Opinion Mining on PM's "Man Ki Baat" using R Programming

Himani Tyagi[#1], Puneet Bhardwaj[*2]
[#]*assistant professor ,Computer Science and Engineering,MDU*
BSAITM,Faridabad,India

Abstract- *The web is a repository that holds valuable, vast and unstructured information about public opinion toward social events, political events, company strategies, marketing campaigns, and product preferences. Many emerging social, geopolitical, and business-related research questions can be answered by analyzing thousands, of responses and reactions (positive, negative ,angry) expressed in various blogs (such as the blogosphere), social media and social networking sites (like YouTube, Facebook), virtual worlds such as tweets (Twitter). The datum generated every day is of high value and various meanings.*

Keywords — *opinion, mining, unstructured, virtual world, blogosphere*

Introduction

Opinion mining is a very interesting and popular technique for extracting people's reaction and responses about a particular incident, product. This technique has application in nearly every field. This paper represents one of the application of oponion mining by analyzing the sentiments of people on PM Narendra Modi's mass interaction program "MAN KI BAAT" through his recent Facebook post using R programming. Reactions are extracted on the post using two methods in this paper. And the results from these methods are compared to obtain the accuracy. The accuracy of result (in terms of percentage of reactions) is also shown.

Technology

R is free, open-source, cross-platform programming environment. It is a very powerful tool for statistical and graphical representation of data. Dislike all other programming languages, R was designed particularly for statistical analysis, which makes it highly suitable for data scientists for analysis.[4]The tools available for carrying out text analysis or data analysis in R make it easy to perform powerful, accurate text analytics using very few simple commands. The key to R's sudden popularity and growth has been its highly populated collection of extension software libraries, known as packages in R's terminology, maintained by R's user community.

Packages

Power of R language can be enhanced by packages which is created by user generally developed using C, C++ and java. For specific statistical method, graphical plots (ggplots), Import/ Export abilities. CRAN is the official repository and it is a network of ftp and web servers maintained and supported by the R community around the world. It is coordinated by the R foundation, and for a package to be published here it has to pass several tests to make sure the publishing package is following CRAN policies. For the analysis of data on Facebook a package name R Facebook is used, for data trimming or cleaning tidytext is there, for Json type of data analysis (text) Rjson is used, sentiment analysis for sentimental analysis. The proposed work is carried out using RStudio Desktop. Features of Rstudio utilized are:-

1) IDE was created specifically for R language.
2) Packages with different functionality.
3) Data viewer.
4) Integrated R support and documentation.

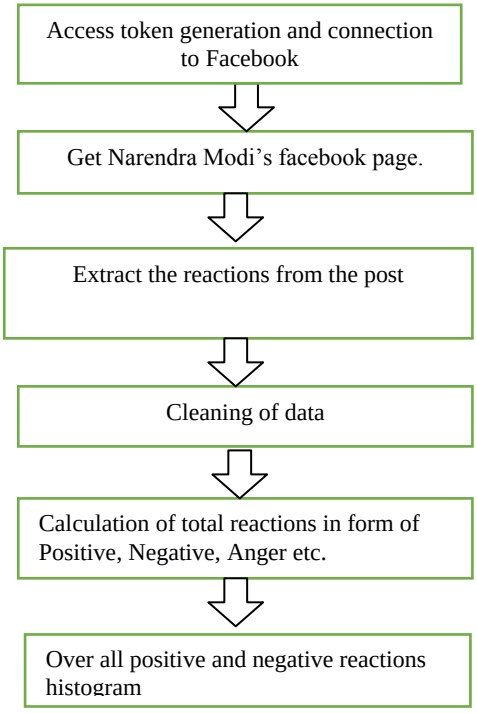


Fig1:process of proposed methodology

Methodology

Finding the access token to connect R to Facebook account.

Access token from Facebook

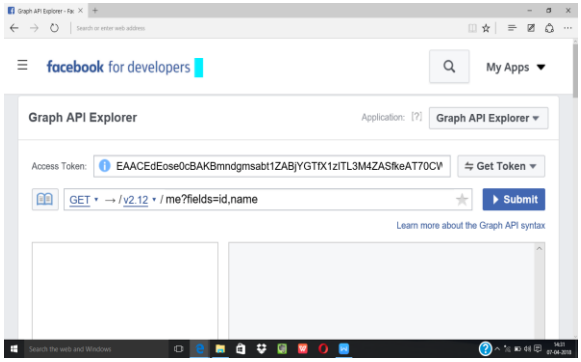


Fig1: showing how to get access token from Facebook

Storing access token in a variable (access token)

access_token<-
EAACEdEose0cBAIXyCEfbabA2aocozzdp42XUiMWlvLTg3Coz8FnPZBMAUfJE1FzGy0ouxewFG6OTscW1bvMC6c16X5j7d6WUmyxD15oqtep9W3YkWyv9z0t3DUo3UMEjfAFJe5NqEp6b2ZCgcoGnaMBdZCX0teLKR8FLBy7npZAtwPi0rZBJ9aZAYHIrzeSwAYFoJDGRpH1pehbAavYTaXZA2P1OyCRmgo34EBcMtrXogZDD

