EasyChair Preprint
№ 2974

Twitter Sentiment Analysis on Coronavirus using Textblob

Chhinder Kaur and Anand Sharma

March 16, 2020

Twitter Sentiment Analysis on Coronavirus using Textblob

Chinder Kaur[1] and Anand Sharma[2]

[1] Research Scholar, UCCA, Guru Kashi University,
Talwandi Sabo, Bathinda, Punjab, India
`chhinderkaur87@gmail.com`
[2] Assistant Professor, UCCA, Guru Kashi University,
Talwandi Sabo, Bathinda, Punjab,India
`andz24@gmail.com`

Abstract. Social networks are the main resources to gather information about people's opinions and sentiments towards different topics and issues. People spend hours daily on social media to share their ideas, opinions, and reactions with others, so in this paper, we analyze the sentiments regarding coronavirus disease(COVID-19) because many peoples of different countries are affected by coronavirus that is very critical issue in these days, so analyze the sentiments of different people's opinion for this disease, we are fetching the twitter streaming tweets related to coronavirus using twitter API and analyze these tweets using machine learning techniques and tools as positive, negative and neutral. In this paper, we run experiments through Python programming on different tweets using twitter API and NLTK library is used for pre-processing of tweets and then analyze the tweets dataset by using Textblob and after that show the interesting results in positive, negative, neutral sentiments through different visualizations.

Keywords: Twitter Sentiment Analysis, Twitter API, TextBlob

1. Introduction

Coronavirus- Jonathan Temte et. Al [24] Coronaviruses are incredibly diverse, found in many animal species, and are commonly encountered in clinical practice during the cold and flu season, yet many primary care clinicians are not familiar with these respiratory pathogens. These are rarely been tested for them, and when these are it's usually when it's been looking for something else. Moreover, there has been no specific treatments for these viruses.

Opinion and sentimental mining- Opinion and sentimental mining are important research areas because due to the huge number of daily posts on social networks, extracting people's opinions is a challenging task. This paper analyzes the sentiments regarding Coronaviruses that are a family of viruses that cause disease in animals. Seven, including the new virus, have made the jump to humans, but most just cause cold-like symptoms, peoples of different countries are affected with coronavirus so in this paper the sentiments of peoples for coronavirus analyzed. About 90 percent of

today's data has been provided during the last two years and getting insight into this large scale data is not trivial [17, 18]. Sentimental analysis has many applications for different domains for example in businesses to get feedbacks for products by which companies can learn users' feedback and reviews on social media to analyze the customer review. Opinion and sentimental mining have been well studied in this reference and all different approaches and research fields have been discussed [10]. There are some works that have been done on sentiment analysis on Facebook [19-23] however in this paper the focus is on the Twitter sentimental analysis. For sentiment analysis firstly understand the text, summarize it and give weight to it whether it is positive, negative or neutral. Two fundamental approaches to extract text summarization-

- Extractive method
- Abstractive method

In the extractive method, words and word phrases are extracted from the original text to generate a summary.

An abstractive method tries to learn an internal language representation and then generates a summary that is more similar to the summary done by a human.

Text understanding is a significant problem to solve. Some machine learning techniques, including various supervised and unsupervised algorithms, are being utilized. There are different approaches to generate a summary. One approach could be to rank the importance of sentences within the text and then generate a summary for the text based on the important numbers. There is another approach called end-to-end generative models. In some domains like image recognition, speech recognition, language translation, and question-answering, the end-to-end method performs better[1].

Wilson, et.al [9] discuss in the phrase level, sentimental analysis system should be able to recognize the polarity of the phrase

Agarwal and et.al [11] discuss Tree kernel and feature-based models have been applied for sentimental analysis in twitter.

SemEval-2017 [12] also shows the seven years of sentimental analysis in twitter tasks. Since tweets in Twitter is a specific text not like a normal text there are some works that address this issue like the work for short informal texts [13]. Sentimental analysis has many applications in news [14].

In this paper, the social network analysis and the importance of it are discussed, then Twitter is discussed as it is a rich resource for sentimental analysis. In the following sections, the abstract of implementation is discussed. The sentiments are shown through positive, negative and neutral on coronavirus using different visualization methods to better understand.

