VaxTweetClassifier: BERT for Dealing with Vaccination Tweets

Shankha Shubhra Das[1], Sohan Choudhury[2], Priyam Saha[3] and Dipankar Das[4]

[1]*Jadavpur University, Jadavpur, Kolkata 700032*
[2]*Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha 751024*
[3]*Jadavpur University, Jadavpur, Kolkata 700032*
[4]*Jadavpur University, Jadavpur, Kolkata 700032*

Abstract

Vaccination is one of the most important prerequisites to fight against all the downsides of various diseases like COVID-19. We propose a modified BERT model, VaxTweetClassifier, to classify tweets into 12 different categories such as Mandatory, Conspiracy, Side-Effect, etc. VaxTweetClassifier is our submitted work for the task Artificial Intelligence on Social Media (AISoMe) in FIRE 2023 [1]. We have built a robust multi-label classifier that categorizes tweets according to the specific concern(s) they are exhibiting. These types of classification tools can help in analyzing the sentiment of the public in general and also Governments of different countries with respect to the acceptance of vaccination. They can get a rough idea about what people are thinking, whether fake news is spreading or not, if the people have a good or bad point of view about the whole process, etc. For this task, we have chosen threshold-based criteria to select the final categories for the aforementioned tweets. The evaluation score of our submitted run is reported in terms of macro-F1 score and Jaccard metric. We achieved a macro-F1 score of 0.54 and a Jaccard score of 0.58.

Keywords

Sentiment Analysis, COVID-19-BERT, Vaccination Tweets, Multi-label Classifier

1. Introduction

Vaccination is emerging as a critical tactic in the continuous fight to protect public health to reduce the dangers and spread of many diseases. Recent events highlight the crucial role that immunization plays in limiting the COVID-19 pandemic's widespread effects. However, the need for mass vaccination goes beyond the current emergency and includes a wider range of disorders like childhood illnesses, recurring flu outbreaks, and more. Nevertheless, the skepticism that clouds the vaccination adoption landscape is a complicated scenario made of political machinations, worries about potential adverse effects, and other nuanced issues. Understanding the various issues that surround vaccines is essential to effectively addressing this complex topic. In this age of digital communication, the breadth of social media appears as

Forum for Information Retrieval Evaluation, December 15-18, 2023, India

shankhasdas07@gmail.com (S. S. Das); sohan2004cc@gmail.com (S. Choudhury); priyam.saha2003@gmail.com (P. Saha); dipankar.dipnil2005@gmail.com (D. Das)
https://www.dasdipankar.com/ (D. Das)

a formidable library of viewpoints, enabling the quick gathering of insights on vaccine-related conversations.

People throughout the world have different perspectives about vaccination, which they communicate on social media channels like Twitter. A single tweet can convey a variety of viewpoints and gently impact the perceptions of others. Some of these tweets may not be in line with the good purpose of vaccination and should not be distributed widely. As a result, there is a pressing need to develop machine-learning algorithms capable of categorizing COVID-19 vaccine-related tweets.

In this research article, section 2 describes the details of this shared task whereas the model-based related works are discussed in section 3. In sections 4 and 5, we have mentioned the dataset and pre-processing steps respectively. In sections 6 and 7, we have provided our proposed methodology and evaluation of the same. Finally, section 8 concludes the paper.

2. Task

The primary goal at hand is to create a strong multi-label categorization system capable of assigning meaningful labels to tweets. These labels are intended to encompass the authors' wide range of vaccine-related concerns. It's worth noting that a single tweet may have many categories, indicating the author's plethora of concerns.

The classification task's taxonomy of concerns covers the following:

1. **Unnecessary:** The tweet implies that immunizations are unneeded or that other treatments are preferable.
2. **Mandatory:** Against compulsory vaccination — According to the tweet, immunizations should not be made mandatory.
3. **Pharma:** Against Big Pharma — The tweet implies that the Big Pharmaceutical firms are only interested in making money, or that the tweet is against such corporations in general due to their past.
4. **Conspiracy:** Deeper Conspiracy — The tweet implies a deeper conspiracy, other than Big Pharma's desire to profit (e.g., vaccinations are being used to track individuals, COVID is a fraud).
5. **Political:** The political aspect of vaccinations — The tweet expresses worry that governments/politicians are using vaccines to further their agenda.
6. **Country:** Country of origin — The tweet is critical of vaccination because of where it was created or manufactured.
7. **Rushed:** Untested / Hasty Process – The tweet raises the worry that the vaccinations have not been adequately tested or that the reported data is incorrect.
8. **Ingredients:** Vaccine ingredients/technology — The tweet shows worry about the vaccine components (for example, fetal cells, chemicals) or the technology utilized (for example, mRNA vaccines can affect your DNA).
9. **Side-effect:** Negative Effects/fatalities — The tweet indicates worry about the vaccinations' negative effects, including fatalities.
10. **Ineffective:** Vaccine is Ineffective — The tweet conveys worry that vaccinations are ineffective and worthless.