Getting page from Facebook

page<-getPage("narendramodi",token = access_token,n=100)

Getting post from the page

post<-
getPost("177526890164_10160207502330165",token = acess_token,reactions = FALSE,n.comments = 2000,likes = FALSE,api = "v2.12")

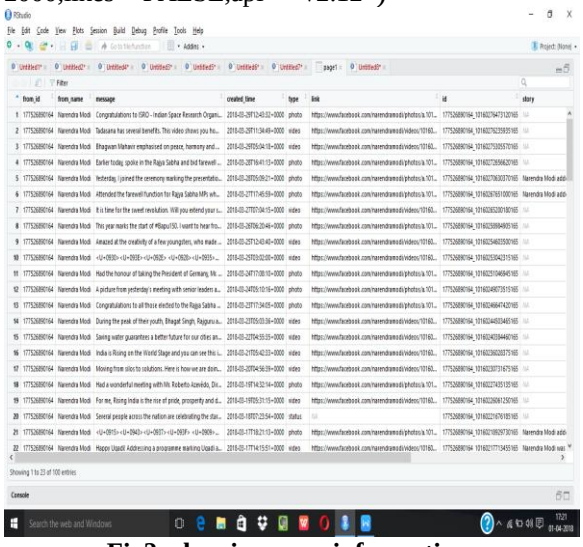


Fig2: showing page information

Getting comment (after removing stickers and other ASCII values character) on post in a text file

comment<-post$comments$message
comment.clean<-iconv(comment,sub = " ",'UTF-8','ASCII')
comment.file<-file("comment.txt")
writeLines(comment.clean,comment.file)

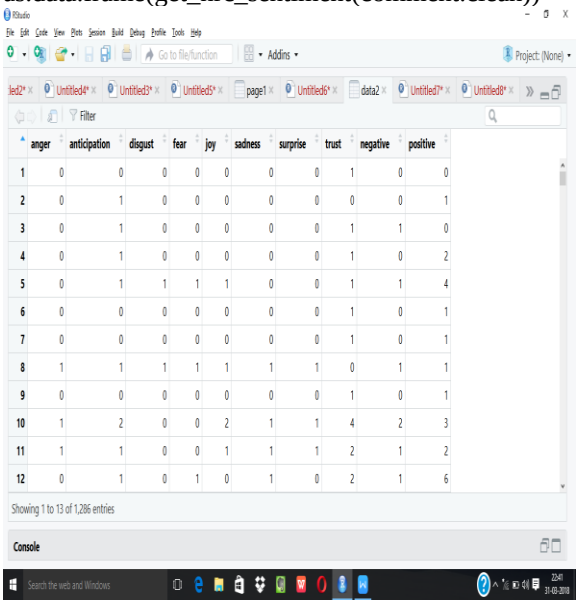
Obtaining reactions on the post

Two methods are proposed for extracting the reactions and performing analysis. On the basis of which it can be shown about a particular opinion that whether it is acceptable or not.

Method 1 for obtaining reactions

In method 1 for obtaining reactions on post, an available package have been used. With the help of available function in the package we have found the reactions on the post

reaction<-
as.data.frame(get_nrc_sentiment(comment.clean))



	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	0	0	0	0	0	0	0	1	0	0
2	0	1	0	0	0	0	0	0	0	1
3	0	1	0	0	0	0	0	1	1	0
4	0	1	0	0	0	0	0	1	0	2
5	0	1	1	1	1	0	0	1	1	4
6	0	0	0	0	0	0	0	1	0	1
7	0	0	0	0	0	0	0	1	0	1
8	1	1	1	1	1	1	1	0	1	1
9	0	0	0	0	0	0	0	1	0	1
10	1	2	0	0	2	1	1	4	2	3
11	1	1	0	0	1	1	1	2	1	2
12	0	1	0	1	0	1	0	2	1	6

Fig3: showing details of reactions

post.positive<-sum(reaction$positive)
post.anticipation<-sum(reaction$anticipation)
post.trust<-sum(reaction$trust)
post.surprise<-sum(reaction$surprise)
post.negative<-sum(reaction$negative)
post.fear<-sum(reaction$fear)
post.anger<-sum(reaction$anger)
post.sad<-sum(reaction$sadness)
post.disgust<-sum(reaction$disgust)

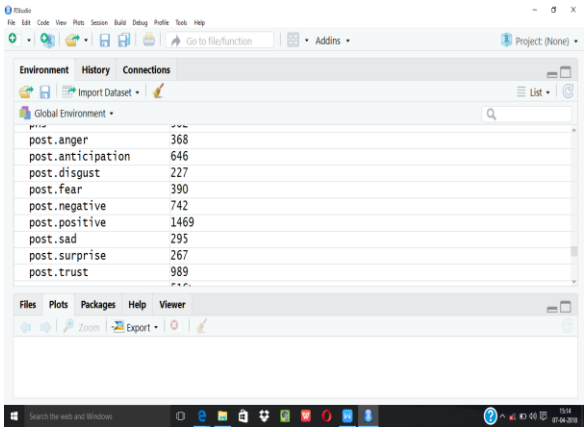


Fig 4: Results after applying above instructions.

