

A BERT-based model for Profiling Irony and Stereotype Spreaders on Twitter

Notebook for PAN at CLEF 2022

Wenbin Lin, Zhongyuan Han*, Jinxi Zhang, Zengyao Li, Guiyuan Cao, Jianhong Yu and Leilei Kong

Foshan University, Foshan, China

Abstract

The PAN 2022 profiling irony and stereotype spreaders on Twitter (IROSTEREO) challenge focuses on determining whether a person is an irony and stereotype spreader. In this paper, we propose a method based on a BERT model and multi-model voting. The task of "Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO)" is considered as a binary classification task by our team, and a BERT model is used to determine whether an author is an irony and stereotype propagator. By using this model, we can achieve the task of "Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO)". In the end, we achieved an accuracy of 0.9333.

Keywords

Irony and Stereotype, BERT, Multi-model voting

1. Introduction

Irony is an ironic tone or writing technique in speech or writing, where the true meaning of the words cannot be understood from the words alone, but in fact, the original meaning is the opposite of the literal meaning and usually needs to be understood in context and context. In real life, there are often irony and stereotype purveyors on social software, and their comments are extremely influential on the social climate, so if software can accurately determine whether what a person says is sarcastic or not and block his comments in time, it can ensure that those comments will not influence the social climate. At PAN'22, reviewers will focus on profiling ironic authors on Twitter[1]. Special emphasis will be given to those authors that employ irony to spread stereotypes. Our goal will be to classify authors as ironic or not depending on their number of tweets with ironic content.

After establishing the goal and after a rigorous discussion among our team, we concluded that the task of "analyzing satirists and stereotypes on Twitter" is a binary classification task, i.e., the goal of the task is for us to determine whether the author is a satirist and stereotypist. Therefore, by analyzing multiple papers, our team felt that using a multi-model voting and BERT model would be a good fit for this task.

For this task, we will use the BERT model to train and predict the dataset and then apply a multi-model voting approach to improve the accuracy of the final results. For this task, our goal is to analyze which authors belong to the irony and stereotype purveyors and which authors do not belong to the irony and stereotype purveyors.

¹CLEF 2022 – Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy

EMAIL: lwblinwenbin630@gmail.com (A. 1); hanzhongyuan@gmail.com (A. 2)(*corresponding author); zhangjinxi7502@gmail.com (A. 3); lzy1512192979@gmail.com (A. 4); kongleilei@fosu.edu.cn (A. 7)

ORCID: 0000-0002-5851-0324 (A. 1); 0000-0001-8960-9872 (A. 2); 0000-0002-2921-1181 (A. 3); 0000-0001-8472-4150 (A. 4); 0000-0002-4636-3507 (A. 7)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org) Proceedings

2. Datasets

After our team's counting, there are 420 authors in the training set this year, and each author has 200 tweets, which are all XML files, so there are 84,000 tweets in total. Also, the folder contains a list of authors and a "truth.txt" file with basic facts. The first column in the "truth.txt" file corresponds to the author ID, and the second column contains the truth tag. We divided the training set into a training set and a test set; the ratio of the training set to the test set is 4:1. Thus, more test data can be obtained.

3. Method

We deem the task as a binary classification task. For each author, each paragraph of sentences is fed into the training model to make predictions, and eventually, an array of predictions with only two elements will be obtained. We believe that the second element of the array can be used as a basis for determining whether a given author is an irony and stereotype purveyor. If the second element of the prediction array has a larger value than the first element, then we can assume that the author is a satirist and stereotype purveyor. If the second element of the prediction array has a smaller value than the first element, then we can assume that the author is not a satirist or stereotype purveyor[2].

3.1. Processing of training set data

The official data is not a straightforward collection of sentences, and the XML file given contains many tags and symbols that are not useful, so we need to do some processing on the official data given.

During the training set data processing, we removed some useless punctuation and symbols such as "#user", "#RT" and "#CD". Then for each author, we cut operation for each document, that is, every 512 words into a paragraph, followed by "\t" and "1" and "o", the width of "\t" is equivalent to the position of 8 spaces, "1" represents the true value of the paragraph is true at this time because the subsequent model is certain to need a parameter representing the true value, so we first put all sentence parameters are set to "1", "o" on behalf of the paragraph belongs to the author, if the last paragraph is less than 512 words, then we add "\t" and "1" and "n" after the paragraph, "n" on behalf of the next paragraph does not belong to the author. The entire paragraph division is shown in Figure 1. Next, we put all the authors of the txt file into a txt file inside, and then apply the 5-fold method to all the authors into 5 parts, for each part of the data are rotated as a test set, and the remaining four parts are used as the training set, so we have 5 sets of training and test sets, and then because there is a random function "random.shuffle(xml_list)" in 5-fold, for each author, it will disrupt the order of all the sentences in it, and then repeat the above operation, i.e., for each part of the data is rotated as the test set, and the remaining four parts are used as the training set, so that we have another 5 sets of training and test sets, after this data processing we have a total of 10 sets of training and test sets.

