COVID-19 vaccine stance classification from tweets

Sumanth Balaji[1], Sagar Joshi[1], Aditya Hari[1], Abhijeeth Singam[1] and Vasudeva Varma[1]

[1]International Institute of Information Technology, Hyderabad

Abstract

Since the discovery and betterment of vaccines for human diseases, Anti-Vaccine rhetoric and resistance have been prevalent in social circles. These sentiments adversely affect the effectiveness of preventing the contraction of deadly contagious diseases, such as COVID-19. With the advent of social media platforms, the expression of anti-vaccine stances has a far greater reach in society. In this paper, we tackle the task of COVID-19 vaccine stance detection to gauge people's receptiveness towards vaccines and subsequently understand the effectiveness of the vaccination drives.

Keywords

COVID-19 vaccine, AntiVax, Stance classification, Twitter analysis

1. Introduction

In this paper, we tackle the task of COVID-19 vaccine stance detection from tweets.

After the introduction of vaccines for COVID, a considerable portion of people displayed skepticism. This was displayed on social media websites such as the micro-blogging website Twitter. To track and analyze public sentiments toward vaccines, we explore the stance expressed by tweets towards the COVID-19 vaccine. Given a singular tweet item, our task is to classify each tweet into one of three categories:

- 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

2. Literature Review

In the following, a short literature review regarding stance detection, particularly during COVID on Twitter, is conducted in order to underline the current state-of-the-art approach. We see the effectiveness of simple pre-processing with Ekphrasis and Bert transformer [1] in detecting the COVID vaccine stance of tweets from [2]. On general COVID misinformation, [3] pretrains on a balanced collection of COVID-related datasets to improve knowledge. The best results are

Forum for Information Retrieval Evaluation, December 9-13, 2022, India
sumanth.balaji@research.iiit.ac.in (S. Balaji); sagar.joshi@research.iiit.ac.in (S. Joshi); aditya.hari@research.iiit.ac.in (A. Hari); abhijeeth.singam@students.iiit.ac.in (A. Singam); vv@iiit.ac.in (V. Varma)

found by making use of a Siamese BERT network to obtain the sentence representation given a pair of a Tweet and the misinformation item to compare with. Recent work towards leveraging unannotated data as well as experimenting with the aspect topics of tweets for vaccine attitude detection was explored by [4]. The paper experiments with a novel semisupervised approach for vaccine attitude detection called VADET.

3. Dataset

The datasets we make use of are the following:

- Cotfas dataset: Cotfas et.al [2] provided a dataset containing stances of tweets towards COVID-19 vaccines, crawled between November-December 2020. We make use of the entire dataset for our training.
- IRMiDis Task dataset: The organizers of IRMiDis Fire 2022 provide a dataset of crawled tweets between March-December 2020 with various vaccine-related keywords. This dataset is also made use of for the training set.
- VADet dataset: Vaccine stance detection dataset introduced by Lixing Zhu et.al [4] consists of 2800 tweet ids annotated for stance on vaccine as well as aspect and opinion spans. We make use of the tweet ids and stance annotations to bolster our training dataset with more samples.

The test dataset consists of 717 tweet ids and text provided by the organizers.

Each of the tweets was hydrated to retrieve tweet text and user description using Hydrator[1]. Multiple instances of tweets or tweet authors/accounts not existing anymore were encountered, causing a loss of around 20% of tweets as the tweet text was not retrievable in these cases.

It is to be noted that vaccine stance correlates strongly with the political leaning of a person([5], [6]). Hence, for our experiments, we also extract the user description (i.e., bio of the author of each tweet). This is because of a hypothesis that keywords in the author's biography might clue the model towards the user's political leaning and, consequently, their stance concerning vaccines.

The size and split of the final dataset used can be seen in Table 1

Train	Validation	Test
5594	988	717

Table 1
Dataset size

4. Experiment

We carry out experiments using the CovidTwitter Bert [7] for encoding tweets. This is a Bert model pretrained on COVID-related Twitter tweets, allowing it to have a better domain

[1]https://github.com/DocNow/hydrator

understanding. For each of the experiments, we evaluate our model on a validation set, and use either F1 or accuracy metric to determine the best performing checkpoint. Table 2 to Table 4 show the results on the validation set using different configurations we tried using the Covid-Twitter-Bert. The best performing configuration was used to evaluate on the test dataset provided by the shared task organizers, the results of which can be seen in Table 5.

