CTC: COVID-19 Tweet Classification using CT-BERT

Shivangi Bithel[1]

[1]*Indian Institute of Technology, Delhi, Hauz Khas, New Delhi, Delhi 110016*

Abstract
CTC is my submitted work to the Information Retrieval from Microblogs during Disasters (IRMiDis) Track at the Forum for Information Retrieval Evaluation (FIRE) 2022. Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. Most people infected with the virus experience a mild to moderate respiratory illness and recover without requiring special treatment. However, some become seriously ill and require medical attention. Vaccines against coronavirus and prompt reporting of symptoms saved many lives during the pandemic. The analysis of COVID-19-related tweets can provide valuable insights regarding the stance of people toward the new vaccine. It can also help the authorities to plan their strategies based on people's opinions about the vaccine and ensure the effectiveness of vaccination campaigns. Tweets describing symptoms can also aid in identifying high-alert zones and determining quarantine regulations. The IRMiDis track focuses on these COVID-19-related tweets that flooded Twitter. I developed an effective classifier for both Tasks 1 and 2. The evaluation score of my submitted run is reported in terms of accuracy and macro-F1 score. I achieved an accuracy of 0.770, a macro-F1 score of 0.773 in Task 1, and an accuracy of 0.820, a macro-F1 score of 0.746 in Task 2. I enjoyed the first rank among other submissions in both the tasks.

Keywords
Sentiment Analysis, COVID-19 Tweets, COVID-Twitter-BERT, Tweet classification

1. Introduction

During the COVID-19 pandemic, the globe has waged its most difficult struggle. The disease was unknown to everyone. It was impossible to determine its exact symptoms. Every time a new variant was identified, it was accompanied by new symptoms. Due to the resemblance of its symptoms to those of the common cold and influenza, this fatal virus has often been misdiagnosed as a cold or the flu. Through social media, many people told their friends and family about their own symptoms or the symptoms of their friends or family members. Not only that, but people tweeted about celebrities and their symptoms to the public. By promptly identifying individuals with COVID-19 symptoms, it is possible to offer them appropriate treatment and prevent the disease's spread.

As the pandemic spread through human contact, it became more difficult to contain. Numerous preventative measures, such as wearing masks and observing a 14-day quarantine, assisted in controlling the spread. There was a rush to develop a vaccine capable of producing the necessary antibodies. The coronavirus vaccine was the only method left that could help

in combating and eradicating infectious illnesses by immunising individuals against viruses. When the vaccine finally came, people began using social media platforms such as Twitter to debate the vaccination as it was being disseminated throughout the world. People had both favourable and negative opinions on the ongoing issues of vaccine advancement, accessibility, effectiveness, and side effects. The government and numerous health groups, such as WHO, would benefit from knowing what people think of the new COVID-19 vaccinations. They could use the insights gleaned from these micro-blogs to develop future initiatives and urge everyone to be fully vaccinated.

Classifying tweets manually is laborious and error-prone. Therefore, there was an urgent need to build machine learning algorithms that can assist us in categorising tweets concerning COVID-19 vaccinations and also tweets that can detect individuals with COVID-19 symptoms. In this paper I present an effective 3-class classifier for classification of COVID-19-related vaccine tweets and a 4-class classifier for classification of COVID-19 symptoms reporting tweets.

2. Task Definition

For Task 1, "Building an effective classifier for 3-class classification on tweets regarding people's stance towards COVID-19 vaccines" and Task 2, "Building an effective classifier for 4-class classification on tweets that can detect tweets that report someone experiencing COVID-19 symptoms", organized as a part of IRMiDis (Information Retrieval from Microblogs during Disasters) Track in the FIRE (Forum for Information Retrieval Evaluation) 2022, I present an effective approach in this paper.

The tweets for Task 1 are classified into 3 classes described below with examples:

- **AntiVax** - the tweet indicates hesitancy (of the user who posted the tweet) towards the use of vaccines.
- **ProVax** - the tweet supports / promotes the use of vaccines.
- **Neutral** - the tweet does not have any discernible sentiment expressed towards vaccines or is not related to vaccines.

