Opinion Mining on PM’s “Man Ki Baat” using R Programming

Himani Tyagi¹, Puneet Bhardwaj²
¹assistant 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 opinion 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. Unlike 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.

Fig1: process of proposed methodology

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

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. Access token generation and connection to Facebook
2. Get Narendra Modi’s facebook page.
3. Extract the reactions from the post
4. Cleaning of data
5. Calculation of total reactions in form of Positive, Negative, Anger etc.
6. Over all positive and negative reactions histogram

Fig1: process of proposed methodology
Methodology

Finding the access token to connect R to Facebook account.

**Access token from Facebook**

Storing access token in a variable (access token)

```r
access_token <- EAACEe0e0cBAIXyCEfbabA2aocozzdpl42XUiMWl6LTg3C0oz8FnPZBMAUUJE1FzGybouxewFG6OTscW1bVM6c16XS7d6WUnyxDr5oqtep9W3YkwYy9z3D0u3UMEjAFe5NqEp6bZCgcoGnaMBdZCX0leLR8FLb7npZAtwP6BCJ9aZAYHIrzeSwAYFoJDCpH1pebAvYTaXZAZ2P10yCRmg034EBCmXxOGZD
```

Getting page from Facebook

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

Getting post from the page

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

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

```r
comment <- post$comments$message
comment.clean <- iconv(comment, sub = " ", "UTF-8", "ASCII")
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 has been used. With the help of available function in the package we have found the reactions on the post

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

### Obtaining reactions on the post

- 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$anger)
- post.sadness <- sum(reaction$sadness)
- post.disgust <- sum(reaction$disgust)
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

```r
post.text <- scan(choose.files(), what = "char", sep = "in")
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)
```

Step 2: to store words of language(english) associated with emotions with their respective sentiment or emotion in a dataframe(let it be temp)

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

Step 3: finding the reactions/emotions or sentiment of people by matching comments with temp dataframe and hence obtaining emotions or sentiments

```r
v1 <- c(1:6515)
v2 <- c(1:13901)
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"
      }
    }
  }
}
```

```r
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 <- subset(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$Freq)
post.2.anticipation <- sum(sub.anticipation$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)
```
values <- c(516, 552, 269, 12, 294, 354, 183, 2, 116)
colors <- c("yellow", "yellow", "yellow", "yellow", "red", "red", "red", "red")
barplot(values, names.arg = label, col = colors)

Fig6: Showing different reactions on post

2. Graphical representation of total positive and total negative reactions

Label2 <- c("total positive", "total negative")
Values2 <- c(1349, 943)
Color2 <- c("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 an 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% is observed.

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

Table1: Result by method 1 and method 2

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


ISSN: 2231-2803 http://www.ijcttjournal.org Page 71
International Conference on Circuit, Power and Computing Technologies [ICCPCT], 2015.


