Emotion Recognition from Covid-19 Related Twitter Posts

Md. Ariful Islam Malik
Department of Computer Science
Boise State University
Boise, ID, USA
mdarifulislammal@u.boisestate.edu

Abstract

Context: Twitter is a social media platform with more than 396 million users worldwide. Due to its large user base and the word limitation of the twitter posts, researcher found it easier for analyzing content underneath the tweets. **Objective:** The main objective of our work is to extract emotional attributes from the twitter posts to gain knowledge about the emotional status of the inhabitants of a particular region. **Method:** For this, we collect twitter posts that is related to covid-19 and perform a transfer learning approach to predict the emotional status of the tweet. We collect our dataset from kaggle and implement a Bert pretrained model to identify emotions on the tweets. We perform evaluation on TweetEval dataset in which our Bert model shows 72% accuracy.

Results: After the experiment, we identify happiness to be more common in the worldwide tweet posts. The second most common emotion that we found is anger.

I. Introduction

Our primary objective with the analysis is to identify how human mind is affected by the coronavirus. To meet our goal we collect use a popular social media platform Twitter. We collect tweets from around the world during the covid-19 period and exploit natural language processing techniques to find out the answers of the following research questions:

- **RQ1:** What is the frequency of different emotion attributes during covid-19?
- **RQ2:** What is level of different emotion in major countries?

The remainder of the paper is organized as follows: Section 2 presents some related works. Section 3 discusses about the dataset and the methodology employed in this work. Section 4 shows the evaluation of our model. Section 5 shows the results. Section 6 discusses about the findings from the work. Section 7 conclude our work. [1]

II. Related Study

In this section we describe the work that are related to emotion recognition on twitter data regarding covid-19. In [4] authors performed emotion detection using the tweets from Indian region. They used an LSTM model to detect emotion in the twitter posts. They were able to detect 5 types of emotions.

In [5] Alhajji et al., perform sentiment analysis on twitter posts that are from Arab region. They uses Naive Bayes as their model to predict sentiment in tweets. They collected twitter dataset from Kaggle,2019 which is consisting of 58000 tweets and are divided into training and test set. They were able to achieve .89 accuracy in their evaluation.

In [6] authors collect twitter posts from February 2, 2020 to March 15, 2020 which is related to Covid-19 English post. They use the combination of unigram and bigram to identify most common topics and and exploit a topic modeling technique in their analysis. As the topic modeling algorithm they use latent dirichlet allocation (LDA). With LDA, they were able to find natural clusters in the tweet posts. Finally they identified 12 topics which are catagorized into four: the origin of the virus;

[1]The code and papers can be found in github repository: https://github.com/arifulmalik/emotion_twitter

its sources; its impact on people, countries, and the economy; and ways of mitigating the risk of infection.

In [7] they uses Lexical Oriented Method (Lexicon). Lexicon identifies some of the word or collection of words to infer the feeling of the sentence. In this case, the whole sentence is mapped by some specific keyword. The problem of this method is, there would be many situation where a single word is used to express different feelings. This case cannot be solved by this Lexicon.

In [8] they collected over 35million tweets using twitter API and after subtraction they finally made dataset with 4 million tweets. All of the tweets were considered only with the covid-19 related hashtags (e.g. ”coronavirus”, ”COVID-19”, ”quarantine”). The duration of the tweets are from March 7 to April 2020. In their study they used NRC Emotion Lexicon, which has 8 primary emotions: anger, anticipation, fear, surprise, sadness, joy, disgust, and trust. With the LDA algorithm they identify each twitter posts as an emotion out of 8 emotion catagory. However, they do not mention any specification of which algorithm they particularly use in the field of natural language processing.

In [9] authors study in order to find the correlation between sentiments and emotions of the people from within 5 countries, such as, Pakistan, India, Norway, Sweden, USA, and Canada in the covid-19 period. In this regard, they uses multiple datasets for their study. The sentiment140 dataset to train the polarity of sentiment assessment classifier. Here, each label has .8 million tweets with a total of 1.6 million tweets. Second, they use emotional tweets dataset for the emotion recognition. This dataset contains six classes with a total of 21000 labeled tweets. As a method, they use deep long short-term memory (DLSTM) for predicting the polarity of the sentiment and emotions from the tweets. The variant of the LSTM model is the FastText model. With this model, they were able to achieve 82% accuracy and 81% F1 score.

