Text Processing and Sentiment Analysis of Twitter Data

Chouliaras Andreas
achouliaras@inf.uth.gr

Pappas Apostolos
apopappas@inf.uth.gr

April 2019

1 Introduction

What is Sentiment Analysis?

Sentiment analysis is the automated process of understanding an opinion about a given subject from written or spoken language. It relates to feelings; attitudes, emotions and opinions.

A person's opinion or feelings are for the most part subjective and not facts. Which means to accurately analyze an individual's opinion or mood from a piece of text can be extremely difficult. With Sentiment Analysis from a text analytics point of view, we are essentially looking to get an understanding of the attitude of a writer with respect to a topic in a piece of text and its polarity; whether it's positive, negative or neutral.

Why Sentiment Analysis?

Sentiment Analysis gives us insights into consumer and overall human behavior, what customers want, what customers like and dislike about the products, what their buying habits are, what their decision process looks like etc. It is also extremely useful in social media monitoring, as it allows us to gain an overview of the wider public opinion behind certain topics.

For example, the Obama administration used sentiment analysis to gauge public opinion to policy announcements and campaign messages ahead of 2012 presidential election. Being able to quickly see the sentiment behind everything from forum posts to news articles means being better able to strategize and plan for the future.

In this project, we will examine three topics using Sentiment Analysis.

2 Our topics

For our first topic we selected the Game Of Thrones series. Game of Thrones is one of the most, if not the most, popular show aired today. At the time of this project completion, we are two weeks before the release of its eighth and final season. So, we expect to notice some very interesting results on the fans' feelings.

Of course, for a Sentiment Analysis of Twitter, we could not exclude the current star in it. Donald Trump, the president of USA, is always a hot topic in Twitter. We would like to see peoples' opinions on President's recent activity.

Our third topic is the Dark Souls franchise. A video game trilogy notorious for its high difficulty. What we expect to see is lots of mixed feelings about the game.

3 Implementation

Here, we present the code of our implementation on the Game Of Thrones topic using the hashtag #gameofthrones. The other two topics are implemented using essentially the same code.

3.1 Data Extraction

After installing and loading the required packages, we use the Twitter Web API to extract using a particular hashtag. We then convert this data to a data frame which makes it much more easier to process.

```
tweets_g <- searchTwitter("#gameofthones", n=1000,lang = "en")

got_tweets <- twListToDF(tweets_g)
got_text<- got_tweets$text
```

3.2 Data Processing

We then need to preprocess our data and remove tabs, blank spaces, links etc.

```
# Convert all text to lower case
got_text<- tolower(got_text)

# Replace blank space (\rt")
got_text <- gsub("rt", "", got_text)

# Remove @UserName
got_text <- gsub("@\\w+", "", got_text)

# Remove punctuation
got_text <- gsub("[[:punct:]]", "", got_text)

# Remove links
got_text <- gsub("http\\w+", "", got_text)

# Remove tabs
got_text <- gsub("[ |\t]{2,}", "", got_text)

# Remove blank spaces at the beginning
got_text <- gsub("^ ", "", got_text)

# Remove blank spaces at the end
got_text <- gsub(" $", "", got_text)

#create corpus
got_tweets.text.corpus <- Corpus(VectorSource(got_text))
```

3.3 Data Cleaning and Word-cloud

We need to clean our tweets by removing stop words. Stop words are commonly used words at any language that do not add useful information when analyzing the text. They can vary from determiners (e.g. the, a, an) to prepositions (e.g. above, across, before)

```
#clean up by removing stop words
got_tweets.text.corpus <- tm_map(got_tweets.text.corpus,
                    function(x)removeWords(x,stopwords()))

#generate wordcloud
wordcloud(got_tweets.text.corpus,min.freq = 10,colors=brewer.pal(8, "Dark2"),
                    random.color = TRUE,max.words = 500)
```

The more a specific word appears in a source of textual data (such as a speech, blog post, or database), The bigger and bolder it appears in the word cloud.

3.4 Sentiment Analysis

We have used the inbuilt sentiment analyzer in R, which uses the NRC sentiment dictionary to calculate the presence of eight different emotions and their corresponding importance in a text.

```
#getting emotions using in-built function
mysentiment_got<-get_nrc_sentiment((got_text))

#calculating total score for each sentiment
Sentimentscores_got<-data.frame(colSums(mysentiment_got[,]))
names(Sentimentscores_got)<-"Score"
Sentimentscores_got<-cbind("sentiment"=rownames(Sentimentscores_got),
                            Sentimentscores_got)
rownames(Sentimentscores_got)<-NULL
```

