MATH 381: Project 2

Eunji Lee / Yuan Qu / Arnav Dubey

November 25, 2017

```
library(markovchain)
```

```
## Package:  markovchain
## Version:  0.6.9.8-1
## Date:     2017-08-15
## BugReport: http://github.com/spedygiorgio/markovchain/issues
```

```
library(SentimentAnalysis)
```

```
##
## Attaching package: 'SentimentAnalysis'
```

```
## The following object is masked from 'package:base':
##
##     write
```

```
dat = readLines("shape_of_you.txt") # read data from txt file.

process_line <- function(line){
   line <- trimws(line) # removes leading and trailing whitespaces.
   line <- tolower(line) # makes all words lower case for standardization.
}

v_process_line <- Vectorize(process_line) # vectorizes the funciton.

sentences <- v_process_line(dat) # applices the funciton to the list.

sentences <- sentences[sentences != ""] # removes empty strings.

# prints the head the different between pairs of initial and final states can be noticed.
head(sentences)
```

```
##                                     The club isn't the best place to find a lover
##                                    "the club isn't the best place to find a lover"
##                                                                So the bar is where I go
##                                                               "so the bar is where i go"
##   Me and my friends at the table doing shots  Drinking fast and then we talk slow
## "me and my friends at the table doing shots  drinking fast and then we talk slow"
##                                   Come over and start up a conversation with just me
##                                  "come over and start up a conversation with just me"
##                                                 And trust me I'll give it a chance now
##                                                "and trust me i'll give it a chance now"
##                                  Take my hand, stop, put Van the Man on the jukebox
##                                  "take my hand, stop, put van the man on the jukebox"
```

```
# Analyze sentiment for each sentence of the song.
sentiment <- analyzeSentiment(sentences)

# extract dictionary-based sentiment according to the QDAP dictionary.
sentiments <- sentiment$SentimentQDAP
```

```r
# converts continuous variable (sentiment scores) to discrete variable.
make_discrete <- function(x) if(x < 0) x <- -1 else if(x > 0) x <- 1 else x <- 0

# vectorizes the funciton
v_make_discrete <- Vectorize(make_discrete)

# coverts sentiments to discrete type
discrete_sentiments <- v_make_discrete(sentiments)

discrete_sentiments
```

```
##  [1]  1  0  1  1  0  0  1  1  1  1 -1  0  0  1  1  1  1 -1  1  0  0  0  1
## [24]  0  1  0  1  0  1  0  1  0  0  1  0  0  0  0  1  1  1  1 -1  0  0  1
## [47]  1  1  1 -1  1  0  0  0  1  0  1  0  1  0  1  0  1  1  1  1  1  1  1
## [70]  1  1  1  1 -1  1  0  0  0  1  1  1  1  1  1  1  1  1  1  0  1
```

```r
# uses markovchain package in R to calculate the matrix probabilites
mcFit <- markovchainFit(data = discrete_sentiments, byrow = FALSE)
mcFit
```

```
## $estimate
## MLE Fit
##  A  3 - dimensional discrete Markov Chain defined by the following states:
##  -1, 0, 1
##  The transition matrix  (by rows)  is defined as follows:
##     -1         0          1
## -1 0.0 0.0000000 0.09433962
## 0  0.4 0.4193548 0.30188679
## 1  0.6 0.5806452 0.60377358
##
##
## $standardError
##            -1         0         1
## -1 0.00000000 0.2828427 0.3464102
## 0  0.00000000 0.1163081 0.1368594
## 1  0.04218996 0.0754717 0.1067331
##
## $confidenceLevel
## [1] 0.95
##
## $lowerEndpointMatrix
##            -1         0          1
## -1 0.00000000 0.0000000 0.03020599
## 0  0.00000000 0.2280450 0.35553152
## 1  0.02494331 0.1777469 0.42821326
##
## $upperEndpointMatrix
##           -1         0         1
## -1 0.0000000 0.8652349 1.0000000
## 0  0.0000000 0.6106646 0.8057588
## 1  0.1637359 0.4260267 0.7793339
##
## $logLikelihood
## [1] -Inf
```

```r
# derive the estimated probabilites fromt the fitted model.
mcSentiment <- mcFit$estimate
```

```r
# gives names to the transition states.
names(mcSentiment) <- c("negative", "neutral", "positive")

# gives a name to the mocel.
name(mcSentiment) <- "Sentiment"
```

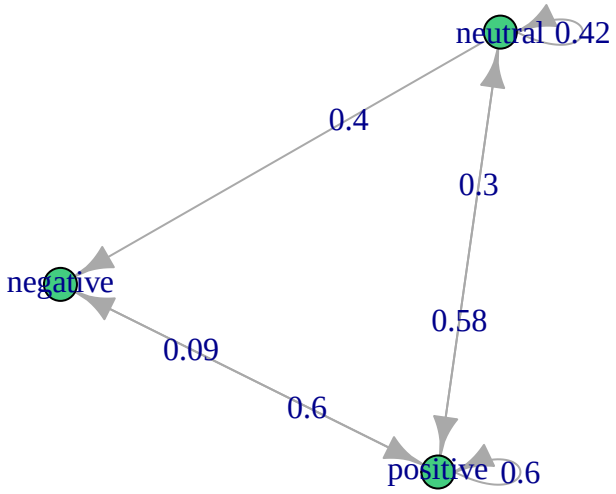
```r
# derives the probablited of negative word after a positive word.
transitionProbability(mcSentiment, "positive", "negative")
```

```
## [1] 0.6
```

```r
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 3.4.2
```

```r
# png(filename="graph.png")
plt <- plot(mcSentiment)
```




```r
plt
```

```
## NULL
```

References:

https://cran.r-project.org/web/packages/markovchain/vignettes/an_introduction_to_markovchain_

package.pdf

https://cran.r-project.org/web/packages/SentimentAnalysis/SentimentAnalysis.pdf