In [10] the authors analyzed the conversation about covid-19 related posts to perceive the evolution of the conversation among five major communities of users namely government and health organizations, news media, politicians, the general public and conspiracy theory supporters. For this, they collected over 2 million tweets in the Dutch location. They use topic modeling algorithm such as LDA to monitor the tweets of the different user groups and found that the primary topic on covid-19 shifted from the virus towards the economy during February and April 2020.

III. Method

A. Dataset

We collected tweets from the kaggle covid-19 tweets dataset [1]. In this dataset, 179108 tweets are collected with 13 attributes in each tweet. The attributes are: user name, user location, user description, user created, user followers, user friends, user favourites, user verified, date, text, hashtags, source, is retweet. Here, text represents the tweet. We consider text, user location for our analysis. Each

B. Model

We use pretrained Bert model from huggingface. The name of the model is ’bert-base-uncased’.

IV. Evaluation

We use Bert model for our analysis. This is a pretrained model and hosted by huggingface [2]. In huggingface the model is named as ’bert-base-uncased’. To instantiate our method, we first create BertClass that enables all the necessary parameters that Bert model requires and then train it with the trained data from the TweetEval dataset [3]. TweetEval combines seven natural language processing tasks in Twitter posts which is done using multi-class tweet classification. The tasks include - irony, hate, offensive, stance, emoji, emotion, and sentiment. All tasks are performed using the same benchmark and each dataset are presented in the same format. It has three splits: train, validation, and test. Our work focuses on the task emotion.