Method 2 for extracting reactions on a particular post

Three steps to obtain reactions on post using method 2.

Step 1: to clean comment taken in text file earlier

```
post.text<-scan(choose.files(),what = "char",sep = "\n")
post.text<-tolower(post.text)
post.split<-strsplit(post.text,"\\W+",perl = TRUE)
post.unlist<-unlist(post.split)
post.table<-table(post.unlist)
post.reaction<-as.data.frame(post.table)
```

Step2: to store words of language(english) associated with emotions with their respective sentiment or emotion in a dataframe(let it be temp)

```
temp<-get_sentiments("nrc")
```

Step3: finding the reactions/emotions or sentiment of people by matching comments with temp dataframe and hence obtaing emotions or sentiments

```
v1<-c(1:6515)
v2<-c(1:13901)
i=1
j=1
for (i in v1) {
 j=1
 for (j in v2) {
  if(post.reaction$post.unlist[i]==temp$word[j]){
   post.reaction$sentiment[i]<-temp$sentiment[j]
   break
  }
  else
  {
   if(j==13901)
   {
    post.reaction$sentiment[i]<-"null"
   }
  }
 }
}
sub.positive<-
subset(post.reaction,post.reaction$sentiment=="positive")
sub.negative<-
subset(post.reaction,post.reaction$sentiment=="negative")
sub.trust`<-
subset(post.reaction,post.reaction$sentiment=="trust")
sub.anticipation<-
subset(post.reaction,post.reaction$sentiment=="anticipation")
sub.joy<-
subset(post.reaction,post.reaction$sentiment=="joy")
sub.anger<-
subset(post.reaction,post.reaction$sentiment=="anger")
sub.fear<-
post.reaction,post.reaction$sentiment=="fear")
sub.surprise<-
subset(post.reaction,post.reaction$sentiment=="surprise")
sub.disgust<-
subset(post.reaction,post.reaction$sentiment=="disgust")
sub.sad<-
subset(post.reaction,post.reaction$sentiment=="sadness")
post.2.positive<-sum(sub.positive$Freq)
post.2.negative<-sum(sub.negative$Freq)
post.2.trust<-sum(sub.trust$Freq)
post.2.anticipation<-sum(sub.anticipation1$Freq)
post.2.anger<-sum(sub.anger$Freq)
post.2.fear<-sum(sub.fear$Freq)
post.2.surprise<-sum(sub.surprise$Freq)
post.2.sad<-sum(sub.sad$Freq)
post.2.disgust<-sum(sub.disgust$Freq)
```

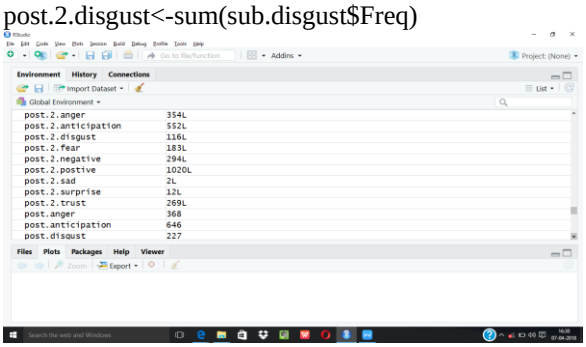


Fig5: Results showing number of reactions on input post

Graphical representation of reactions or sentiments

1.Graphical representation of various sentiments

```
label<-
c("pos","ant","trust","sur","neg","ang","fear","sad","dis")
```

```
values<-c(516,552,269,12,294,354,183,2,116)
color<-
c("yellow","yellow","yellow","yellow","red","red","
red","red","red")
barplot(values,names.arg = label,col=color)
```