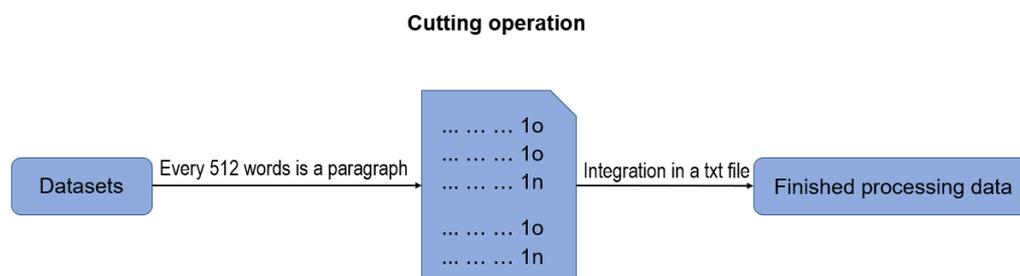


Figure 1: The process of slicing the datasets

3.2. Processing of BERT model and multi-model voting

We trained the above 10 datasets using the popular BERT model. At the end of each epoch of training, the model obtained at that point is predicted for the corresponding test set, and when the prediction results are the best so far, the model is saved. A prediction method is by predicting each segment individually and getting the respective corresponding conformity, and finally calculating their arithmetic mean, when the value is greater than 0.5, then the authors are considered then we consider the authors to be sarcastic and stereotypical authors, and label these authors with the corresponding prediction labels, and get an array of labels about these author categories. A "1" label indicates that the author is an author of that category, and a "0" label indicates that the author is not an author of that category. After the 10 models are trained in multiple rounds by the above method, the authors are predicted by the same prediction method, and the results are voted by adding the 10 labels directly and setting a threshold value, when the number of votes for an author is greater than this threshold value, we consider the author as an irony and stereotypical author, otherwise, the author is not.

The structure of the whole model is shown in the Figure 2 below.

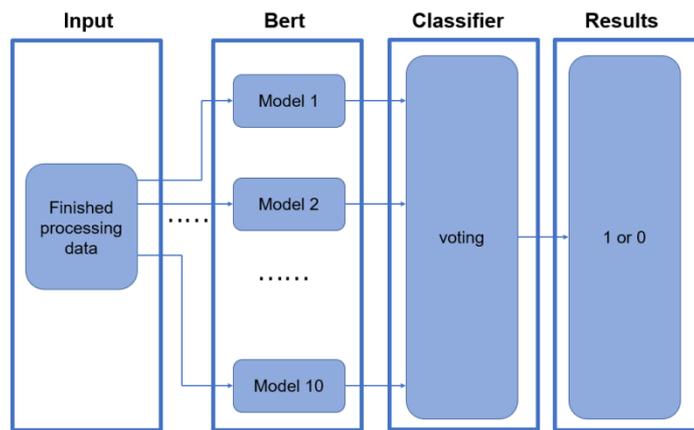


Figure 2: Architecture of the whole model

4. Results

The predictions of the 10 models were counted and further differentiated by voting, with 1 vote when the model predicts a satirical author for one of the authors, and no vote otherwise. Then a number of votes is set as a boundary, and only when the number of votes is more than this value can this author be considered a satirical author, and this is taken as the final prediction result. The results of the experiment are shown in the table below.

Table 1

Accuracy under different parameters

Votes	Accuracy
0	0.9444444444444444
1	0.9523809523809523
2	0.9603174603174603
3	0.9761904761904762
4	0.9761904761904762
5	0.9761904761904762
6	0.9761904761904762
7	0.9682539682539683
8	0.9444444444444444
9	0.8809523809523809

After the best model was selected, this model was next used to evaluate the validation set, and the results are shown in the table below.

Table 2

Validation set evaluation results

Data set	Task.F1
Test set	0.9333

These are the results of our team's work on this year's "Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO)" task. Finally, the top three teams had scores of 0.9944, 0.9778 and 0.9722 respectively. Our team achieved an accuracy of 0.9333 and was ranked 32nd.

5. Conclusion

In this paper, we propose an approach based on a BERT model and multi-model voting. The task of "Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO)" is considered as a binary classification task by our team, and is analyzed using a BERT model and a multi-model voting approach for correlation. Using this approach, the task of "Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO)" can be implemented. In the end, we achieved a score of 0.9333 in this method.

6. Acknowledgements

This work is supported by the Natural Science Foundation of Guangdong Province, China (No. 2022A1515011544).

7. References

- [1] Ortega-Bueno R., Chulvi B., Rangel F., Rosso P. and Fersini E. Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO) at PAN 2022. In: CLEF 2022 Labs and Workshops, Notebook Papers, CEUR-WS.org.
- [2] Zhijie Zhang, Zhongyuan Han, Leilei Kong, Xiaogang Miao, Zeyang Peng, Jieming Zeng, Haojie Cao, Jinxi Zhang, Ziwei Xiao, and Xuemei Peng. Style Change Detection Based On Writing Style Similarity—Notebook for PAN at CLEF 2021. In CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.
- [3] Devlin J., Chang M.W., Lee K., et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, 1: 4171-4186