

zk15120170770 at FakeDeS 2021: Fake news detection based on Pre-training Model

Kun Zhao^[0000-0002-4615-3657], Siyao Zhou^[0000-0002-5583-098X], and Weihua Li*

Yunnan University, Yunnan, P.R. China

* Corresponding author: 1129515873@qq.com

Abstract. With the rapid development of Internet technology, the network makes the transmission of information no longer limited due to the long distance. All kinds of information can be transmitted conveniently and quickly on the Internet. People can view and send all kinds of information around the world with only a mobile network device. A lot of things happen in the world every day, so this information contains a lot of news information, and people can understand what is happening in the world through news information. However, there is much fake news information mixed in this news information. This fake news will interfere with our cognition and judgment. Therefore, we need more attention to distinguish the authenticity of this news item. This also poses certain challenges to our work tasks. In this paper, we describe the method used for the Fake News Detection in Spanish in IberLEF 2021. We fine-tuned the XLM-Roberta pre-training model based on the data sets provided by the host, Spanish, and obtained good results. The F1 score of our model in Spanish tasks reached 0.7053 and ranking seventh.

Keywords: Fake News Detection · IberLEF 2021 · Pre-training Model · XLM-Roberta.

1 Introduction

The rapid development of the Internet has promoted the application of social media, and social media has gained a lot of popularity. The miniaturization and convenience of mobile terminals have led to rapid growth in the use of social media in the past few years. Many people around the world communicate through social media. The information flowing on the Internet at all times is huge and incalculable. In addition to the bridge of social media, language is an indispensable part of our mutual communication. Therefore, language is an important part of communication, and equality, diversity, and inclusion (EDI) are very important to people. Social media such as Twitter, YouTube, and Facebook are some of the important media for information dissemination and communication in the

IberLEF 2021, September 2021, Málaga, Spain.

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

world today. These platforms have a large number of users who conduct various exchanges and release various information on these platforms. Many news media also use these platforms or self-built platforms to release various news information. This makes the Internet flooded with all kinds of news information, this news information is very complex, there is much fake news information, these fake news have some unnecessary interference to our lives, which requires us to be able to identify news the authenticity of the information. The variety and complexity of news information pose a major challenge for us to identify the authenticity of a news item.

To help complete the identification of the authenticity of a certain news item, it is necessary to establish an efficient and accurate system. IberLEF 2021[13] is committed to promoting the equality, diversity, and inclusiveness of language technology, and provides a shared mission for the authenticity of news item. The task established a Spanish data set based on news item collected from the Mexican network. This is a classification task that classifies news item into "True" and "Fake".

To solve this task[11], we use XLM-Roberta, a pre-trained language model based on Transformer. Compared with other methods, this model has some unique advantages. We only need to perform less pre-processing on the pre-training model, and then we can achieve better results for downstream classification tasks, which cannot be achieved by other methods. In addition, the pre-trained model supports fine-tuning for specific tasks.

The rest of the paper is organized as follows: In Section 2 we describe the datasets in detail. Section 3 and describes the model approach we used. We describe our experiments and results in Section 4. Finally, Section 5 gives the conclusion.

2 Data description

The data set[12] used for this task was provided by the organizers of IberLEF 2021, who presented the news corpus in Spanish. The news corpus were collected from January to July of 2018 and all of them were written in Mexican Spanish. The corpus has 971 news collected from different sources.

LANGUAGE DATASET		TOTAL	TRUE	FAKE
Spanish	Train	676	338	338
	Dev	295	153	142
	Test	572		

Table 1. Dataset statistics

We present the statistics for the dataset in Table 1. For a given comment, we need to divide it into the following two categories:

- **True:** A news article is true if there is evidence that it has been published on reliable sites.
- **Fake:** A news article is fake if there is news from reliable sites or specialized websites in the detection of deceptive content that contradicts it or no other evidence was found about the news besides the source.

The statistical data show that the proportion of training set and development set in the number of fake and true news is quite balanced.

3 Methodology

This section describes the deep learning model and architecture[1] that we use to identify the authenticity of news text in this task.