2. Social Network Analysis

Social network analysis is the study of people's interactions and communications on different topics and nowadays it has received more attention. Million of people give their opinion on different topics on a daily basis on social media like Facebook,Twitter,Instagram, etc. It has many applications in different areas of research from social science to business [3]. Twitter nowadays is one of the popular social media which according to the Statista in [4] currently has over 300 million accounts. Twitter is a rich source to learn about people's opinions and sentimental analysis [2].

Each tweet is important to determine the sentiment of the tweet whether is it positive, negative, or neutral. Another challenge with twitter is only 140 characters are the limitation of each tweet which causes people to use phrases and words which are not in language processing. Recently Twitter has extended the text limitations to 280 characters per each tweet.

- **Twitter Sentimental Analysis-** Social networks/Social media is a rich platform to learn about people's opinions and sentiment regarding different topics, they can communicate and share their opinion daily basis on social media including Facebook ,Twitter, and Instagram,etc. The sentiment-aware systems these days have many applications from business to social sciences. Since social networks, especially Twitter, contains small texts and people may use different words and abbreviations which are difficult to extract their sentiment by current Natural Language processing systems easily, therefore some researchers have used deep learning and machine learning techniques to extract and mine the polarity of the text [15]. Some of the top abbreviations are FB for Facebook, B4 for before and so on. Therefore sentimental analysis for short texts like Twitter's posts is challenging [8].

3. Design and Implementation

This technical research paper reports the implementation of the Twitter sentiment analysis, by using the Twitter API. Tweepy is a library of Twitter API for fetching the tweets directly from Twitter that are post by different people. The streaming tweets are fetched and then saved into CSV files to sentiment analysis. There are great works and tools focusing on text mining on social networks. In this research project, the wealth of available libraries has been used. The approach to extract sentiment from tweets is as follows:

1. Import Tweepy for creating the connection with Twitter API
2. Fetch tweets as dataset and then save into CSV file.
3. pre-processing of tweets by removing the stop words,punctuations,#tags, etc..
4. Tokenize each word in the dataset and save it into the dataset.
5. For each word, compare it with positive,negative and neutral sentiments word in the dictionary. Then increment the positive,negative and neutral count.
6. Finally, based on the positive,negative and neutral count, we can get the result percentage about sentiment to decide the polarity. Researchers have done different sentimental analyses on Twitter for different purposes for example the work designed by Wang, et.al [5] is a real-time twitter sentimental analysis of the coronavirus. In this paper shows the sentimental analysis algorithm at a high level. As can be seen in the algorithm, reserchers have different procedures to connect the twitter API,

4

fetch the tweets, tweet cleaning or remove stop words and punctuation marks, classify tweets which means get the polarity of the tweet, and finally return the results.

- Implementation In this paper, python is used to implement the sentimental analysis. Some packages have utilized including Tweepy and Textblob.

The required libraries are installed by following commands:
• pip install tweepy
• pip install textblob

And also download the dictionary by running the following command on anaconda command prompt :
 python -m textblob.download_corpora.

The Textblob is a python library for text processing and it uses NLTK(Natural Language ToolKit) for natural language processing [6]. Corpora is a large and structured set of texts which is required for analyzing tweets.

```
In [151]: import tweepy
          import pandas as pd
          import matplotlib.pyplot as plt
          pd.options.display.max_columns = 50
          pd.options.display.max_rows= 50
          pd.options.display.width= 120
```

```
In [152]: access_token = "                        "
          access_token_secret = "                        "
          consumer_key = "            "
          consumer_secret = "                    "
          auth = tweepy.OAuthHandler(consumer_key=consumer_key, consumer_secret=consumer_secret)

          api = tweepy.API(auth)
```

Figure 1. create connection with Twitter API by using tweepy module and access tweets through keys and tokens.

nnect to Twitter using APIs to by using Tweepy and fetch the latest tweets related to coronavirus, we need to create an account on twitter and define an application. Users need to go to the apps.twitter.com/app/new and generate the API keys, Due to the security reasons the API keys are not is shown in the figure.

