TSIA team at FakeDeS 2021: Fake News Detection in Spanish Using Multi-Model Ensemble Learning

Zhengyi Guan[1][0000−0002−6265−7193]

School of Information Science and Engineering Yunnan University, Yunnan, P.R. China
1941028528@qq.com

Abstract. Fake news has become a hotly debated topic in journalism. This paper describes our contribution of the TSIA team in the Fake News Detection in Spanish Shared Task of IberLEF 2021. We regard this task as a binary classification task. We mainly propose three model architectures based on the pre-trained model BETO and XLM-RoBERTa-Large. We first fine-tuned the Spanish pre-trained model BETO and then we chose the multi-language pre-trained model XLM-RoBERTa-Large to replace BETO and fine-tune it, including the addition of CNN for feature extraction. Finally, our system achieves best F1-score of 0.6860 by hard voting, which ranks 10th out of 21 teams on the final leaderboard. Our score is only 0.0806 worse than the best score on the leaderboard.

Keywords: Fake News Classification · Natural Language Processing · XLM-RoBERTa-Large· Ensemble.

1 Introduction

This goal of Fake News Detection in Spanish Shared Task at IberLEF 2021 [4] [7] aims to help users detect and filter out potentially deceptive news in social networks. As we all know, social networks offer platforms in which information and articles may be shared without fact-checking or moderation. Moderating user-generated content on social media presents a challenge due to both volume and variety of information posted. In particular, highly partisan fabricated materials on social media, fake news, is believed to be an influencing factor in recent elections [1]. Misinformation spread through fake news has attracted significant media attention recently and current approaches rely on manual annotation by third parties [5] to notify users that shared content may be untrue. Social media information may not only represent a lot of negative emotions(terrorism, political elections, advertisement, satire, among others), but also show the particularity that the people can decide to show or hide their identity. The task of

IberLEF 2021, September 2021, Málaga, Spain.

detecting fake news is defined as the prediction of the chances of a particular news article being deceptive [12]. The conventional solution to this task is to ask professionals such as journalists to check claims against evidence based on previously spoken or written facts. However, it is time-consuming and expensive. For example, it is hard for editors to judge whether a piece of news is real or not. As the Internet community and the speed of the spread of information are growing rapidly, automated fake news detection on Internet content has gained interest in the Artificial Intelligence research community. The goal of automatic fake news detection is to reduce the human time and effort to detect fake news and help us stop spreading it. The task of fake news detection has been studied from various perspectives with the development in subareas of Computer Science, such as Machine Learning (ML), Data Mining (DM), and NLP [8]. Besides the fact that most of the previous works done in these two tasks, namely aggressiveness detection and fake-news detection, are for English, little research has been done for Spanish using the most recent NLP techniques such as deep learning approaches [16]. In this paper, We use popular techniques in natural language processing to solve the problem of identifying fake news in Spanish.

The remainder of the paper is structured as follows: a brief analysis on related work is performed in section 2, followed by a description of the datasets and details on the methods employed for detection of fake news in Section 3. Section 4 outlines the evaluation process and results, while conclusions and future work are drawn in section 5.

2 Related work

For datasets in different languages, it brings challenges to fake news detection. In recent years, researchers have done a lot of research on fake news detection on English datasets. And due to the impact of Covid-19, many competitions have issued tasks on fake news detection. Such as SemEval 2021 Task [1]released the detection of toxic text span, HASOC 2020 [2] issued the challenge of hate speech and offensive content identification in Indo-European languages and CONSTRAINT 2021' task [3], about hostility detection in Hindi. All these show that the detection of fake news has always been a fiery challenge. Hence, the researches on the detection of fake news in Spanish in social media is also valuable. This is also helpful for the detection of Covid-19 information in Spanish social media.

