indicating their tweet refers to that particular topic. [1] An example of twitter's effective social awareness usage is the emergence of the topic of "#GivingTuesday", which aims to tackle the massive consumer purchase during sales events like Thanksgiving and Cyber Monday by asking people to donate to charity. This successful hashtag helped in raising $10.1 million donations.

With nearly 320 million active users, it is important to analyze what people talk about in their tweets as it can provide huge insight about their emotions, feelings and thoughts. Also, it can be assumed that most users would be concise and to the point in order to make most of their tweets as tweets are limited to containing 140 characters only. Sentiment analysis is one tool to observe tweets by classifying them into positive and negative sentiments. It is usually done by using a lexicon of words that are considered to have positive and negative meaning and then the

text provided is filtered and compared with those references to measure sentiment. The code has been written in "R programming language", which is primarily used for data science and analytics because of the easily available machine learning algorithms. A detailed explanation of sentiment analysis and the code will be explained in the later sections.

2. Methodology

The workflow of the sentiment analysis is as follows:

Figure .1 Sentiment Analysis Workflow

2.1 Collection of Tweets

In order to collect the tweets, the system first authenticates itself with the twitter application by using a consumer and secret consumer key. These keys are generated when an application is created using a Twitter Developer account.

wordcloud_hc.R | wordcloud_dt.R | hillary_tweet.R | test_twitter.R | donaldTrump_tweets.R

Source on Save | Run

```
1  api_key <- "sample_api_key"
2
3  api_secret <- "sample_secret_api_key"
4
5  access_token <- "sample_access_token"
6
7  access_token_secret <- "sample_secret_token key"
8
9  setup_twitter_oauth(api_key,api_secret,access_token,access_token_secret)
10
11 tweets = searchTwitter("#POTUS", n=500) ###search for tweets with the hashtag POTUS
12
```

Figure 2. Obtaining Twitter Authentication

The “searchTwitter” method is called, which collects the latest tweets (500 in this case) which include the particular search term (“#POTUS).

2.2 Filtering the Tweets

Now, observing the tweets shows that they contain various characters/symbols and emotion depicting icons such as “@, #, ☺, ☹”. These symbols should be first removed as they do not help us in analyzing the text however, emojis can tell us about the emotion being expressed but the system is only looking at the text.

1	Whichever #candidate for #POTUS you like, do your own research: https://t.co/6OMjubGpyl #blog
2	Whichever #candidate for #POTUS you like, do your own research: https://t.co/p1oNwQhfyP #blog
3	@I_AmAmerica @speedprayers and its like we've not had the best yet to contend for the #POTUS but certainly #Trump isn't my kind
4	RT @nendaas1: #HillaryClinton needs help getting in and out of a car #HillarysHealth getting worse by the day unfit to be #POTUS https://t.â¦
5	#JOEBiden always threatening impeachment to Rs, but his own boss #POTUS has broken the law over and over and Joe is silent @HeyTammyBruce
6	RT @pastormarkburns: On #TrumpForceOne last night with our next #POTUS, @realDonaldTrump! With @DanScavino, & Steve! #TrumpTrain https://t.â¦
7	RT @DanScavino: .@RealBenCarson speaks in New Hampshire & introduces our next #POTUS, @realDonaldTrump. #TrumpTrain #MAGA #USA https://t.coâ¦
8	RT @DanScavino: .@RealBenCarson speaks in New Hampshire & introduces our next #POTUS, @realDonaldTrump. #TrumpTrain #MAGA #USA https://t.coâ¦
9	Moderators should #AskTheRobotQuestion of 2016's #POTUS candidates. What is their stance on workforce #automation? https://t.co/SFK6BIopn0
10	RT @DanScavino: .@RealBenCarson speaks in New Hampshire & introduces our next #POTUS, @realDonaldTrump. #TrumpTrain #MAGA #USA https://t.coâ¦
11	RT @pastormarkburns: On #TrumpForceOne last night with our next #POTUS, @realDonaldTrump! With @DanScavino, & Steve! #TrumpTrain https://t.â¦
12	The latest History Happenings ! https://t.co/yqtdJZL0pU #nps100 #potus
13	What a great speech by our future #POTUS #imwithher #HillaryClinton
14	Gloves off! I'm so proud of our future #POTUS @HillaryClinton for calling out #Trump and the #altright #ImWithHer https://t.co/4UuiPdtBgo
15	RT @DanScavino: .@RealBenCarson speaks in New Hampshire & introduces our next #POTUS, @realDonaldTrump. #TrumpTrain #MAGA #USA https://t.coâ¦
16	Less"big enough"more:Confident enough to speak publicly. â¢Normalization of #Trump's behavior â¢8 years of BLK #POTUS https://t.co/ZqshiEe7iK

