



An Automatic Emotion Analysis of Real Time Corona Tweets

A. Kalaivani¹(✉) and R. Vijayalakshmi²

¹ Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India

² IBM, Bangalore, India

Abstract. Emotion Analysis from text is a recent research field originated from Sentiment Analysis. The Sentiment Analysis identify the sentiments and classify them to positive, neutral, or negative sentiments based on the text. Emotion Analysis focus to detect and recognize emotions through the text expression which are anger, disgust, fear, happiness, sadness and surprise. The real time applications of emotion analysis can be widely applied in software engineering, website customization, education, and gaming domains. Emotion Analysis can be done by gathering social media data such as twitter, reviews, blogs. Emotion Analysis can be done on the public opinion on a particular product or on a particular topic.

In this paper we are focusing to do emotion analysis of the public on real time recent corona tweets. The tweets are collected through twitter application interface. The collected real tweets are applied to pre-processing strategies to remove inconsistent and redundant factors and the pre-processed tweets are visualized through word cloud. The emotions scores are obtained using nrc_sentiment dictionary which contains basic emotions and both sentiments of positive or negative. The emotions score levels of the public are identified from the Corona Tweets and depicted through graphical analysis.

Keywords: Emotion analysis · COVID-19 tweets · Word cloud · nrc_sentiment dictionary

1 Introduction

Sentiment Analysis applies natural language processing to a text documents to understand the sentiments of the text. The outcome of Sentiment classification will classify the text into positive sentiments, negative sentiments or neutral sentiments. Sentiment polarity can be applicable to a whole textual document, sentence and word. The sentiment analysis can be applied to marketing to identify customer trends and then to medical to find the opinion of the patients on a particular disease. Twitter analysis can be used to identify the impact of the education system in the society. Twitter Analysis can be used to track trends and also expanded to disseminate health information during viral epidemics.

Coronavirus disease (COVID-19) is a global infectious disease originated in China during year end of 2019. COVID-19 started spreading across the world and people

are still in pandemic situation. The symptoms of COVID-19 virus will have a mild to moderate respiratory illness and can be recovered easily at the initial stage without special treatment. Patients infected by COVID-19 virus are aged people with medical health issue related to heart problems, diabetics patient, respiratory irregularity and cancer patient have a chance of adaptability to this disease easily. Prevention Mechanism can be carried out to slow down COVID transmission by informing public about COVID-19 virus and also conduct awareness program. Simple mechanism is to protect oneself is by wash your hands frequently or using an sanitizer to keep yourself free. The COVID-19 virus spreads through personal contacts of the infected person with cough or cold by means of saliva droplets or nose discharge.

Our human being can interpret the sentiments of writing or speech. Based on the contextual understanding user can classify the sentiment. Humans are gifted with the capability to interpret the tone of a piece of writing. Consider the sentence: “My flight”s been delayed. Brilliant!” Most of them quickly interpret the person as sarcastic, but having a delayed flight is not a good experience. By applying this contextual understanding user can classify the sentence into negative sentiments. Without contextual understanding, a machine look at the sentence “brilliant” and identify sentiment as positive. In this paper we are collecting some real time data on corona and we are going to show what people are thinking about Corona in terms of sentiment analysis.

Sentimental analysis is a resource available for any organization used to analyze and enhance their products and services. Mobile Usage is found with all communities and before purchase of any mobile we can analyze feedback from mobile users through online tweets. Online Mobile tweets from customers of their mobile products can be



Fig. 1. Basic emotions classification

used for conducting a sentiment analysis. In this paper, the online mobile tweets are collected and processed to identify the emotion scores of mobiles users using R tool. The online mobile tweets are extracted and pre-processed and then classified the tweets into positive and negative sentiments and also their emotion scores are identified for the customers. The basic emotion classification system for online mobile tweets are depicted in Fig. 1.

The organization of the paper flow is as follows: Sect. 2 describes the researchers work on this domain and Sect. 3 describes system methodologies on process of gathering twitter data to emotion classification. The implementation of the proposed work in a stage by stage way is discussed in Sect. 5 and the paper is concluded with the contributions of the proposed work and future enhancement.

