Irony Detection in the Portuguese Language using BERT

Shengyi Jiang[1,2], Chuwei Chen[1], Nankai Lin[1(✉)], Zhuolin Chen[1] and Jinyi Chen[1]

[1] School of Information Science and Technology, Guangdong University of Foreign Studies, China
[2] Guangzhou Key Laboratory of Multilingual Intelligent Processing, Guangdong University of Foreign Studies, Guangzhou
neakail@outlook.com

Abstract. In this article, we report the solution of the team BERT 4EVER for the Irony Detection in the Portuguese language task in IberLeF 2021, which aims to identify irony news articles or tweets in the Portuguese language spread via digital media. We propose the BERT (Bidirectional Encoder Representations from Transformers) model to tackle the problem. In addition, we adopt weight loss and ensemble learning to improve the generalization capability. Experimental results as well as the leading position of our team on the task leaderboard demonstrate the effectiveness of our method in the field of news.
Keywords: BERT, Irony Detection, Portuguese.

1 Introduction

Irony refers to the use of words contrary to the original meaning to express the meaning, which is a form of figurative language with strong emotional color. The irony detection is a key-challenge in various tasks involving Natural Language Processing (NLP). In the field of Opinion Mining, for instance, Luís Sarmento et al. [1] noted the role of irony in minimizing the error when discriminating negative from positive opinions.

Table 1. The statistics of the dataset used in this task.

Field	News	Tweets
Irony	7222	12736
No Irony	17272	2476
All	18494	15212

IberLEF 2021 proposes the task "Irony Detection in the Portuguese language" [**Erro! Fonte de referência não encontrada.**], which aims at encouraging more work to address the problem of identifying the presence of irony in texts (tweets and news)

IberLEF 2021, September 2021, Málaga, Spain.

written in Portuguese. The distribution of the dataset is shown in Table 1. Our team, BERT 4EVER, also participates in this task and achieves the first rank in the field of news. In this report, we will review our solution to this task, namely, the BERT model aided by weight loss and ensemble learning.

2 Related Work

Most of the research studies in Irony Detection focus on English language [3]. Christos Baziotis et al. [4] presented an ensemble of two different deep learning models: a word- and a character-level deep LSTM for capturing the semantic and syntactic information of tweets respectively, which ranked 1st for both subtasks in SemEval-2018 Task 3 “Irony detection in English tweets” [4]. Barbieri et al. proposed a new evaluation framework named TWEETEVAL consisting of seven heterogeneous Twitter-specific classification tasks. In particular, they provided a strong set of baselines as starting point and compared different language modeling pre-training strategies, which established a relatively perfect evaluation system of English irony detection [5].

Recently, the NLP community also focuses on other languages for the need to develop linguistic and computational resources, which has spawned a multitude of irony detection competitions for other languages, such as Arabic [6], Spanish [7], and Italian [8].

Portuguese is a low-resource language, which limits the amount of research done for this language. Freitas et al. proposed a set of patterns that might suggest ironic/sarcastic statements by observing a corpus constituted by tweets [9]. In particular, they developed special clues for irony detection, through the implementation and evaluation of a set of patterns. Fabio Ricardo Araujo da Silva [10] proposed a Convolutional Neural Network (CNN) adapted for automatically detecting irony/sarcasm in Brazilian Portuguese, which was trained and tested by datasets from Twitter obtained by the author and from thirds. For the reason that there is no website corpus in Portuguese corpus (only the corpus from social network twitter), Gabriel Schubert Marten et al. [11] developed a corpus in the Portuguese language to sarcasm and irony detection task.

3 Method

As shown in Figure 1, based on the BERT model, we train three kinds of models with different strategies. In the prediction stage, we fuse the prediction results of the three models for each field (news/twitter). These three strategies are as follows:

(1) We fine-turn the BERT model separately for the training set in each field.

(2) On the basis of (1), we adopt the Loss Weight strategy for the training set in each field to solve data imbalance.

(3) We combine the data from these two fields together and fine-tune the BERT model, so as to make use of the information from the other field to assist classification and improve the generalization ability of the model.

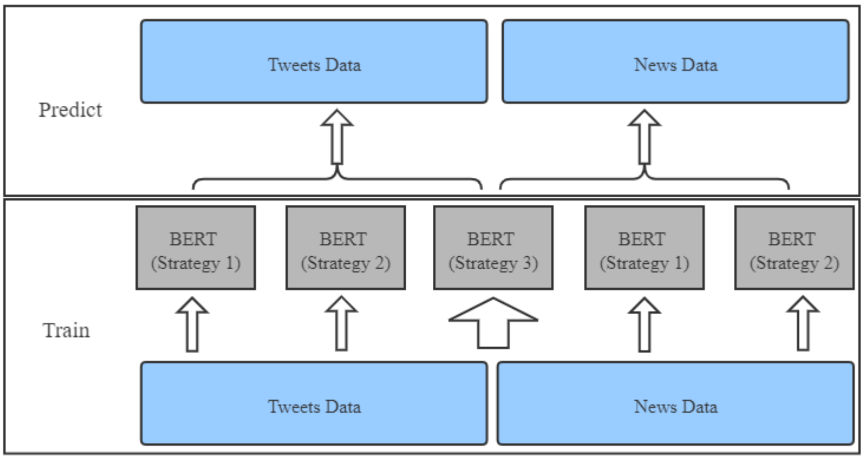

Fig. 1. The framework of our method.