11. **Religious:** Religious grounds – The tweet is anti-vaccines for religious grounds.
12. **None:** No specific reason stated in the tweet, or some reason other than the given ones.

3. Related Work

People frequently utilize social media platforms such as Twitter for several objectives, including sharing their ideas and feelings about COVID-19, vaccination, and related initiatives. Extraction of useful information from textual tweets is a key aspect of social computing. Basic machine learning approaches such as the Naive-Bayes classifier, Linear classifier, and Support Vector Machine are insufficient for such difficult problems. Advanced deep learning approaches, such as Long Short Term Memory (LSTMs) and Bidirectional Recurrent Neural Networks (RNNs) have shown great success in text categorization tasks. Recent advances in natural language processing have resulted in the introduction of models such as BERT (Bidirectional Encoder Representations from Transformers) [2] and its specialized variations such as BERTweet [3] and RoBERTa [4].

3.1. BERT

BERT, which stands for Bidirectional Encoder Representations from Transformers, is a machine learning breakthrough for natural language processing (NLP). BERT is a flexible model developed by Google AI Language researchers in 2018 that can handle 11+ basic NLP tasks, including sentiment analysis and named entity identification. Historically, computers have struggled with language understanding. NLP, a synthesis of linguistics, statistics, and Machine Learning, tries to fill this need. Before BERT, NLP tasks necessitated the use of specialized models. BERT revolutionized NLP with its unified approach and exceptional performance by succeeding at various tasks [5].

3.2. BERTweet

BERTweet is a game-changing advancement in natural language processing designed exclusively for English Tweets. BERTweet, which is distinguished by its architectural closeness to BERT-base as proposed by Devlin et al. in 2019, is trained using the RoBERTa pre-training approach described by Liu et al. in 2019. Extensive testing confirms BERTweet's superiority against tough competitors, including RoBERTa-base and XLM-R-base, as clarified by Conneau et al. in 2020. Notably, BERTweet outperforms prior state-of-the-art models in three critical Tweet NLP tasks: part-of-speech tagging, named-entity identification, and text categorization. This achievement highlights BERTweet's pioneering role in improving the efficacy of NLP in the Twitter conversation space.

3.3. RoBERTa

RoBERTa, a BERT refinement, improves critical hyperparameters by removing the next-sentence pretraining target and increasing mini-batch size and learning rates. Language model pretraining has resulted in significant performance increases, but comparing different techniques is difficult due to different datasets and computing costs. This work replicates the pretraining of BERT, indicating that BERT was under-trained and capable of outperforming succeeding models. The top model obtains state-of-the-art performance in GLUE, RACE, and SQUAD, highlighting the significance of previously missed design decisions and raising concerns about recent advancements. The researchers have made their models and code public for review.

4. Dataset

We were given a training dataset from a previous paper titled ”CAVES: A Dataset to facilitate Explainable Classification and Summarization of Concerns towards COVID Vaccines” (the current version is available on Arxiv [6], while the original article is available on the ACM site [7]. The dataset contains 9921 tweets that were anti-vaccination for COVID-19 in 2020-21 [8]. Human annotators labeled each tweet with one or more labels.

As a test set, we were given 500 tweets without labels. It goes beyond conversations regarding COVID-19 vaccinations to include debates about other types of vaccines such as the MMR vaccine and the Flu vaccine.

4.1. Trends in the dataset

- The dataset is divided into 12 categories.
- With 3,805 occurrences, ’side-effect’ is the most common category.
- With 64 instances, ’Religious’ is the least frequent category.
- ’Pharma,’ ’Ineffective,’ ’Rushed,’ and ’Mandatory’ are moderately common.
- ’None’ is related to 629 items, suggesting that the data is neutral or unlabeled.
- The categories ’Conspiracy’ and ’Political’ imply talks on potentially contentious issues.
- ’Pharma’ and ’Side-effect’ are major health-related categories.
- The ’Mandatory’ and ’Country’ sections hint at social effect considerations.
- ’Ineffective,’ ’Rushed,’ ’Unnecessary,’ and ’Conspiracy’ are examples of unfavorable attitudes.