Figure 3. Unfiltered Tweets containing #POTUS

Similarly, the next step is to remove the commonly used words in the english language such as “the”, “a”, these are known as stop words. We can also use stemming to reduce words to their original form but that might take away the emotions.

1	HillaryClinton Enough is Enough wUr HillaryLies Insulted Millions AmericansClintonCorruptFoundationrealDonaldTrump TRUMP for POTUS
2	RT DanScavino RealBenCarson speaks in New Hampshire amp introduces our next POTUS realDonaldTrump TrumpTrain MAGA USAhttpstco
3	Hillary sez RT affinity for LSE scoundrels like Farage indicts Putin Real culprits are oligarchs Weaken Putin amp POTUS the same
4	Lets get going dammit They have laser in my eyes everydayU need to fathom betteramphave a better barometer potus MichelleObama joebiden
5	RT DanScavino RealBenCarson speaks in New Hampshire amp introduces our next POTUS realDonaldTrump TrumpTrain MAGA USAhttpstco
6	RT pastormarkburns On TrumpForceOne last night with our next POTUS realDonaldTrump With DanScavino amp Steve TrumpTrain httpst

Figure 4. Filtered Tweets

```
#filtering tweets regex method gsub():
tweets = laply(tweets,function(t)t$getText())
#to remove emojis
tweets <- iconv(tweets, 'UTF-8', 'ASCII')
#removing punctuation
tweets = gsub('[[:punct:]]', '', tweets)
tweets = gsub('[[:cntrl:]]', '', tweets)
tweets = gsub('\\d+', '', tweets)
#convert to lower case:
tweets = tolower(tweets)

#creating a corppus
corpus = Corpus(VectorSource(tweets))

#removing stopwords
corpus = tm_map(corpus, removeWords, stopwords(kind="en"))

```

Figure 5. Code for Filtering Tweets

2.3 Adding Sentiment

Finally, we compare each word of the filtered tweets, with the lexicon of positive and negative words provided by [2].

