

Emotional Analysis using Twitter Data during Pandemic Situation: COVID-19

Amrita Mathur Purnima Kubde Sonali Vaidya

Department of Information Technology, Department of Information Technology, Department of Information Technology,
St. Francis Institute of Technology, St. Francis Institute of Technology, St. Francis Institute of Technology, Mumbai Mumbai
Mumbai

amritamathur@sfit.ac.in purnimakubde@sfit.ac.in sonalivaidya@sfit.ac.in

Abstract— During the crisis situation caused due to COVID-19 disease, managing mental health and psychological well-being is as important as physical health of people. As web based life is broadly utilized by individuals to communicate their feeling and supposition, our framework utilizes Twitter information posted by individuals during this emergency circumstance to dissect the feelings of individuals. For processing the cleaned data NRC Word-Emotion Association Lexicon (have aka EmoLex) is used. NRC Word-Emotion Association Lexicon is a list of English with real-valued scores of intensity for eight basic emotion words ns (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust). The text content of tweeter dataset created by fetching tweets across the world have classified into basic emotions like anger, anticipation, disgust, fear, joy, sadness, surprise and trust. This analysis can be used by authorities to understand the mental health of the people and can take necessary measures to decide on policies to fight against coronavirus which is affecting the social well-being as well as economy of the whole world.

Keywords—COVID-19;coronavirus; emotions; Twitter; tweets

I. INTRODUCTION

COVID-19, an infectious disease caused by a newly discovered coronavirus which has brought the whole world to halt. Corona viruses are a large family of viruses which may cause illness in animals or humans. At this time, there are no specific vaccines or treatment for COVID-19. In such a situation, the best way to prevent and slow down transmission is to be well informed about the COVID-19. During this pandemic situation managing mental health and psychological well-being is as important as physical health. Understanding public opinion and emotions can help the Ministry of Health and family Affairs to take necessary actions. In this fast growing digital world, social networking services like Twitter, Facebook, Instagram are becoming a very popular medium of

communication and a medium to express oneself. The five key characteristics of social media: collectivity, connectedness, completeness, clarity and collaboration lend itself to be used increasingly to support crisis management function. The twitter data is used because of its accessibility. It is easy to use both for sharing the information as well as collecting it. Twitter provides unprecedented information to our lawmakers as well as the general public. Twitter, a microblogging website is widely used by the public to share information and express their opinion through tweets. It also provides users to express their emotions related to any prevalent topic. Tweets posted by twitter users during crisis situations can be used by media to investigate real-time public opinion. It also helps authorities to gain insight for quickly deciding the assistance policies.[1]

Our method aims at fetching the tweets related to COVID-19 and performing emotional analysis of each tweet posted by the twitter users. It aims at classifying the tweets into positive and negative tweets and further classified into six basic emotions given by Ekman i.e joy, sadness, anger, fear, disgust and surprise[2]. Twitter data is chosen for emotional analysis of people as tweets contain a large number of opinionated text. One of the main characteristics of tweeter which has attracted many researchers is text length limit which can be maximum of 140 characters and relevance of content, this makes the task of classifying tweets is a much easier task than classifying longer documents like blogs [3] .

The rest of this paper is organized as follows. Section II gives an insight of related work in said domain. Section III introduces the dataset. Section IV and V explains the emotional analysis method and experimental results. Lastly paper concludes in Section VI

II. LITERATURE REVIEW

Emotion analysis or sentiment analysis is becoming a wide area of research in different applications. In our proposed work, twitter data related to COVID -19 is analyzed. Many researchers have worked on emotional analysis on data collected from their local microblogging site. Wang et al [4]have performed an emotional analysis method on COVID-19 data collected from Sina Weibo ,a microblogging site in China using Support Vector Machine, naïve Bayes and Random Forest classifiers. Mansur ALP et al [2] classified Turkish Tweets into basic emotions using Deep Neural Network. In our proposed method analysis is based on the data