2 Literature Review

Many researchers contributed for their research work on pre-processing online tweets and twitter classification and further work also expanded on emotion score. Researchers concentrated much on stop phrase removal to achieving higher accuracy, formation of word cloud to identify frequent words, frequency count of accuracy for number of words present in the word cloud and finally sentiment emotions of overall tweets. Vijayalakshmi R, and A.Kalaivani, [1] proposed a brief information about word cloud formation and also pre processing techniques for apple mobile tweets finally accuracy of sentiment emotion score are depicted through bar graph.

Naramula Venkatesh and A.Kalaivani [2], proposed a preprocessing techniques for mobile tweets and formed word cloud visualization on apple mobile data sets and finding the frequency of the words. Bhattacharajee [3] et al. proposed a preprocessing algorithm for noise reduction based on lexicon. The Cosine Similarity Algorithm is proposed to classify the sentiment comment into a five point scale of -2 (highly negative) to $+2$ (highly positive). Ghag and shah [4] produced a research work by using movie document datasets to analyze the effect of stopwords removal on sentiment classification models. The proposed improved algorithm to produced better accuracy than traditional classifier. The survey is carried out with the classifier based on term weighting technique.

S. Rill et al. in [5], invented a system *PoliTwi* to detect emerging political topics in Twitter rather than other standard information channels. The identified *Top Topics* are shared via different channels towards wider publicity. The topics are compared with Google Trends and observed topics emerged in Twitter than in Google Trends. Finally, these topics can be used as a knowledge bases for concept-level sentiment analysis.

F. H. khan [6] et al. proposed an hybrid algorithm for twitter feeds classification. The proposed method applies contributed on multiple pre-processing steps before sent to the classifier. Proposed techniques overcomes the previous limitations and achieves higher accuracy when compared to the state-of-art techniques.

Rehab Duwairi, Mahmoud El-Orfali [7] analyzed the role of text pre-processing, feature selection and representation and classification using support vector machines. The level of accuracy achieved is improved when compared to the existing literature work. Duwairi and El-orfali [8] discussed sentiment analysis related to Arabic text. The sentiment analysis was investigated for Arabic text datasets with multiple classifiers

of SVM, Naive Bayes and K-Nearest Neighbour. The experimental results shows that selection of preprocessing strategies on the input tweets increases the performance of the classifiers.

E. Haddi, X. Liu, and Y. Shi [9] demonstrated the role of text pre-processing in sentiment analysis. The experimental results focussed on appropriate feature selection and representation, sentiment analysis accuracies using support vector machines (SVM) achieved improved performance. Alexander Pak, Patrick Paroubek [10] performed Sentiment Analysis using microblogging. It is a popular communication tool among Internet users. The proposed work automatically collect a corpus for both sentiment analysis and opinion mining purposes. A textual analysis is performed on the collected corpus and a sentiment classifier is built to determine different types of sentiments. Proposed techniques are efficient and performed better than state of art techniques. The research currently focussed on English language and in future can be expanded for any other languages. The primary issues discussed in literature review are classification accuracy such as most of the tweets with a very high percentage as neutral. The other issues to be considered by the researchers in future work are data sparsity and sarcasm.

3 System Methodology

Datas are gathered from users through web analytics tools which are independent, semi based and unreadable manner. From twitter API we are collecting some real time data on corona what people tweet about the virus and how they are protecting themselves from covid-19 and then the collected data is converted in the form of.csv files. The twitter data collected from the users is unstructured, incomplete, noisy and inconsistent. The data processing strategies are applied to discover knowledge records.

The general data pre-processing steps are removal of lowercase, punctuation, number, URL, special characters and expression. In the raw text stop words removal eliminates noise from text by removing words such as “the,” “and” and “a”. Tokenization and Visualization are effective method to discover abstract thoughts and express information in the raw text. The outcomes of Sentiment analysis are represent in the form of Graphs, Histograms and Matrices. The most famous representation are Interactive Maps and Word Cloud. Visualization presentation are used in multimedia, medicine, education, engineering and technological applications. The words with biggest size is most frequently used and with much less length are least used.

