Tweets Sentimental Analysis

Yatharth Malik

December 17, 2016

Synopsis

The analysis is aimed at knowing the sentimental behavior of people for US Election's candidates Donald J. Trump and Hillary Clinton by analysing the sentiments of their tweets. In the analysis, we would be using Microsoft's Text Analytics API in R to calculate the sentiment score of tweets. Here 1 indicates positive sentiment while 0 indicates negative sentiment.

Loading the required libraries

```
library(twitteR)
library(httr)
library(ggplot2)
library(dplyr)
library(httr)
library(jsonlite)
```

Setting up twitter API

```
consumer_key = "TiSNgNuLqwJ6zizM6uuC1ggOg"
access_token = "443542132-SM0FVcmbabPyTFmf8Fi0BfCCfV5Nbyblm39hT2sB"
consumer_token = "qrwe5VNUevtrU0t9lvlv0z6w9J97PNosHQ9O8Orn9yFf0EdsYZ"
acess_secret = "3fEcym0sP3UtnhkhIVvgEGf5zbC646S9B1dDSaFL9doi2"

setup_twitter_oauth(consumer_key,consumer_token,access_token,acess_secret)
```

```
## [1] "Using direct authentication"
```

Retrieving tweets

```
number_of_tweets = 200         # Number of tweets to be extracted
trump = searchTwitter("Donald Trump+@realDonaldTrump",n = number_of_tweets ,lang = "en")
clinton = searchTwitter("Hillary Clinton + @HillaryClinton",n = number_of_tweets ,lang = "en")
trumpdf = twListToDF(trump)
clintondf = twListToDF(clinton)

# Extracting only tweets from data frame
trumpdf = select(trumpdf,text)
clintondf = select(clintondf,text)
```

Removing duplicates and cleaning tweets

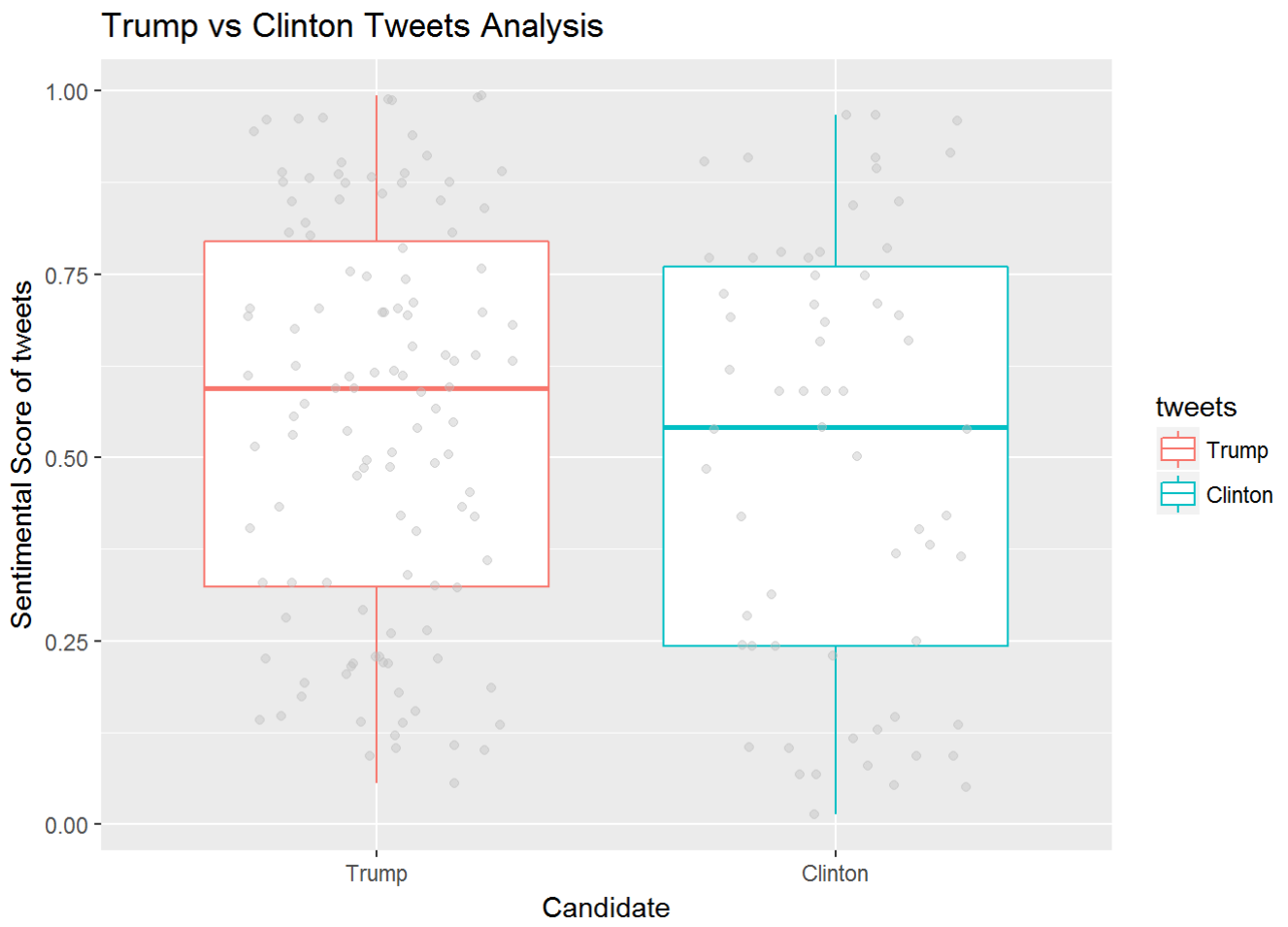
```
# Trump tweets
trump_unique = as.data.frame(trumpdf[!duplicated(trumpdf),],stringsAsFactors = F)
names(trump_unique) =  c("text")
trump_unique$text = gsub("(@|$|#)\\w+","",trump_unique$text)
trump_unique$text = gsub("http\\S+","",trump_unique$text)
trump_unique$text = gsub("RT","",trump_unique$text)
trump_unique$text = gsub("[[:punct:]]","",trump_unique$text)
trump_unique$text = gsub("\\d+\\w+","",trump_unique$text)

# Clinton tweets
clinton_unique = as.data.frame(clintondf[!duplicated(clintondf),],stringsAsFactors = F)
names(clinton_unique) =  c("text")
clinton_unique$text = gsub("(@|$|#)\\w+","",clinton_unique$text)
clinton_unique$text = gsub("http\\S+","",clinton_unique$text)
clinton_unique$text = gsub("RT","",clinton_unique$text)
clinton_unique$text = gsub("[[:punct:]]","",clinton_unique$text)
clinton_unique$text = gsub("\\d+\\w+","",clinton_unique$text)
```

