

Social Media Sentimental Analysis During Mumbai Flood

S. Gomathi¹, J. Pooja², K.J. Sowbarnika³

¹Assistant Professor, Department of Computer Science(PG),
PSGR Krishnammal College for Women,
641004 Coimbatore, India
gomathisri@psgrkw.ac.in

Abstract. This paper has been deep-rooted about the analysis of Mumbai Flood. The analysis of the data about the disaster and emotions in Mumbai has been implemented in the form of a chart using R Programming. The reach or awareness of the Mumbai flood among the people around the world is mainly by social media such as Twitter has been presented in this paper. The mentality of the people during the disaster has been analyzed using R-Programming.

Keywords: Flood, Social media, Mumbai, data, R programming, hash tag, emotions, Twitter

1. Introduction

Communication is a source of sharing or exchanging of our thoughts, ideas, information, visuals, message or even signals. Social media is being a good source of communication, it could help reduce pressure, social support is being provided to those who are being affected by any natural disasters such as Floods, wildfires, Earthquake or Tsunami etc. And additionally social Media could provide essential information required in order to deal or cope with natural disasters.

The wide active users in social media, made us to consider as an important source of disaster management and also it's being a close Real-Time communication channels used to share information during the critical times of Natural Disaster and crisis. Social media analytics is the process of gathering and analyzing data from the users of social networks such as Facebook, Instagram, Twitter and more [1]. During disaster there will be shortage of situational information management and no coordination in relief efforts. Using pattern matching methods, all the data gathered from the social media users are been matched with situations during disaster and analysed. By this way social media plays a vital role during disaster to help the needy hands. At the time of disaster a volunteered girl said: *"I know my street, but I don't know what is happening in the rest of multiple streets in my own area. If someone needs help, how will I know and where should I start?"* In this paper sentimental analysis, word count, Text analysis concept are used to analyse the twitter data and represent it using bar plot and pie chart.

2. Literature Overview

Twitter performs a significant part in spreading of facts during natural disasters. The message posted on Twitter.com is known as tweet which accept only 140 characters. Even if most tweets contain texts, URLs, video, pictures, vines and GIFs can be inserted in a tweet. Tweets comprise elements known *hashtags*, which might be words that capture the topic of the tweet or act as tag for the tweet. Hashtags are generally prefixed with the aid of the '#' symbol. They will additionally deliver feelings (#happy, #sad, #emotional) or a happening (#CSKVsMumbaiIndians) or it can be anything. The symbol "@" accompanied through text represents the Usernames or handles of individual who post. Any user can direct a message to some other user by means of including the handle, with the '@' symbol. Retweet ('rt' in brief) is a tweet by a user A that has been shared via user B to all or any of B's followers. Thus, retweet is a method of determining how widespread the tweet is [4].

2.1 Background: Mumbai Flood

Mumbai, India had experienced record-breaking rains and flooding since early July. Beginning on 20 July, the rain intensified again creating more flooding and inflicting thousands of individuals to be stranded and trapped. Heavy rains lashed out in Mumbai. The city struggled with severe traffic jams and water-logging,

whereas the people scrambled to reach their homes [3]. The native train services were stopped, whereas 9 flights were diverted at the city airport. The IMD had forecasted fairly widespread rain in the region. Social Medias are playing a significant role in saving several people`s life.

The below list are the hashtags that was popularly used during the Mumbai Floods,

- #mumbaimonsoon
- #mumbairainlive
- #mumbairains
- #mumbai
- #mumbaitrain

Various hashtags for Mumbai floods are observable. These tags are the evident of the disaster, other than the disaster itself.

To Collect the Twitter messages R Programming was been used. At the time of Disaster, a lot of variety of hashtags were created that are associated with Mumbai Flood. And hence, The hashtags #mumbairains, #mumbairainlive and #mumbai were used to extract related tweets. The collected data was from timeline of tweets that are posted between 05-July-2019 to 05- August-2019. The Dataset consists of 1000 tweets. It provided some Twitter attributes in dataset like tweet ID, favoured, text, replytoSN created favouriteCount, replyTOSID, truncated, screenName, statusSource isRetweet, retweetCount, latitude longitude retweeted, The whole dataset was been downloaded using R and separated the unique tweets from the from the retweets based on the retweet status information by reanalyzing the dataset. However extracted the unique tweet-Id for retweets by processing the Twitter data.

