Detecting COVID-19 Vaccine Stance and Symptom Reporting from Tweets using Contextual Embeddings

Akshit Bansal[1], Rohit Jain[1] and Jatin Bedi[1]

[1]*Department of Computer Science and Engineering, Thapar Institute of Information and Technology, Patiala, Punjab, India - 147001.*

Abstract

Determining sentiments of the public with regard to COVID-19 vaccines is crucial for nations to efficiently carry out vaccination drives and spread awareness. Hence, it is a field requiring accurate analysis and captures the interest of many researchers. Microblogs from social media websites such as Twitter sometimes contain colloquial expressions or terminology difficult to interpret making the task a challenging one. In this paper, we propose a method for multi-label text classification for the track of "Information Retrieval from Microblogs during Disasters (IRMiDis)" presented by the "Forum of Information Retrieval Evaluation" in 2022, related to vaccine sentiment among the public and reporting of someone experiencing COVID-19 symptoms. The following methodologies have been utilised: (i) Word2Vec and (ii) BERT, which uses contextual embedding rather than the fixed embedding used by conventional natural language models. For Task 1, the overall F1 score and Accuracy are 0.503 and 0.529, respectively, placing us fourth among all the teams, while for Task 2, they are 0.740 and 0.790, placing us second among all the teams who submitted their work. Our code is openly accessible through GitHub. [1].

Keywords

Text classification, COVID-19, Twitter, Transformer, IRMiDiS

1. Introduction

We all have just faced a life-threatening world pandemic, and the only long-term remedy seems to be through society-scale vaccination drives. Researchers from all over the world have created a number of vaccines to combat the COVID-19 virus. Covaxin, Covishield, and Spikevax are some of the well-known vaccines. Many people, however, are dubious about the use of vaccines for a number of reasons, including the politicisation of the issue and the hasty development of vaccines. There is reluctance among some people to get vaccinated since they believe that vaccines have not been tested and there have not been enough trials to make sure they are safe. To make the vaccination drive a success, it is imperative to understand the sentiments of the public towards vaccines. This task can be accomplished using data from social networking websites like Twitter because most of us use social media to communicate our views and opinions in today's world. In Task 1, data from social media and microblogging websites like Twitter was obtained to understand people's sentiments on vaccinations. The tweets are classified into three categories: "ProVax", "AntiVax", "Neutral" based on their willingness to get vaccinated as

[1]https://github.com/RohitJain001/FIRE-Conference-BERT-Natural-Language-Preprocessing

Forum for Information Retrieval Evaluation, December 9-13, 2022

abansal2_be19@thapar.edu (A. Bansal); rjain1_be19@thapar.edu (R. Jain); jatin.bedi@thapar.edu (J. Bedi)

analysed from their tweets.

Using tweets gathered between February 2 and March 15, 2020, Abd-Alrazaq et. al. [1] used topic modelling and sentiment analysis to identify the critical conversation subjects and sentiments surrounding COVID-19. However, the use of colloquialisms, lack of structure and having to analyse a large amount of data makes the task quite difficult for humans to accomplish. As a result, the classification of tweets about vaccination is a challenging NLP task. Shamrat et al. [2], used NLP to preprocess the data and then applied the KNN classification algorithm to classify the tweets for three different vaccines into “Negative”, “Positive” and “Neutral”. Multiple hashtags were utilised by Jia Xue et al. [3] to collect data for experiments from Twitter. Then the LDA machine learning algorithm was employed to do sentiment analysis on this data. Fear was discovered to be an important characteristic in the discussion of COVID-19.

In task 2 of the “Information Retrieval from Microblogs during Disasters” track, we specifically explore if tweets that report someone experiencing COVID-19 symptoms can be automatically identified. It focuses on classifying these tweets into four categories: “Primary Reporting”,” Secondary Reporting”, “Third-party Reporting” and “Non-Reporting”. Our team has contributed to the problem as part of the shared task effort by cleaning the tweets and trying various NLP models including BERT and Word2Vec to classify the tweets.