In our proposed system, the dataset chosen for sentiment analysis is real time corona dataset. The datasets is applied to pre-processing strategies to remove inconsistent and redundant factors. Our proposed pre-processing techniques involves elimination of punctuations, special characters, digits, escaping HTML characters. Further the dataset is fine-tuned by applying removal of stop words, removal of URLs and removal of expressions. The pre-processed data visualization is represented as word cloud with the frequency of the key words. Finally the tweets are classified into emotions based on nrc-sentiment dictionary and descriptive analysis for the emotions in the form of graph. In Real-world the data may contain unreadable formats which lack in trends, unpolished, disorder and noisy data with errors. Data processing is the best pre-processing techniques to resolve and the proposed block diagram is shown in Fig. 2.



Fig. 2. Proposed system for automatic emotion analysis

a. Data Collection

Customers are free to express their comments on public forums like blogs, discussion boards and reviews. Public opinions are collected on private or public social network sites like Facebook and Twitter. Opinions and feelings are expressed in terms of vocabulary, context writing, short forms and slang. The data collected through public forums or social network are unstructured and huge disorganized data. The manual analysis of sentiment data is virtually impossible, so in our proposed work we used “R” tool for the efficient data analysis.

b. Text Preparation

The data collected through public forums should be filtered to extract the data for data analysis. Text preparation is done by eliminating non-textual content in the data collection. After the text preparation process the relevant data alone exist which can be used for further data analysis.

c. Sentence Classification

Preprocessed sentence are examined for subjectivity and objectivity expressions. Each sentence of the tweets are examined for subjectivity and objective expressions. Sentences with subjective expressions are retained and that which conveys objective expressions are discarded. The various computational techniques used for identifying subjective sentences are unigrams, lemmas, negation.

d. *Sentiment Scores*

The subjective sentence identified are further classified into two groups as positive and negative. Sentiment Analysis plays a vital role to analyze and categorize the sentence into positive and negative tweets and the emotion scores are also calculated.

4 System Implementation

The steps to connect R and twitter API to extract Tweets on COVID-19 are

1. Make a Twitter account with the mobile number.
2. Create first Twitter app from this link -<http://apps.twitter.com>
3. Snap on Create New App. Pick a name for your app and give a concise depiction to your application and give your profile link.
4. Snap on “Create your Twitter application”. On the off chance that your application is made and it should look like this as shown in the Fig. 3.



Fig. 3. Creation of Twitter APP

5. Open your application and go to “Keys and Access Tokens” to learn your Consumer Key (Programming Interface Key) and Consumer Secret (API Secret) key as shown in Fig. 4.

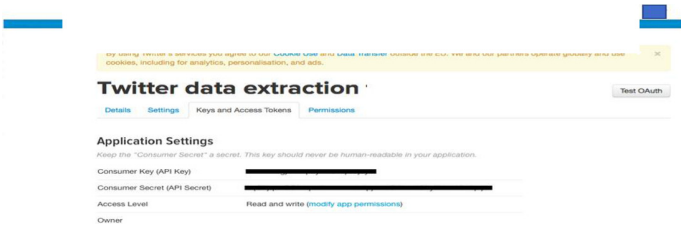


Fig. 4. Twitter data extraction -“Keys and Access Tokens”

6. If you’re doing this for the primary time then you’ve got to scroll down on an equivalent keys and access tokens page and generate your Access tokens as shown in Fig. 5.

R Studio is installed using the following steps:

1. Install necessary packages and load the libraries as shown in Fig. 6. These packages are important to install as they permit R interface to associate with twitter and offers validation to outsider applications.


```
[1] RT @sardesairajdeep: Good PM stayed away from expected China bashing: leave that to generals! Good he reiterated need to be corona consciou...
[2] RT @abbygov: bro I looked up how long the Black Plague lasted in an attempt to reassure myself about corona and that shit lasted 800 years...
[3] RT @surnell: #Coronil Ok Ok Not a cure for #Corona It is immunity booster that serves the same purpose. Let us buy it in big numbers now
[4] RT @abbygov: bro I looked up how long the Black Plague lasted in an attempt to reassure myself about corona and that shit lasted 800 years...
[5] @elonmusk @RationalEtienne @TimothyBuffett @WhatsupFranks @kimbal There is so little solid information about this -. https://t.co/2eSbeYSE0x
[6] RT @nitinsehra10: On National Doctors Day today.. \nI salute all the doctors of this nation.. \n\nYou are taking so much pain to save us.. \n\nA...
[7] @lagosfather boss the corona virus is real actually. It's hard saying this but I just recovered. so please stay safe
```