The detection of fake news is the same as other text classification problems in natural language processing. The most important thing is to find suitable features to represent sentences. The task is to assign predefined categories to a given text sequence. Many work has shown that pre-trained models on large corpora are beneficial for text classification and other NLP tasks, which can avoid training new models from scratch. Since 2013, people have proposed some word embedding approaches such as word2vec [6] and glove [9]. However, because their

[1] https://sites.google.com/view/toxicspans
[2] https://hasocfire.github.io/hasoc/2020/
[3] http://lcs2.iiitd.edu.in/CONSTRAINT-2021/

word embeddings are all in the same space, they can not express the role of polysemy. In other words, they are non-contextual embedding, they can not capture the high-level concepts of sentences, such as semantics and context [13]. Later, someone proposed the ELMo [10] model to solve this problem. Compared with word2vec and glove, ELMo captures contextual information and not just individual information of words. In word2vec, the vector representations of words are completely consistent in different contexts, but ELMo is optimized for this [17]. More recently, pre-trained language models have shown to be useful in learning common language representations by utilizing a large amount of unlabeled data: such as OpenAI GPT [2] and BERT [3]. BERT is based on a multi-layer bidirectional Transformer [15] and is trained on plain text for masked word prediction and next sentence prediction tasks. Since BERT is suitable for English and the dataset of this competition is Spanish, which also added Covid-19 related data for English. We finally choose BETO[4] and a multi-language pre-trained model—XLM-RoBERTa-Large [5] as our pre-trained model. And we fine-tuned this two pre-trained models, submited three Runs and made a hard voting on the three Runs finally.

3 Data and Methods

3.1 Dataset

The dataset used in the model are all provided by the organizer. There are 676 training set and 295 development set. The corpus consists of news compiled mainly from Mexican web sources: established newspaper websites, media companies websites, special websites dedicated to validating fake news, and websites designated by different journalists as sites that regularly publish fake news. The corpus contains the following information [11]:

- **Category**: Fake / True.
- **Topic**: Science / Sport / Economy / Education / Entertainment / Politics, Health / Security / Society.
- **Source**: The name of the source media.
- **Headline**: The title of the news.
- **Text**: The complete text of the news.
- **Link**: The URL where the news was published.

Since the corpus contain different labels, in order to increase the learning ability of the model. We added "Category" and "Topic" column to the "Text" column. We did not use the label—"Link". This does improve the learning ability of the model, but it also leads to the poor generalization ability of the model. In addition, we did simple data preprocessing, such as: we strip emojis from the training set, and we deleted the link of website, etc.

[4] https://github.com/dccuchile/beto
[5] https://huggingface.co/xlm-roberta-large

3.2 Fine-tuned of BETO and XLM-RoBERTa-Large

Pre-trained and fine-tuning architecture is already a popular method for text classification. Our system used BETO and XLM-RoBERTa-Large as the pre-trained model, and we provided three runs with ensemble. They are:

- **Run 1**: Fine-tuned of BETO
- **Run 2**: XLM-RoBERTa-Large
- **Run 3**: XLM-RoBERTa-Large + CNN

BETO is similar to BERT. They all have 12 hidden layers. BETO is a BERT model trained on a big Spanish corpus. BETO is of size similar to a BERT-Base and was trained with the whole word masking technique. Representing each word in the sentence as a vector, which includes word embedding and character embedding. The character embedding is initialized randomly. The word embedding is usually imported from a pre-trained word embedding file. All embeddings will be fine-tuned during training. For the Run 1, as is shown in Fig. 1, P_O is the pooler output of BETO, HO is hidden-state of the first token of the sequence(CLS token) at the output of the hidden layer of the model. Then, we concatenate P_O and HO of the last three hidden layers into the classifier after obtaining P_O.