This command is run two times and fetch a total of 3000 tweets and then combined in one CSV file to analyze the dataset after preprocessing/cleaning 2058 tweets are used for sentiment analysis.

```
In [153]: results = []
          for tweet in tweepy.Cursor(api.search, q="#cronavirus", lang = 'en').items(2000):
              results.append(tweet)

          print (len(results))
```

Figure 2. fetch tweets related to cronavirus from #cronavirus hashtag. run this command two times and fetch tweets for analysis

The tweets are extractedon coronavirus disease according to the country and source for a better understanding of tweets.
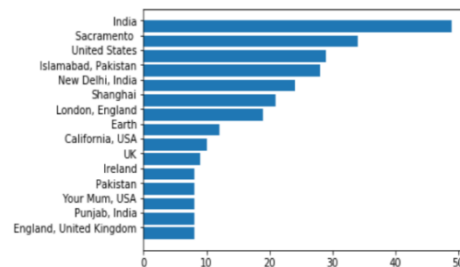

Figure A shows the no. of tweets that are posted from different countries, here mostly tweets regarding coronavirus posted by india
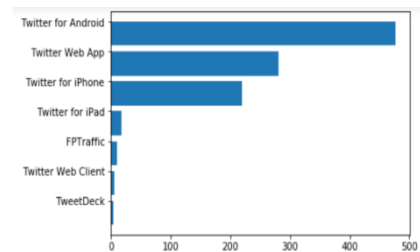

Figure B shows the different sources where tweets was posted , in this fig. mostly tweets are posted through Twitter for Andriod

In fig.A shows the different countries which share the opioin regarding coronavirus and this figure shows most of the tweets from india.

In fig.B shows the tweets are gathering from different sources and this figure shows most of the tweets from Twitter from Andriod.

The tweets are also shown into a word cloud to visualize the stopwords,in which word cloud coronavirus dead,worldwide panic,dead approx words are visualized.
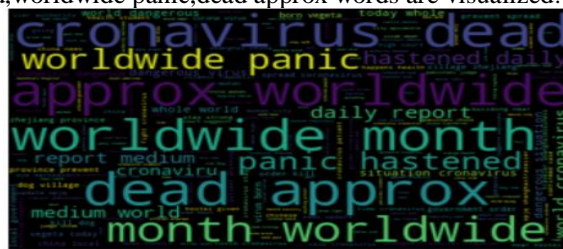

Figure C shows the different words regarding coronavirus using wordcloud

In fig. C shows the different words related to coronavirus using word cloud.

- The result of 'Coronavirus' is shown by positive,negative and neutral as based on 2058 tweets from Twitter that is posted by peoples of different countries. Textblob is used to analyze the tweets.

Sentiment Analysis

```
In [156]: from textblob import TextBlob

In [157]: df['polarity'] = df['Lemmatized'].apply(lambda x: TextBlob(x).sentiment[0])
          df['subjectivity'] = df['Lemmatized'].apply(lambda x: TextBlob(x).sentiment[1])
          df.head()
```

Figure 3. use Textblob to analysis the polarity and subjectivity

```
#show the dataframe
df[['cleanedtweet','polarity','subjectivity','Analysis']]
```

Out[171]:

	cleanedtweet	polarity	subjectivity	Analysis
1	combining face detection camera thermal camera effective protect people https opvusc	0.600000	0.800000	Positive
2	greetings respect please read consider news says https wrtijbxbsb	0.000000	0.000000	Neutral
5	news coronavirus reaches afghanistan first case confirmed herat province cronavirus outbreak	0.325000	0.666667	Positive
7	iran covered coronavirus boost election turnout cover increasing public anxiety risking https	0.000000	0.066667	Neutral
12	airlines suspended china flights coronavirus outbreak flights china iran continue https ktyhi paqv	0.000000	0.000000	Neutral
13	airlines suspended china flights coronavirus outbreak flights china iran continue https dncimrtrhd	0.000000	0.000000	Neutral
14	stand firm iron brothers pakistani student wuhan appnews cronavirus https emvlprfiu	-0.200000	0.400000	Negative
15	johnson blame everything heard started cronavirus also laugh	0.300000	0.100000	Positive