Fig. 9. Tweets after downloading

```
write.table(s_tweets, "/Users/parumohan/Desktop/COVID-19.csv",
           append = T, row.names = F, col.names = T, sep = ",")
```

Fig. 10. Tweets conversion into .CSV

From that csv file around 5–10 tweets are taken for data pre processing of tweets. The tweets are applied to tokenization in different structure, single tokenizers can include number of words present in sentences and shown in Fig. 11. The tweets upper case characters are converted into lowercase which are shown in Fig. 12.

```
[1] rt @sardesairajdeep: good pm stayed away from expected china bashing: leave that to generals! good he reiterated need to be corona consciou...
[2] rt @abbygov: bro i looked up how long the black plague lasted in an attempt to reassure myself about corona and that shit lasted 800 years...
[3] rt @surnell: #coronil ok ok not a cure for #corona it is immunity booster that serves the same purpose. let us buy it in big numbers now
[4] rt @abbygov: bro i looked up how long the black plague lasted in an attempt to reassure myself about corona and that shit lasted 800 years...
[5] @elonmusk @rationaletienne @timothybuffett @whatsupfranks @kimbal there is so little solid information about this -. https://t.co/2esbeyse0x
[6] rt @nitinsehra10: on national doctors day today.. \n i salute all the doctors of this nation.. \n\n you are taking so much pain to save us.. \n\n a...
[7] @lagosfather boss the corona virus is real actually. it's hard saying this but i just recovered. so please stay safe
```

Fig. 11. Tweets after tokenization

```
[1] rt @sardesairajdeep: good pm stayed away from expected china bashing: leave that to generals! good he reiterated need to be corona consciou...
[2] rt @abbygov: bro i looked up how long the black plague lasted in an attempt to reassure myself about corona and that shit lasted 800 years...
[3] rt @surnell: #coronil ok ok not a cure for #corona it is immunity booster that serves the same purpose. let us buy it in big numbers now
[4] rt @abbygov: bro i looked up how long the black plague lasted in an attempt to reassure myself about corona and that shit lasted 800 years...
[5] @elonmusk @rationaletienne @timothybuffett @whatsupfranks @kimbal there is so little solid information about this -. https://t.co/2esbeyse0x
[6] rt @nitinsehra10: on national doctors day today.. \n i salute all the doctors of this nation.. \n\n you are taking so much pain to save us.. \n\n a...
[7] @lagosfather boss the corona virus is real actually. it's hard saying this but i just recovered. so please stay safe
```

Fig. 12. Tweets converted to lower case

The tweets are further applied with the pre-processing and the tweets upper case characters are converted into lowercase which are shown in Fig. 12. Pre-processed tweets after the removal of punctuation mark are shown in Fig. 13 and removal of numbers are shown in Fig. 14. The pre-processed tweets after the removal of stop words are shown

in Fig. 15 and removal of URL are shown in Fig. 16. Further, they are processed to remove whitespaces and the output of the tweets after removal of white space are shown in Fig. 17.

```
[1] rt sardesai@rajdeep good pm stayed away from expected china bashing leave that to generals good he reiterated need to be corona consciou..
[2] rt abbygov bro i looked up how long the black plague lasted in an attempt to reassure myself about corona and that shit lasted 800 year
s..
[3] rt surnell coronil ok ok not a cure for corona it is immunity booster that serves the same purpose let us buy it in big numbers now
[4] rt abbygov bro i looked up how long the black plague lasted in an attempt to reassure myself about corona and that shit lasted 800 year
s..
[5] Elonmusk rationaletienne timothybuffett whatsupfranks kimbal there is so little solid information about this ... httpstco2esbeyseox
[6] rt nitinsehra10 on national doctors day today \n i salute all the doctors of this nation \n\n you are taking so much pain to save us \n\n
a..
[7] lagosfather boss the corona virus is real actually its hard saying this but i just recovered so please stay safe
```