2.2 Implementation

```
//Authentication

library(twitter)
consumer_key<- "qFdpgtn5hhBklJmnmvfiEocVcY"
consumer_secret<"1126p4UPfwg9JmOLP59pENf2qMlaJFuNVobjbiVtHbUnG1d8tHR"
access_token<"1121288243473870848NJhS1S79ybQHDGusgrL2eB5YwUutAE"
access_secret<"yIGSt294Cv8iZCdHeb7xiv57iLIOYWemNoVgItMCkAG5M"

//setup

setup_twitter_oauth(consumer_key,consumer_secret, access_token,
access_secret)
tw<-searchTwitter('#MumbaiRainlive',n = 1000,lang='en')
d <-twListToDF(tw)

//read/write

write.csv(d,"Z:/mumbairains.csv",row.names = FALSE)
```

Thus the dataset gets downloaded in the D folder of the system.

3. Clustering

Clustering is the way of combining a set of items in this type of manner, that items within the identical set known as a cluster and more like to every other than to those in other sets. Its key job is analyzing data removal, and approach for statistical records evaluation, utilized in several areas with bioinformatics, information retrieval, records compression, and computer graphics [5].

3.1 K-Means Clustering

In K-means, the amount of clusters are decided. The rule then generates k file clusters in an approach that confirms the inside cluster distances from every cluster members to the center of mass of the cluster is decreased [5].

4. Text Mining

Text mining is the technique of exploring and analyze huge amount of free data by software that can name concept, patterns, keywords and other attributes in the data [2].

4.1 Text Preparation

The subsequent stages has been completed to clean the data and arrange it for additional analysis.

Eliminating punctuation: Punctuation marks are not needed for the evaluation of text therefore they are eliminated.

Eliminating numbers: To identify each message TweetIDs are generated which are numbers. Numbers are not needed for evaluation of text, therefore they are eliminated.

Eliminating stop words: Stop words are words that are generally used in every single sentence such as 'is', 'but', 'shall', 'by' etc, which have no analytical meaning. These words have been eliminated by matching the body with the stop words list inside the tm package of R. Expletives are additionally eliminated.

Eliminating URLs & links: Several messages are enclosed with hyperlinks to webpages and videos to somewhere else on the Internet. Thus regular expressions are used to remove it.

Stripping whitespace: Words that have additional spaces at the starting, middle or end are directed to a regular expression which eliminates the extra space and holds only the words.

Checking for impure characters: Once the modifications were made the corpus is ben checked. Then some URLs were left, because of the removal of whitespaces, numbers and punctuations. Regular expressions were used to eliminate them.

4.2 Word Frequencies and Associations

After cleaning, word cloud are plotted to compare the most used terms. The subsequent observations are made:

- Few common words repeat more frequently than others: such as 'stay safe', 'food' etc. These are on spot reactions and responses to the issues.
- Some common terms such as 'area', 'sadak', 'flood', 'rain', 'road' etc. that give information about the situation in some regions.
- Words like 'pray for mumbai', 'hopeforbest' are used, in texts that principally deliver hope or sympathy.
- There have been tweets about news reviews covering the crisis, with words like media , channel, etc.

```
csvdata<-read.csv("Z:/mumbaiRain.csv")
```

```
//download library file.
```

```
//tm-text mining
```

```
library(tm)
floddata<-iconv(csvdata$text,to = 'utf-8')
floddata<-Corpus(Vector Source(floddata))
inspect(floddata[1:5])
floddata<-tm_map(floddata, tolower)
inspect(floddata[1:5])
floddata<-tm_map(floddata, remove Punctuation)
inspect(floddata[1:5])
floddata<-tm_map(floddata, remove Numbers)
inspect(floddata[1:5])
cleanset floddata<-tm_map( floddata,removeWords,stopwords('english'))
inspect(cleanset floddata[1:5])
```


5. Sentimental Analysis

Sentimental Analysis is the method of figuring out whether text is good, bad or neutral. It's additionally called opinion mining or sentiment mining.

