Emotion Classification for Spanish with XLM-RoBERTa and TextCNN

Suidong Qu[1][0000−0002−7274−5891], Yanhua Yang[2][0000−0003−1508−4318], and Qinyu Que[3][0000−0001−6688−7896]

[1] Yunnan University,Yunnan,P.R.China
icat@mail.ynu.edu.cn
[2] Yunnan University,Yunnan,P.R.China
thomasduke2008@gmail.com
[3] Yunnan University,Yunnan,P.R.China
K1ky0@pm.me

Abstract. On social networking platforms, users usually cannot perceive each other's speech tone and facial expressions, making the emotions conveyed by users on the Internet not clear enough. This task is Spanish emotion detection and evaluation from EmoEvalEs@IberLEF 2021. Specifically, the task includes dividing the emotions expressed on Twitter into one of the following seven categories: Anger, Disgust, Fear, Joy, Sadness, Surprise or Others. Our team (team name is Dong) first use XLM-Roberta for embedding. Then we input the word vector into Transformer Encoder for the secondary extraction of features, and then input the result into TextCNN. Using TextCNN's ability to capture local features, our model could extract high-level features such as text semantics, word order, and context. Finally, the output of the model is input into the fully connected layer for classification. Our model rank 14th in this task. The weighted-averaged F1 is 0.5570, and the accurcy is 0.5368.

Keywords: Emotion classification · TextCNN · XLM-RoBERTa.

1 Introduction

With the popularization of the mobile Internet, it is easier for people to express their opinions and emotions on online social media platforms. At the same time, due to the popularity of Twitter, short texts with emotional inclination that people publish on Twitter have the characteristics of wide spreading and fast speed. These texts can quickly produce an important and far-reaching impact on social development. With the passage of time, the number of emotionally inclined text content in the mobile Internet has explosively increased, and these data have made the task of emotional classification in text a hot research topic. Emotion classification is an important task in various basic tasks of emotion analysis. The

IberLEF 2021, September 2021, Málaga, Spain.

task of emotion classification needs to find emotional texts in subjective texts and analyze them [1]. Different from the traditional classification method that analyzes the objective content of the text, emotion classification is to extract some viewpoint information from the text. So far, a large number of researchers have proposed many natural language processing models based on deep learning to realize emotion classification. This task [2] is Spanish emotion detection and evaluation from EmoEvalEs@IberLEF 2021 [3].

2 Related Work

In the past 10 years, text-based emotion analysis has been a hot research field that has attracted much attention from researchers in psychology, sociology, and computer science. The emotion analysis of the text is the study of how to dig out the subjective sentiment tendency of the user from the text [4]. In 2013, Mikolov et al. [5] proposed the Word2vec model. This model maps high-dimensional vectors to low-dimensional spaces, and proposes a natural language analysis method based on word vectors. In 2014, Kim et al. [6] proposed a model for processing text classification tasks in natural language processing. Because the Word2vec model lacks the ability to detect local features of text, the model includes a convolutional neural network (TextCNN) that can extract local features of text. In 2014, Jeffrey Pennington et al. [7] proposed a model (Glove) that combines the advantages of global matrix factorization and local context windows, and learns word vectors by mapping words to global vectors. In 2018, Peters [8] proposed an embedded language model (ELMo) based on a special two-way LSTM structure, which can learn the usage and expression of words in text in a fixed direction. In 2018, OPEN AI proposed a generative pre-trained transformer model (GPT) based on a one-way Transformer structure [9]. This model uses a unique self-attention mechanism in Transformer to enable the model to extract sentences. Thanks to the multi-head attention mechanism of the Transformer model, the pre-training methods of many natural language processing models have been improved accordingly. In 2018, Google researchers proposed a deep two-way code converter (BERT) [10] based on the Transformer encoding structure. This model uses the MLM (Masked Language Modeling) strategy to achieve two-way language learning in the pre-training process. Later, the RoBERTa model [11] released by Facebook improved the original pre-training method of the BERT model, and further improved the training effect of related models. In 2019, researchers from the GoogleBrain team proposed an XLNet model based on an autoregressive language model pre-training method [12]. The XLNet model can learn bidirectional semantic information in the text, and can be competent for many natural language processing tasks. We participate in this task for Spanish and propose a method that includes XLM-RoBERTa [13], Transformer Encoder and TextCNN models. It can integrate the advantages of each model to enhance the effectiveness of the classification of emotions.