Figure 4. Analyze the positive, negative and neutral result of different tweets

List of positive and negative tweets are showing below

```
47) cute sisters married chinese litle https ztqddqsh
48) almighty allah super power created land please help human living organisms
49) finally relief first anti crona drug approved cronavirus favilavir cronaout
50) healthy looking women sudden collapse street going stop soon serious reperc
51) think investment health care well important single disease break supply cha
52) food delivery corona patient hospital china simply amazing mind blowing pow
53) take crona virus outbreak instead mass casualty event caused newly installe
54) clearly know https ocqccags


107) mondaymotivaton todayindoom iran angry sanction tensions grow turkey syria cronaviru
s pruoearj
108) today whole world dangerous situation cronavirus dangerous virus born vegeta
109) stimulus pledge chinese suggests situation regarding cronavirus much much worse htt
110) silly comparison chernobyl happened junior engineer mistake https rythv
111) cronavirus scary know scarier virus made china control popu
112) pray china support difficult times cronavirus https bala
113) theeyesofdarkness noval published dear unbelievable cronavirus https hxvl ctst
```

Figure 5 postive and negative tweets

Here the percentage of tweets is also calculated using Textblob.

```
In [203]:   # get the percentage of postive tweets
            ptweets = df[df.Analysis == 'Positive']
            ptweets = ptweets['cleanedtweet']

            round( (ptweets.shape[0] / df.shape[0]) *100 , 1)

Out[203]:   24.0

In [202]:   # get the percentage of negative tweets
            ntweets = df[df.Analysis == 'Negative']
            ntweets = ntweets['cleanedtweet']

            round( (ntweets.shape[0] / df.shape[0]) *100 , 1)
Out[202]:   32.1

In [204]:   # get the percentage of neutral tweets
            neutweets = df[df.Analysis == 'Neutral']
            neutweets = neutweets['cleanedtweet']

            round( (neutweets.shape[0] / df.shape[0]) *100 , 1)
Out[204]:   43.9
```

Figure 6 percentage of positive,negative and neutral tweets

Here show the sentimental analysis results based on Coronavirus.

The result is visualized by different plot methods using matplotlib that is the most popular library in python to the visualization of a result like a bar chart, histrogram,pie chart, etc. as shown in figure 3, illustrates. As can be clearly seen in the table and diagram the percentage of neutral tweets are significantly high.
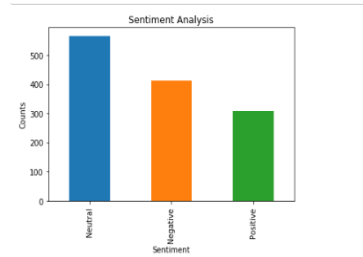

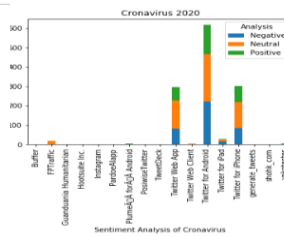**Figure 7 positive,negative and neutral tweets**

Figure 8

In fig. 7 shows the positive,negative and neutral results, In fig. 5 shows the result of positive,negative and neutral according to the different sources where tweet is getting.

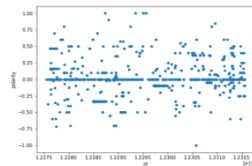The polarity and subjectivity of the tweets is also visualized.

Figure 9

In fig. 9 shows the polarity of the weets , in this figure shows the different polarity.

This figure shows the result of tweets for 'coronavirus(Covid 2019)' based on 2058 tweets from Twitter.

Positive tweets percentage   24.0%
Negative tweets percentage   32.1%
Neutral tweets percentage   43.9%
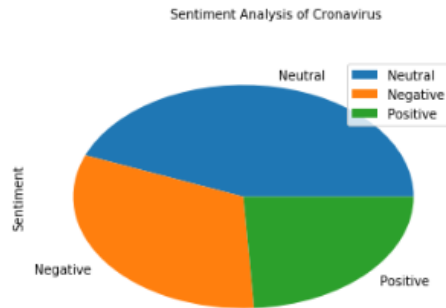
8

Sentiment Analysis of Cronavirus

Figure 10

In fig. 10 shows the result in pie chart through positive,negative and neutral, in this figure positive 24.0%, negative 32.1%, neutral 43.9% shows in different colours to clearly visulations.