We have made the following contributions in context of the problem of the shared tasks:

- A transformer model i.e. BERTl which uses bidirectional word embeddings as opposed to the fixed embeddings used by traditional Natural Language Processing models is implemented to accurately identify the target tweets class.
- Word2Vec was also implemented as a part of the effort. The results showed that models using contextual embeddings give relatively better results when compared to models using fixed embeddings.

2. Related Work

On the basis of tweets and facebook postings, several analyses of the public’s opinions on the COVID-19 vaccination have been made. Samuel, et al. [3] use Naive Bayes and Logistic Regression techniques to classify tweets based on public sentiment insights from the COVID-19 dataset. They used a set of 24000 tweets based on news about COVID-19 vaccines in India and the Naive Bayes model provided a classification accuracy of 91% while logistic regression achieved an accuracy of 74%. The SAB-LSTM model developed and used by Kumar et al. [4] is an original bidirectional LSTM extension (SAB-LSTM)-based multi-class sentiment analysis model.

About 4.5 million tweets from the COVID-19 pandemic were the subject of a research in Europe. Multilingual language embeddings were used to do sentiment analysis using neural networks [5]. To identify anti-vaccination tweets, the performance of Bidirectional long short-term memory networks with pre-trained GLoVe embeddings (Bi-LSTM) and the BERT model were evaluated by Quyen et al. [6] in comparison to traditional machine learning techniques like Support Vector Machine (SVM) and Naive Bayes. Their results show an accuracy of 91.6% and a F1 score of 95.5% with BERT. Bi-LSTM model performance showed an accuracy of 89.8% and F1 score of 45.5% . Support Vector Machine and K Nearest Neighbour machine learning models are employed by Adamu et al. [7] to categorise and assess the tweets into three categories: “positive, “negative,” and “neutral”.

3. Dataset

3.1. Task 1

The dataset provided contained 4392 tweets for training and 717 tweets for testing. Every tweet in the dataset was labelled as one of the 3 stances related to vaccine sentiment expressed in the tweet, as explained below:

1. **AntiVax** - The tweet expresses reluctance on the part of the user who tweeted it to get vaccinated against COVID-19.
2. **ProVax** - The tweet endorsing or promoting the use of vaccinations.
3. **Neutral** - The tweet does not clearly reflect any opinion about vaccines or has nothing to do with vaccines.

Table 1
Proportion of each target class in the Dataset for Task-1

Label Category	Class Count	Dataset %age
Provax	1676	38.16%
Antivax	1081	24.61%
Neutral	1635	37.23%
Total	**4292**	100%

3.2. Task 2

The dataset provided contained 1574 tweets for training and 400 tweets for testing. Every tweet in the dataset was labelled as one of four classes, which identifies reporting of COVID19, as explained below:

Table 2
Proportion of each target class in the Dataset for Task-2

Label Category	Class Count	Dataset %age
Non-Reporting	814	51.72%
Third-Party Reporting	196	12.45%
Primary Reporting	437	27.76%
Secondary Reporting	127	8.07%
Total	**1574**	100%

1. **Primary Reporting** - The tweet’s submitter is outlining personal symptoms.

2. **Secondary Reporting** - The user is outlining the symptoms of a friend, family member, neighbour, or recent acquaintance.
3. **Third-party Reporting** - A third party or famous person's symptoms are being referenced by the user.
4. **Non-Reporting** - The user is not claiming that anyone is displaying COVID-19 symptoms and is instead utilising symptom-words in a different context. This collection of tweets only offers general information on COVID-19 symptoms; no specific people who may be experiencing these symptoms are mentioned.

4. Methodology

The tweets contain some colloquial expressions and terminology which might be difficult to interpret; because of which, it could be challenging to determine the stance of a tweet. Hence, an efficient method is to be established so these kinds of tasks can be accurately performed.