Fig. 13. Tweets after removal of punctuations

```
[1] rt sardesai@rajdeep good pm stayed away from expected china bashing leave that to generals good he reiterated need to be corona consc
iou..
[2] rt abbygov bro i looked up how long the black plague lasted in an attempt to reassure myself about corona and that shit lasted year
s..
[3] rt surnell coronil ok ok not a cure for corona it is immunity booster that serves the same purpose let us buy it in big numbers now
[4] rt abbygov bro i looked up how long the black plague lasted in an attempt to reassure myself about corona and that shit lasted year
s..
[5] Elonmusk rationaletienne timothybuffett whatsupfranks kimbal there is so little solid information about this ... httpstcoesbeyseox
[6] rt nitinsehra national doctors day today \n i salute all the doctors of this nation \n\n you are taking so much pain to save us \n
\na..
[7] lagosfather boss the corona virus is real actually its hard saying this but i just recovered so please stay safe
```

Fig. 14. Tweets after removal of numbers

```
[1] rt sardesai@rajdeep good pm stayed away expected china bashing leave generals good reiterated need corona conscio
u..
[2] rt abbygov bro looked long black plague lasted attempt reassure corona shit lasted years..
[3] rt surnell coronil ok ok cure corona immunity booster serves purpose let us buy big numbers now
[4] rt abbygov bro looked long black plague lasted attempt reassure corona shit lasted years..
[5] Elonmusk rationaletienne timothybuffett whatsupfranks kimbal little solid information ... httpstcoesbeyseox
[6] rt nitinsehra national doctors day today \n salute doctors nation \n\n taking much pain save us \n\n..
[7] lagosfather boss corona virus real actually hard saying just recovered please stay safe
```

Fig. 15. Tweets after removal of stop words

The preprocessed tweets are applied to tokenization, where the tokens are extracted from the input tweets. The frequency of the words extracted are computed and depicted using bar graph as shown in Fig. 18. Word cloud is the visual representation of the tokenized words which are depicted in Fig. 19.

The final preprocessed tweets are applied to src_dictionary to identify the emotion scores and also to classify the tweets as positive or negative. The various emotions showed by the tweets are anger, disgust, fear joy, surprise and so on. The sentiment emotion score for COVID-19 Tweets are shown in Fig. 20.

- [1] rt sardesairajdeep good pm stayed away expected china bashing leave generals good reiterated need corona conscio u...
- [2] rt abbygov bro looked long black plague lasted attempt reassure corona shit lasted years...
- [3] rt surnell coronil ok ok cure corona immunity booster serves purpose let us buy big numbers now
- [4] rt abbygov bro looked long black plague lasted attempt reassure corona shit lasted years...
- [5] elonmusk rationaletienne timothybuffett whatsupfranks kimbal little solid information ... httpstcoesbeyseox
- [6] rt nitinsehra national doctors day today \n salute doctors nation \n\n taking much pain save us \n\n...
- [7] lagosfather boss corona virus real actually hard saying just recovered please stay safe

Fig. 16. Tweets after removal of URL

- [1] rt sardesairajdeep good pm stayed away expected china bashing leave generals good reiterated need covid conscio u...
- [2] rt abbygov bro looked long black plague lasted attempt reassure covid shit lasted years...
- [3] rt surnell coronil ok ok cure covid immunity booster serves purpose let us buy big numbers now
- [4] rt abbygov bro looked long black plague lasted attempt reassure covid shit lasted years...
- [5] elonmusk rationaletienne timothybuffett whatsupfranks kimbal little solid information ...
- [6] rt nitinsehra national doctors day today salute doctors nation taking much pain save us ...
- [7] lagosfather boss covid virus real actually hard saying just recovered please stay safe