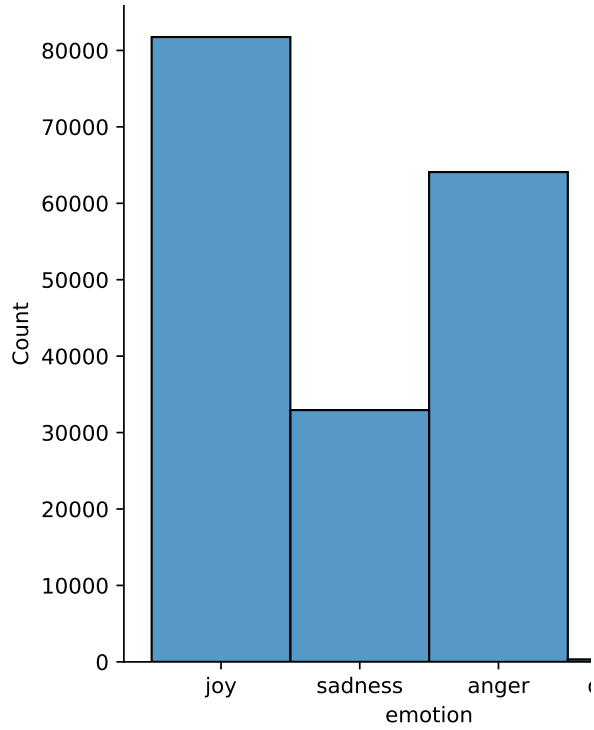

Fig. 1. Comparison of Emotions during Covid-19

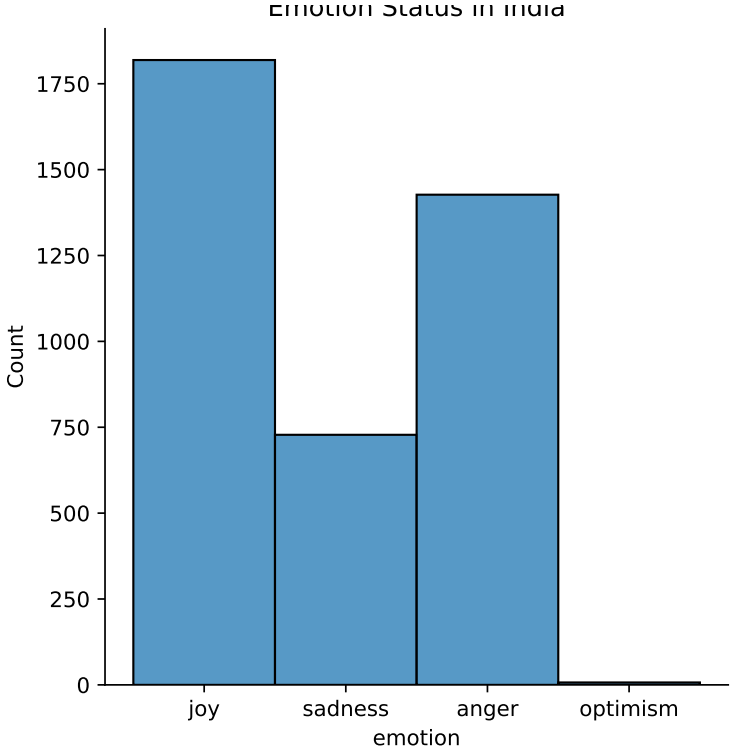

Fig. 2. Emotion status of India

A. *Procedure*

B. *Task*

C. *Metrics & Baseline*

V. Results

In this section we discuss about the results that we got from the experiment. Figure 1 shows the five different emotions all over the world during covid-19. Here, we see the emotion 'joy', which refers to happiness dominates over the others. The second highly observed emotions among people is anger. Surprisingly sadness are very less frequent during covid-19 period.

Figure 2 refers to the emotional status of indian people during covid-19. Here, emotion 'joy' overshadowed the other emotions. But the anger is much more prominent than the sadness.

Figure 3 is the emotional status of the USA region. Here also, happiness dominates over the anger and sadness.

Figure 4 illustrates the emotional status of England. Here also the happiness shows a pick and anger follows the happiness and sadness is observed half as common as anger.

VI. Discussion

A. *Challenges*

A challenge we find working in this project is collecting datasets. Most of the streaming link from Twitter API does not work because Twitter closes some streaming end. But, they do not give a notice on the original streaming home page. implications?

B. *Limitations*

Although, our dataset represents a wide range of covid-19 related post and we carried out a reasonable evaluation method, however, it is impossible to cover all the aspects that completes the analysis. In this section, we list some of the many limitations that we encounter in our work.