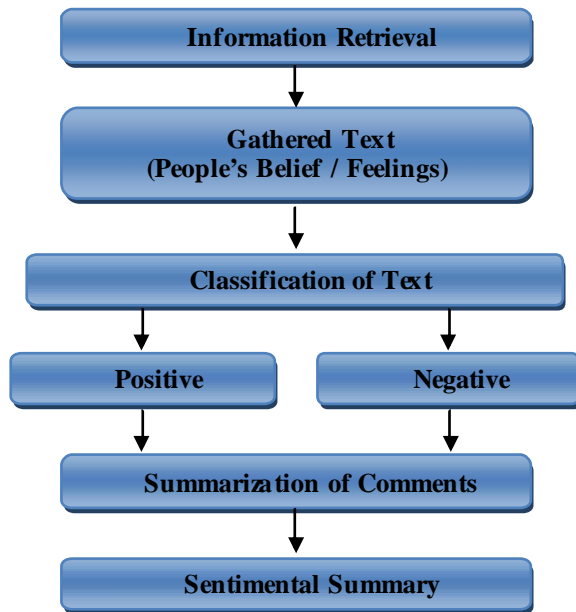


Fig 3. Flow Chart 1 –Process Involved

```
fld<-read.csv("Z:/mumbaiRain.csv")
library(syuzhet)
library(lubridate)
library(ggplot2)
library(scales)
library(reshape2)
library(dplyr)
tweets<- iconv(vj$text, to = 'utf-8')

//Obtain sentiment scores
//nrc-non-resource commercial

s <- get_nrc_sentiment(tweets)
head(s)
tweets[4]
get_nrc_sentiment('delay')

//Bar plot

barplot(colSums(s),
las = 2,
col = rainbow(10),
ylab = 'Count',
main = 'Sentiment Scores for Mumbai Disaster Tweets')
```

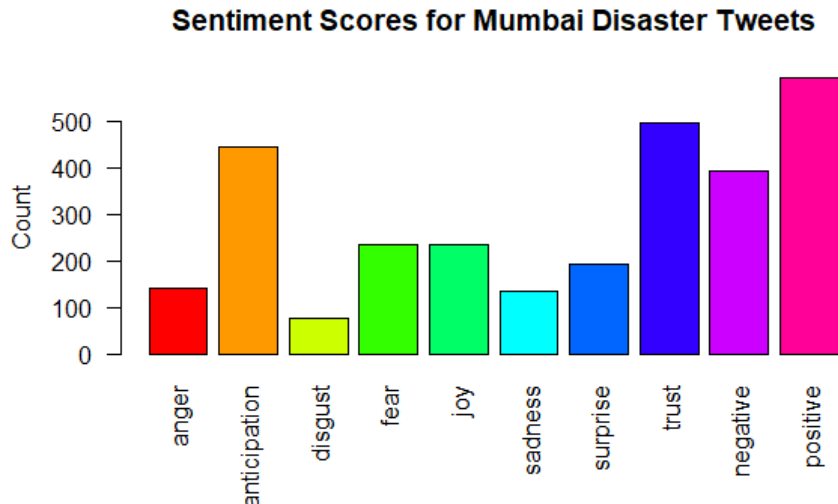


Fig 4. Chart 2 – Sentiment Score for Mumbai Disaster Tweets

6. Challenges and Problems

The following points are the challenges and problems faced during the research.

- Retweets includes facts that a lot of users find related. The objectivity of this significance to the disaster at hand is difficult, if no longer impossible to measure.
- In the dataset, there is no hashtags for many tweets. These tweets can also be tremendously related to the crisis but because of the absence of hashtags it may be neglected.
- Twitter has character-restriction of all tweets. So, users truncate or brief words to forms that is simply understandable to people, but hard for a machine. E.g., ‘afctedppl’ which clearly means ‘affected people’ for a human, but not for a machine.

7. Result

During disaster there were lots of tweets been posted and retweets thus here is the analysis of the retweet counts of the twitter data.

Table 1. Analysis of the retweet counts of the twitter data

Twitter Messages(tweets)	Absolute Number	Percentage
Total tweets	8192	100%
Retweets	7325	89.41%
Original tweets of Retweets	552	6.73%
Tweets never got retweeted	315	3.84%
Mean Retweet	2730.67	33.3%

As shown in *Table 1* only 6.73% tweets are original tweets and most of our data consists of “retweets” which constitutes 89.41% of the total data gathered. That is because the original tweets got retweeted many times, at an average retweet rate of 33.3.

