sentiment analysis NLP

LANGAT ERICK

2023-08-22

```
# This R program is designed and test for whatsapp chats(group or individual).
#You can export the chats from whatsapp to your email ID and download it to your PC and work on it.
# It is defined by the kaggle/python docker image: https://github.com/kaggle/docker-python
library(bookdown)
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(scales)
library(reshape2)
library(tm)
```

```
## Loading required package: NLP
```

```
##
## Attaching package: 'NLP'
```

```
## The following object is masked from 'package:ggplot2':
##
##     annotate
```

```
library(SnowballC)
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
library(RColorBrewer)
library(stringr)
library(syuzhet)
```

```
##
## Attaching package: 'syuzhet'

## The following object is masked from 'package:scales':
##
##     rescale
```

```
library(dplyr )
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
suppressPackageStartupMessages(require(tidyverse))
```

```
#get the data from whatsapp chat
text <- readLines("C:/Users/langa/OneDrive/Desktop/R PROGRAMMING PRACTICE/GBcomments.csv", n = 10000)
```

```
#Convert the file to UTF-8 encoding
s2 <- iconv(text, "UTF-8", "ASCII", sub = "")
text <- s2
head(s2)
```

```
## [1] "video_id,comment_text,likes,replies"
## [2] "jt2OHQhOHoQ,\"It's more accurate to call it the M+ (1000) because the price is closer than calli
## [3] "jt2OHQhOHoQ,\"To be there with a samsung phone\\n\",1,0"
## [4] "jt2OHQhOHoQ,\"Thank gosh, a place I can watch it without having to be at HD... my speed doesnt s
## [5] "jt2OHQhOHoQ,\"What happened to the home button on the iPhone X? *****Cough****copying Samsung***
## [6] "jt2OHQhOHoQ,\"Power is the disease. Care is the cure. Keep caring for yourself and others as bes
```

```
#let us create the corpus
docs <- Corpus(VectorSource(text))
head(docs)
```

```
## <<SimpleCorpus>>
## Metadata:  corpus specific: 1, document level (indexed): 0
## Content:  documents: 6
```

```
#clean our chat data
trans <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs <- tm_map(docs, trans, "/")
```

```
## Warning in tm_map.SimpleCorpus(docs, trans, "/"): transformation drops
## documents
```

```
docs <- tm_map(docs, trans, "@")
```

```
## Warning in tm_map.SimpleCorpus(docs, trans, "@"): transformation drops
## documents
```

```
docs <- tm_map(docs, trans, "\\|")
```

```
## Warning in tm_map.SimpleCorpus(docs, trans, "\\|"): transformation drops
## documents
```

```
docs <- tm_map(docs, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(docs, content_transformer(tolower)):
## transformation drops documents
```

```
docs <- tm_map(docs, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(docs, removeNumbers): transformation drops
## documents
```

```
docs <- tm_map(docs, removeWords, stopwords("english"))
```

```
## Warning in tm_map.SimpleCorpus(docs, removeWords, stopwords("english")):
## transformation drops documents
```

```
docs <- tm_map(docs, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(docs, removePunctuation): transformation drops
## documents
```

```
docs <- tm_map(docs, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(docs, stripWhitespace): transformation drops
## documents
```

```
docs <- tm_map(docs, stemDocument)
```

```
## Warning in tm_map.SimpleCorpus(docs, stemDocument): transformation drops
## documents
```