We use a method which focuses on learning the semantic relationships between the sentence tokens of tweets and map them with the given label. The work was seen as an extension of a multi label phrase classification problem. Task 1 required the prediction of three labels linked to public opinion on vaccines, and Task 2 required the prediction of four labels pertaining to the reporting of someone who had COVID-19 symptoms. After analysis of data during text Exploratory Data Analysis, the model's accuracy was improved with effective stop-word removal including removal of all special characters and website URLs.

We used these preprocessed tweets to train specialised algorithms such as Word2vec and BERT (as shown in Figure 1) to efficiently handle sentence classification. The process of a deep learning-based transformers technique was applied. It helped in getting the correct interpretation of the tweets and efficiently classifying them as one of the labels. For getting optimised results from the BERT model, we tuned the hyperparameters including the max length limit of the tokens, the training batch size parameter to determine how many tweets were processed at once before the model was updated, the validation batch size and the learning rate of the model. We chose the max length parameter to be 256, train and valid batch size to be 32 and the learning rate is 10-5. We chose BCE with Logits Loss in Pytorch, a combination of a Sigmoid layer and BCE Loss into a single class which is numerically stable, as our loss function and kept saving the checkpoint at each epoch if the accuracy improved. Using the best saved model that we obtained after running the epochs, the tweets were classified into respective labels.

In Word2Vec methodology (as shown in Figure 1), initially a unique word dictionary was created using the Keras Tokenizer() and the text_to_sequence() function was used to get a list of lists where each value represent the index of that word in a unique word dictionary. Now as every row has a different dimension, so the pad_sequences() method is used to add paddings and generate an embedding matrix. The embedding matrix was fed to embedding layers in Keras as weights. The input dimensions of the embedding layer is the total number of unique words, and the output dimensions are the embedding vector dimensions. We received the word embeddings as the output from the Keras embedding layer. Furthermore, the Long Short Term Memory, Recurrent Neural Network Model was trained with the word embeddings and output