Text categorization[2] has been a focus of research for years as social media has become more popular. In the past, people used SVM[3] and LR classifiers[4] for sentiment analysis. In recent years, text classification technology is mainly implemented by bag-of-words (BOW[5]), recursive neural network (RNN[6]), and Word embedding[7]. In these years of research, the RNN model has achieved good results in emotion analysis tasks. As the research went on, we found that using the pre-training model worked much better than the previous methods.

We chose the XLM-Roberta[8] model based on Transformer[9] and trained the task on the corpus in Spanish languages. XLM-Roberta is mainly composed of bidirectional Transformers, using a dynamic tuning Masking mechanism different from Bert[10]. It is adjusted based on Bert, using the larger batch size and longer training sequence, and performing the pre-training of the next sentence of Bert. XLM-Roberta delivers better downstream mission performance compared to Bert.

The model structure is shown in figure 1 below. The hidden layer of the XLM-Roberta model is 768 dimensions and has 12 Transformer encoder layers. Since a bi-directional Transformer cannot remember time-series information, we add the [CLS] token to the beginning of the input text to indicate whether it is used for a classification task, and the [SEP] token is used as a separator between sentences or a marker at the end of a sentence. Then, after the computation of the neural network, the [CLS] token we get from XLM-Roberta's output is treated as an aggregated representation of the entire text. It is passed as input to a full connection layer, and the Softmax activation function is used by the deep neural network for classification. Thus, the XLM-Roberta Society predicts whether a news item can be classified as "True" or "Fake".

4 Experiments and results

In this section, we describe the methods used to preprocess the data and to train our model on processed text datasets.

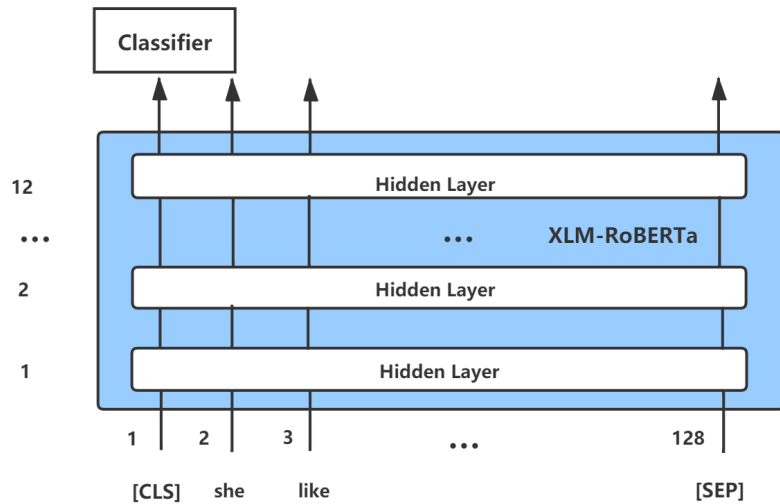


Fig. 1. The structure of XLM-Roberta model

4.1 Data preprocessing

Since the data set was captured directly from Mexican web sources, the original data contained a variety of unnecessary features that would affect the performance of model training, and we processed the text before entering it into the deep learning model. As a pre-processing step, to clean up the data by retaining the important information in the data and deleting the non-important information, we performed the following steps:

- The following data preprocessing is performed on the news item so that the data set achieves good results in the downstream classification and detection tasks on the XLM-Roberta model. There is some information that we don't need in the news item. This redundant information will interfere with our detection. Removing them will improve the performance of the classifier.
- Translate the emoji into a textual description of the corresponding emotion.
- Convert the text to lowercase.
- Remove words that have no emotional meaning.
- Remove all URLs.
- Remove excess Spaces.

4.2 Experimental settings

For the implementation of the model, we used the Transformer library provided by HuggingFace. The Huggingface Transformers package is a Python library