3 Data and Resources

The data set [14] used in this mission is provided by EmoEvalEs@IberLEF 2021, which is derived from online tweets based on global events in different fields in April 2019. The data set is divided into training, development, and testing parts, in which the hashtag is replaced by the keyword "HASHTAG". Our task is to divide the tweets in the data set into one of Anger, Disgust, Fear, Joy, Sadness, Surprise or Others, where Others represents neutral or no emotion.

4 System Description

4.1 Data Preprocessing

We remove punctuation marks, emojis, empty characters and other special symbols in the data set. Removing them can reduce the difficulty of training the model without affecting the classification results. Considering that this task is a text-based seven classification task, our model inserts a [CLS] identifier before the sentence for text classification. Then the model predicts what the words are obscured based on the remaining words. As the training process progresses, the tokens that are masked in the sentence are not exactly the same each time. The model will gradually adapt to different mask positions, thereby learning multiple semantic features.

4.2 Model Description

Our system includes XLM-RoBERTa, Transformer Encoder, and TextCNN models. During the training process, we train the weights of XLM-RoBERTa, TextCNN, Transformer Encoder and final classification layer. The RoBERTa model inherits some characteristics of the BERT model, and it expresses the input sentence as a word vector, a sentence vector, and a position vector. Then RoBERTa optimizes the pre-training method in terms of model structure and data processing, which uses more training resources, more training data, larger batch-size and longer training time. The RoBERTa model uses continuous full-sentences and doc-sentences as the input of the model and removes the NSP loss. In this task, we use Hugging Face's implementation of XLM-RoBERTa model, which inherits the XLM training method and draws on the ideas of RoBERTa, making our model more suitable for cross-language training tasks in this task. In this layer, our network layer has 12 layers, the hidden layer 768 layers, and the number of self-attention heads is 12.

The XLM-RoBERTa model converts the input labeled corpus into corresponding feature maps, which integrate global feature information. We consider that not only the amount of parameters in this layer is huge, but also the amount of parameter changes is small. This situation is likely to cause the model to overfit and make the final classification effect unsatisfactory. Therefore, we use the output of the XLM-RoBERTa model as the input of the Transformer Encoder

layer, which is to use the encoder to perform secondary feature extraction on the information of the previous layer. Since the parameters in the Transformer Encoder layer are much smaller than the parameters of the previous model, we find that its parameters changed a lot during the training process and are more sensitive to changes in the input data. This method effectively alleviates the over-fitting phenomenon of the results and enhances the generalization ability of the entire model. The Transformer Encoder layer we designed contains only one Transformer encoding block.

The output of the Transformer Encoder layer has global features of the text. We use it as the input of TextCNN to reduce parameters and capture the local features. After obtaining the local features of the text, we input them into the max-pooling layer and the fully connected layer (softmax) to obtain the classification results. The TextCNN in this method includes a one-dimensional convolutional layer.

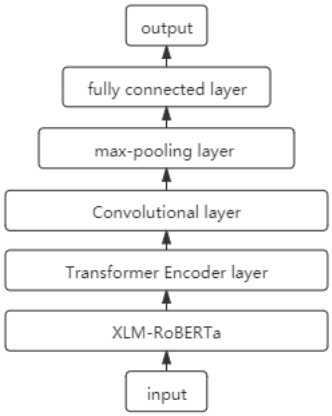

Fig. 1. Our system includes XLM-RoBERTa, Transformer Encoder, and TextCNN models.

5 Hyper-parameters and Results

In this task, we use PyTorch to implement our model. We use Hugging Face's implementation of XLM-RoBERTa model. The optimization method of the model is Adam algorithm and cross entropy loss function. The activation function in the convolutional layer is ReLU. We fine-tuned the model many times, and the hyperparameter settings of the final model are shown in Table 1.

Table 1. Hyper-parameters

hyper-parameters	Values
learning rate	1e-5
per gpu train batch size	64
Gradient accumulation steps	16
num train epochs	12
max seq length	100

The evaluation indicators of this classification task are accuracy and the weighted-averaged versions of precision, recall, and F1. The final results of our model are shown in Table 2. Our model ranks 14th in this task, the weighted-averaged F1 is 0.5570 (Top team is 0.7170), the accurcy is 0.5368 (Top team is 0.7276).

