

# A Multi-Model Voting Ensemble Classifier based on BERT for Profiling Irony and Stereotype Spreaders on Twitter

Notebook for PAN at CLEF 2022

Haojie Cao, Zhongyuan Han\*, Zhenwei Mo, Zengyao Li, Ziwei Xiao, Zijian Li and Leilei Kong

Foshan University, Foshan, China

## Abstract

For the Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO) task, the goal is to determine whether its author spreads irony and stereotypes through a Twitter feed in English. Profiling irony and stereotype spreaders on Twitter is regarded as a classification task by us. This paper presents our classifier based on BERT and Multi-Model Voting. On top of training our model with BERT, we use multi-model voting and dynamically adjust two hyperparameters to improve the accuracy of our model. Finally, the accuracy of our classifier on test data is 0.9389.

## Keywords

Irony and Stereotype Spreader, BERT, Multi-Model Voting, Classifier

## 1. Introduction

Irony is a way in which language expresses the opposite of its literal meaning metaphorically and subtly. Someone sometimes uses it to mock or scorn a victim, which puts him or her at risk of being hurt. Stereotypes are often used when discussing controversial issues. At PAN'22 [1], The task Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO) [2] was set. The goal is to determine whether its author spreads irony and stereotypes through a Twitter feed in English.

We propose a multi-model voting ensemble classifier based on BERT to solve this task. We trained three models with BERT and used three models to predict and vote whether an author is an irony and stereotype spreader. Suppose the number of votes obtained by an author is greater than or equal to 2. In that case, we consider an author an irony and stereotype spreader, and vice versa, we consider an author not to be an irony and stereotype spreader.

## 2. Method

We abbreviate irony and stereotype spreader to Irony. This part describes our classifier for the Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO) task. In section 2.1. the entire framework is described. Section 2.2. introduces the Data Processing. And in section 2.3. the input and training are described. Section 2.4. Classifier and Output describes how we tune our classifier to improve the accuracy of our Classifier for Stereotype Spreaders.

### 2.1. Model

---

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

EMAIL: caohaojie0322@163.com (A. 1); hanzhongyuan@gmail.com (A. 2) (\*corresponding author); mozhenwei45@163.com (A. 3); lzy1512192979@gmail.com (A. 4); kongleilei@fosu.edu.cn@mail.com (A. 7)

ORCID: 0000-0002-8365-168X (A. 1); 0000-0001-8960-9872 (A. 2); 0000-0002-5722-7551 (A. 3); 0000-0001-8472-4150 (A. 4); 0000-0002-4636-3507 (A. 7)

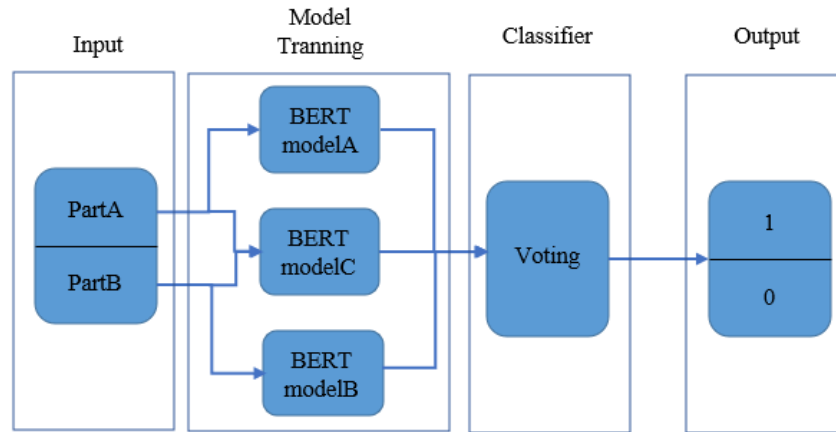


© 2022 Copyright for this paper by its authors.

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

CEUR Workshop Proceedings (CEUR-WS.org) Proceedings

Our team uses BERT and multi-model voting to approach the task, and in Figure 1, we give the architecture of the whole model. We divide the data into two parts, one for PartA and the other for PartB, and then feed the data from PartA and PartB into BERT for training to obtain BERT modelA and BERT modelB. In addition, we feed all the data into BERT together to obtain BERT modelC. Finally, we use the three models to vote on the data to determine whether an author is an Irony and Stereotype Spreader.



**Figure 1:** Architecture of the whole model

## 2.2. Data Processing

The dataset [3] was provided by PAN'22, containing 420 authors. In the dataset, each author corresponds to an XML file that stores the 200 tweets of the author. Of the 420 authors, half were irony and stereotype spreaders, while the other half were not.

