

Narendra_Modi_twitter_sentiment.R

jas

Mon Sep 12 16:20:24 2016

```
rm(list=ls())
library(twitterR)
library(ROAuth)
library(RCurl)

## Loading required package: bitops

library(stringr)
library(tm)

## Loading required package: NLP

library(ggmap)

## Loading required package: ggplot2

##
## Attaching package: 'ggplot2'

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

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:twitterR':
##
##      id, location

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

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

library(plyr)

## -----
```

```

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first,
## then dplyr:
## library(plyr); library(dplyr)

## -----

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##      arrange, count, desc, failwith, id, mutate, rename, summarise,
##      summarize

## The following object is masked from 'package:twitterR':
##
##      id

library(wordcloud)

## Loading required package: RColorBrewer

library(httr)

##
## Attaching package: 'httr'

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

#Setting the working directory
getwd()

## [1] "D:/R/Project/Twitter - Sentiment analysis"

drive <- "D:/R/Project/Twitter - Sentiment analysis"
setwd(drive)

# Set API Keys
api_key <- "XjyXrCI8hgnvC3b6qbDbLeUYr"
api_secret <- "y8zf4iadjQbNlXiUh6HrwJtH9UMXQsIuiiNavyJewpTouCpFaG"
access_token <- "774697724511940609-mdu0odwgHkT9vkw7BprRGiqHnI7Xc0p"
access_token_secret <- "3Ir1UcxoZNbMMt5HlGW3s1zsP0o1xT76WStqeDgxm6jA2"
setup_twitter_oauth(api_key, api_secret, access_token, access_token_secret)

## [1] "Using direct authentication"

# Getting latest tweets on Narendra Modi - Prime minister of india
modi_tweets <- searchTwitter('Narendra+Modi', n=3000)

# Loop over tweets and extract text

```

```

tweet_feed = laply(modi_tweets, function(t) t$getText())

# Dictionary to separate out positive and negative words trait
dir(drive)

## [1] "Hillary_tweets.csv"
## [2] "Iphone 7 twitter sentiment1.R"
## [3] "Narendra Modi twitter sentiment.R"
## [4] "Narendra Modi_tweets.csv"
## [5] "Narendra Modi_tweets1.csv"
## [6] "Narendra Modi_tweets2.csv"
## [7] "Narendra_Modi_twitter_sentiment.docx"
## [8] "Narendra_Modi_twitter_sentiment.R"
## [9] "Narendra_Modi_twitter_sentiment.spin.R"
## [10] "Narendra_Modi_twitter_sentiment.spin.Rmd"
## [11] "negative-words.txt"
## [12] "positive-words.txt"
## [13] "Rplot.pdf"

good = scan('positive-words.txt',what='character', comment.char=';')
bad = scan('negative-words.txt',what='character', comment.char=';')

# Add a few twitter-specific negative phrases
bad_text = c(bad, 'wtf', 'epicfail', 'douchebag', 'idiot')
good_text = c(good, 'upgrade', ':)', '#iVoted', 'voted')

#scoring the tweet texts based on how many "good" and "bad" words show up
score.sentiment = function(sentences, good_text, bad_text, .progress='none')
{
  scores = laply(sentences, function(sentence, good_text, bad_text) {

    # clean up sentences with R's regex-driven global substitute, gsub():
    sentence = gsub('[:punct:]', '', sentence)
    sentence = gsub('[:cntrl:]', '', sentence)
    sentence = gsub('\\d+', '', sentence)

    #to remove emojis
    sentence <- iconv(sentence, 'UTF-8', 'ASCII')
    sentence = tolower(sentence)

    # split into words. str_split is in the stringr package
    word.list = str_split(sentence, '\\s+')

    # sometimes a list() is one level of hierarchy too much
    words = unlist(word.list)

    # compare our words to the dictionaries of positive & negative terms
    pos.matches = match(words, good_text)
    neg.matches = match(words, bad_text)
  })
}

```

```

# match() returns the position of the matched term or NA
# we just want a TRUE/FALSE:
pos.matches = !is.na(pos.matches)
neg.matches = !is.na(neg.matches)

# and conveniently enough, TRUE/FALSE will be treated as 1/0 by sum():
score = sum(pos.matches) - sum(neg.matches)

return(score)
}, good_text, bad_text, .progress=.progress )

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

```

```

# Call the function and return a data frame
df <- score.sentiment(tweet_feed, good_text, bad_text, .progress='text')

```

```

##
|
|
|
|
|=
|
|=
|
==
|
==
|
==
|
===
|
===
|
====
|
====
|
====
|
=====
|
=====
|
=====
|
=====