Finally we use ggplot2 to plot our bar plot:

```
library(ggplot2)
#plotting the sentiments with scores
ggplot(data=Sentimentscores_got,aes(x=sentiment,y=Score))+
            geom_bar(aes(fill=sentiment),stat ="identity")+
            theme(legend.position="none")+xlab("Sentiments")+
            ylab("scores")+
            ggtitle("Sentiments of people behind the tweets on
            Game of Thrones")
```
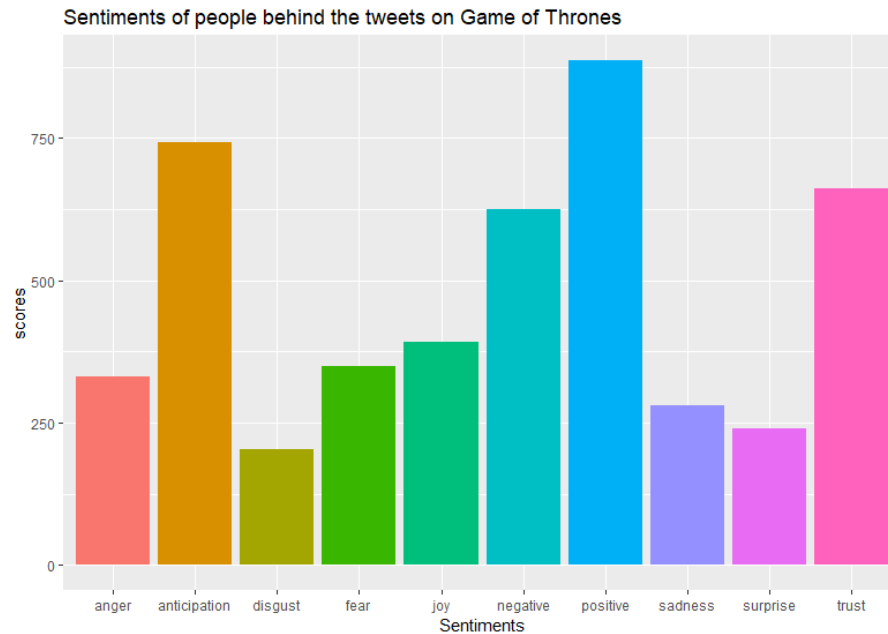
4 Results

4.1 Word-clouds

Firstly. let's take a look at what our word-clouds say.

4.2   Sentiment Graphs
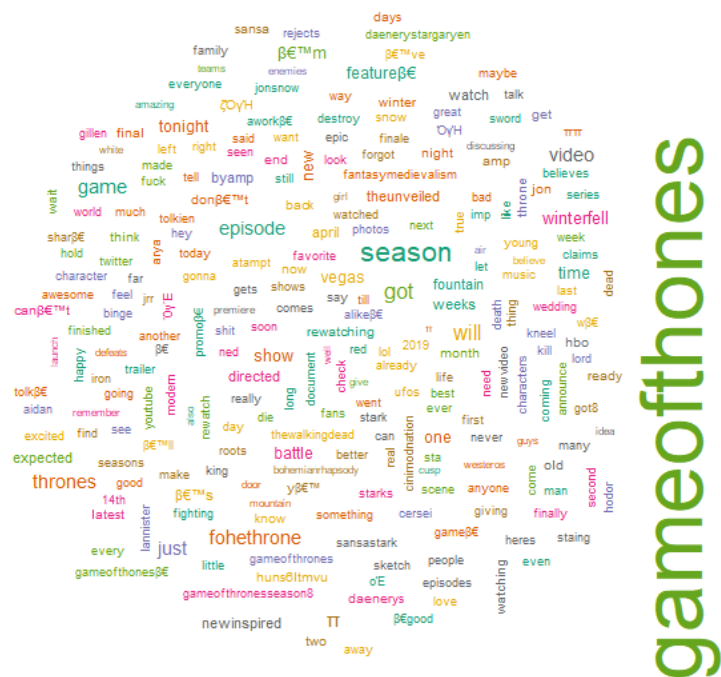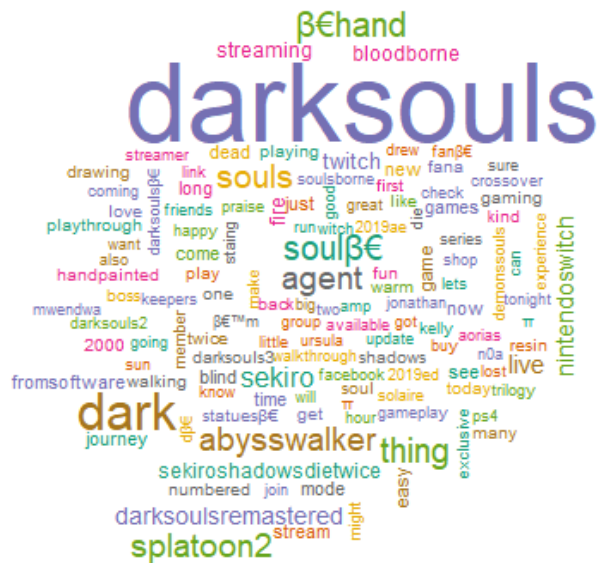
Sentiments of people behind the tweets on Game of Thrones

We can confirm Game of Thrones popularity with its high positive and trust sentiment values The anticipation is high and it can be explained if we consider we are two weeks before the release of the final season. Its high negative value probably comes from disagreements about various characters. People have different opinions about who they want to see succeed and that may rises negativity.

