Notebook for PAN at CLEF 2022:Profiling Irony and Stereotype Spreaders on Twitter

Wang Bin[1], Ning Hui[1,*]

[1]Harbin Engineering University, 145 Nantong St, Harbin, 150000, China

Abstract

Twitter is currently one of the most widely used social media platforms. The second task posted for the PAN @ CLEF 2022 competition aimed to profile irony and stereotype spreaders on Twitter. This paper reports on the methodology we used in this competition. We tried different variants of BERT models based on transformer and Auto-Keras as the classifier. As a result, our best accuracy on the test set was 93.89%, and the last submission was 93.33%.

Keywords

IROSTEREO, word embedding, BERT, AutoML

1. Introduction

Twitter is currently one of the most widely used social media platforms. Users can exchange real-time information on the platform about various topics or current news. Hundreds of millions of Twitter users generate huge amounts of tweets every day [1]. As information rapidly interacts and spreads, some messages can be harmful to certain groups [2]. Such harm can become a social problem when irony and stereotypes are spread widely without checking.

The second task released for the PAN @ CLEF 2022 challenge [3] considers profiling irony and stereotype spreaders on Twitter [4]. The performance of participant's system will be ranked by accuracy. All systems should be submitted through the TIRA platform [5]. The task was aimed at identifying users as irony and stereotype spreaders based on 200 tweets from Twitter users. This will help reduce the identification effort and help address the problem of avoiding the spread of harmful information more effectively than identifying tweets one by one.

For the task we tried to extract word embeddings with transformer-based [6] pre-trained models and classify the data with AutoML model.

In Section 2, we present some related work on this task. In Section 3, we illustrate our approach to the data and our method of training, and then we show our results in Section 4. Finally, we state the conclusions we have drawn in Section 5.

*Corresponding author

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

sgomw@hrbeu.edu.cn (W. Bin); ninghui@hrbeu.edu.cn (N. Hui)
0000-0002-6711-5887 (W. Bin)

2. Related work

The Transformer architecture is widely used for irony and sentiment detection. [7] embedded pre-trained Twitter words in context by using the Transformer architecture, they also studied and interpreted how the multi-head self-attention mechanisms are specialized on detecting irony by means of considering the polarity and relevance of individual words and even the relationships among words. [8] propose a neural network methodology that builds on a recently proposed pre-trained transformer-based network architecture, which is further enhanced with the employment and devise of a recurrent convolutional neural network (RCNN). Some approaches continue to augment the results using word embeddings in pre-trained transformers. Transformer-based Deep Intelligent Contextual Embeddings (T-DICE) and attention-based BiLSTM were proposed by [9]. [10] proposed a network that combines the use of word embeddings from XLNET, multichannel CNNs, and an attention mechanism that includes automatic weight adjustment to effectively improve the results of sentiment recognition.

Automated machine learning (AutoML) is a promising solution for building a DL system without human assistance and is being extensively studied[11]. AutoML model under the AutoGluon framework was also used in the pan2021 competition to complete the task [12]. [13] demonstrate it is possible today to automatically discover complete machine learning algorithms just using basic mathematical operations as building blocks. [14] conduct various evaluations of the tools on many datasets, in different data segments, to examine their performance, and compare their advantages and disadvantages on different test cases. [15] introduce an open, ongoing, and extensible benchmark framework which follows best practices and avoids common mistakes to comparing different AutoML systems.

3. Data and Methodology

3.1. Dataset

The organizers provided a dataset containing tweets only in English. In total, tweets from 600 Twitter users were collected. Each user had 200 tweets. Of these users, 420 were labeled as irony and stereotype spreaders and were used for training. The remaining 180 users were not annotated and were used for testing.

3.2. Pre-processing

We performed a simple pre-processing of the text data in order to reduce noise. Firstly remove some meaningless characters and punctuation. Then replace 'USER' with '@USER' and 'URL' with 'HTTPURL'. Finally replace the emoji with text using the emoji package in python.

