

# Análise de sentimentos em Redes Sociais: uma visão do que as pessoas ao redor do mundo dizem a respeito de Bolsonaro

Mario Peres

October 31, 2021

% !TEX encoding = UTF-8 Unicode

O objetivo é capturar dados da rede social Twitter relacionados ao tema Bolsonaro, e realizar análise de sentimentos. Depois da coleta e tratamento dos dados, serão geradas visualizações do tipo nuvem de palavras, bem como o cálculo de score de sentimento acompanhado de um boxplot e histograma para facilitar a interpretação dos resultados.

Vale ressaltar que esse projeto teve como principal objetivo a prática das ferramentas de coleta e análise de dados do Tweeter, e portanto, não tem interesse em promover ou redimir o presidente da República Jair Bolsonaro, mas apenas fazer uma análise de sentimento dos últimos 10.000 tweeters, contento o tema Bolsonaro, e escritos em inglês.

Então vamos carregar os pacotes que vamos utilizar, incluindo o script utils.R, o qual possui algumas funções de limpeza que vamos utilizar nesse projeto.

```
# Carregando os pacotes
```

```
library(twitterR)
library(httr)
library(knitr)
library(rmarkdown)

library(SnowballC)
library(tm)
```

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

```
##
```

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

```
## The following object is masked from 'package:httr':
```

```
##
```

```
##      content
```

```
library(RColorBrewer)
library(wordcloud)
```

```
library(ggplot2)
```

```
##
```

```
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:NLP':
```

```
##
```

```
##      annotate
```

```
library(stringr)
library(plyr)
```

```
##
## Attaching package: 'plyr'
## The following object is masked from 'package:twitterR':
##
##      id
# Carregando a biblioteca
source('utils.R')

options(warn=-1)
```

Agora vamos fazer o processo de autenticação no Tweeter. Aqui voce deve inserir suas informações da sua conta de desenvolvedor no tweeter.

```
key <- "enter your key" secret <- "enter your secret" token <- "token" tokensecret <- "enter your tokensecret"
```

Agora vamos inserir o seguinte comando:

```
setup_twitter_oauth(key, secret, token, tokensecret)
```

Capturando e visualizando as 6 primeiras linhas dos tweets que mencionam Bolsonaro.

```
# Coleta de tweets que mencionam Bolsonaro
tema_Bol <- "Bolsonaro"
qtd_tweets <- 10000
lingua <- "en"
tweets_Bol = searchTwitter(tema_Bol, n = qtd_tweets, lang = lingua)

# Visualizando as primeiras linhas dos objetos tweets
head(tweets_Bol)
```

```
## [[1]]
## [1] "saramaffio: RT @mwillersqc: Very proud to be a part of the All Rise group and supporter of @The
##
## [[2]]
## [1] "eddiecamilo: @jairbolsonaro https://t.co/Zs7oLFmFot BRAZIL LOVES BOLSONARO! @ValdoCruz @octavio
##
## [[3]]
## [1] "congeebrain: important clarification as it seems MTL is causing a lot of misunderstanding: \n\n
##
## [[4]]
## [1] "KareemfanWSR: @JWillows1997 @AndyGraham22 Yup, the cities are backing the races. Bolsonaro want
##
## [[5]]
## [1] "danyloria: RT @michaelreid52: Latin America was a leader in the fight against climate change un
##
## [[6]]
## [1] "eddiecamilo: @MarcosQuezado1 https://t.co/Zs7oLFmFot BRAZIL LOVES BOLSONARO! @ValdoCruz @octavi
```

Aqui vamos fazer alguns processo de limpeza, como remover pontuação, converter os dados para letras minúsculas e remover as stopwords (palavras comuns do idioma inglês, neste caso).

```
# Tratamento (limpeza, organização e transformação) dos dados coletados
tweetlist_Bol <- sapply(tweets_Bol, function(x) x$getText())
tweetlist_Bol <- iconv(tweetlist_Bol, to = "utf-8", sub="")
```

```

tweetlist_Bol <- limpaTweets(tweetlist_Bol)
tweetcorpus_Bol <- Corpus(VectorSource(tweetlist_Bol))
tweetcorpus_Bol <- tm_map(tweetcorpus_Bol, removePunctuation)
tweetcorpus_Bol <- tm_map(tweetcorpus_Bol, content_transformer(tolower))
tweetcorpus_Bol <- tm_map(tweetcorpus_Bol, function(x)removeWords(x, stopwords()))