```

	0%
	1%
	1%
	2%
	2%
	3%
	4%
	4%
	5%
	5%
	6%
	7%
	7%
	8%

=====		8%
=====		9%
=====		10%
=====		10%
=====		11%
=====		12%
=====		12%
=====		13%
=====		13%
=====		14%
=====		15%
=====		15%
=====		16%
=====		16%
=====		17%
=====		18%
=====		18%
=====		19%
=====		19%
=====		20%
=====		21%
=====		21%
=====		22%
=====		22%
=====		23%

=====	24%
=====	24%
=====	25%
=====	25%
=====	26%
=====	27%
=====	27%
=====	28%
=====	28%
=====	29%
=====	30%
=====	30%
=====	31%
=====	32%
=====	32%
=====	33%
=====	33%
=====	34%
=====	35%
=====	35%
=====	36%
=====	36%
=====	37%
=====	38%
=====	38%

=====	39%
=====	39%
=====	40%
=====	41%
=====	41%
=====	42%
=====	42%
=====	43%
=====	44%
=====	44%
=====	45%
=====	45%
=====	46%
=====	47%
=====	47%
=====	48%
=====	48%
=====	49%
=====	50%
=====	50%
=====	51%
=====	52%
=====	52%
=====	53%
=====	53%

=====	54%
=====	55%
=====	55%
=====	56%
=====	56%
=====	57%
=====	58%
=====	58%
=====	59%
=====	59%
=====	60%
=====	61%
=====	61%
=====	62%
=====	62%
=====	63%
=====	64%
=====	64%
=====	65%
=====	65%
=====	66%
=====	67%
=====	67%
=====	68%
=====	68%

=====	69%
=====	70%
=====	70%
=====	71%
=====	72%
=====	72%
=====	73%
=====	73%
=====	74%
=====	75%
=====	75%
=====	76%
=====	76%
=====	77%
=====	78%
=====	78%
=====	79%
=====	79%
=====	80%
=====	81%
=====	81%
=====	82%
=====	82%
=====	83%
=====	84%

=====	84%
=====	85%
=====	85%
=====	86%
=====	87%
=====	87%
=====	88%
=====	88%
=====	89%
=====	90%
=====	90%
=====	91%
=====	92%
=====	92%
=====	93%
=====	93%
=====	94%
=====	95%
=====	95%
=====	96%
=====	96%
=====	97%
=====	98%
=====	98%
=====	99%

```

|
|=====| 99%
|
|=====| 100%

df$name <- "Narendra Modi"
write.csv(df, 'Narendra Modi_tweets.csv', row.names = FALSE)

# Cut the text, just gets in the way
names(df)

## [1] "score" "text" "name"

plotdat <- df[c("name", "score")]
write.csv(plotdat, 'Narendra Modi_tweets1.csv', row.names = FALSE)

# Remove neutral values of 0
str(plotdat)

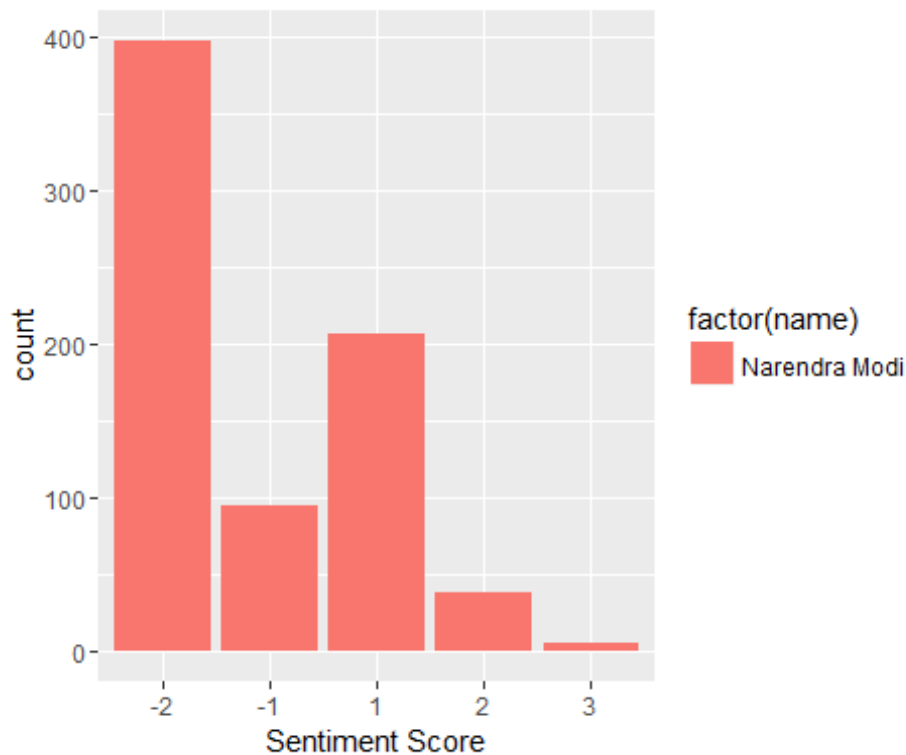
## 'data.frame': 3000 obs. of 2 variables:
## $ name : chr "Narendra Modi" "Narendra Modi" "Narendra Modi" "Narendra
Modi" ...
## $ score: int 0 0 0 0 0 0 0 0 0 0 ...

plotdat1 <- plotdat[!plotdat$score == 0, ]

# Remove anything less than -3 or greater than 3
plotdat1 <- plotdat1[!plotdat1$score > 3, ]
plotdat1 <- plotdat1[!plotdat1$score < (-3), ]

# Nice little quick plot
library(ggplot2)
qplot(factor(score), data=plotdat1, geom="bar",
       fill=factor(name),
       xlab = "Sentiment Score")

