

Emotion Analysis for Spanish Tweets: The Model based on XLM-RoBERTa and Bi-GRU

Yuanchi Qu¹[0000-0002-0971-1795], Shuangjun Jia²[0000-0001-8315-5662], and Yanjie Zhang³[0000-0001-7356-6791]

¹ Yunnan University, Yunnan, P.R. China
qychi@foxmail.com

² Yunnan University, Yunnan, P.R. China
2858044698@qq.com

³ Yunnan University, Yunnan, P.R. China
jyzhangisme@163.com

Abstract. Our team (the team-name is qu) participate in the classification task of EmoEvalEs@IberLEF 2021. The task requires to classify the emotion of the Spanish tweets (anger, disgust, fear, joy, sadness, surprise, others). To solve the problem, we propose a model (XLM-RoBERTa combined with Bidirectional Gated Recurrent Units). In our model, the accuracy is 0.449879 and the averaged precision is 0.618833. In the final evaluation results, our team ranked 15th.

Keywords: Emotion Analysis · BiGRU · Spanish · Classification.

1 Introduction

With the development of society and technology, the number of users on the network has increased sharply, and communication on Internet has become an important way of communication in modern society [1]. The comments made on social media have a huge impact on the mood of users. Emotional classification of the speech on social media can help the government monitor people's emotional changes and trends of public opinion, and this even can avoid the occurrence of vicious and false incidents [2]. Therefore, it has great significance to detect and classify users' emotions expressed on social media.

Since most of the tweets' messages do not have intonation and expression. So it is a difficult task to analyze the emotion of users on social media. To encourage the development of this field, EmoEvalEs@IberLEF 2021 [3] released a new task: Emotion detection and evaluation for Spanish. The goal is to classify the emotion of tweets into 7 categories (anger, disgust, fear, joy, sadness, surprise, others). This task has three main challenges: (1) Most of the tweets are very short sentences, so the context cannot be analyzed effectively. (2) Tweets are

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).

not formal text, and there are likely some unnormal factors such as misspelling of words and emoticons. (3) Different from the two-classification task, this task is a seven-classification task, which has more stringent requirements for model performance.

This article uses a model combining XLM-RoBERTa [4] with Bi-GRU [5] to complete the emotion classification of Spanish. Experimental results show that this model can effectively improve the performance of emotion classification. In the process of model training, we use the training data provided by the competition official. The paper's structure of the rest is as follows: The second part is related research work on the analysis of emotions. The third part is about the processing of the data set and the description of the model (XLM-RoBERTa and Bi-GRU). The fourth part is the analysis of the experimental results and the summary of the shortcomings. The fifth part is the conclusion of the paper. And the sixth part is the acknowledgements of the paper.

2 Related Work

In recent years, in the fields of NLP (Natural Language Processing) and DM (Data Mining), researches about emotion analysis have always existed. Scholars at home and abroad are mainly doing researches and learning from emotional dictionaries, Machine Learning, and Deep Learning. The keys of the analysis method based on the emotion dictionary are to use the emotional dictionary and to judge the users' emotional tendency through a series of calculations (semantic correlation). This method requires to extract the features manually. For "Big Data", the method can no longer meet daily needs.

To make up for the shortcomings of emotional dictionary, the papers compared SVM (Support Vector Machines), CRF (Conditional Random Field) and other methods [6], and finally concluded that SVM is generally more advantageous. Zampieri et al. [7] proposed a emotion classification based on matrix projection, normalized vector algorithm and KNN (K-Nearest Neighbor) algorithm. Experiments prove that the accuracy of the algorithm is high, and the time of classification is greatly reduced. Comparison of Machine Learning and emotional dictionary methods, Machine Learning no longer needs to manually label tags, and it no longer needs to build emotional dictionaries. And Machine Learning is more refined in feature extraction and semantic analysis. However, some features of text still need to be manually annotated, and they are not fully automated.

For the shortcomings of Machine Learning, the methods based on Deep Learning can solve it well, and it has been proved that Deep Learning is more prominent in emotion analysis than Machine Learning. Kim [8] proposed to use CNN (Convolutional Neural Networks) training model to complete the emotion analysis.