Sentimental analysis of tweets

```
# Trump
trump_unique$language = "en"
trump_unique$id = seq.int(nrow(trump_unique))
trump_request = trump_unique[c(2,3,1)]
trump_size = nrow(trump_request)

trump_request_json = toJSON(list(documents = trump_request))
response = POST("https://westus.api.cognitive.microsoft.com/text/analytics/v2.0/sentiment",
                body = trump_request_json,
                add_headers(.headers = c("Content-Type"="application/json",
                "Ocp-Apim-Subscription-Key"="5155d7c8316641f9986d24dc562da8df")))

content_trump = content(response)
output_trump = data.frame(matrix(unlist(content_trump),nrow = trump_size,byrow = T),stringsAsFac
tors = F)
names(output_trump) = c("Sentiment_Score","ID")
output_trump$tweets = as.factor("Trump")
output_trump$Sentiment_Score = as.numeric(output_trump$Sentiment_Score)


# Clinton
clinton_unique$language = "en"
clinton_unique$id = seq.int(nrow(clinton_unique))
clinton_request = clinton_unique[c(2,3,1)]
clinton_size = nrow(clinton_request)

clinton_request_json = toJSON(list(documents = clinton_request))
response = POST("https://westus.api.cognitive.microsoft.com/text/analytics/v2.0/sentiment",
                body = clinton_request_json,
                add_headers(.headers = c("Content-Type"="application/json",
                                "Ocp-Apim-Subscription-Key"="5155d7c8316641f9986d24dc562da8d
f")))

content_clinton = content(response)
output_clinton = data.frame(matrix(unlist(content_clinton),nrow = clinton_size,byrow = T),string
sAsFactors = F)
names(output_clinton) = c("Sentiment_Score","ID")
output_clinton$tweets = as.factor("Clinton")
output_clinton$Sentiment_Score  = as.numeric(output_clinton$Sentiment_Score)
```

Plotting the result

```
final = rbind(output_trump,output_clinton)
g = ggplot(final,aes(tweets,Sentiment_Score)) +
    geom_boxplot(aes(colour = tweets)) +
    geom_jitter(position = position_jitter(width = 0.3),colour = "grey ", alpha = 0.4) +
    labs(title = "Trump vs Clinton Tweets Analysis") +
    labs(x = "Candidate", y = "Sentimental Score of tweets")
print(g)
```



Summary of sentimental score of Trump's Tweets

```
print(summary(output_trump$Sentiment_Score))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.05566 0.32400 0.59410 0.55380 0.79400 0.99400
```

Summary of sentimental score of Clinton's Tweets

```
print(summary(output_clinton$Sentiment_Score))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01339 0.24260 0.54180 0.50490 0.76050 0.96680
```