Sentiment Analysis

Riccardo Ruta

7/5/2022

Contents

Inspect the Syuzhet dictionary

```
head(get_sentiment_dictionary(dictionary = "nrc", language = "italian"),10)
```

```
##       lang         word sentiment value
## 1  italian         abba  positive     1
## 2  italian     capacità  positive     1
## 3  italian sopra citato  positive     1
## 4  italian     assoluto  positive     1
## 5  italian  assoluzione  positive     1
## 6  italian    assorbito  positive     1
## 7  italian   abbondanza  positive     1
## 8  italian   abbondante  positive     1
## 9  italian   accademico  positive     1
## 10 italian    accademia  positive     1
```

```
tail(get_sentiment_dictionary(dictionary = "nrc", language = "italian"),10)
```

```
##          lang         word sentiment value
## 13892 italian    testimone     trust     1
## 13893 italian meraviglioso     trust     1
## 13894 italian       parola     trust     1
## 13895 italian        culto     trust     1
## 13896 italian        degno     trust     1
## 13897 italian          wot     trust     1
## 13898 italian     bramosia     trust     1
## 13899 italian         <NA>     trust     1
## 13900 italian      zelante     trust     1
## 13901 italian        gusto     trust     1
```

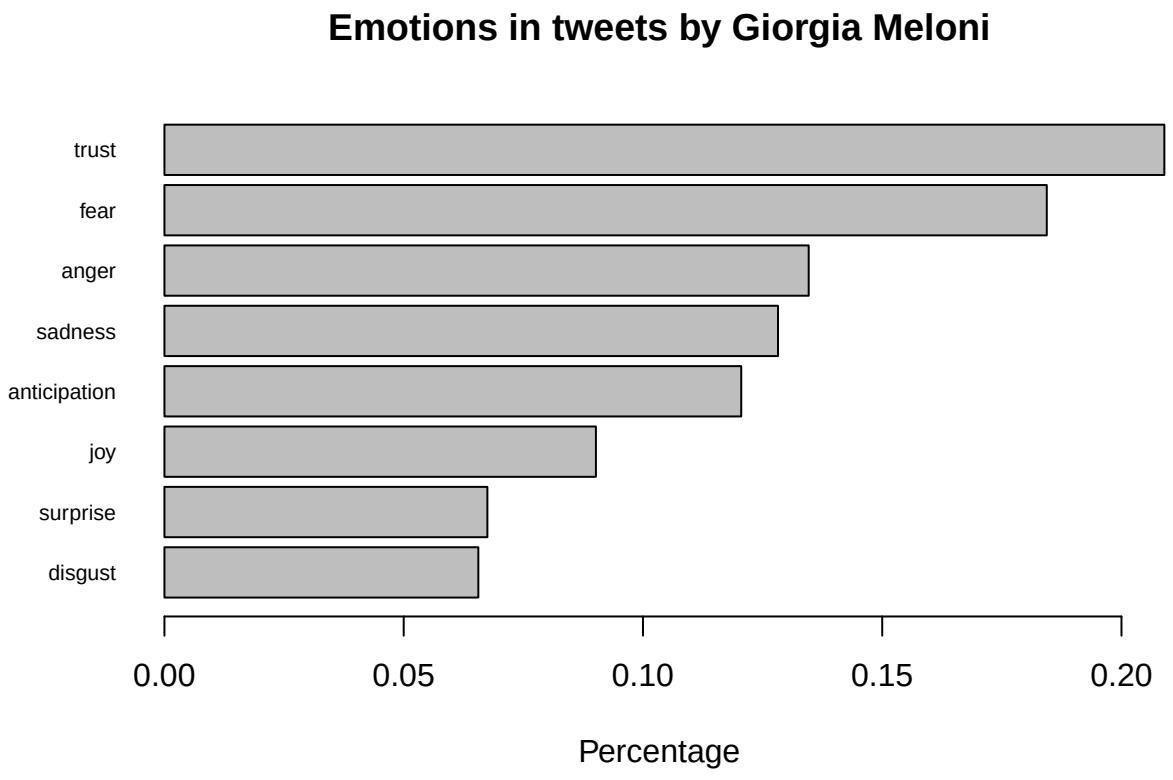
Giorgia Meloni

Filter the corpus

```
politician_dataset <- dataset %>% filter(nome  %like% "MELONI")
politician <- corpus_subset(corpus, nome %like% "MELONI")
nrc_data <- get_nrc_sentiment(politician_dataset$tweet_testo, language="italian")
```

```
## Warning: `spread_()` was deprecated in tidyr 1.2.0.
## Please use `spread()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

```
barplot(
  sort(colSums(prop.table(nrc_data[, 1:8]))),
  horiz = TRUE,
  cex.names = 0.7,
  las = 1,
  main = "Emotions in tweets by Giorgia Meloni", xlab="Percentage"
)
```



Plot the wordcloud of emotions

```
all <- c(
  paste(politician_dataset$tweet_testo[nrc_data$anger > 0], collapse=" "),
  paste(politician_dataset$tweet_testo[nrc_data$anticipation > 0], collapse=" "),
  paste(politician_dataset$tweet_testo[nrc_data$disgust > 0], collapse=" "),
  paste(politician_dataset$tweet_testo[nrc_data$fear > 0], collapse=" "),
  paste(politician_dataset$tweet_testo[nrc_data$joy > 0], collapse=" "),
  paste(politician_dataset$tweet_testo[nrc_data$sadness > 0], collapse=" "),
  paste(politician_dataset$tweet_testo[nrc_data$surprise > 0], collapse=" "),
  paste(politician_dataset$tweet_testo[nrc_data$trust > 0], collapse=" ")
)

# clean the text

# function to make the text suitable for analysis
clean.text = function(x)
{
  # tolower
  x = tolower(x)
  # remove rt
  x = gsub("rt", "", x)
  # remove at
  x = gsub("@\\w+", "", x)
  # remove punctuation
  x = gsub("[[:punct:]]", "", x)
  # remove numbers
  x = gsub("[[:digit:]]", "", x)
  # remove links http
  x = gsub("http\\w+", "", x)
  # remove tabs
  x = gsub("[ |\t]{2,}", "", x)
  # remove blank spaces at the beginning
  x = gsub("^ ", "", x)
  # remove blank spaces at the end
  x = gsub(" $", "", x)
  return(x)
}

all <- clean.text(all)

# create corpus
corpus_viz <- Corpus(VectorSource(all))

# create term-document matrix
tdm <- TermDocumentMatrix(corpus_viz)

# convert as matrix
tdm <- as.matrix(tdm)

# add column names
colnames(tdm) <- c('anger', 'anticipation', 'disgust', 'fear', 'joy', 'sadness', 'surprise', 'trust')
```

```r
# Plot comparison wordcloud
layout(matrix(c(1, 2), nrow=2), heights=c(1, 4))
par(mar=rep(0, 4))
plot.new()
text(x=0.5, y=0.5, 'Emotion Comparison Word Cloud for tweets from Giorgia Meloni')
comparison.cloud(tdm, random.order=FALSE,
                 colors = c("#00B2FF", "red", "#FF0099", "#6600CC", "green", "orange", "blue", "brown"),
                 title.size=1.5, max.words=80)
```

Emotion Comparison Word Cloud for tweets from Giorgia Meloni