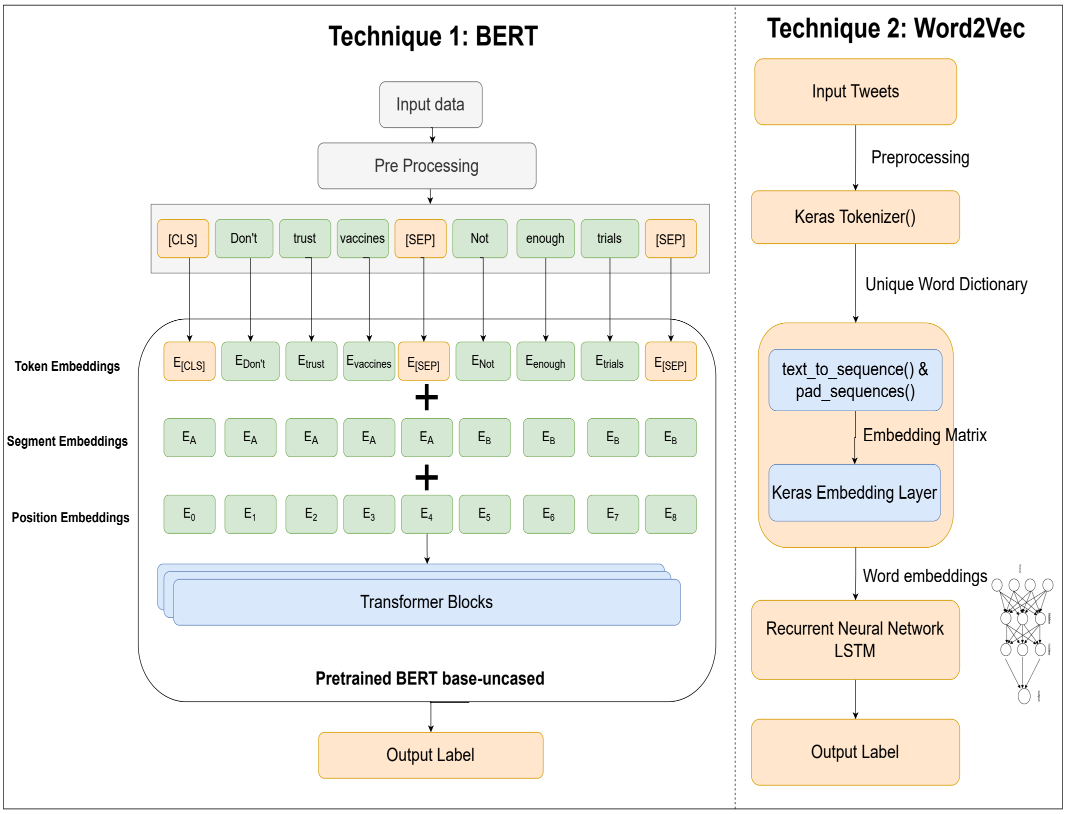

Figure 1: Working methodology of the techniques applied for Task-1 and Task-2

was evaluated.

4.1. Data Preparation

Before training the model, the data was thoroughly cleaned for the right text analysis and to remove noise. For preparation, we used stop-word removal and tokenization. The words were initially divided up into smaller parts called "tokens." The text was changed to lowercase, the urls were taken out, and words like "the", "a", "and", "is", and "will"—common stopwords in English—were removed. This was done since none of these help to the goal of text classification.

4.2. Modelling

BERT [8] is a bidirectional language model that utilises the transformer architecture to learn the contextual relationships between words. We make use of pre-trained models that have been officially released. BERT accepts either a single sentence or a text pair as input. The special categorization token [CLS] comes first in each sequence, then WordPiece tokens for the first text, a separator token [SEP], and (not mandatorily) WordPiece tokens for the second text. BERT uses positional embeddings along with token embeddings to indicate where a token is in the sequence. During training, BERT employs the Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) objectives.

Word2Vec: Mikolov [9] proposed the Word2vec word embedding technique in 2013 for word expression, which takes into account the context and meaning of words in a text. It includes continuous bag-of-words (CBOW) and skip-gram learning methods. It is employed in numerous investigations, including emotion analysis, emotion categorization, and event detection because word2vec's cosine similarity of word vectors contains the meaning of words in the document more than other word embedding approaches do. Convolution Neural Networks performs exceptionally well for categorization in the field of natural language processing.

5. Experimentation and Results

The test data was tested using the proposed methodology and multiple runs were submitted. We experimented with various models including BERT and Word2Vec; and also tested for varying number of epochs and learning rates for the BERT model. After a thorough analysis of the test dataset processed by our proposed methods, the best results were obtained on using BERT; with the optimal hyperparameters coming out to be 15 training epochs and a learning rate of 10^{-5}. As mentioned earlier, multiple runs were done using different learning rates and number of epochs, but BERT outperformed Word2Vec every single time and generated optimal results. BERT performs better as compared to Word2Vec since it uses a bidirectional transformer. As compared to traditional language models like OpenAI GPT and Word2Vec which take the previous 'n' tokens and predict the next token, BERT trains in such a manner which takes into account both the previous and the next tokens making it deeply bidirectional. Word2Vec uses fixed word embeddings, not taking into account its neighbouring tokens, which hinders its ability to capture the contextual understanding of the sentence.

As mentioned in Table 3, training and validation accuracies for Task-1 turned out to be 0.986 and 0.942 respectively with BERT and in case of Word2Vec, training and validation accuracies were 0.582 and 0.524 respectively. Similarly, as mentioned in Table 4, we achieved a training accuracy of 0.991 and validation accuracy of 0.975 with BERT for task 2. With Word2Vec, the training accuracy was 0.534 and validation accuracy was 0.480.