An example for each class of tweets has been given below:

- **AntiVax Tweet:** *"Let all politicians and their families be the first to take it. And then lets see how they are doing in 6 months or less. These vaccines take years to make and hopefully get it right! No way this has taken that long to make so i won't be getting one and never will https://t.co/PcFL4NXNZM"*
- **ProVax Tweet:** *"Good News: Pfizer COVID-19 vaccine 90 percent effective in phase 3 https://t.co/cXb4WUZ0VV"*
- **Neutral Tweet:** *"Great thread by @nataliexdean about today's Moderna vaccine news https://t.co/IVqszNrFxm"*

The tweets for Task 2 are classified into 4 classes described below with examples:

- **Primary Reporting** - The user (who posted the tweet) is reporting symptoms of himself/herself.

- **Secondary Reporting** - The user is reporting symptoms of some friend / relative / neighbour / someone they met.
- **Third-party Reporting** - The user is reporting symptoms of some celebrity / third-party person.
- **Non-Reporting** - The user is not reporting anyone experiencing COVID-19 symptoms, but talking about symptom-words in some other context. This class includes tweets that only give general information about COVID-19 symptoms, without specifically reporting about a person experiencing such symptoms.

An example for each class of tweets has been given below:

- **Primary Reporting Tweet:** *"Wondering if I should get tested for covid.. I have had this cough for 2 weeks now, not getting better or worse, also runny nose and headaches.. Just in case..."*
- **Secondary Reporting Tweet:** *"@cdngarbageman Omg David. Me too!! My sister in law just recovered from Covid. It took her two weeks it was like a very mild flu. My brother has a mild cough but tested negative. Its very very serious."*
- **Third-party Reporting Tweet:** *"#Recent #TamilNaduCoronaupdate 18 months old child dead due to corona. Was admitted at Viluppuram government medical College and hospital on 26/06/2019 with symptoms of cough fever breathlessness and was found to be #Corona positive. https://t.co/1NnCkAG9ya"*
- **Non-Reporting Tweet:** *"@trumpwarrior45 Dry cough, shortness of breath, and fever are what to look for. If you have a mucus cough, stuffy/runny nose, that is just a cold. Still a coronavirus, but not COVID-19. Just be mindful of symptoms."*

3. Related Work

Users publish information on micro-blogs such as Twitter for a variety of reasons, including to express their opinions on Coronavirus, inform their connections about their health, report symptoms and cautions of themselves or others they know. People discuss about the COVID-19 vaccines, and vaccination campaigns in large number before getting their dose. The extraction of information from these textual tweets is a common application of social computing. Traditional machine learning techniques such as Naive-Bayes classifier, Linear classifier, Support Vector Machine, and Deep neural techniques such as Long Short Term Memory (LSTMs) and Bidirectional RNN are very effective for text classification. The most current language models for natural language processing are **BERT** (Bidirectional Encoder Representations from Transformers) [1] and its domain-specific version **CT-BERT** (COVID-Twitter-BERT) [2]. **VaccineBERT** [3] is a BERT based model, which performs the task of tweet classification over COVID-19-related vaccine tweets.

4. Dataset

The training dataset provided for the track 1 contains 4392 tweets. 2792 tweets were extracted from [4] on the stance of people towards COVID-19 vaccine crawled between November-December 2020 and remaining 1600 tweets were crawled between March-December 2020 and

were annotated by crowdworkers for the three labels. It contains tweet-texts along with the tweet IDs and the classes. The test dataset contains 500 tweets with tweet IDs and tweet-texts only.

The dataset shared for task 2 contains English tweets from February 2020 - June 2021, crawled using keywords related to COVID-19 symptoms (e.g., 'fever', 'cough'). The training dataset contains 1574 tweet-texts along with the tweet IDs, classified into four classes by human workers. The test dataset contains 400 tweets with tweet IDs and tweet-texts only.