collected from twitter data across the world. Emotional analysis approach using social media is used by various researchers during natural disasters and for tracking pandemics. Hyo Jin Do et al [1] used emotional analysis technique to understand the behavior of humans and the characteristics of the socio culture system during epidemic outbreak in Korea. Vasileios Lampos et al [5] have used an approach of analyzing the twitter data to monitor the diffusion of pandemic in people. Rashid Kamal et al [6] proposed a crowd source sensing technique using twitter data and word model which can classify the emotions and proved to be more efficient when analysis is done based on location. In many twitter dataset available, location from which tweet is sent is not mentioned. In such situations word models help in accurate classification of emotions. Twitter provides the service to people to share their emotions, Kiichi Tago et al [7] proposed the influence of emotional behavior on user relationships based on Twitter data. Felipe Taliar [8] performed emotional analysis on data collected from facebook and studied facebook reactions classified them into positive, negative and neutral and examined if they are related to Ekman's emotion classification. Mondher Bouazizi et al [9] proposed a multiclass emotion classification using feature extraction. In their research they have used a tool called SENTA. Harshvardhan Achrekar et al [10] proposed a method to predict influenza epidemic situations in the real world using Twitter data.

III. PROPOSED SYSTEM

Our proposed system aims to fetch tweets from dataset and after applying preprocessing classify into positive and negative tweets and further classify into six emotions using lexical-oriented method using R. Following is the flow of our process:

A. COVID-19 pandemic

The 2019-20 Coronavirus pandemic is an ongoing pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It is declared as a Public Health Emergency of International concern by the World Health Organization. COVID-19 is reported in 185 countries and territories among which India is also one of the countries which are badly affected by this deadly disease. As per data collected from the official website of Ministry of Health and Family Welfare, statistics of COVID-19 in India is presented in Fig 1. It shows the increase in the number of active cases (black), cured (green), deaths (red) and migrated patients between 22nd January to 15th April 2020 in India. As per the news reported by WHO, there is a large increase in the number of stroke patients during this pandemic situation. It is very important to manage mental health and psychological well-being along with physical health.

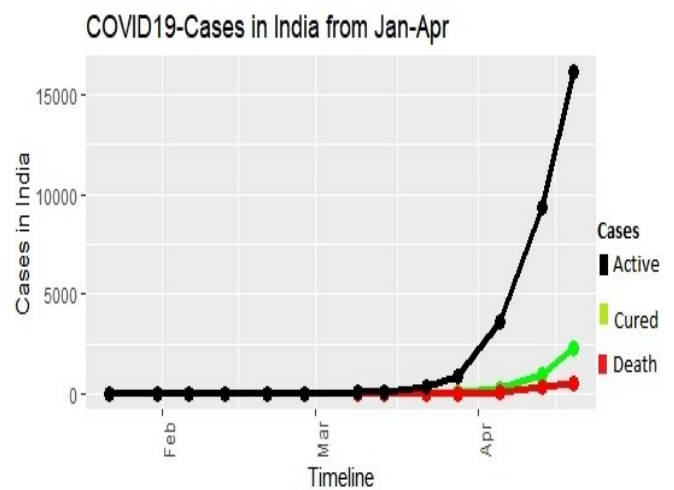


Fig.1 COVID-19 cases in India

B. Data Source

- Tweets can be fetched in Real Time also by using TwitterR used in R which provides access to Twitter API. For experimentation purpose, Dataset used includes Tweets related to COVID-19 between 22nd January 2020 and 15th April 2020 used for classification is downloaded from TweetBinder website [11]. Data collected consists of around 30,000 tweets in English language from all over the world.

IV. EMOTIONAL ANALYSIS METHOD

Two common approaches used for mining of text emotions are:

C. Lexical Oriented Method/Lexicon:

These methods consider a word or collection of words to make inferences about the feelings. Such approaches are described as keyword based approaches. In this method emotion detection on related keywords are used. Keywords are matched with predefined emotional keywords.

D. Machine Learning based Method:

This method uses Machine Learning algorithms, use training data to build models for prediction of emotion conveyed in data.

In this paper, public emotions about COVID-19 is analyzed by exploring tweets using lexical oriented method. The flow of our method is given in Fig.2. Text content of twitter dataset collected from web resource [11] is cleaned using various preprocessing commands. The cleaned data is further analyzed to classify in different emotions given by Ekman.