Experiment	F1	Accuracy
$Preprocessed_{accuracy}$	0.55	0.74
$Unprocessed_{accuracy}$	0.65	0.80
$Preprocessed_{loss}$	0.55	0.74
$Unprocessed_{loss}$	0.65	0.80

Table 2
With and without preprocessing

Experiment	F1	Accuracy
$Description_{accuracy}$	0.55	0.74
$NoDescription_{accuracy}$	0.65	0.75
$Description_{loss}$	0.23	0.53
$NoDescription_{loss}$	0.65	0.80

Table 3
With and without author description

Experiment	F1	Accuracy
$NoDescription_{unprocessed+loss}$	0.79	0.85

Table 4
After Hyperparameter tuning

Experiment	F1	Accuracy
$NoDescription_{unprocessed+loss}$	0.528	0.582

Table 5
Test results

4.1. Model objective metric

In each of the experiments, we investigated to see if maximizing accuracy works as a better objective or minimizing the cross entropy loss works better for the classification task. The models optimized on either of the metrics are noted with the corresponding subscript (*accuracy* or *loss*) in the result tables.

4.2. Preprocessing

In order to see the effect of preprocessing or cleaning of twitter text, we ran the experiments on preprocessed as well raw, unprocessed data. For preprocessing, we experimented with preprocessing of tweets through Ekphrasis[2] which is a tool that allows for tokenization, word normalization, word segmentation (for splitting hashtags), and spell-checking on Twitter data. The results of models trained on preprocessed and unprocessed data can be seen in Table 2.

4.3. Using User Description

We run experiments to see if the corresponding Twitter author's description helps the model detect the author's stance towards the COVID vaccine in their tweet. The results of models trained on Description and NoDescription can be seen in Table 3.

For all the experiments, training is done using the HuggingFace Trainer API for 4 epochs with batch size 4 and learning rate $1.5e^{-5}$.

5. Analysis

We find the following:

- Preprocessing of tweets does not contribute to improvement in metrics of accuracy and F1. This could be caused due to the loss of semantic information neccessary for the classification which is present in raw tweets, or the relatively small amount of tweet data available for training, hence not being sufficient for the model to learn a better correlation from cleaned tweet text.
- Presence of the author's Twitter description does not contribute to improvements in performance, as the model seems to be inefficient on addition of that information. This demonstrates the tweet-specific (and not user-oriented) nature of the task.
- Loss turned out to be a better metric to optimize rather than accuracy, as training models for reduction of loss yielded better performance than training for an increase in accuracy.

The best model's performance in our validation dataset can be seen in Table 4. The performance of this model in the shared task test dataset can also be seen in 5.

6. Future Work

Future work can explore methods for utilizing unsupervised methods in order to tackle the lack of data available on COVID vaccine stance. More experimentation should also be undertaken to take into account user-specific information such as location, follower network between users of the website, demographic, etc. Metrics of tweets such as likes, retweets, and shares can also be utilized to gauge the reception of tweets by users of the website.

[2] https://github.com/cbaziotis/ekphrasis

7. Conclusion

This paper describes the COVID vaccine stance detection system entered in Task 1 of IRMiDis 2022. Our method decomposes and experiments on the addition of different external information and achieves the 3rd best performance in the shared task with an accuracy of 0.582 and a macro-F1 score of 0.528.

References

[1] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018. URL: https://arxiv.org/abs/1810.04805. doi:10.48550/ARXIV.1810.04805.

[2] 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.

[3] Y. Hou, P. van der Putten, S. Verberne, The covmis-stance dataset: Stance detection on twitter for covid-19 misinformation, 2022. URL: https://arxiv.org/abs/2204.02000. doi:10.48550/ARXIV.2204.02000.

[4] L. Zhu, Z. Fang, G. Pergola, R. Procter, Y. He, Disentangled learning of stance and aspect topics for vaccine attitude detection in social media, 2022. URL: https://arxiv.org/abs/2205.03296. doi:10.48550/ARXIV.2205.03296.

[5] M. A. Milligan, D. L. Hoyt, A. K. Gold, M. Hiserodt, M. W. Otto, Covid-19 vaccine acceptance: influential roles of political party and religiosity, Psychology, Health & Medicine 0 (2021) 1–11. URL: https://doi.org/10.1080/13548506.2021.1969026. doi:10.1080/13548506.2021.1969026. arXiv:https://doi.org/10.1080/13548506.2021.1969026, pMID: 34407721.

[6] G. Pennycook, J. McPhetres, B. Bago, D. G. Rand, Beliefs about covid-19 in canada, the united kingdom, and the united states: A novel test of political polarization and motivated reasoning, Personality and Social Psychology Bulletin 48 (2022) 750–765. URL: https://doi.org/10.1177/01461672211023652. doi:10.1177/01461672211023652. arXiv:https://doi.org/10.1177/01461672211023652, pMID: 34180276.

[7] M. Müller, M. Salathé, P. E. Kummervold, Covid-twitter-bert: A natural language processing model to analyse COVID-19 content on twitter, CoRR abs/2005.07503 (2020). URL: https://arxiv.org/abs/2005.07503. arXiv:2005.07503.