Sentiments of people behind the tweets on Donald Trump
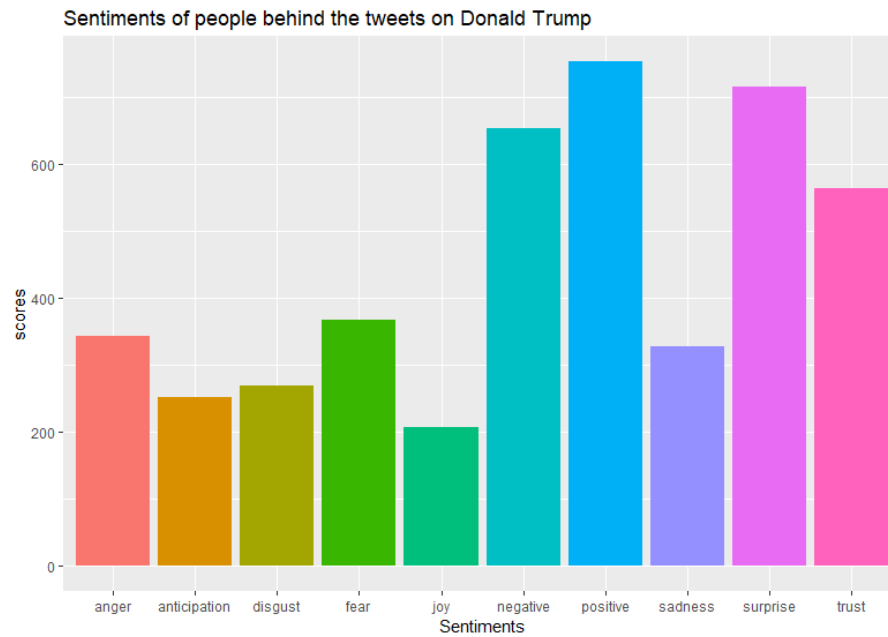
Feelings about the US President, Donald Trump, are also mixed. We have many positive and negative feelings. What we immediately noticed, is the high amount of surprise and trust. After some research, we found out that President Trump announced some changes in his health care policy, surprising the Republicans.

Sentiments of people behind the tweets on Dark Souls
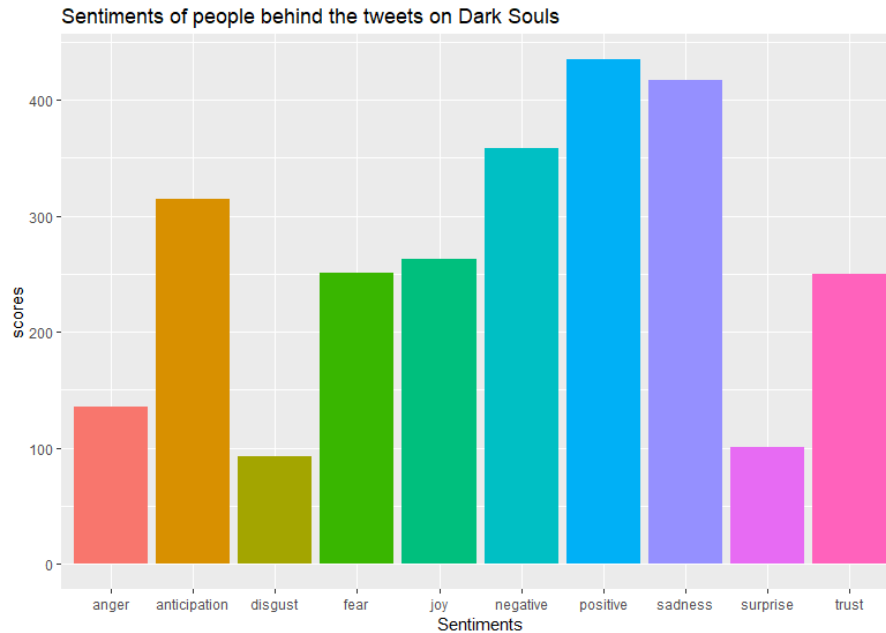
We see that this notorious hard game franchise creates both negative and positive feelings. Although most people seems to like the games we see that the sadness value is very high. These feelings of sadness along with anticipation, fear and joy probably project the increased difficulty of the games. This seems reasonable if we consider that the player feels negative emotions while losing over and over again, but the joy is also big when they finally succeed.

Some of these feelings may be due to the fact that their most recent game was released about two weeks ago.

5   Conclusion

Sentiment analysis can be seen as a very powerful tool for the right people. We can now monitor, and even control, the public opinion on several matters. It can benefit a very wide spectrum of aspects such as businesses, political campaigns and others.

As we saw, people are anticipating the new season of Game of thrones, are very surprised by Donald Trump's recent health care policy change, and really hate dying in Dark souls.