Fig. 17. Tweets after removal of whitespace

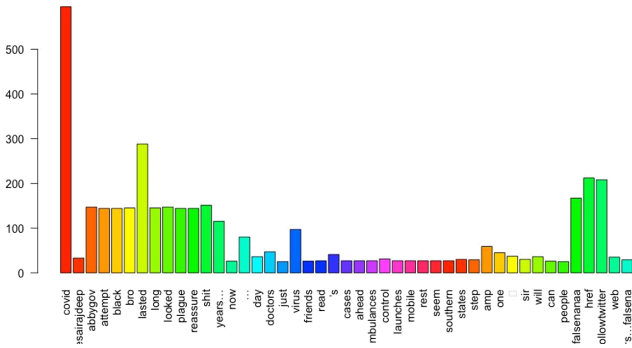


Fig. 18. Bar plot for frequent words



Fig. 19. Word cloud formation

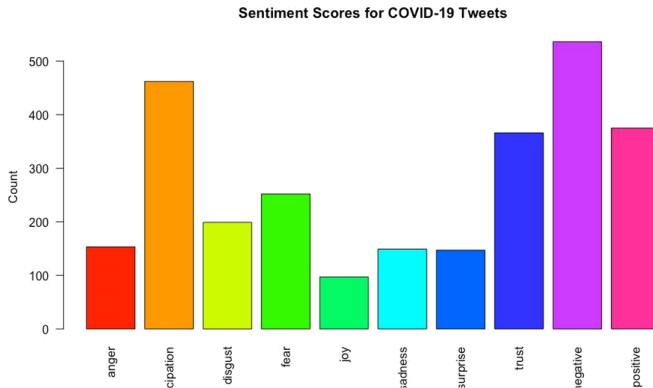


Fig. 20. Emotion scoring on COVID-19 tweets

5 Conclusion

Sentiment Emotion Scoring for COVID-19 tweets shows the negativity as high because all people think that is very dangerous and no medicine was found still. People are afraid of that diseases so the emotion scoring for negativity is high. People who are recovered from covid have tweeted on how to take self-care and their experience on covid. Future scope of the proposed work can be related to any other pandemic diseases or on products to give direction for the customers. Emotion scoring identified for the whole tweets based on words and in future we can classify sentiments for individual tweets based on the emotion score of individual tweets and can further analyse the data at an extreme point of view.

References

1. Vijayalakshmi, R., Kalaivani, A.: Sentiment emotion scoring for Apple mobile tweets. Test Eng. Manag. **82**, 6756–6763 (2020)

2. Kalaivani, A., Venkatesh, N.: Word cloud for online mobile phone tweets towards sentiment analysis. *Int. J. Eng. Adv. Technol.* **8**(6), 2249–8958 (2019)
3. Bhattacharjee, S., Das, A., Bhattacharya, U., Parui, S.K., Roy, S.: Sentiment analysis using cosine similarity measure. In: *IEEE 2nd International Conference Recent Trends in Information Systems (ReTIS)*, pp. 27–32 (2015)
4. Ghag, K., Shah, K.: Comparing analysis of effect of stop words removal on sentiment classification. In: *IEEE International Conference on Computer Communication and Control*, pp. 2–7 (2015)
5. Rill, S., Reinel, D., Scheidt, J., Zicari, R.: Early detection of emerging Politics topic on twitter and the impact on concept-level sentiment analysis. *Knowl.-Based Syst.* **69**, 24–33 (2014)
6. Khan, F.H., Bashir, S., Qamar, U.: TOM: Twitter opinion mining Frame work using hybrid classification scheme. *Decis. Support Syst.* **57**(1), 245–257 (2014)
7. Duwairi, R., El-Orfali, M.: A study of the effects of preprocessing strategies on sentiment analysis for Arabic text. *J. Inf. Sci.* 215–221 (2014)
8. Duwairi, R., Elorfali, M.: A study of the effects of preprocessing strategies on sentiment analysis for Arabic text. *J. Inf. Sci.* 1–13 (2013)
9. Haddi, E., Liu, X., Shi, Y.: The role of text pre-processing in sentiment analysis. *Procedia Comput. Sci.* **17**, 26–32 (2013)
10. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: *Proceedings of the International Conference on Language Resources and Evaluation*, pp. 17–23 (2010)