```

Nessa etapa, vamos gerar uma nuvem de palavras (wordCloud) para o politico Bolsonaro para verificar a relação entre as palavras que ocorrem com mais frequência.

```

# Gerando uma nuvem palavras
pal2 <- brewer.pal(8,"Dark2")

wordcloud(tweetcorpus_Bol,
  min.freq = 2,
  scale = c(5,1),
  random.color = F,
  max.word = 50,
  random.order = F,
  colors = pal2)

```



Agora vamos proceder com a análise de sentimentos. Construímos uma função (chamada sentimento.score) para verificar cada item do conjunto de dados e compara com as listas de palavras fornecidas, e a partir daí calcular o score de sentimentos, sendo positivo, negativo ou neutro.

```

# Criando uma função para avaliar o sentimento
sentimento.score = function(sentences, pos.words, neg.words, .progress = 'none')
{

  # Criando um array de scores com lapply
  scores = lapply(sentences,
    function(sentence, pos.words, neg.words)
    {

```

```

sentence = gsub("[[:punct:]]", "", sentence)
sentence = gsub("[[:cntrl:]]", "", sentence)
sentence = gsub('\\d+', '', sentence)
tryTolower = function(x)
{
  y = NA

  # Tratamento de Erro
  try_error = tryCatch(tolower(x), error=function(e) e)
  if (!inherits(try_error, "error"))
    y = tolower(x)
  return(y)
}

sentence = sapply(sentence, tryTolower)
word.list = str_split(sentence, "\\s+")
words = unlist(word.list)
pos.matches = match(words, pos.words)
neg.matches = match(words, neg.words)
pos.matches = !is.na(pos.matches)
neg.matches = !is.na(neg.matches)
score = sum(pos.matches) - sum(neg.matches)
return(score)
}, pos.words, neg.words, .progress = .progress )

scores.df = data.frame(text = sentences, score = scores)
return(scores.df)
}

```

Carregando as listas de palavras positivas e negativas (essas listas acompanham este projeto).

```

# Mapeando as palavras positivas e negativas
pos = readLines("data/palavras_positivas.txt")
neg = readLines("data/palavras_negativas.txt")

```

Obtendo o texto dos tweets, carregados anteriormente, e criando dois objetos para armazenar o número de sentenças para cada político bem como para armazenar os textos dos tweets relacionados ao ‘Bolsonaro’.

```

# Obtendo os textos dos tweets
txt_Bol = sapply(tweets_Bol, function(x) x$getText())

# Vetor com o número de tweets para Bolsonaro
number_Tweet = length(txt_Bol)

```

Nessa etapa, usamos a função (sentimento.score) para calcular o score de sentimento para o Bolsonaro. Depois criamos mais três novas variáveis, uma com o nome ‘Bolsonaro’, outra com score positivo e outra com score negativo. Finalmente, visualizamos as primeiras linhas da tabela de dados atualizada.

```

# Aplicando função para calcular o score de sentimento
scores = sentimento.score(txt_Bol, pos, neg, .progress = 'text')

```

```

## | |

# Calculando o score para Bolsonaro
scores$txt_Bol = factor(rep("Bolsonaro", number_Tweet))
scores$muito.pos = as.numeric(scores$score >= 1)
scores$muito.neg = as.numeric(scores$score <= -1)