The Facebook AI team released XLM-RoBERTa in November 2019 as an update of its original XLM-100 model. They are all transformer-based language models, all rely on the mask language model target, and they can handle texts in 100 different languages. Compared to the original version, the biggest update of XLM-RoBERTa is a significant increase in the amount of training data. The commonly used crawler datasets that have been cleaned and trained occupy up to 2.5tb of storage space. It is several orders of magnitude larger than the Wiki-100 corpus used to train its previous version, and this expansion is especially noticeable in languages with fewer resources. XLM-RoBERTa-Large adds 12 hidden layers on the basis of XLM-RoBERTa. Therefore, the network structure of XLM-RoBERTa-Large is much more complicated, and the number of pre-trained layers is deeper. For the Run 2, wo just add a classifier after the XLM-RoBERTa-Large(Note: we did not give the architecture of Run2). For the Run 3, as is shown in Fig. 2, we add CNN before P_O is sent to the classifier. Firstly, we got pooler output (P_O), P_O is the pooler output of XLM-Roberta-Large. It is obtained by its last layer hidden state of the first token of the sequence (CLS token) further processed by a linear layer and a tanh activation function. Then, we let P_O go through a three-layers CNN (including convolution and pooling). Finally, input this two-dimensional vector into a linear classifier to do a binary classification.

3.3 Ensemble learning

We use the multi-model ensemble learning approach to get a stable system that performs well in all aspects. We further use hard voting to determine the final category, whose main idea is to vote for a speech by the classification results

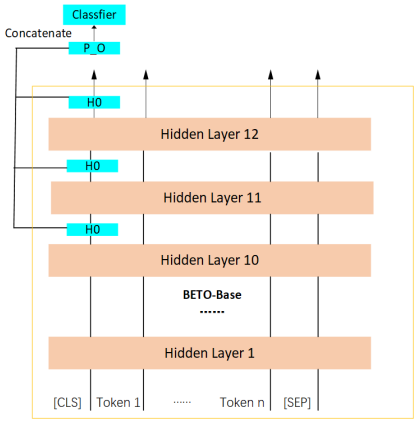

Fig. 1. Model for Run 1

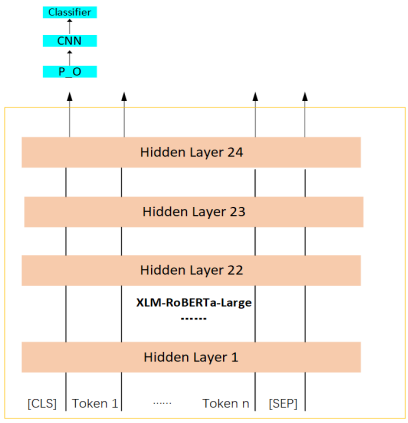

Fig. 2. Model for Run 3

of each model and the minority obeys the majority. Thus, our final predition result integrates the models of Run 1, Run 2 and Run 3 by ensemble learning. The experimental results in the next chapter verify the effectiveness of ensemble learning.

4 Experiments and Results

4.1 Hyper-parameters settings

In this work, our models were implemented based on Pytorch [6]. Our experiments were run on Google Colab [7]. The GPU is Tesla P4. The batch size is 32. Our hidden layer state of BETO and XLM-RoBERTa-Large by setting the output hidden states was True. We used the adam optimizer and the learning rate of three Runs was 5e-5. The three models were trained in 30 epochs. For the Run 3, we used three convolutional layers. The number of convolution kernels is 256. The activation function is Relu. The pooling layer uses maximum pooling.

4.2 Criteria evaluation and results

We mainly used F1-score to evaluate our model. The criteria evaluation of F1-score is as follows:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}, F1 = \frac{Precision * Recall * 2}{Precision + Recall}$$

The result is shown in Table 1.

[6] https://pytorch.org/
[7] https://drive.google.com/drive/my-drive

Table 1. Result of three Runs on development set and test set

	Development set			Test set		
Run	Accuracy	F1	Recall	Accuracy	F1	Recall
Run 1	0.9392	0.9389	0.9390	-	-	-
Run 2	0.9593	0.9593	0.9595	-	-	-
Run 3	0.9695	0.9695	0.9698	-	0.6252	-
Ensemble	-	-	-	-	0.6860	-

4.3 Result analysis

From the data in Table 1, it can be seen that the three Runs on the development set all obtain good results, which the F1-score of Run 3 is the best. This shows that CNN is helpful in this task. Therefore, we choose the XLM-RoBERTa-Large + CNN architecture to predict the final test set. The result using this model on the test set is 0.6252. Finally, we submitted the result of ensembling the three Runs by hard voting. The final best result on the test set is 0.6860, which shows that ensemble learning strengthens the learning ability of multiple classifiers.