5. Pre-processing

The tweets in the dataset have been processed by removing irrelevant expressions in order to improve the efficiency of word embeddings produced by the BERT model. Our tweets contain links, emojis, punctuation marks, and Twitter embeds(images or GIFs) that may affect the performance of the BERT model.

We have used the following processing pipeline for cleaning the tweets:

1. **Conversion of words to lowercase:** All the words in the tweet are converted to lowercase to enhance the classification process
2. **Conversion of emojis to words:** Emojis and emoticons in tweets significantly contribute to the context of the tweet and can be valuable for sentiment analysis. We converted them into the :emojiname: format by using the ‘emoji’ library [9]. The colons have been removed by the following process.

3. **Removal of non-alphanumeric characters:** We used Python's regular expression or 're' library to remove irrelevant characters such as colons, semi-colons, brackets, hashtags, etc. We also got the raw words by removing the colons from the processed emojis.
4. **Removal of URLs:** The URLs present in the tweets have been removed by using the regular expression library.

Let us consider the following tweet:

Input: #VAERS 16y♀#Pfizer died pulmonary embolism suspected #Cardiacarrest #vaccinedeaths #adolescent Symptoms: Cardiac arrest, Circulatory collapse, Computerised tomogram thorax abnormal, Death, Lung assist device therapy, Pulmonary embolism https://t.co/dTAd-jik8gO

Output: vaers 16y **female sign** pfizer died pulmonary embolism suspected cardiacarrest vaccinedeaths adolescent symptoms cardiac arrest circulatory collapse computerized tomogram thorax abnormal death lung assist device therapy pulmonary embolism

6. Methodology

6.1. Model

BERT-BASE-UNCASED [10]: It is a transformers-based model that is pre-trained on a very large data corpus in English and it is self-supervised. The uncased model was trained on lowercase text and this reduces the model's complexity and makes it easier to train and use. It was pre-trained with methods:

1. **Masked Language Modelling (MLM):** A sentence is taken and the model masks 15% of the sentence randomly. Then the sentence is run through the model which predicts the masked words.
2. **Next Sentence Prediction (NSP):** This pre-training method helps the BERT model to understand longer-term dependencies across sentences. Two sentences are provided as input and the BERT model predicts the sequence of the sentences.

6.2. Experimental Setup

The tweets are first tokenized and the labels are converted into a binary matrix which can be fed into the model for classification. We used sklearn's MultiLabelBinarizer for this task.

For instance, in our current task, we have 12 labels for tweet classification. The labels in the training will be converted into a 1x12 matrix.

The tokenized tweets have three parts - input_ids, attention_mask, and segment_ids. The input_ids include the raw tokenized tweets, and attention_masks are used to distinguish between the relevant input tokens and the tokens that need to be ignored. Finally, the segment_ids help the model to distinguish between different parts of the input token such as different sentences.

The aforementioned input nodes are fed into a keras layer which embeds and encodes the input tokens into a dense vector. Our next layers consist of two dense layers one having 128 units and the next having 64 units. These dense layers process the embeddings produced by the Keras layers and produce a new set of features.

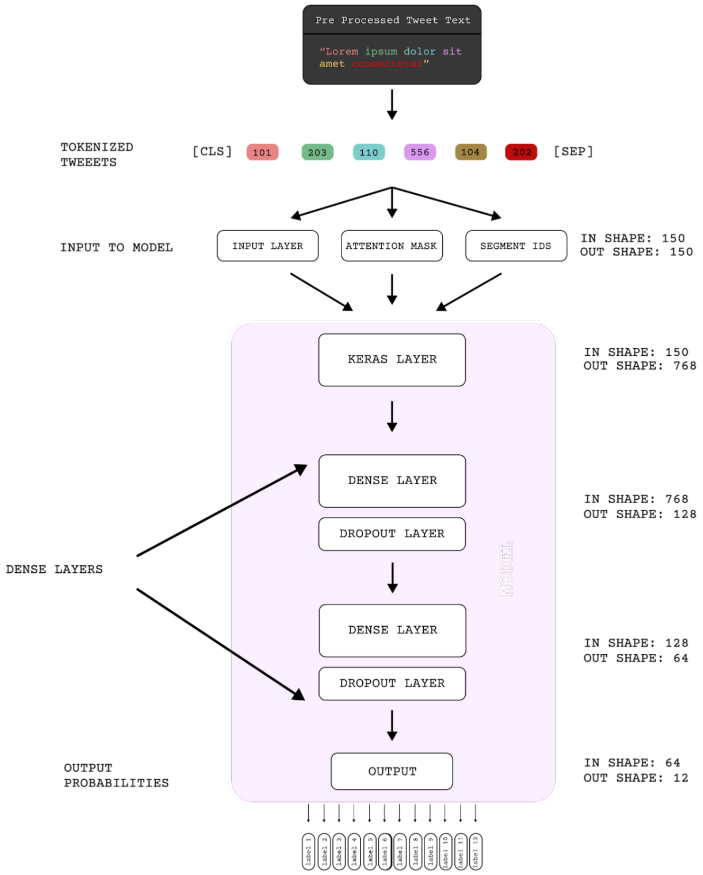

Figure 1: The Graphical Representation of The Model

The output layer, also a dense layer, has 12 units as we have 12 labels for classification. It takes the features produced by the two dense layers as input and produces a set of probabilities, one for each label.