4. CONCLUSION

In this research paper, the opinion of different peoples of different countries on coronavirus the affect various countries has been discussed. The main focus of this paper is on Twitter ,Twitter API and have implemented the python programming language and write code in Jupternotebook to implement the sentimental analysis as positive,negative and neutral. The results are shown by using Matplotlib library. It has been realized that the neutral sentiments are significantly high which shows there is a need to improve Twitter sentiment analysis.

REFERENCES

1. https://www.elsevier.com/connect/coronavirus-information-center 2020/27/2

2. Boguslavsky, I. Semantic Descriptions for a Text Understanding System. In Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue"(2017) (pp. 14-28) "(2017).

3. Pak, A., & Paroubek, P. Twitter as a corpus for sentiment analysis and opinion mining. In LREC (Vol. 10) (2010, May)..

4. Scott, J. Social network analysis: developments, advances, and prospects. Social network analysis and mining, 1(1), 21-26(2011)..

5. Statista, https://www.statista.com/statistics/282087/number-ofmonthly-active-twitter-users/(2017)

6. Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S.. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In Proceedings of the ACL 2012 System Demonstrations (pp. 115-120). Association for Computational Linguistics(2012, July)..

7. TextBlob, https://textblob.readthedocs.io/en/dev/2017

8. Pang, B., & Lee, L. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1- 135(2008)..

9. Dos Santos, C. N., & Gatti, M. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts (2014, August)

10. Wilson, T., Wiebe, J., & Hoffmann, P. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the conference on human language technology and empirical methods in natural language processing(pp. 347-354). Association for Computational Linguistics(2005, October)..

11. Liu, B. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167(2012)..

12. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. Sentiment analysis of twitter data. In Proceedings of the workshop on languages in social media (pp. 30-38). Association for Computational Linguistics(2011, June)..

13. Rosenthal, S., Farra, N., & Nakov, P. SemEval-2017 task 4: Sentiment analysis on Twitter. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)(pp. 502-518) (2017)..

14. Kiritchenko, S., Zhu, X., & Mohammad, S. M. Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research, 50, 723-762(2014)..

10

15. Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., ... & Belyaeva, J. Sentiment analysis in the news. arXiv preprint arXiv:1309.6202(2013)..

16. Poria, S., Cambria, E., & Gelbukh, A. Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 2539- 2544) (2015).

17. Ortigosa, A., Martín, J. M., & Carro, R. M. Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior, 31, 527-541(2014)..

18. Bagheri, Hamid, and Abdusalam Abdullah Shaltooki. "Big Data: challenges, opportunities, and Cloud-based solutions." International Journal of Electrical and Computer Engineering 5.2: 340(2015):.

19. Bagheri, Hamid, Mohammad Ali Torkamani, and Zhaleh Ghaffari. "Multi-Agent Approach for facing challenges in Ultra-Large Scale systems." International Journal of Electrical and Computer Engineering 4.2: 151(2014):.

20. Ortigosa, A., Martín, J. M., & Carro, R. M. Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior, 31, 527-541(2014)..

21. Feldman, R. Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82-89(2013)..

22. Dasgupta, S. S., Natarajan, S., Kaipa, K. K., Bhattacherjee, S. K., & Viswanathan, A. Sentiment analysis of Facebook data using Hadoop based open source technologies. In Data Science and Advanced Analytics (DSAA), 2015. 36678 . IEEE International Conference on (pp. 1-3). IEEE(2015).

23. Trinh, S., Nguyen, L., Vo, M., & Do, P. Lexicon-based sentiment analysis of Facebook comments in the Vietnamese language. Recent developments in intelligent information and database systems (pp. 263-276). Springer International Publishing(2016).

24. Haddi, E., Liu, X., & Shi, Y.The role of text pre-processing in sentiment analysis. Procedia Computer Science, 17, 26-32(2013)..

25. Coronavirus-https://www.elsevier.com/connect/coronavirus-information-center 2020/02/02.