3.1 BERT

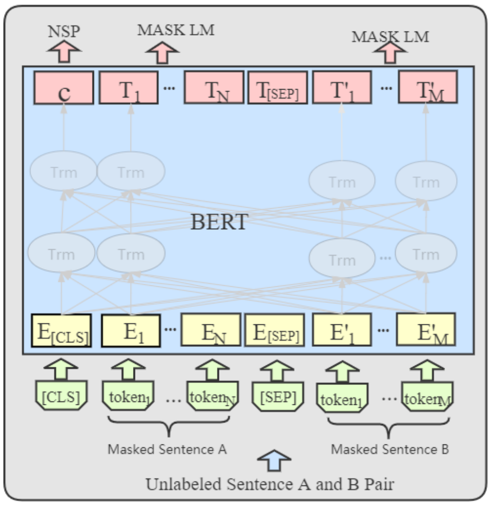

Fig. 2. BERT Model.

BERT (Bidirectional Encoder Representations for Transformers) is designed to pre-train deep bidirectional representations from unlabeled text by jointly building models both left and right context in all layers [12]. The BERT model structure is shown in Figure 2.

It consists of two pretrain tasks, namely Mask Language Model (MLM) and Next Sentence Prediction (NSP) :(1) MLM is defined as masking some words in the input sequence, and then predicting the masked words according to the context; (2) NSP refers to predicting whether the second sentence is the follow-up (next sentence) of the first sentence.

3.2 Weight Loss

We can see that in the field of news, there are more data that are not ironic, while in the field of tweets, there are more data that are ironic. In this paper, we adopt a class weight adjustment method [13] to tackle the problem of data with class imbalance. Assuming that the given labels $c = (c_1, c_2, \ldots, c_k)$, the class weight w_i of the i-th label is calculated as follows.

$$w_i = log_{mu}\left(\frac{c_i}{\sum_{i=1}^{n} c_i}\right) \tag{8}$$

where the value of mu is e.

3.3 Model Fusion

We train multiple models through multiple strategies. Each model predicts the test data separately. For each sample x, the predicted probability of the model is

$$y = \left[\frac{pt_1 + pt_2 + pt_3}{3}, \frac{pn_1 + pn_2 + pn_3}{3}\right]$$

in which pt_i is the "no irony" probability of i-th BERT model, pn_i is the "irony" probability of i-th BERT model.

4 Result

Based on five-fold cross-validation, we reported the result of BERTs model with three strategies and other machine learning algorithms. We used Transformers[2] library and Pytorch[3] library as backend to construct BERT-based models and scikit-learn[4] to construct machine learning models. The BERT[5] model we used was pre-trained by Souza et. al [14]. When using other machine learning algorithms, we selected the text features with TFIDF. We used Bacc (Balanced Accuracy) as the evaluation indicator and the results were shown in Table 2. We could see that the performance of most algorithms was excellent, and the BERT models with three strategies were more than 0.99 on the validation set.

[2] https://github.com/huggingface/transformers
[3] https://github.com/pytorch/pytorch
[4] https://github.com/scikit-learn/scikit-learn
[5] https://huggingface.co/neuralmind/bert-base-portuguese-cased

Table 2. The results of our model based on five-fold cross-validation.

Field	Model	Bacc
Tweets	BERT (Strategy 1)	**0.9992**
	BERT (Strategy 2)	**0.9992**
	BERT (Strategy 3)	0.9977
	KNN	0.9599
	Random Forest	0.9497
	Decision Tree	0.9947
	SVM	0.9959
	Naive Bayes	0.8849
News	BERT (Strategy 1)	**0.9906**
	BERT (Strategy 2)	0.9900
	BERT (Strategy 3)	0.9883
	KNN	0.9393
	Random Forest	0.9497
	Decision Tree	0.8923
	SVM	0.9822
	Naive Bayes	0.8397

Due to the limitation of the number of contest submissions, we submitted three results, the results of the test set were shown in Table 3 and Table 4. In the field of news, we respectively submitted the result of BERT (Strategy 1), the result of BERT (Strategy 3), and the fusion result of three BERT models. It can be seen that when multiple models were fused, the performance dropped instead. Among the three strategies, the strategy 3 with the worst performance on the validation set had the best performance on the test set. This was because the model used more data for training, which increased the generalization capability of the model to some extent. The result reached the best performance in this evaluation competition. However, in the field of tweets, our models were overfitting to the training data. Although the fusion strategy had brought a certain improvement, the result of our test set was still very low, the Bacc value was only 0.4975.

