

YETI at FakeDeS 2021: Fake News Detection in Spanish with ALBERT

Hongxin Luo

School of Information Science and Engineering Yunnan University, Yunnan, P.R.
China
1104792873@qq.com

Abstract. This paper explains our participation in the IberLEF2021 shared task 7: Fake News Detection Task. The goal of this task is to analyze a corpus of Spanish News and determine the authenticity of its content. The threat of false information is designed to negatively affect people, by disseminating information that does not match the facts, so that users can accept biased or erroneous information. Therefore, fake news becomes particularly important. For this task, this paper mainly discusses different methods of fake news detection. We chose the ALBERT Model. On this basis we made a simple modification to the upper structure of the ALBERT model. In the end, our system got 63.16 % F1 score in the task. Although our proposal did not reach the best, it provides a new idea for fake news detection.

Keywords: ALBERT · Fake News Classification · Natural Language Processing · Deep-learning.

1 Introduction

Many years ago, the main channels for us to obtain news and information were television and newspapers. In recent years, with the rise of the mobile Internet, more and more people are choosing to obtain news information from social media. But the quality of news on social media is far lower than traditional media. Since anyone can easily publish a news article on social media, the quality of articles on social media is uneven, and there are even a lot of fake news.

The threat of false information is designed to negatively influence people and deliberately persuade users to accept biased or erroneous information. Therefore, detecting fake news on social media becomes particularly important. This is the mission of IberLEF2021 [3], the forum aims to encourage research on social media content analysis in Spanish [1][10]. In this work, we explored the task of fake news detection in IberLEF2021 from the perspective of deep learning [3]. This task can be regarded as a binary classification in Spanish. The corpus consists of news

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).

compiled mainly from Mexican web sources: established newspaper websites, media company websites, special websites dedicated to validating fake news, websites designated by different journalists as sites that regularly publish fake news. The news was collected from January to July of 2018 and all of them were written in Mexican Spanish [10]. There are a total of 971 news items in the corpus.

We used several different neural network models for comparison, such as convolutional neural network (TextCnn) [12], fast text classifier (fastText) [4] and a little BERT for self-supervised learning of language representations model (ALBERT) [7]. For the given data set in the task, we found that ALBERT performed best on our validation set. Therefore, we accomplish this task by using the ALBERT model.

The rest of this paper is organized as follows. Chapter 2 briefly introduces related work. Chapter 3 introduces our method in detail, including the description of the data set, data preprocessing and architecture. Chapter 4 outlines the evaluation process. Finally, Chapter 5 summarizes our work.

2 Related Work

IberLEF is an Iberian language evaluation forum for NLP tasks. In the 2020 version of fake news detection, participants have proposed a variety of methods, from traditional machine learning to deep learning, such as BoW, n-grams, Neural Networks, Transformers, etc. [1][8][13]. According to the author’s analysis, the best results are obtained using the Supervised Autoencoder (SAE) method, which is a neural network that learns the representation (encoding) of the input data and then learns to reconstruct the original input [14]. They use three different types of features as input representation: word n-grams, char n-grams and BETO encodings [14]. In the previous version, they used the supervised automatic coding method to get good results [1].

Detecting fake news on social media is becoming more and more important. To build an effective classifier, one of the most important problems is to find suitable input features. Generally, there are two types of features that are widely used: one is a surface feature, such as n-grams [8], and the other is a word representation trained by a neural network, such as skip-grams. General classifiers use traditional machine learning methods, such as support vector machines, random forests, logistic regression, etc., to train for different types of tasks. In many NLP tasks, it is effective to use pre-trained word embeddings to extract features [16][15]. The word embedding model is extracted from a shallow neural network, which requires the neural network to be obtained by training a large amount of text data, it can learn the contextual representation of words, such as skip-grams and GloVe [9]. But these word embeddings are learned from all possible words, which makes the word embedding may cover up the nuances of semantics. However, transformer-based language models, such as OpenAI Generative Pre-trained Transformer (GPT) [11] and BERT [2] have been extended to a depth of as much as 12 layers. ALBERT uses techniques such

as parameter sharing and matrix decomposition to greatly reduce model parameters [7]. ALBERT can greatly improve the level of language models. It can learn a good feature representation for words by running an unsupervised language representation learning algorithm based on a massive corpus. The so-called self-supervised learning means that there is no human Supervised learning running on labeled data. Compared with ELMo and GPT, the pre-trained ALBERT model has achieved good results in a series of NLP tasks [7].