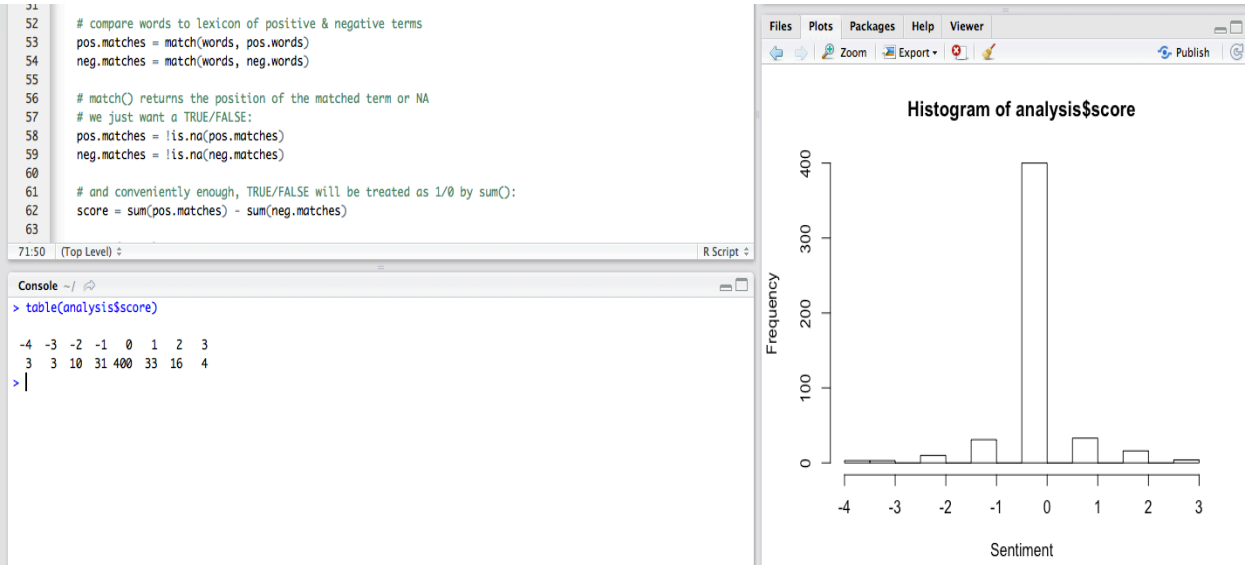

Figure 6. Histogram Depiction Sentiments of Tweets

3. Results

Due to the lack of time, only one experiment is conducted, which is as follows:

1. The data set contains 10000 tweets which have been collected by searching twitter for the tags “#HillaryClinton” and “#DonaldTrump”.
2. The tweets have been collected over the past one week, this is due to the limitation in the Twitter API.
3. In order to categorize the data, the “syuzhet” package [3] has been used. In particular, the “get_nrc_sentiment” method has been called.

```
> get_nrc_sentiment("This is a very pleasant surprise")
  anger anticipation disgust fear joy sadness surprise trust negative positive
1     0            1       0    1   2       0        2     1        0        2
> 
```