Table 3. The results of our model on the final news test set.

Method	Bacc	Acc	F1	Precision	Recall
BERT (Strategy 1)	0.9107	0.9000	0.8800	0.8148	**0.9565**
BERT (Strategy 3)	**0.9215**	**0.9133**	**0.8943**	**0.8397**	**0.9565**
BERT (Merge)	0.9063	0.8967	0.8755	0.8134	0.9478

Table 4. The results of our model on the final tweets test set.

Method	Bacc	Acc	F1	Precision	Recall
BERT (Strategy 1)	0.4959	0.4067	**0.5782**	0.4080	**0.9919**
BERT (Merge)	**0.4975**	**0.4100**	0.5776	**0.4088**	0.9837

5 Conclusion

In the Irony Detection in the Portuguese language task in IberLeF 2021, we train three kinds of models with different strategies based on the BERT model. Experimental results as well as the leading position of our team on the task leaderboard in the field of news demonstrate the effectiveness of our method. However, in the field of tweets, our models are overfitting to the training data. In the future, we will try to solve the problem of overfitting in tweets field in order to achieve better results on the Irony Detection in the Portuguese language task.

Acknowledgements

This work was supported by the Key Field Project for Universities of Guangdong Province (No. 2019KZDZX1016), the National Natural Science Foundation of China (No. 61572145) and the National Social Science Foundation of China (No. 17CTQ045). The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

References

1. Sarmento, L., Carvalho P., Silva M. J., et al.: Automatic creation of a reference corpus for political opinion mining in user-generated content. In: Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion, pp. 29-36 (2009).
2. Corrêa, U. B., dos Santos, L. P., Coelho, L., de Freitas, L. A.: Overview of the IDPT Task on Irony Detection in Portuguese at IberLEF 2021. Procesamiento del Lenguaje Natural, vol. 67, (2021).
3. Joshi, A., Bhattacharyya, P., Carman, M. J.: Automatic sarcasm detection: A survey. ACM Computing Surveys (CSUR) 50(5), 1-22 (2017).
4. Baziotis, C., Athanasiou, N., Papalampidi, P., et al.: NTUA-SLP at SemEval-2018 Task 3 Tracking Ironic Tweets using Ensembles of Word and Character Level Attentive RNNs. CORR (2018).
5. Barbieri, F., Camacho-Collados, J., Neves, L., et al.: TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification. CORR (2020).
6. Ghanem, B., Karoui, J., Benamara, F., et al.: IDAT AT FIRE2019 Overview of the Track on Irony Detection in Arabic Tweets. In: Proceedings of the 11th Forum for Information Retrieval Evaluation, pp. 10-13 (2019).
7. Ortega-Bueno, R., Rangel, F., Hernández, F. D., et al.: Overview of the task on irony detection in Spanish variants. In: Proceedings of the Iberian languages evaluation forum (IberLEF 2019), co-located with 34th conference of the Spanish Society for natural language processing (SEPLN 2019), CEUR, vol. 2421, pp. 229-256 (2019).
8. Cignarella, A. T., Frenda, S., Basile, V., et al.: Overview of the evalita 2018 task on irony detection in italian tweets (ironita). In: Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2018), CEUR, vol. 2263, pp. 1-6 (2018).

9. de Freitas, L. A., Vanin, A. A., Hogetop, D. N., Bochernitsan, M. N., and Vieira, R.: Pathways for irony detection in tweets. In: Proceedings of the 29th Annual ACM Symposium on Applied Computing, pp. 628-633 (2014).
10. DA SILVA, F. R. A.: Detecção de ironia e sarcasmo em língua portuguesa: Uma abordagem utilizando deep learning. (2018).
11. Marten, G. S., and de Freitas, L. A.: The Construction of a Corpus for Detecting Irony and Sarcasm in Portuguese. Brazilian Journal of Development 7(5), 47973-47984 (2021).
12. Devlin, J., Chang, M. W., Lee K., and Toutanova K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, In: Proceedings of NAACLHLT 2019, pp. 4171-4186. (2019).
13. Wang, L., Lin X., and Lin, N.: Research on pseudo-label technology for multi-label classification. In: 16th International Conference on Document Analysis and Recognition, (2021).
14. Souza, F., Nogueira, R., and Lotufo, R.: BERTimbau: Pretrained BERT Models for Brazilian Portuguese. In: Brazilian Conference on Intelligent Systems, pp. 403-417. (2020).