Table 3
Training and Validation Accuracy on Task-1

Task 1	Training Accuracy	Validation Accuracy
BERT	0.991	0.975
Word2Vec	0.534	0.480

Table 4
Training and Validation Accuracy on Task-2

Task 2	Training Accuracy	Validation Accuracy
BERT	0.986	0.942
Word2Vec	0.582	0.524

Table 5
Testing Results for Task 1

Task 1	Method	Macro F1 Score	Accuracy
Run 1	BERT with 15 epochs	0.503	0.529
Run 2	BERT with 20 epochs	0.482	0.510
Run 3	Word2Vec	0.401	0.402

Table 6
Testing Results for Task 2

Task 2	Method	Macro F1 Score	Accuracy
Run 1	BERT with 18 epochs	0.740	0.790

For Task-1, Word2Vec gave an accuracy of 0.402 and a macro F1 score of 0.401 on the test dataset (shown in Table 5). Observing the results of Task-1, when we compare the best run of BERT (with 15 epochs) which gave a F1 score of 0.503, with the run with 20 epochs which gave a macro F1 score of 0.482, we can observe a direct improvement of 4.35% has been achieved. The result for the model with 20 epochs is low in comparison since the model is a victim of over-fitting. As mentioned in table 4, for task 2, we attained an accuracy of 0.790 and a macro F1 score 0.740 on a test dataset using BERT with 18 epochs which again turned out to be better than Word2Vec due to its ability to better understand the contextual meaning of the sentence.

6. Conclusion

Task 1 of the challenge aimed to classify tweets based on sentiments for vaccination against COVID-19 into three categories, namely "ProVax", "AntiVax" and "Neutral". For Task 2, our aim was to classify whether a tweet was reporting symptoms of COVID-19 into four categories - "Primary Reporting", "Secondary Reporting", "Third-Party Reporting" and "Non-Reporting". After preprocessing the tweets and implementing stop word removal, we applied the BERT model which makes use of contextual embeddings to classify the tweets. In addition, we also implemented the Word2Vec model which makes use of fixed embeddings for labelling the tweets. We conclude from our experiments, which were carried out on corpora of around 4300 tweets for Task1 and around 1600 tweets for Task2, that BERT is the superior algorithm for analysing and classifying tweets based on COVID-19 and is significantly more capable of capturing the underlying meaning in the best way possible. Multiple labels in the corpus make it challenging for different models to perform well. However, the task was made easier to tackle by the use of deep learning techniques and the bidirectional embeddings of BERT. In further work, we aim to augment our current work with better hyperparameter modelling (RoBERTa, ERNIE) and optimization techniques such as Bayesian Optimization and Population based Training.

References

[1] A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi, Z. Shah, et al., Top concerns of tweeters during the covid-19 pandemic: infoveillance study, Journal of medical Internet research 22 (2020) e19016.

[2] F. Shamrat, S. Chakraborty, M. Imran, J. N. Muna, M. M. Billah, P. Das, O. Rahman, et al., Sentiment analysis on twitter tweets about covid-19 vaccines using nlp and supervised knn classification algorithm, Indonesian Journal of Electrical Engineering and Computer Science 23 (2021) 463–470.

[3] J. Samuel, G. Ali, M. Rahman, E. Esawi, Y. Samuel, et al., Covid-19 public sentiment insights and machine learning for tweets classification, Information 11 (2020) 314.

[4] D. A. Kumar, A. Chinnalagu, Sentiment and emotion in social media covid-19 conversations: Sab-lstm approach, in: 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), IEEE, 2020, pp. 463–467.

[5] A. Kruspe, M. Häberle, I. Kuhn, X. X. Zhu, Cross-language sentiment analysis of european twitter messages duringthe covid-19 pandemic, arXiv preprint arXiv:2008.12172 (2020).

[6] Q. G. To, K. G. To, V.-A. N. Huynh, N. T. Nguyen, D. T. Ngo, S. J. Alley, A. N. Tran, A. N. Tran, N. T. Pham, T. X. Bui, et al., Applying machine learning to identify anti-vaccination tweets during the covid-19 pandemic, International journal of environmental research and public health 18 (2021) 4069.

[7] H. Adamu, M. J. B. M. Jiran, K. H. Gan, N.-H. Samsudin, Text analytics on twitter text-based public sentiment for covid-19 vaccine: A machine learning approach, in: 2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), IEEE, 2021, pp. 1–6.

[8] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[9] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781 (2013).