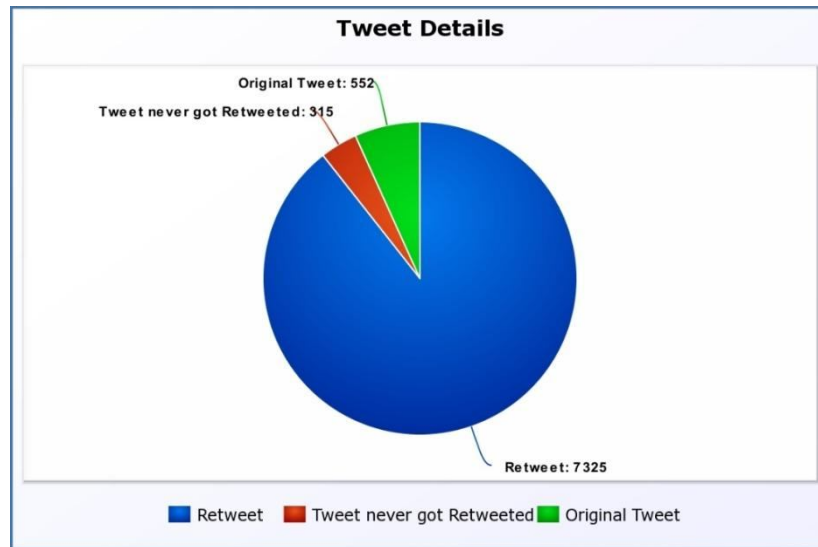


Fig. 5 Chart 3 – Tweet Count

The below bar graph demonstrates the active Twitter users participated in relief efforts. (Note: These evidences are factual only with the reference to the tweets gathered, if all the tweets on the topic were evaluated it may change.)

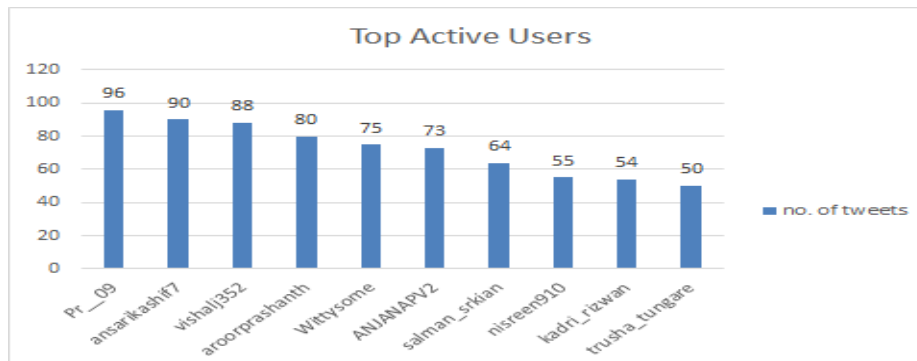


Fig 6. Chart 4 – Most active users during Mumbai flood.

Limitations:

The study didn’t consider pictures, news report, other social media reviews that would have caused extra awareness to the issue and among the whole set of tweets only 1000 tweets were been taken for analysis.

Conclusion:

The system for data analysis over the Twitter public focusing on three different problems Data Collection, Sentimental Analysis, Word Count, Text-analysis has been discussed in this paper. The charts and tables shows the analysis of data based on Mumbai flood data set obtained from Twitter. R programming language is used to visualize the data.

References:

1. Kouloumpis, Efthymios, Theresa Wilson, and Johanna Moore. "Twitter sentiment analysis: The good the bad and the omg!" *Fifth International AAAI conference on weblogs and social media*. 2011.
2. Bala, Myneni Madhu, K. Navya, and P. Shruthilaya. "Text mining on realtime Twitter data for disaster response." *Int. J. Civ. Eng. Technol* 8.8 (2017): 20-29.
3. Pandey, N. and Natarajan, S., 2016, September. How social media can contribute during disaster events? Case study of Chennai floods 2015. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1352-1356). IEEE.
4. Kaur, H. J., & Kumar, R. (2015, May). Sentiment analysis from social media in crisis situations. In *International Conference on Computing, Communication & Automation* (pp. 251-256). IEEE.
5. Sangameswar, M. V., M. Nagabhushana Rao, and S. Satyanarayana. "An algorithm for identification of natural disaster affected area." *Journal of Big Data* 4.1 (2017): 39.