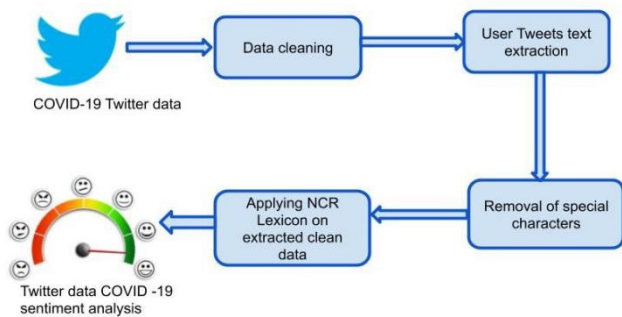


Fig.2 Emotional analysis with the help of COVID-19 Twitter data

For our experimentation reason, R measurable programming is utilized due to its thickly populated assortment of expansion programming libraries, known as bundles. One of the primary focal points of performing content examination in R is that it is frequently conceivable and generally simple to switch between various bundles or to consolidate them. The following steps were used:

- **Data Collection:** There are many ways of data collection for sentiment analysis on social media. Social media websites like twitter have user tweets for the public and that are accessible for analysis. One can directly import tweets from twitter website. Twitter’s API allows to perform complex queries like pulling every tweet about a certain topic and have used a dataset available on twitter binder. Twitter binder has extracted more than 560 MM COVID-19 related tweets. Raw Data related to COVID-19 is imported in the form of CSV file form TwitterBinder website [11]. The text section of the information which is the thing that have to process comprise of lot of special character and superfluous information which don't require so expelling of such uncommon characters is done as a piece of preprocessing.
- **Data Preprocessing:** A code in R is written to remove tabs, blank spaces, links etc. Further data is preprocessed to remove stop words like determiners (e.g the, a, an), conjunctions etc. Hash tags which are generally used by twitter users are also removed from data for analysis purpose. It is observed that many social media users make use of special characters like “#”(hashtag) and “@” at in their tweets. which need to be removed for lexical processing with dictionary words.
- **Sentiment Analysis:** The sentiment analysis is based on lexical oriented method where keywords extracted form tweets are further compared with words list associated with emotions. There are many libraries, dictionaries and packages available in R to evaluate emotions prevalent in a text. The tidy text and text data packages have word –to- emotion evaluation repositories. Three of the general purpose lexicons are Bing, AFINN and NRC (from the text data package). For handling the cleaned information the inbuilt sentiment analyzer is utilized in R, which utilizes the NRC emotion dictionary reference to figure the

nearness of six emotions. A complete score is determined for every supposition and plotted utilizing ggplot2 library.

V. RESULTS AND ANALYSIS

Using our emotional analysis method the tweets related to COVID-19 have classified into basic emotion categories. As it is illustrated in Fig 3. The number of positive and negative tweets is almost equal. Another classification of emotion shows more number of tweets are related to trust which shows that people have a lot of positivity trust to fight against COVID-19 and policies taken by authorities across the globe. Another majority of tweets shows fear, as the number of patients with COVID-19 is continuously increasing and is spreading at a fast pace across the globe. Till today no proper vaccine and treatment is available so many people have expressed fear. Quite a few tweets also shows sadness.

- **Performance Analysis:** To evaluate performance of our system, accuracy of classification of emotions is calculated. For calculation of accuracy 10% of random samples of data set were analyzed. A comparison was done between classifications of random tweet done by the system manual classification done based on word Emotion Association Library.

$$\text{Percentage accuracy} = \frac{\text{Tweets correctly classified}}{\text{Total Tweets}}$$

Using above formula, accuracy obtained is around 80% which is slightly better than accuracy obtained in classification of Korean data [1]. Further the scope of classification was only Korean data in [1]. Whereas our system considered data across the world and size of the dataset considered for our proposed system is much bigger than used in classification of Korean data [1]

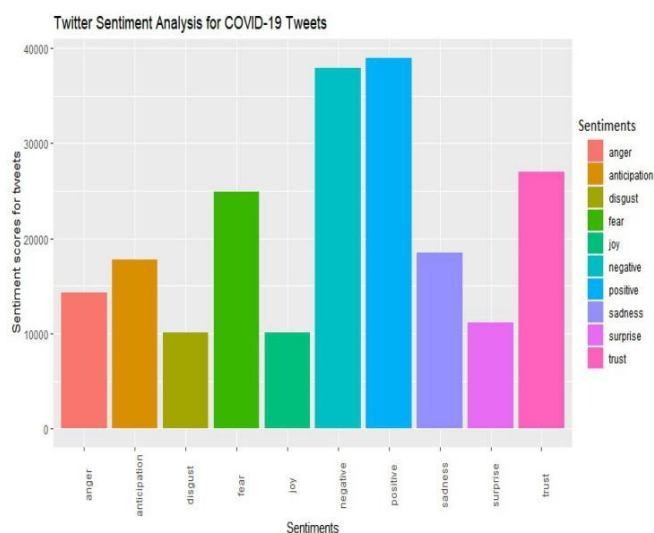


Fig 3. Classification of tweets into emotions

VI. CONCLUSION

In this paper, Twitter Data is analyzed across the world to understand the mental health of people during the COVID-19