3 Methodology

3.1 Datasets

The Spanish news corpus was collected from January to July 2018, all written in Mexican Spanish. They are news aggregated from several online sources: existing newspaper websites, media company websites, websites that specialize in verifying fake news, and websites designated by different reporters to publish fake news on a regular basis. There are 971 news items in the aggregated corpus. The news includes 9 different types of news topics, making the corpus as balanced as possible. The number of fake news and real news is also roughly balanced. In the data set, 676 pieces of data are used as the training data set and 295 pieces are used as the validation data set. The ratio of the training set to the validation set is about 7:3 [1].

3.2 Pre-processing

Although deep learning methods can learn the main features from the data, the output performance of the model also depends on the expected quality of the input training data set and 295 pieces are used as the validation data set. The ratio of the training set to the validation set is about 7:3 [6]. Data preprocessing can remove the noise data in the input data to improve the performance of the model. For the model we used, we have performed the basic preprocessing of the data as follows:

1. Convert the input text to lowercase.
2. Remove punctuation marks.
3. Delete numeric characters.
4. Delete the stop-words.

We removed information that was not useful for model extracting features. We used the Natural Language Toolkit (NLTK) to complete the stop-word removal step. In the experiment, we use 5-fold cross-validation to control the order of data in each batch. For each fold of the data set, the input data format is [CLS]+sentence+[SEP]([CLS], which are used to separate each sample, [SEP], which are used to separate different sentences in the sample). The pre-trained model is loaded from the ALBERT-base-V2 model [7]. In V2 version, ALBERT apply 'no dropout', 'additional training data' and 'long training time' strategies to all models. ALBERT-base is trained for 10M steps and other models for 3M steps.

3.3 ALBERT

The research trend in the NLP field is to use larger and larger models to obtain better performance, the depth of the network can improve the results of the model [11][2]. Research on ALBERT shows that blindly stacking model parameters may reduce the effect., and memory and training speed will also be hindered. ALBERT solves this problem by designing a Lite BERT architecture, which has fewer parameters than the traditional BERT architecture [7]. ALBERT is “A Lite” version of BERT, a popular unsupervised language representation learning algorithm. ALBERT uses parameter-reduction techniques that allow for large-scale configurations, overcome previous memory limitations, and achieve better behavior with respect to model degradation. ALBERT uses parameter sharing, matrix decomposition and other technologies to greatly reduce model parameters, and at the same time replaces NSP (Next Sentence Prediction) Loss with SOP (Sentence Order Prediction) Loss to improve the performance of downstream tasks. The reduction of parameters can make training faster [7]. The structure of ALBERT is basically the same as BERT, and there are three specific improvements. Embedding layer parameter factorization, cross-layer parameter sharing, NSP task is changed to SOP task. We use albert to fine-tune on the training data set.

3.4 Method

In classification task, the output of ALBERT-Base(pooler output) 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. Because the pooler output cannot summarize the input semantic content well, and studies have shown that the top layer of Bert can learn richer semantic information features [11]. We try to modify the ALBERT model to obtain rich semantic information features. We pass Spanish News directly to the ALBERT model, and we concatenate H_0 (H_0 is hidden-state of the first token of the sequence (CLS token) at the output of the hidden layer of the model.) of the last three hidden layers into the classifier. We call this method ALBERT_Classifier.