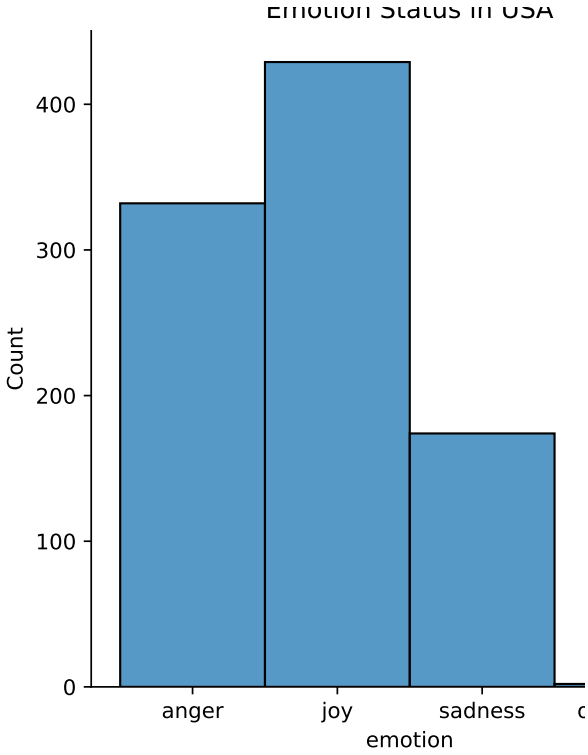

Fig. 3. Emotion status of USA

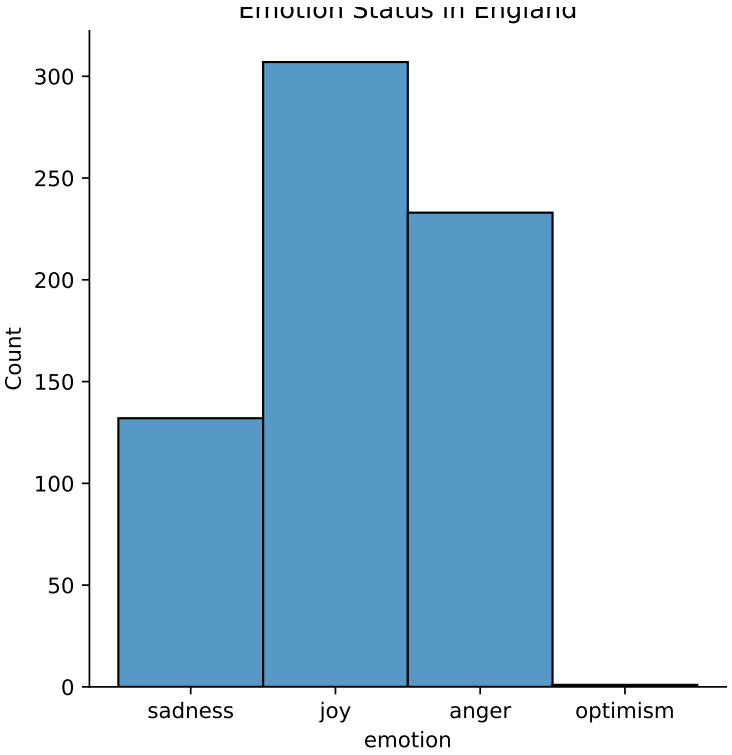

Fig. 4. Emotion status of England

- We perform evaluation on TweetEval dataset, which is widely used in scientific research. But we could not show how closely relates this dataset to the actual tweet posts for our case.
- We only consider English language in our study. But, the region that we include in our study has different language speaking people. Thus, the study does not fully cover the population of those regions.
- We only take Twitter posts into our study. But, many other social media platform are more popular than the Twitter. Thus, we do not get all the information that happened in that specific region.
- We use Bert model for our study. But, some other model are more accurate and more focused on the Twitter datasets.

VII. Conclusion

In this paper we use Twitter posts of over 179000 and analyze using the pretrained Bert model to identify the emotion status of the people around the world. We also show emotional level for the three most frequent countries: India, USA, England. Whether, our work does only cover English language, the future research can play more impact if region specific languages are also considered. Moreover, the study only limits the use of Twitter posts, where other popular social media platform should be a good choice for the analysis.

References

[1] https://www.kaggle.com/datasets/gpreda/covid19-tweets?resource=download
[2] https://huggingface.co/docs/transformers/model_doc/bert
[3] https://huggingface.co/datasets/tweet_eval
[4] D. Kumar and U. Bansal, "Emotion Sentiment Analysis of Indian Twitter-Data of COVID-19 After Lockdown," 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), 2021, pp. 421-426, doi: 10.1109/ICSCCC51823.2021.9478166.
[5] M. Alhajji, A. Al Khalifah, M. Aljubran, and M. Alkhalifah, "Sentiment Analysis of Tweets in Saudi Arabia Regarding Governmental Preventive Measures to Contain COVID-19," 2020.

[6] A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi, Z. Shah, Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study, J. Med. Internet Res. 22 (4) (2020) e19016, http://dx.doi.org/10.2196/19016

[7] A. Mathur, P. Kubde and S. Vaidya, "Emotional Analysis using Twitter Data during Pandemic Situation: COVID-19," 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 845-848, doi: 10.1109/ICCES48766.2020.9138079.

[8] Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, Zhu T Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach J Med Internet Res 2020;22(11):e20550 URL: https://www.jmir.org/2020/11/e20550 DOI: 10.2196/20550

[9] A. S. Imran, S. M. Daudpota, Z. Kastrati and R. Batra, "Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets," in IEEE Access, vol. 8, pp. 181074-181090, 2020, doi: 10.1109/ACCESS.2020.3027350.

[10] Boris Marinov, Jennifer Spenader, and Tommaso Caselli. 2020. Topic and Emotion Development among Dutch COVID-19 Twitter Communities in the early Pandemic. In Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media, pages 87–98, Barcelona, Spain (Online). Association for Computational Linguistics.