```



frequency of negative + positive comments histogram

#WordCloud
`library(tm)`

Create corpus

```
df1 <- read.csv('Narendra Modi_tweets.csv')  
feed1 = unlist(df1$text)  
write.csv(feed1, 'Narendra Modi_tweets2.csv', row.names = FALSE)
```

```
corpus=Corpus(VectorSource(feed1))
```

Convert to Lower-case
`corpus=tm_map(corpus, tolower)`

Remove stopwords
`corpus=tm_map(corpus, function(x) removeWords(x, stopwords()))`

convert corpus to a Plain Text Document
`corpus=tm_map(corpus, PlainTextDocument)`

```
set.seed(1234)  
col=brewer.pal(6, "Dark2")
```

```

wordcloud(corpus, min.freq=25, scale=c(2,1),rot.per = 0.25,
          random.color=T, max.word=200, random.order=F,colors=col)

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : @economictimes could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : https://t.co could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : khauratna could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : @mediacooks: could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : #rio2016 could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : <u+0939><u+0948> could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : baloch could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : malik could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : modi's could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : https://t.co/dgr96r0jox could not be fit on page. It will not be
## plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : @adityarajkaul: could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : @economictimes: could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : official could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : page could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : rights could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : watch could not be fit on page. It will not be plotted.

## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =
## 0.25, : .@narendramodi could not be fit on page. It will not be plotted.

```

```
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : <u+092f><u+0939> could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : https://t.co/tyzifn2jnt could not be fit on page. It will not be  
## plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : modi. could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : kapil sharma could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : watch: could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : @omthanvi: could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : <u+091c><u+0901><u+091a><u+0924><u+0940> could not be fit on page.  
## It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : <u+0928><u+0939><u+0940><u+0902>, could not be fit on page. It  
will  
## not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : <u+092c><u+093e><u+092c><u+093e>! could not be fit on page. It  
will  
## not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : <u+092d><u+093e><u+0937><u+093e>? could not be fit on page. It  
will  
## not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : <u+0930><u+093e><u+092e> could not be fit on page. It will not be  
## plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : <u+0930><u+0947>! could not be fit on page. It will not be  
plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : https://t.co/pxgjnqryll could not be fit on page. It will not be  
## plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : @narendramodi_in: could not be fit on page. It will not be  
plotted.
```

```
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : alleged could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : @narendramodi177: could not be fit on page. It will not be  
plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : https://t.co/3kcgcr2edn could not be fit on page. It will not be  
## plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : https://t.co/y6isuccrxf could not be fit on page. It will not be  
## plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : powerful could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : video could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : army could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : https://t.co/ could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : human could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : protests could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : interesting could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : karachi could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : organisation could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : times could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : <u+0915><u+094b> could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : impact could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : reacted could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : @fatima_baluch could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : @theindiantalks: could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : article could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : thinking could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : @kirenrijiju: could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : winning could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : back could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : shah could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : <u+0915><u+0947> could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : govt could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : never could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## = 0.25, : #transformingindia could not be fit on page. It will not be  
## plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : @arvindkejriwal could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : <u+0914><u+0930> could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : <u+0915><u+0940> could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : <u+0938><u+0947> could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : animal could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : karnataka could not be fit on page. It will not be plotted.
```



```
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : may could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : won could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : #narendramodi could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : https://t. could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : development could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : new could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(corpus, min.freq = 25, scale = c(2, 1), rot.per =  
## 0.25, : taken could not be fit on page. It will not be plotted.
```