5. Methodology

5.1. Pre-processing

Following [5] and [3], I pre-processed the tweets in order to improve the quality of word embeddings produced by CTC. Tweets generally contain unique lexicons like *HASHTAGS*, *@USER*, *HTTP-URL* and *EMOJIS* which without pre-processing, often reduce the performance of the model. Thus, we used the following data cleaning pipeline as part of pre-processing the tweets in the dataset:

- converted words to their lower case
- carefully removed stopwords such as "a", "an", "the", etc.
- converted emoticons to words using python's 'emoji' library (https://pypi.org/project/emoji/).
- expanded contractions to text using python's 'contractions' library (https://pypi.org/project/contractions/).
- removed non-alphanumeric characters like brackets, colon, semi-colon, @, etc.
- remove URLs from the text using regular expression.

5.2. Model

I experimented with the following transformer-based model:

- **BERT:** It stands for Bidirectional Encoder Representations from Transformers. BERT makes use of Transformer, an attention mechanism to learn contextual relations between words (or sub-words) in a text. Thus the textual representations generated by BERT are very powerful and generalize well to solve many NLP tasks.
- **CT-BERT:** COVID-Twitter-BERT is a transformer-based model, pretrained on a large corpus of Twitter messages on the topic of COVID-19 collected during the period from January 12 to April 16, 2020. CT-BERT is optimised to be used on COVID-19 content, in particular social media posts from Twitter. This model showed a 10–30% marginal improvement compared to its base model, BERT-large, on five different specialised datasets.
- **VaccineBERT:** VaccineBERT is the best performing vaccine tweet classification model from FIRE 2021, IRMiDis Track Task 2. It uses CT-BERT, fine-tuned over the shared training dataset for the classification output.

Fine-tuned CT-BERT model, similar to VaccineBERT performed best on the Validation set.

5.3. Experimental Setup

I first shuffled the training data, then split it into training and validation sets in the ratio 90:10 such that the percentage of instances of each class were preserved in both sets. Both training and validation instances were pre-processed, as explained in section 5.1. The resulting training data was used for fine-tuning CT-BERT[2], similarly as done by VaccineBERT[3] while validation data was used for evaluation. In order to prevent overfitting, I used early stop monitoring the validation loss with patience value 3.

5.4. Prediction

For the prediction over the available test data, I used the fine-tuned CT-BERT model as a text classification model to generate the embeddings for the tweet and then further predicted the probability scores of each tweet against all three classes in Task 1 and all four classes in Task 2. The class having the maximum probability was reported as the predicted class for that tweet. The final prediction file containing the Tweet ID and the predicted class was submitted as run for Task 1 and 2.

6. Results and Discussion

IRMiDis Track results are evaluated using overall accuracy and the macro-F1 score as metrics. The result of my submitted automated run for Task 1 and 2 is shown in Table 1. CTC got the 1st rank among other submissions for both the tasks.

Task	Team_ID	Accuracy	macro-F1 score	Rank
1	Data@IITD	0.770	0.773	1
2	Data@IITD	0.820	0.746	1

Table 1
Result of Task 1 and 2

7. Conclusion and Future Work

In this work, I propose a simple but effective approach to COVID-19 Tweet Classification Task based on Covid-Twitter-BERT, a transformer-based model pre-trained on a large corpus of COVID-19-related tweets. The experimental results showed that my solution achieved an accuracy of 0.770, a macro-F1 score of 0.773 in Task 1, and an accuracy of 0.820, a macro-F1 score of 0.746 in Task 2. CTC is ranked in the first place in Information Retrieval from Microblogs during Disasters (IRMiDis) track at FIRE 2022. For future work, we can experiment with ensembling learning to improve the model accuracy and its robustness.

References

[1] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018. URL: http://arxiv.org/abs/1810.04805.

[2] M. Müller, M. Salathé, P. E. Kummervold, Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter, arXiv preprint arXiv:2005.07503 (2020).

[3] S. Bithel, S. Verma, Vaccinebert: Bert for covid-19 vaccine tweet classification, 2021, pp. 1199–1203. URL: http://ceur-ws.org/Vol-3159/T8-1.pdf.

[4] L.-A. Cotfas, C. Delcea, I. Roxin, C. Ioanăş, D. S. Gherai, F. Tajariol, The longest month: Analyzing covid-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement, IEEE Access 9 (2021) 33203–33223. doi:10.1109/ACCESS.2021.3059821.

[5] S. Bithel, S. S. Malagi, Unsupervised identification of relevant prior cases, 2021. arXiv:2107.08973.