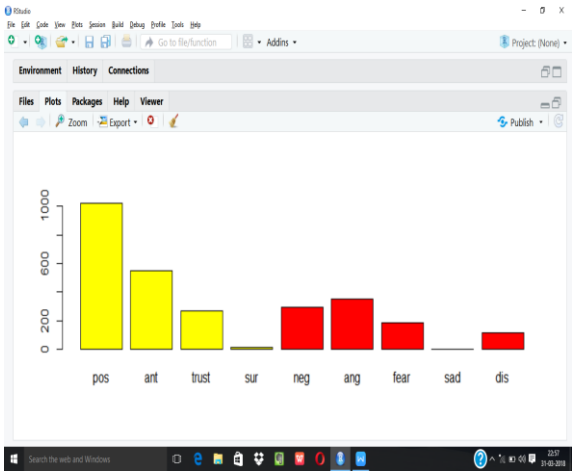


Fig6: Showing different reactions on post

2. Graphical representation of total positive and total negative reactions

```
Label2<-("total positive", "total negative")
Values2<-(1349,943)
Color2<-("yellow", "red")
Barplot(values2,names.arg = label2,col=color2)
```

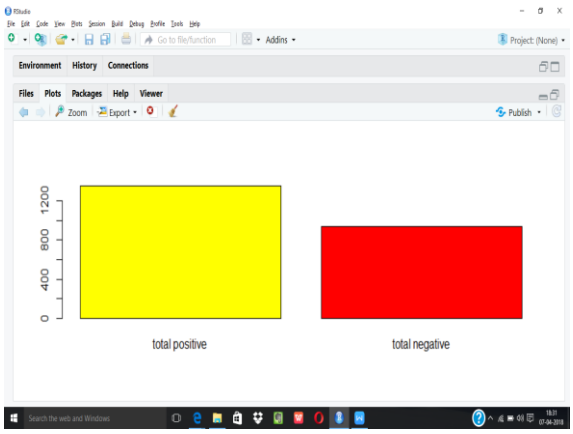


Fig7: Histogram of total positive and negative reactions

Comparison of Methods (Accuracy Calculation)

Results obtained from both the methods are compared and this comparison gives us a idea about accuracy of results. The basis of comparison is percentage of results obtained. The results of comparison suggest that our data is highly accurate. Only a marginalised error of less than 4% in observed.

Table1: Result by method1 and method 2

Sentiments	METHOD 1 Sentiments OD1	METHOD 2
Positive	1469	516
Anticipation	646	552
Trust	989	269
Surprise	267	12
Negative	742	294
Anger	368	354
Fear	390	183
Sad	295	2
Disgust	227	116
Total	5393	2298
Total positive sentiment(postive+anticipation+surprise+trust)	3371 (62.50%)	1349 (58.70%)
Total negative sentiments(negative+anger+fear+sad+disgust)	870 (33.83%)	943 (41.0%)

Conclusion

The sentiment analysis of PM Narendra Modi's mass interaction program "MAN KI BAAT" using his recent Facebook post on the same is done using r programming. People expressed their opinion in large numbers and it is seen that 62% reactions of the people are positive and only 33 % are negative reactions. So it can be concluded that prime minister program to interact people through "MAN KI BAAT" is a successful program and creating a space in people's mind. This program can be used to create awareness about social and economic problems face by people. Hence it can be advised or concluded that this program "MAN KI BAAT" should be continued to have interaction with people in masses. The paper represents an algorithm for converting data into a useful form and extracting useful information that suits our requirements.

References

[1] Xing Fang, Justin Zhan, "Sentiment analysis using product review data, Journal of Big Data, 16 June 2015.

[2] Fiaidhi J, Mohammed O, Mohammed S, Fong S, Kim TH. Opinion mining over Twitter space: Classifying tweets programmatically using the R Approach. IEEE. 978-1-4673-2430-4/12.

[3] T. R. Foundation, "R: What is R?,". [Online]. Available: https://www.r-project.org/about.html

[4] Pooja Khanna , Sachin Kumar , Sumita Mishra, Anant Sinha "sentiment analysis: an approach to opinion mining from Twitter data using R",IJARCS, Volume 8, No. 8,September-October 2017.

[5] Blog on term frequency and inverse document frequency: http://www.tfidf.com.

[6] Xin Chen, Student Member, IEEE, Mihaela Vorvoreanu, and Krishna Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences",IEEE transactions on learning technologies,Vol. 7,No.3,July-September 2014.

[7] Y.Sandeep et al, "Methodological Study of Opinion Retrieval Techniques for Twitter Social Network",

International Conference on Circuit, Power and Computing Technologies [ICCPCT],2015.

[8] FissehaBerhane, "Sentiment Analysis on Donald Trump using R and Tableau ", R news and tutorials contributed by (600) R bloggers , 2 January 2016.

[9] Shruti Kohli, Himani Singal, "Data Analysis with R", IEEE/ACM 7th International Conference on Utility and Cloud Computing,2014.

[10] Article on Twitter Sentiment Analysis training dataset:http://thinknook.com/twitter-sentiment-analysistraining-corpus-dataset-2012-09-22.

[11] G.Vinodhini and R.Chandrasekaran 'sentiment Analysis and opinion minning: A survey', International Journal vol.2, 6,2012.