Table 2. Evaluation results

Accuracy	0.5368
Precision	0.6537
Recall	0.5368
F1	0.5570

Table 3. The results of different models on the development set

	Accuracy	Averaged F1 score
XLM	0.5056	0.4951
XLM-RoBERTa	0.6274	0.6188
Our model	0.6369	0.6471

The results of other models are shown in Table 3. We use accuracy and F1 to evaluate the performance of our model. Compared with XLM-RoBERTa, our model has improved its accuracy by nearly 1% and its F1 by nearly 3% on the development set. Although our model may use the TextCNN layer to capture local features, the final effect is very limited. We believe that there may be two reasons for this result. On the one hand, the training set data distribution is unbalanced. In particular, the number of "fear" is very small compared to other tags, and the number of "others" accounts for nearly half of the total amount of data. This makes the model biased towards the side with more data during training. On the other hand, the total amount of data in the training set is relatively small. We train the weight of the model by increasing the number of training epochs. But in order to prevent the occurrence of over-fitting, we can only choose a compromise.

6 Conclusion

In this article, we propose a classification model to solve the task of Spanish emotion detection and evaluation. According to our experimental process, we have considered the problem of less training data for this task and imbalanced data distribution. Due to the huge amount of parameters of the XLM-RoBERTa model, different downstream tasks require different fine-tuning strategies, and this situation requires a lot of data to support. We have tried to use convolutional neural networks to capture the local features of the text and reduce the model parameters to make the classification effect better, but the results are not ideal. Our weighted-averaged F1 is 0.5570 and accurcy is 0.5368. This result shows that our model has certain limitations. For example, the performance on this data set is not outstanding, and it does not perform well on the more refined data set. In view of these disadvantages, we can consider weighting the loss of each category. In addition, we plan to use K-fold cross-validation for model fine-tuning, and then find the hyperparameter values that make the model's generalization performance optimal.

Acknowledgements

We would like to thank the organizers for organizing this task and providing data support, and thank the review experts for their patience. Finally, we would like to thank the school for supporting our research.

References

1. Routray, P., Swain, C.K., Mishra, S.P.: A survey on sentiment analysis. International Journal of Computer Applications **76**(10), 1–8 (2013)
2. Plaza-del-Arco, F.M., Jiménez-Zafra, S.M., Montejo-Ráez, A., Molina-González, M.D., Ureña-López, L.A., Martín-Valdivia, M.T.: Overview of the EmoEvalEs task on emotion detection for Spanish at IberLEF 2021. Procesamiento del Lenguaje Natural **67**(0) (2021)
3. Montes, M., Rosso, P., Gonzalo, J., Aragón, E., Agerri, R., Álvarez-Carmona, M.Á., Álvarez Mellado, E., Carrillo-de Albornoz, J., Chiruzzo, L., Freitas, L., Gómez Adorno, H., Gutiérrez, Y., Jiménez-Zafra, S.M., Lima, S., Plaza-de Arco, F.M., Taulé, M. (eds.): Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021) (2021)
4. Gao, W., Sebastiani, F.: From classification to quantification in tweet sentiment analysis. Social Network Analysis and Mining **6**(1), 1–22 (2016)
5. CHURCH, Ward, K.: Word2vec. Natural Language Engineering **23**(01), 155–162 (2017)
6. Jaderberg, M., Simonyan, K., Vedaldi, A., Zisserman, A.: Reading text in the wild with convolutional neural networks. International Journal of Computer Vision **116**(1), 1–20 (2016)
7. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1532–1543 (2014)

8. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., 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). vol. 1, pp. 2227–2237 (2018)
9. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. vol. 30, pp. 5998–6008 (2017)
10. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding (2018)
11. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach (2019)
12. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., Le, Q.V.: Xlnet: Generalized autoregressive pretraining for language understanding (2019)
13. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V.: Unsupervised cross-lingual representation learning at scale (2019)
14. Plaza del Arco, F.M., Strapparava, C., Urena Lopez, L.A., Martin, M.: EmoEvent: A multilingual emotion corpus based on different events. In: Proceedings of the 12th Language Resources and Evaluation Conference. pp. 1492–1498. European Language Resources Association, Marseille, France (May 2020), https://www.aclweb.org/anthology/2020.lrec-1.186