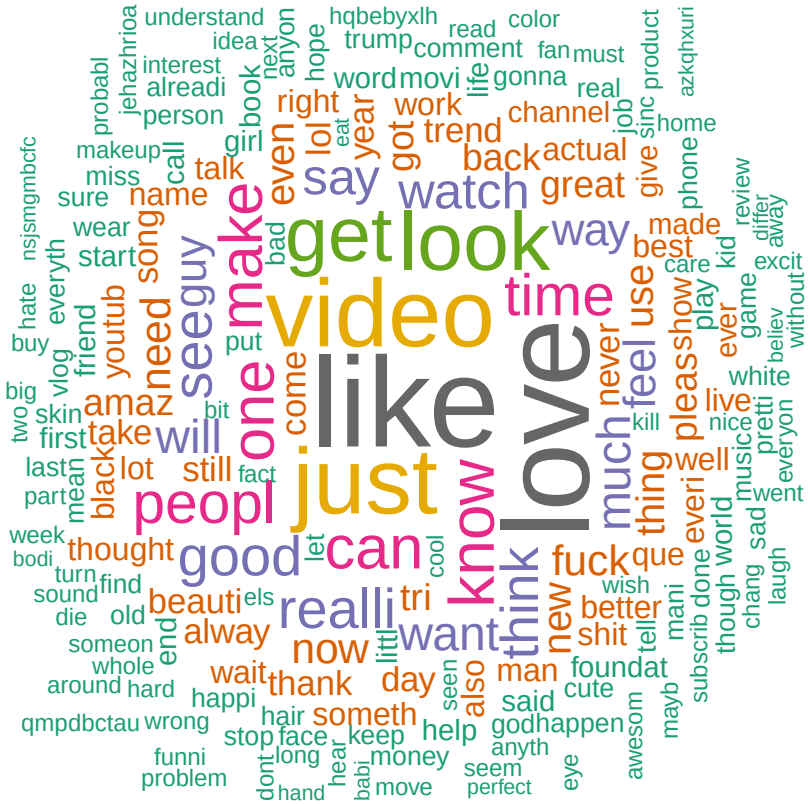
```
#create the document term matrix
dtm <- TermDocumentMatrix(docs)
mat <- as.matrix(dtm)
v <- sort(rowSums(mat),decreasing=TRUE)
```

```r
#Data frame
data <- data.frame(word = names(v),freq=v)
head(data)
```

```
##         word freq
## like   like 1073
## love   love 1019
## video video  774
## just   just  751
## look   look  641
## get     get  578
```

```r
#generate the wordcloud
set.seed(1056)

wordcloud(words = data$word, freq = data$freq, min.freq = 1,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
```




```r
#fetch sentiment words from texts
Sentiment <- get_nrc_sentiment(as.character(text))
head(Sentiment)
```

```
##   anger anticipation disgust fear joy sadness surprise trust negative positive
```