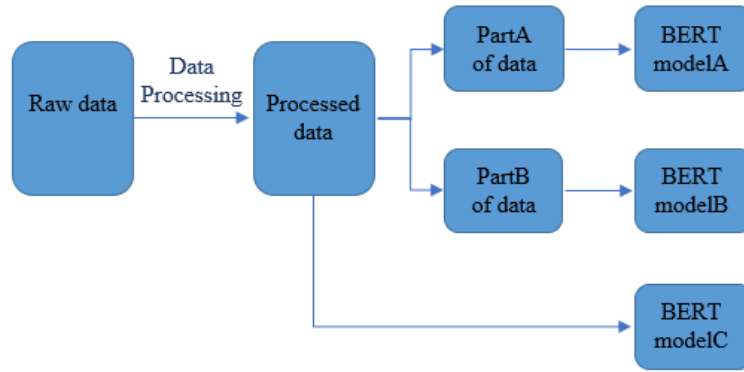
In the dataset provided by PAN'22, each author's tweets are stored in an XML file. We remove tags like #USER from the tweets to ensure accuracy.

If an author is labeled as Irony, we assume that every tweet this author posts is Irony. If an author is labeled as not Irony, we assume that every tweet this author posts is not Irony. We label each tweet in the training dataset based on such rules and feed these tweets and their corresponding labels into the BERT model for training.

## 2.3. Input and Model Training

After processing the training dataset according to the method mentioned in 2.2. Data Processing, we obtain the processed training dataset. Among the 420 authors in the dataset, the first 210 authors are all Irony, and the last 210 are not Irony. We take the 106th to 315th authors in the dataset as a subset, and the remaining 1st to 105th authors and 316th to 420th authors as another subset. This way, we can get two subsets with the same number of Irony authors and not Irony authors. We denote these two subsets as PartA and PartB. We feed each tweet from PartA and its corresponding label into BERT for training and obtain BERT modelA. Similarly, we feed each tweet in PartB and its corresponding label into BERT for training and obtain BERT modelB. Later we will use the hard voting strategy to determine whether an author is Irony. In order to prevent a tie vote, in addition to BERT modelA and BERT modelB, we feed each tweet and its corresponding label from the whole training dataset into BERT for training, and finally, we get BERT modelC. The architecture of input and model training is shown in Figure 2.

The trained BERT model is capable of scoring each tweet. The score of a tweet is represented by (Irony, not\_Irony). For example, a tweet with a score of (0.8, 0.2) means an 80% probability that the statement is an Irony tweet and a 20% probability that it is not.



**Figure 2:** Architecture of input and model training

## 2.4. Classifier and Output

In our classifier, we set two hyperparameters, one of which is `passProbability`, representing the threshold for the tweet to be Irony. We set the tweet to be Irony only if the value of the Irony part of the tweet's score is greater than the `passProbability` and greater than the value of the not Irony part.

Each author has posted 200 tweets, and by feeding each tweet and its label into the BERT model, we get 200 scores for each author. To determine whether the author is an Irony author, we set a hyperparameter `passNumber`, representing the threshold for the author to be Irony. When the number of Irony tweets is greater than the `passNumber`, we consider the author to be an Irony.

To improve our accuracy, we also use multi-model voting. Input and Model Training mentions that BERT modelA is trained with the dataset of PartA, and BERT modelB is trained with the dataset of PartB. We use BERT modelA to evaluate the PartB dataset on whether an author is Irony or not, and compare the evaluation results with the PartB dataset to obtain the accuracy of BERT modelA. BERT modelB does the same thing. In the evaluation process, we keep adjusting the `passProbability` and `passNumber` of the classifiers corresponding to the two BERT models so that the accuracy of the two BERT models is maximized. The optimal values of the two hyperparameters are found by analyzing the hyperparameters of BERT modelA and BERT modelB. The analysis process for the optimal values of the two hyperparameters is described in 3. Results. The whole dataset is fed into BERT for training to get a BERT modelC, and we set the `passProbability` of the classifier corresponding to the BERT modelC to 0.6 and the `passNumber` to 100.

Now, we get three models in total. We use a hard voting strategy to determine whether an author is Irony. We use each model to vote for an author, and if the author receives more than or equal to 2 votes, then we consider the author to be Irony, and we consider the author to be non-Irony otherwise.

## 3. Results

In the end, as shown in Table1, the accuracy of the top three teams was 0.9944, 0.9778, and 0.9722, respectively. Our model achieves an accuracy of 0.9389 and is ranked 27th. [2] We attribute this accuracy to two things that we have done. First, we set two hyperparameters and dynamically adjusted both of them, allowing the classifier corresponding to each model to achieve high accuracy. The second is that we used multi-model voting, which allows us to more accurately predict whether an author is an irony and stereotype spreader.