But the results of our model on the test set are not the most competitive. This may be because we did not do a better job of data augmentation(DA), which leads to the poor model generalization. We need to allow limited data to produce value equivalent to more data without substantial increase in data. Therefore, we need to put more effort in data processing and augmentation.

5 Conclusions and future work

In this paper, we describe our strategy to classify fake and real text in Spanish document. In our three systems, we used transformers based pre-trained models, BETO, XLM-RoBERTa-Large and XLM-RoBERTa-Large adding CNN. Our proposals show to be competitive for this specific task. However, we must also further test and improve our model, because our results are 0.0806 worse than the best F1-score. So we still have a lot of work to do in the future.

In the future, We should first try to fine-tune the appropriate parameters of the model, because we have not done too many attempts to fine-tune the parameters. Then, future development directions include exploring other related datasets for fake news fields. Also, We just did ensemble learning for the prediction results of the three models. We need to try more integrated learning methods. And we have too few ensemble models, We need to explore more models that are as competitive as others. In addition, advanced error analysis techniques, such as feature importance or model explainability, could also be used to improve the model's performance [14].

References

1. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. Journal of Economic Perspectives **31**(2), 211–236 (2017)

2. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Amodei, D.: Language models are few-shot learners (2020)
3. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. CoRR **abs/1810.04805** (2018), http://arxiv.org/abs/1810.04805
4. Gómez-Adorno, H., Posadas-Durán, J.P., Bel-Enguix, G., Porto, C.: Overview of fakedes task at iberlef 2020: Fake news detection in spanish. Procesamiento del Lenguaje Natural **67**(0) (2021)
5. Heath, A.: Facebook is going to use snopes and other fact-checkers to combat and bury 'fake news' (2016)
6. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. Computer Science (2013)
7. Montes, M., Rosso, P., Gonzalo, J., Aragón, E., Agerri, R., Álvarez-Carmona, M.Á., Mellado, E.Á., de Albornoz, J.C., Chiruzzo, L., Freitas, L., Adorno, H.G., Gutiérrez, Y., Zafra, S.M.J., Lima, S., de Arco, F.M.P., Taulé, M. (eds.): Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021). CEUR Workshop Proceedings, 2021
8. Oshikawa, R., Qian, J., Wang, W.Y.: A survey on natural language processing for fake news detection (2018)
9. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Conference on Empirical Methods in Natural Language Processing (2014)
10. Peters, M., Neumann, M., Iyyer, M., Gardner, M., Zettlemoyer, L.: Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers) (2018)
11. Posadas-Durán, J.P., Gómez-Adorno, H., Sidorov, G., Escobar, J.J.M.: Detection of fake news in a new corpus for the spanish language. Journal of Intelligent & Fuzzy Systems **36**(5), 4869–4876 (2019)
12. Rubin, V.L., Conroy, N.J., Chen, Y.: Towards news verification: Deception detection methods for news discourse. In: Hawaii International Conference on System Sciences (2015)
13. Sun, C., Qiu, X., Xu, Y., Huang, X.: How to fine-tune bert for text classification? (2020)
14. Tanase, M.A., Zaharia, G.E., Cercel, D.C., Dascalu, M.: Detecting aggressiveness in mexican spanish social media content by fine-tuning transformer-based models. In: MEX-A3T at IberLEF 2020: Authorship and aggressiveness analysis in Twitter: case study in Mexican Spanish (2020)
15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. CoRR **abs/1706.03762** (2017), http://arxiv.org/abs/1706.03762
16. Villatoro-Tello, E., Ramírez-De-La-Rosa, G., Kumar, S., Parida, S., Motlicek, P.: Idiap and uam participation at mex-a3t evaluation campaign. In: IberLEF2020 (2021)
17. Zhang, Y., Shen, D., Wang, G., Gan, Z., Carin, L.: Deconvolutional paragraph representation learning. In: NIPS (2017) (2017)