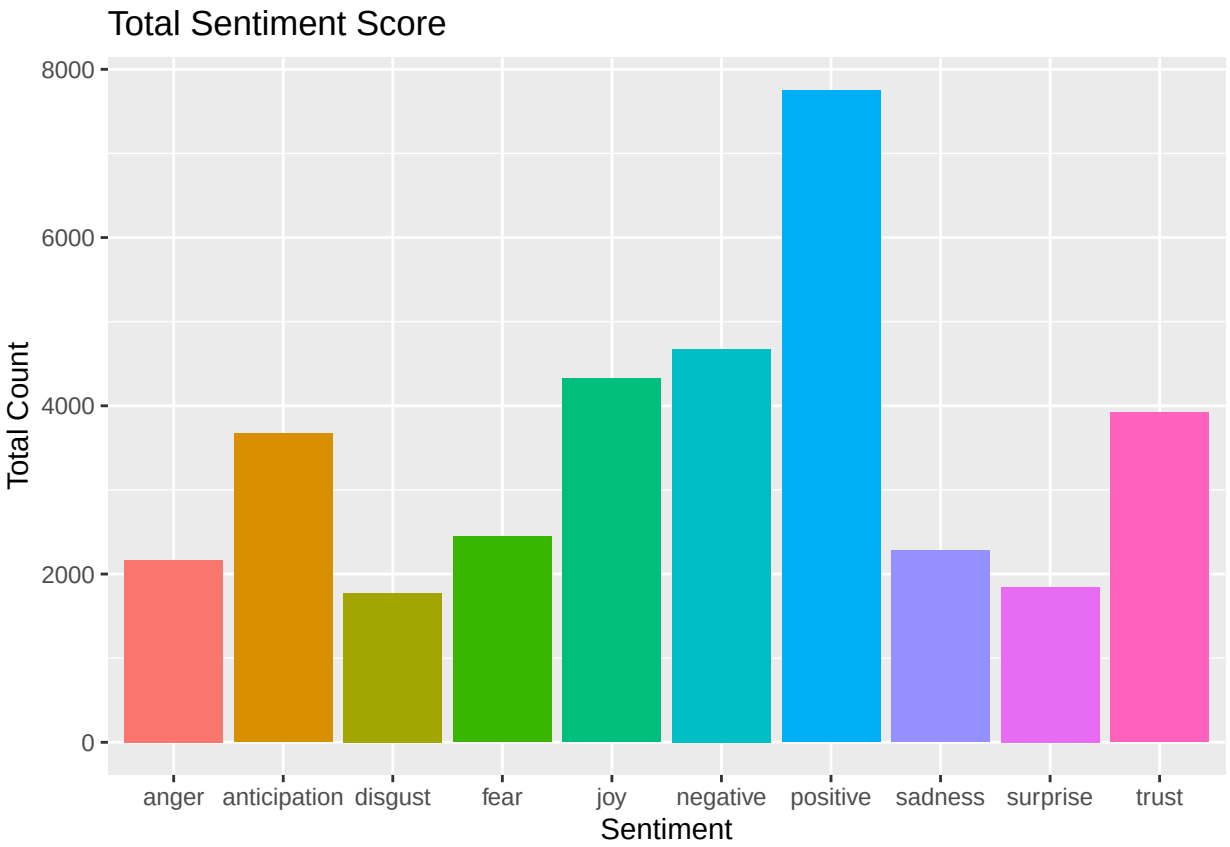
```
## 1     0            0       0    0   0       0        0     0        0        0
## 2     0            0       0    0   0       0        0     1        0        1
## 3     0            0       0    0   0       0        0     0        0        0
## 4     0            1       0    1   0       0        0     0        0        0
## 5     0            0       1    0   0       0        0     0        1        0
## 6     1            0       1    1   0       1        0     0        1        0
```

```
text <- cbind(text,Sentiment)
head(text)
```

```
##
## 1                                                                                     video_
## 2          jt2OHQhOHoQ,"It's more accurate to call it the M+ (1000) because the price is closer thar
## 3                                                                        jt2OHQhOHoQ,"To be there
## 4                 jt2OHQhOHoQ,"Thank gosh, a place I can watch it without having to be at HD... my
## 5                  jt2OHQhOHoQ,"What happened to the home button on the iPhone X? *****Cough****coj
## 6 jt2OHQhOHoQ,"Power is the disease. Care is the cure. Keep caring for yourself and others as best as
##   anger anticipation disgust fear joy sadness surprise trust negative positive
## 1     0            0       0    0   0       0        0     0        0        0
## 2     0            0       0    0   0       0        0     1        0        1
## 3     0            0       0    0   0       0        0     0        0        0
## 4     0            1       0    1   0       0        0     0        0        0
## 5     0            0       1    0   0       0        0     0        1        0
## 6     1            0       1    1   0       1        0     0        1        0
```

```
#count the sentiment words by category
TotalSentiment <- data.frame(colSums(text[,c(2:11)]))
names(TotalSentiment) <- "count"
TotalSentiment <- cbind("sentiment" = rownames(TotalSentiment), TotalSentiment)
rownames(TotalSentiment) <- NULL
```

```
#total sentiment score of all texts
ggplot(data = TotalSentiment, aes(x = sentiment, y = count)) +
  geom_bar(aes(fill = sentiment), stat = "identity") +
  theme(legend.position = "none") +
  xlab("Sentiment") + ylab("Total Count") + ggtitle("Total Sentiment Score")
```


Total Sentiment Score
Total Count
8000
6000
4000
2000
0
anger
anticipation
disgust
fear
joy
negative
positive
sadness
surprise
trust
Sentiment