In order to prevent overfitting in the training process, dropout layers are used after every dense layer while fine-tuning the model.

Figure 1 shows the graphical representation of our model.

6.3. Prediction

We used the fine-tuned BERT model to generate the embeddings for the tweets and to generate the probability scores of each tweet against the twelve classes. For a tweet, the classes having probabilities greater than the threshold i.e., 0.25 were treated as the predicted classes for that tweet. The final prediction file containing the tweet IDs and predicted classes was submitted in the requested format.

Run file	Summary of Methodology	Macro-f1	Jacc
run1.csv	Overall model is the same as discussed above and for this run, We set the learning rate at 1e-4	0.54	0.58
run3.csv	Overall model is the same as discussed above and for this run, We set the learning rate at 1e-6	0.5	0.55
run2.csv	Overall model is the same as discussed above and for this run, We set the learning rate at 1e-8	0.4	0.5

Table 1
Result of Task AISoMe

7. Evaluation

The AISoMe task involves categorizing data into multi-classes and multi-labels. There are two measures to evaluate the performance: the macro-F1 score and the Jaccard Coefficient, which are applied to 12 unique categories. Table 1 shows the results of our submission for this challenge. The VaxTweetClassifier placed 20th out of all entries, with a macro-F1 score of 0.54 and a Jaccard Coefficient score of 0.58.

8. Conclusion and Future Work

In this work, we used VaxTweetClassifier, a transformer-based model that was pre-trained on a large dataset of anti-vaccination tweets related to COVID-19. Our major goal is to categorize tweets into 12 unique groups. Surprisingly, our results show that the transformer-based model outperforms classic natural language processing classifiers like Naive Bayes, Logistic Regression, and Support Vector Machine. This advantage is due to the transformer model's ability to produce more expressive word embeddings, which results in improved performance on the task at hand.

Given that transformer-based models have a strong thirst for data, we recommend exploring data augmentation approaches to improve our model's performance. In addition, we propose the use of adversarial training to improve the model's resilience and robustness. These advancements may improve our categorization accuracy even further.

References

[1] S. Poddar, M. Basu, K. Ghosh, S. Ghosh, Overview of the fire 2023 track:artificial intelligence on social media (aisome), in: Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation, 2023.

[2] J. Devlin, K. L. Ming-Wei Chang, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding (2018). URL: https://arxiv.org/abs/1810.04805.

[3] T. V. Dat Quoc Nguyen, A. T. Nguyen, Bertweet: A pre-trained language model for english tweets (2020). URL: https://arxiv.org/abs/2005.10200.

[4] Y. Liu, N. G. Myle Ott, M. J. Jingfei Du, O. L. Danqi Chen, L. Z. Mike Lewis, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach (2019). URL: https://arxiv.org/abs/1907.11692.

[5] B. Muller, Bert 101 - state of the art nlp model explained (2022). URL: https://huggingface.co/blog/bert-101.

[6] S. Poddar, N. G. Azlaan Mustafa Samad, Rajdeep Mukherjee, S. Ghosh, Caves: A dataset to facilitate explainable classification and summarization of concerns towards covid vaccines (2022). URL: https://arxiv.org/abs/2204.13746.

[7] S. Poddar, N. G. Azlaan Mustafa Samad, Rajdeep Mukherjee, S. Ghosh, Caves: A dataset to facilitate explainable classification and summarization of concerns towards covid vaccines (2022). URL: https://dl.acm.org/doi/abs/10.1145/3477495.3531745.

[8] S. Poddar, A. M. Samad, R. Mukherjee, N. Ganguly, S. Ghosh, Caves: A dataset to facilitate explainable classification and summarization of concerns towards covid vaccines, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 3154–3164.

[9] T. Jalilov, Emoji for python (2014). URL: https://github.com/carpedm20/emoji.

[10] 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). URL: https://huggingface.co/bert-base-uncased.

A. Online Resources

The sources for the ceur-art style are available via

- GitHub,
- Overleaf template.