Figure 7. Get_NRC_Sentiment Example

4. The results are:

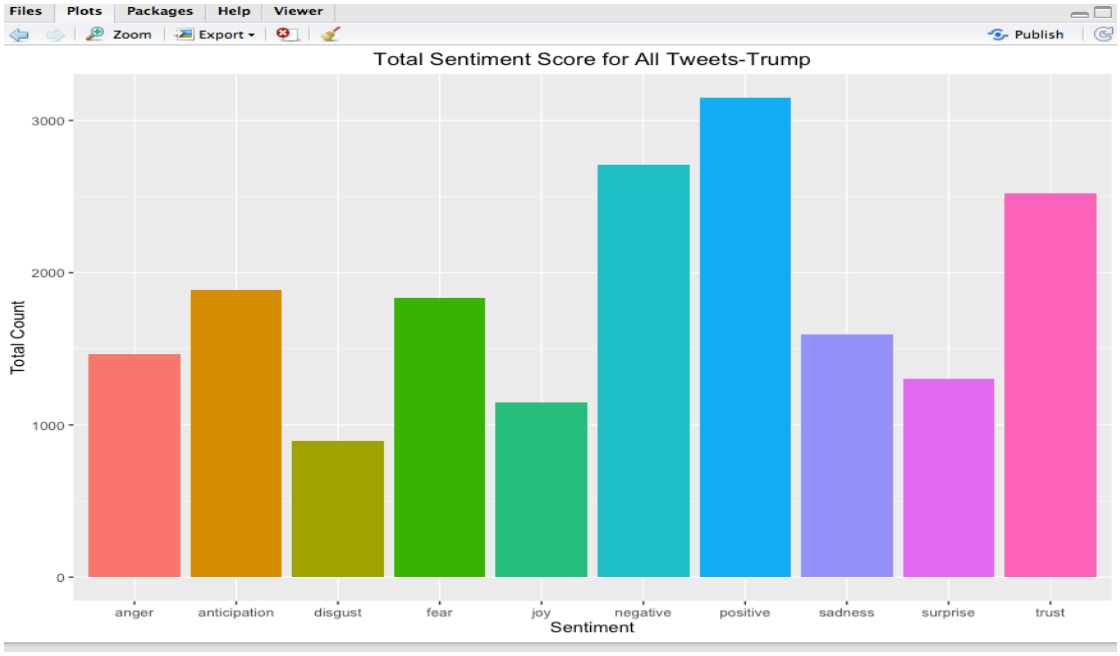

Figure 8. Emotions Expressed in #DonaldTrump

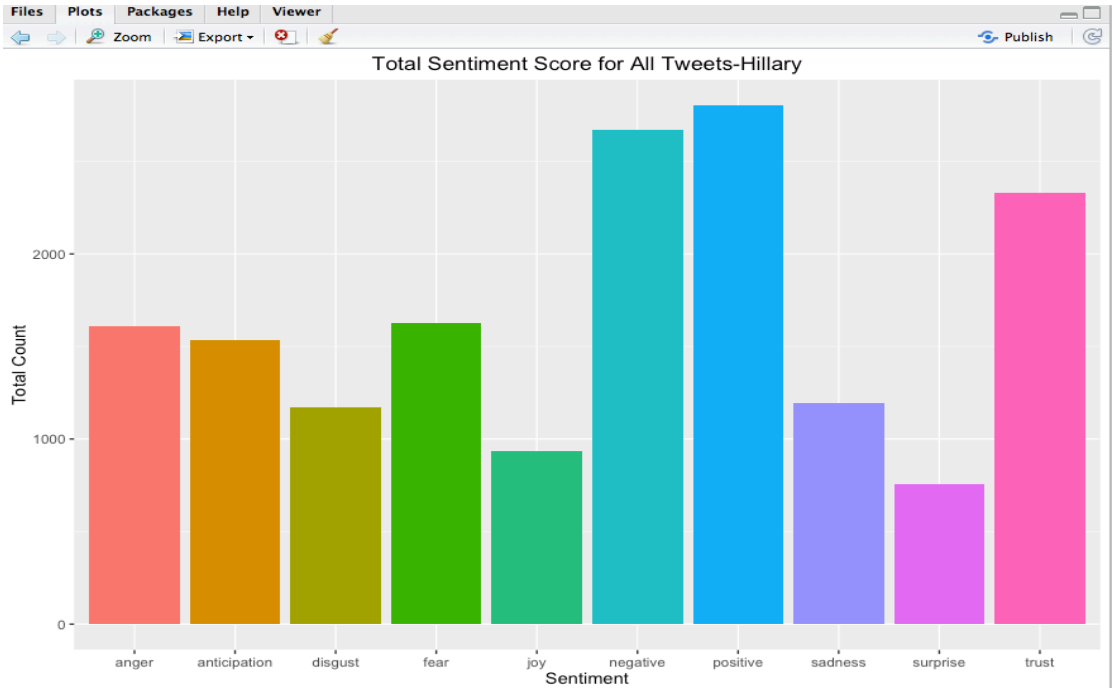

Figure 9. Emotions Expressed in #HillaryClinton

4. Conclusion and Discussion

The paper shows how sentiment analysis can be used on tweets to observe the sentiments and emotions displayed in those tweets. The advantages of this system are that tweets can be easily be obtained, filtered and visualized with the defined packages. Also, by analyzing the latest tweets, it gives the most recent trends and opinions about the concerned topic.

However, there seems to be a few limitations to this model, first of all, the package used to determine the emotions is not able to distinguish between a phrase and the negative of it. For example,

```
Console ~/
> get_nrc_sentiment("I am looking forward to the elections this year")
  anger anticipation disgust fear joy sadness surprise trust negative positive
1     0            0       0    0   0       0        0     0        0        1
> get_nrc_sentiment("I am not looking forward to the elections this year")
  anger anticipation disgust fear joy sadness surprise trust negative positive
1     0            0       0    0   0       0        0     0        0        1
>
```

The second sentence clearly expresses a negative emotion but the “not” is not taken into consideration. Similarly, the model does not take into account the entire meaning of the sentence. Instead of deciding the positivity or negativity of sentence by the meaning, it breaks the sentence into words and then decides the sentiment.

```
> get_nrc_sentiment("I hope this guy dies")
  anger anticipation disgust fear joy sadness surprise trust negative positive
1     0            1       0    0   1       0        1     1        0        1
>
```

In general, it is possible to note the emotion direction by considering the sentiments but there does not exist enough evidence to narrow down to a particular emotion being expressed in the sentence.

5. Bibliography and Reference

1. Sichynsky, T. (2016, March 21). These 10 Twitter hashtags changed the way we talk about social issues. Retrieved August 25, 2016, from https://www.washingtonpost.com/news/the-switch/wp/2016/03/21/these-are-the-10-most-influential-hashtags-in-honor-of-twitters-birthday/
2. Jeffreybreen/twitter-sentiment-analysis-tutorial-201107. (2011, July 11). Retrieved August 25, 2016, from https://github.com/jeffreybreen/twitter-sentiment-analysis-tutorial-201107/blob/master/data/opinion-lexicon-English/negative-words.txt
3. Jockers, M. (2016, April 27). *Introduction to the Syuzhet package*. Retrieved August 26, 2016, from https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html