4 Results

Task 7 is to detect fake news, and fake news detection solution will be ranked by the F1 measure on the “Fake” class.

In our work, the implementation of all models is based on TensorFlow, and the pre-trained models are cased. Due to the limitation of personal GPU memory, the batch size and max-seq-length in the fine-tuning stage were adjusted according to the memory capacity in order to achieve the best results. The optimizer used in the model in this experiment is Adam [5]. Table 1 shows the hyperparameters of each model on the validation data set of the fake news detection task and the results of the model on the validation data set.

Table 1. The hyper-parameters of our model and the results on the validation set.

Model Architecture	Hyperparameters	Stopwods	ACC	F1
ALBERT	max-seq-length=512	Yes	0.7627	0.7602
	train-batch-size=16	No	0.7762	0.7758
ALBERT_Classifier	warmup-step=100	Yes	0.8169	0.8154
	learning-rate=1e-5	No	0.8203	0.8199

Table 1 shows the results of the model used in this paper on the validation set. We can see from the results in the above table that the result of our modified ALBERT model is better than the ALBERT model. Compare to know that our modification to the ALBERT model has made it richer in semantic content and has a certain improvement. In the experiment, we also found that deleting stop-words can improve the classification effect of fake news to a certain extent. We input the processed data into the model, use the training data set data for training, and use the test data set to predict the results of the model. We combine the four results through hard voting to get the result. We can output the final prediction result according to the proportion of the prediction results of all the predictors. We use the absolute majority voting method to predict the result, that is, if a prediction result has more than half of the votes, the prediction is the label; otherwise, the category involved is randomly predict. The final F1 result obtained in our model for this task is 63.16%.

5 Conclusions

In this paper, we introduce the method of participating in the sharing task of Spanish fake news detection organized by IberLEF2021. We propose to modify the upper structure of the ALBERT model and use the ALBERT-Base-V2 pre-trained model for training. The experiment uses 5-fold cross-validation. Finally, we get the final result through hard voting. In future work, we hope to explore more effective data preprocessing methods and use data augmentation to make the model perform better, and improve our results in the next IberLEF competition.

References

1. Aragón, M., Jarquín, H., Gómez, M.M.y., Escalante, H., Villaseñor-Pineda, L., Gómez-Adorno, H., Bel-Enguix, G., Posadas-Durán, J.: Overview of mex-a3t at iberlef 2020: Fake news and aggressiveness analysis in mexican spanish. In: Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain (2020)
2. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)

3. 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)
4. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759 (2016)
5. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2014)
6. Kotsiantis, S.B., Kanellopoulos, D., Pintelas, P.E.: Data preprocessing for supervised leaning. *International Journal of Computer Science* **1**(2), 111–117 (2006)
7. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R.: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942 (2019)
8. Liu, S., Demirel, M.F., Liang, Y.: N-gram graph: Simple unsupervised representation for graphs, with applications to molecules. arXiv preprint arXiv:1806.09206 (2018)
9. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. pp. 1532–1543 (2014)
10. Posadas-Durán, J.P., Gomez-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)
11. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. *OpenAI blog* **1**(8), 9 (2019)
12. Rakhlin, A.: Convolutional neural networks for sentence classification. GitHub (2016)
13. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. arXiv preprint arXiv:1706.03762 (2017)
14. Villatoro-Tello, E., Ramírez-de-la Rosa, G., Kumar, S., Parida, S., Motlicek, P.: Idiap and UAM participation at mex-a3t evaluation campaign. In: *Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF)*, Malaga, Spain (2020)
15. Wang, J., Peng, B., Zhang, X.: Using a stacked residual lstm model for sentiment intensity prediction. *Neurocomputing* **322**, 93–101 (2018)
16. Wang, J., Yu, L.C., Lai, K.R., Zhang, X.: Community-based weighted graph model for valence-arousal prediction of affective words. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **24**(11), 1957–1968 (2016)