```

```
head(scores)
```

```
##
## 1 RT @mwillersqc: Very proud to be a part of the All Rise group and supporter of @ThePlanetVS. Let
## 2 @jairbolsonaro https://t.co/Zs7oLFmFot BRAZIL LOVES BOLSONARO! @ValdoCruz @octavio_guedes @JoeBi
## 3 important clarification as it seems MTL is causing a lot of misunderstanding: \n\nthis says "fora
## 4 @JWillows1997 @AndyGraham22 Yup, the cities are backing the races. Bolsonaro
## 5 RT @michaelreid52: Latin America was a leader in the fight against climate change until recently,
## 6 @MarcosQuezado1 https://t.co/Zs7oLFmFot BRAZIL LOVES BOLSONARO! @ValdoCruz @octavio_guedes @JoeBi
## score txt_Bol muito.pos muito.neg
## 1 2 Bolsonaro 1 0
## 2 1 Bolsonaro 1 0
## 3 0 Bolsonaro 0 0
## 4 0 Bolsonaro 0 0
## 5 0 Bolsonaro 0 0
## 6 1 Bolsonaro 1 0
```

O número total de score positivos e negativos são calculados para estimar o `global_score`, que representa a proporção de scores positivos em relação ao total. Em outras palavras, a porcentagem de sentimentos positivos com relação a Bolsonaro.

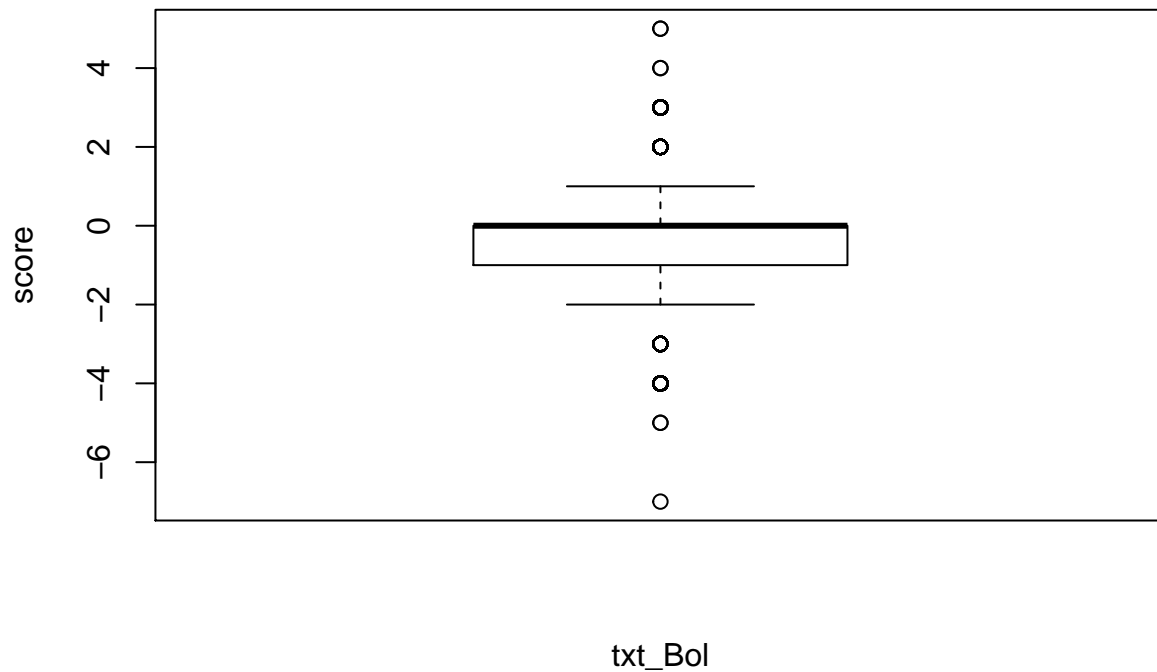
```
# Calculando o total
numpos = sum(scores$muito.pos)
numneg = sum(scores$muito.neg)

# Calculando a pontuação global
global_score = round( 100 * numpos / (numpos + numneg) )
global_score
```

```
## [1] 37
```