3.3. Word Embedding

The transformer architecture has been widely used since it was proposed. Its performance in natural language understanding and natural language generation has surpassed previous alternative neural models [16].

We use pre-trained models provided by the hugging face community for word embedding extraction. Three pre-trained models, ConvBERT [17], BERT-Large [18], and BERTweet [19], were used for training. We first tried to extract word embeddings from the last hidden layer of these models. Then BERTweet, which performed best in the training set with a five-fold cross-validation, was selected and we tried to use its last four layers for extracting.

3.4. Classification

AutoML emerged with the aim of reducing the heavy development cost and automating the machine learning pipeline. Many AI companies have open sourced their AutoML tools. [20] propose a novel framework enabling Bayesian optimization to guide the network morphism for efficient neural architecture search. The framework develops a neural network kernel and a tree-structured acquisition function optimization algorithm to efficiently explores the search space. The proposed method is wrapped into an open-source AutoML system, namely Auto-Keras. We used the AutoML tool under the Auto-Keras framework as a classifier.

Since it is difficult to analyze and predict all tweets of a user at once, we tried to split the all tweets of one user and predict them separately. Then a majority vote is used to predict whether a user is an irony and stereotype spreader.

4. Results

As a result, the accuracy of extracting the word embeddings in the last hidden layers of the ConvBERT, BERT-Large, and BERTweet correspond to 94.05%, 94.52%, and 94.76% in the five-fold cross-validation of the training set, respectively. While the accuracy of extracting the word embeddings of the last four layers of the BERTweet model corresponds to 95.48% (see Table 1).

Table 1
The accuracy on the training set corresponding to different models

Model	Accuracy
ConvBERT	94.05%
BERT-Large	94.52%
BERTweet	94.76%
BERTweet(4 layers)	95.48%

In the end, based on the feedback from the organizers, our best accuracy on the unlabeled test set was 93.89% and the last submission was 93.33%. It is slightly lower than the accuracy obtained on the training set.

The accuracy of these results far exceeds the accuracy of the similar task for PAN @ CLEF 2021 [21]. Considering that the training set in the dataset provided by PAN @ CLEF 2022 contains more than twice the total number of tweets provided by PAN @ CLEF 2021we believe that differences in the dataset contribute to this effect.

5. Conclusions

Transformer-based models have been very widely and effectively used in NLP tasks. The use of AutoML techniques reduces the cost of experiments. We have tried different models in this task with relatively simple methods and low cost, then we have obtained relatively good results. However, as the amount of data increases, many traditional methods and more complex models can also perform very well. We believe that a broader dataset may be more helpful when dealing with the identification of Twitter users in practice.

6. Acknowledgments

Thanks to all the conference organizers, especially Reynier Ortega and Magdalena Anna Wolska for their efforts!

References

[1] S. Noor, Y. Guo, S. H. H. Shah, M. S. Nawaz, A. S. Butt, Research synthesis and thematic analysis of twitter through bibliometric analysis, International Journal on Semantic Web and Information Systems (IJSWIS) 16 (2020) 88–109.

[2] M. Appel, S. Weber, Do mass mediated stereotypes harm members of negatively stereotyped groups? a meta-analytical review on media-generated stereotype threat and stereotype lift, Communication Research 48 (2021) 151–179.

[3] J. Bevendorff, B. Chulvi, E. Fersini, A. Heini, M. Kestemont, K. Kredens, M. Mayerl, R. Ortega-Bueno, P. Pezik, M. Potthast, F. Rangel, P. Rosso, E. Stamatatos, B. Stein, M. Wiegmann, M. Wolska, E. Zangerle, Overview of PAN 2022: Authorship Verification, Profiling Irony and Stereotype Spreaders, and Style Change Detection, in: M. D. E. F. S. C. M. G. P. A. H. M. P. G. F. N. F. Alberto Barron-Cedeno, Giovanni Da San Martino (Ed.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Thirteenth International Conference of the CLEF Association (CLEF 2022), volume 13390 of *Lecture Notes in Computer Science*, Springer, 2022.