that not only contains pre-trained XLM-Roberta but also provides pre-trained models for various NLP missions. As the implementation environment, we use the PyTorch library, which supports GPU processing. The XLM-Roberta model runs on an NVIDIA RTX 3080 graphics card with 24GB of video storage. Based on our experiment, We use stratified 5-fold cross-validation with 42 random seeds for the training set, and stratified sampling ensures that the proportion of samples in each category of each fold data set remains unchanged. For the XLM-Roberta, we use the pre-trained model, which contains 12 layers. We trained our classifier using Adam optimizer with a learning rate of 2e-5 and CrossEntropy loss. The dropout set to 0.1, the epoch and maximum sentence length is 10 and 512, respectively. If the maximum sentence length exceeds 512, it will cause overflow, stop reading subsequent text data, the model will truncate the data set, and then proceed to the next step. In order to save GPU memory, the batch size was set to 8, and the gradient steps were set to 4, so that each time a sample is input, the gradient is accumulated 4 times, and then the back-propagation update parameters are performed. We extract the hidden layer state of XLM-Roberta by setting the output hidden States as true. In the process of fine-tuning and sequence classification, we use the HuggingFace libraries to provide the RobertaForSequenceClassification module.

4.3 Result

In this work, we will present the results of our submitted evaluations. In this experiment, we participated in the tasks of Spanish languages provided by IberLEF 2021, and the results were evaluated by the task organizer of IberLEF 2021, using a weighted average F1-score as the evaluation standard. The results are shown in Table 2 below. In this task, we got a score of macro F1-0.7053, ranking 7th on the leaderboard. The results are shown in Table 2. The results reported by the organizers showed that the competition among participating teams was very intense, and our best performance in the Spanish language task was an F1 score of 0.7053, which gave us 7th place.

User	Language	Rank	F1-score
zk15120170770	Spanish	7th	0.7053

Table 2. The results of our model on the official test set.

As can be seen from the results, our method works quite well in the Spanish news item dataset. This may be because the XLM-Roberta model is pre-trained in the multi-language dataset, so the work effect is so good in the fake news detection in Spanish.

5 Conclusion

This paper introduces the general idea and specific plan of Team zk15120170770 in IberLEF 2021. As users grow in diversity and number, online platforms must support multiple languages. In the competition, we used the Transformer pre-training model XLM-Roberta to complete Fake News Detection in Spanish. The performance of our system was very competitive and we achieved good results. In the future, we hope to extend our system to more languages and increase the number of tasks it can perform. We will also explore the use of other pre-training models and make comparative analyses.

References

1. Goodfellow I, Bengio Y, Courville A, et al. Deep learning[M]. Cambridge: MIT press, 2016.
2. Sebastiani F. Machine learning in automated text categorization[J]. ACM computing surveys (CSUR), 2002, 34(1): 1-47.
3. Platt J. Sequential minimal optimization: A fast algorithm for training support vector machines[J]. 1998.
4. Jianqiang Z, Xiaolin G. Comparison research on text pre-processing methods on twitter sentiment analysis[J]. IEEE Access, 2017, 5: 2870-2879.
5. Wallach H M. Topic modeling: beyond bag-of-words[C]//Proceedings of the 23rd international conference on Machine learning. 2006: 977-984.
6. Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization[J]. arXiv preprint arXiv:1409.2329, 2014.
7. Levy O, Goldberg Y. Dependency-based word embeddings[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2014: 302-308.
8. Conneau A, Khandelwal K, Goyal N, et al. Unsupervised cross-lingual representation learning at scale[J]. arXiv preprint arXiv:1911.02116, 2019.
9. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. arXiv preprint arXiv:1706.03762, 2017.
10. Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
11. H. Gómez-Adorno, J.-P. Posadas-Durán, G. Bel-Enguix, C. Porto, Overview of fakedestask at iberlef 2020: Fake news detection in spanish., *Procesamiento del Lenguaje Natural*67 (2021)
12. Posadas-Durán J P, Gomez-Adorno H, Sidorov G, et al. Detection of fake news in a new corpus for the Spanish language[J]. *Journal of Intelligent & Fuzzy Systems*, 2019, 36(5): 4869-4876.
13. Manuel Montes, Paolo Rosso, Julio Gonzalo, Ezra Aragón, Rodrigo Agerri, Miguel Ángel Álvarez-Carmona, Elena Álvarez Mellado, Jorge Carrillo-de-Albornoz, Luis Chiruzzo, Larissa Freitas, Helena Gómez Adorno, Yoan Gutiérrez, Salud María Jiménez Zafra, Salvador Lima, Flor Miriam Plaza-de-Arco and Marionna Taulé (eds.): *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021)*, CEUR Workshop Proceedings, 2021.