pandemic situation through emotion analysis and classified it into basic emotions. As seen on the official websites of WHO and the Health Ministry, a separate tab is created for behavioral and psychological health which includes various audios and videos to keep people mentally healthy. By using our analysis, authorities can better understand the mental health of the people and update the content accordingly. Analysis is done on data across the world as although people live in different countries, having different physical and cultural environments, they still have the same basic emotions.

The scope of paper is restricted to only text content in twitter. But if closely view the social media content, many people are also using various emojis and graphics to convey their emotions. Further the system can be extended to consider such graphics and emojis in analyzing the sentiments

REFERENCE

- [1] H. J. Do, C. G. Lim, Y. J. Kim, and H. J. Choi, "Analyzing emotions in twitter during a crisis: A case study of the 2015 Middle East Respiratory Syndrome outbreak in Korea," 2016 Int. Conf. Big Data Smart Comput. BigComp 2016, pp. 415–418, 2016, doi: 10.1109/BIGCOMP.2016.7425960.
- [2] M. A. Tocoglu, O. Ozturkmenoglu, and A. Alpkocak, "Emotion Analysis from Turkish Tweets Using Deep Neural Networks," IEEE Access, vol. 7, pp. 183061–183069, 2019, doi: 10.1109/ACCESS.2019.2960113.
- [3] A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," ACM Comput. Surv., vol. 49, no. 2, 2016, doi: 10.1145/2938640.
- [4] L. Li *et al.*, "Characterizing the Propagation of Situational Information in Social Media during COVID-19 Epidemic: A Case Study on Weibo," IEEE Trans. Comput. Soc. Syst., vol. 7, no. 2, pp. 556–562, 2020, doi: 10.1109/TCSS.2020.2980007.
- [5] V. Lampos and N. Cristianini, "Tracking the flu pandemic by monitoring the social web," 2010 2nd Int. Work. Cogn. Inf. Process. CIP2010, pp. 411–416, 2010, doi: 10.1109/CIP.2010.5604088.
- [6] R. Kamal, M. A. Shah, C. Maple, M. Masood, A. Wahid, and A. Mehmood, "Emotion classification and crowd source sensing: A lexicon based approach," IEEE Access, vol. 7, pp. 27124–27134, 2019, doi: 10.1109/ACCESS.2019.2892624.
- [7] K. Tago and Q. Jin, "Influence analysis of emotional behaviors and user relationships based on Twitter data," Tsinghua Sci. Technol., vol. 23, no. 1, pp. 104–113, 2018, doi: 10.26599/TST.2018.9010012.
- [8] F. T. Giuntini *et al.*, "How do i feel? Identifying emotional expressions on facebook reactions using clustering mechanism," IEEE Access, vol. 7, pp. 53909–53921, 2019, doi: 10.1109/ACCESS.2019.2913136.
- [9] M. Bouazizi and T. Ohtsuki, "Multi-class sentiment analysis on twitter: Classification performance and challenges," Big Data Min. Anal., vol. 2, no. 3, pp. 181–194, 2019, doi: 10.26599/bdma.2019.9020002.
- [10] H. Achrekar, A. Gandhe, R. Lazarus, S. H. Yu, and B. Liu, "Predicting flu trends using twitter data," 2011 IEEE Conf. Comput. Commun. Work. INFOCOM WKSHPs 2011, pp. 702–707, 2011, doi: 10.1109/INFOCOMW.2011.5928903.
- [11] <https://www.tweetbinder.com/blog/twitter-atase/>, dated 15th April 2020
- [12] https://www.who.int/health-topic/coronavirus/#tab=tab_1, dated 4th April 2020
- [13] <https://main.mohfw.gov.in/date> : 4th April 2020