[4] O.-B. Reynier, C. Berta, R. Francisco, R. Paolo, F. Elisabetta, Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO) at PAN 2022, in: CLEF 2022 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2022.

[5] M. Potthast, T. Gollub, M. Wiegmann, B. Stein, TIRA Integrated Research Architecture, in: N. Ferro, C. Peters (Eds.), Information Retrieval Evaluation in a Changing World, The Information Retrieval Series, Springer, Berlin Heidelberg New York, 2019. doi:10.1007/978-3-030-22948-1_5.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

[7] J. Á. González, L.-F. Hurtado, F. Pla, Transformer based contextualization of pre-trained word embeddings for irony detection in twitter, Information Processing & Management 57 (2020) 102262.

[8] R. Alexandros Potamias, G. Siolas, A.-G. Stafylopatis, A transformer-based approach to irony and sarcasm detection, arXiv e-prints (2019) arXiv–1911.

[9] U. Naseem, I. Razzak, P. Eklund, K. Musial, Towards improved deep contextual embedding for the identification of irony and sarcasm, in: IJCNN, 2020.

[10] C.-T. Yang, Y.-L. Chen, Dacnn: Dynamic weighted attention with multi-channel convolutional neural network for emotion recognition, in: 2020 21st IEEE International Conference on Mobile Data Management (MDM), IEEE, 2020, pp. 316–321.

[11] X. He, K. Zhao, X. Chu, Automl: A survey of the state-of-the-art, Knowledge-Based Systems 212 (2021) 106622.

[12] T. Anwar, Identify Hate Speech Spreaders on Twitter using Transformer Embeddings Features and AutoML Classifiers—Notebook for PAN at CLEF 2021, in: G. Faggioli, N. Ferro, A. Joly, M. Maistro, F. Piroi (Eds.), CLEF 2021 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2936/paper-153.pdf.

[13] E. Real, C. Liang, D. So, Q. Le, AutoML-zero: Evolving machine learning algorithms from scratch, in: H. D. III, A. Singh (Eds.), Proceedings of the 37th International Conference on Machine Learning, volume 119 of *Proceedings of Machine Learning Research*, PMLR, 2020, pp. 8007–8019. URL: https://proceedings.mlr.press/v119/real20a.html.

[14] A. Truong, A. Walters, J. Goodsitt, K. Hines, C. B. Bruss, R. Farivar, Towards automated machine learning: Evaluation and comparison of automl approaches and tools, in: 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), 2019, pp. 1471–1479. doi:10.1109/ICTAI.2019.00209.

[15] P. Gijsbers, E. LeDell, J. Thomas, S. Poirier, B. Bischl, J. Vanschoren, An open source automl benchmark, arXiv preprint arXiv:1907.00909 (2019).

[16] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al., Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations, 2020, pp. 38–45.

[17] Z.-H. Jiang, W. Yu, D. Zhou, Y. Chen, J. Feng, S. Yan, Convbert: Improving bert with span-based dynamic convolution, Advances in Neural Information Processing Systems 33 (2020) 12837–12848.

[18] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, CoRR abs/1810.04805 (2018). URL: http://arxiv.org/abs/1810.04805. arXiv:1810.04805.

[19] D. Q. Nguyen, T. Vu, A. T. Nguyen, Bertweet: A pre-trained language model for english tweets, arXiv preprint arXiv:2005.10200 (2020).

[20] H. Jin, Q. Song, X. Hu, Auto-keras: An efficient neural architecture search system, in: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2019, pp. 1946–1956.

[21] F. Rangel, P. Rosso, G. L. D. L. P. Sarracén, E. Fersini, B. Chulvi, Profiling Hate Speech Spreaders on Twitter Task at PAN 2021, in: A. J. M. M. F. P. Guglielmo Faggioli, Nicola Ferro (Ed.), CLEF 2021 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2021.