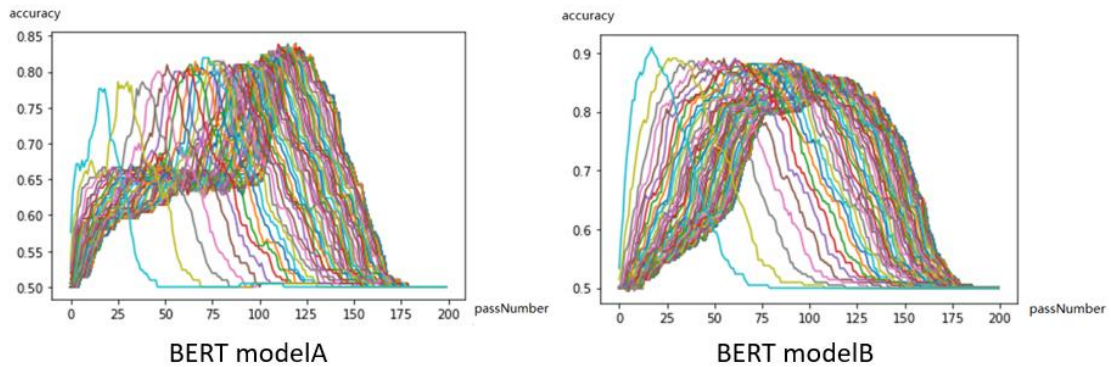
**Table 1**

The results of the top three teams and our team

POS	TEAM	ACCURACY
1	wentaoyu	0.9944
2	harshv	0.9778
3	edapal	0.9722
27	caohaojie	0.9389

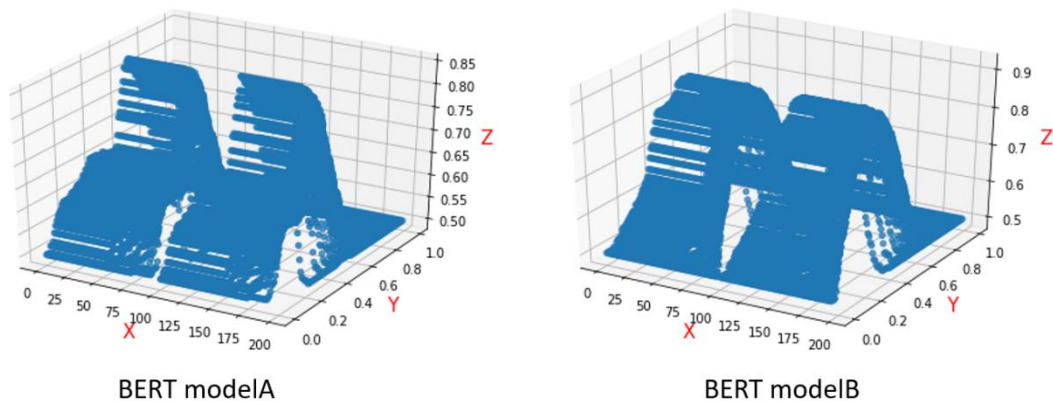
We analyzed the two hyperparameters of BERT modelA and BERT modelB so that the optimal values of the two hyperparameters are found.

As shown in Figure 3, we plot the variation of passProbability in BERT modelA and BERT modelB with the increase of passNumber for different values of accuracy, where the values of passProbability are 0.01,0.02,0.03.... .0.99,1.00, a total of 100 values. By observation, we can see that more curves are reaching the peak when the passNumber is between 100-125, so we know that the best value of passNumber is between 100-125.



**Figure 3:** The curve of accuracy with passNumber for different values of passProbability

As shown in Figure 4, we plotted the scatter plots of BERT modelA and BERT modelB with passNumber as the x-coordinate, passProbability as the y-axis, and accuracy as the z-axis, respectively. We observed that passProbability peaked at more points at 0.5-0.6.



**Figure 4:** Scatter plot with passNumber as x-axis, passProbability as y-axis, and accuracy as the z-axis

Based on the above observation, we know that the passProbability should be set to 0.6, and the passNumber should be set to 100.

## 4. Conclusion

This paper describes our team's experiments to solve the task Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO). We consider this task to be a classification task. In our experiments, we first trained three different models via BERT. To improve the accuracy of our models, we analyzed two hyperparameters, and finally, we obtained the optimal values of the two hyperparameters for each BERT model.

## 5. Acknowledgements

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

## 6. References

- [1] Janek Bevendorff, Berta Chulvi, Elisabetta Fersini, et al. Overview of PAN 2022: Authorship Verification, Profiling Irony and Stereotype Spreaders, and Style Change Detection, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Thirteenth International Conference of the CLEF Association (CLEF 2022), Springer, 2022.
- [2] Ortega-Bueno Reynier, Chulvi Berta, Rangel Francisco, Rosso Paolo and Fersini Elisabetta, Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO) at PAN 2022, in: CLEF 2022 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2022.
- [3] Ortega-Bueno Reynier, Chulvi Berta, Rangel Francisco, Rosso Paolo and Fersini Elisabetta, PAN 22 Author Profiling: Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO), 2022. URL: [https://zenodo.org/record/6514916#.Yos\\_yXZBxD9](https://zenodo.org/record/6514916#.Yos_yXZBxD9)