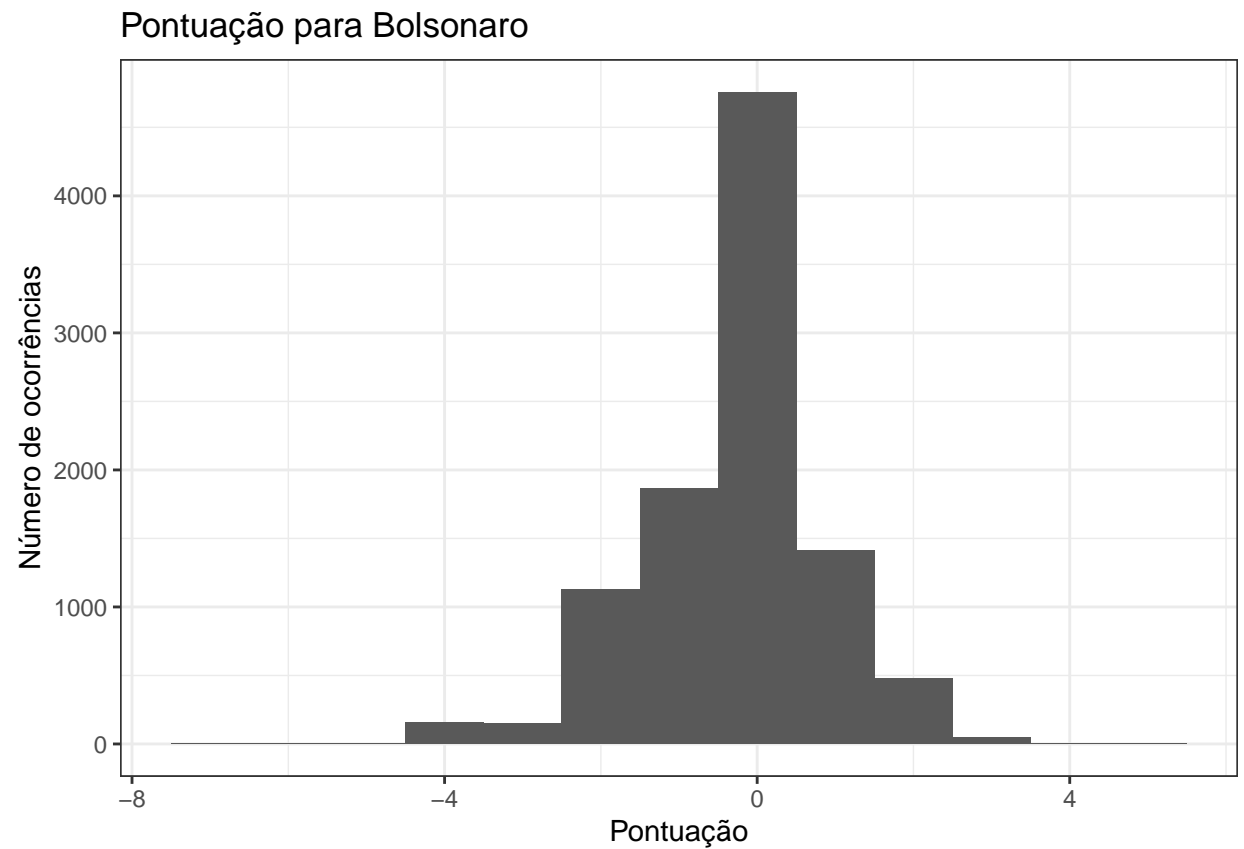
Gerando um boxplot para visualização do score do Bolsonaro. Por essa imagem, observa-se que a mediana do score encontra-se na região da neutralidade.

```
# Gerando um boxplot
boxplot(score ~ txt_Bol, data = scores)
```



Um histograma com a distribuição do score do Bolsonaro também foi gerado para facilitar a visualização dos resultados.

```
# Gerando um histograma
ggplot(data= scores) +
  geom_histogram(aes(x=score), binwidth = 1) +
  xlab("Pontuação") +
  ylab("Número de ocorrências") +
  ggtitle("Pontuação para Bolsonaro") +
  theme_bw()
```



Finalmente, a média do score do Bolsonaro foi calculada.

```
meansSentimentPolitician <- data.frame(MeanBolsonaro = mean(scores[scores[,3] == 'Bolsonaro',]$score))
meansSentimentPolitician
```

```
## MeanBolsonaro
## 1 -0.2678
```