Zhu et al. [9] proposed a model based on LSTM to solve the problem of emotion classification. The papers [10, 11] proposed the BERT pre-training model and used the Transformer to train massive corpus. The papers [12, 13] proposed

that they used the BERT model to extract the semantic feature from the comments, and then input the acquired features into the LSTM model for tendency classification. The accuracy is also improved compared with the BERT model. To further improve the performance of the model, we propose a model (XLM-R combined with Bi-GRU).

3 Methodology and Data

3.1 Data description

In this mission, we use the official data set provided by EmoEvalEs@IberLEF 2021 [14]. There are a total of 8223 data (tweets written in Spanish). There are 844 items in the development set, 1656 items in the test set, and 5723 items in the training set. The labels in the training set are anger, disgust, fear, joy, sadness, surprise, others [15]. Their data volume is not evenly distributed.

3.2 Model Description

In order to choose a better model, we get the scores of three different models on the development set under the same conditions. As shown in Table 1 below.

In the Table 1, we observe that XLM-R combined with Bi-GRU model has more advantages than other models. LSTM network or GRU network can effectively deal with the long-range dependence problem of length sequences, and are widely used in natural language processing. However, the two networks only consider the correlation between sequences in one direction, and do not consider the possible correlation of future information, so they have great limitations in the application of strong round-trip correlation.

Table 1. Evaluation results of different models on development set

Model	Optimization	Accuracy	Averaged F1 score
XLM-R	sgd	0.4452	0.4535
	adam	0.4746	0.4059
XLM-R + LSTM	sgd	0.5156	0.5342
	adam	0.5556	0.5725
XLM-R + Bi-GRU	sgd	0.5812	0.6031
	adam	0.5949	0.6034

This paper uses the XLM-R model to extract the features of the training set and then inputs the acquired features into the Bi-GRU model to extract the emotional features of comments. Finally, we classify the emotion tendency by the Softmax method [16]. The specific process is shown in Figure 1.

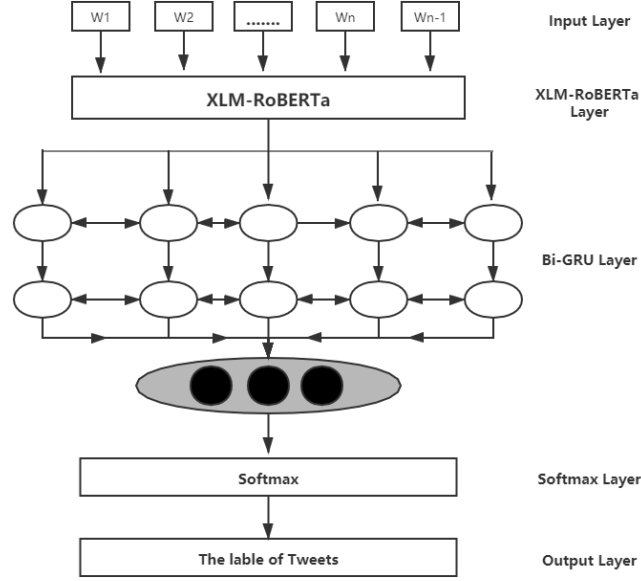


Fig. 1. The flow chart of emotion classification based on XLM-RoBERTa combined with Bi-GRU model

In the GRU, there are only two gates: update gate and reset gate. The facts have proved that compared with LSTM, GRU reduces the parameters and improves the efficiency of the model. The specific structure is shown in Figure 2.

In Figure 2, the Z_t represents the update gate and the r_t represents the reset gate. The Z_t can determine the renewal of the cell state and the r_t decides to write some information into \hat{h}_t . The smaller the number of r_t , the less information to be written into the candidate set, which means that the amount of previously useful information is less. Z_t is the update gate, which determines to update the previous information or not. It can discard useless information to avoid the problem of long-term dependence. Z_t is the logic gate when the model renews the activation. The equation is as follows:

$$Z_t = \sigma(\omega_z \cdot [h_{t-1}, x_t]) \quad (1)$$

In the Equation 1, σ is the activation function (Sigmoid) and the ω is the weight matrix.

The r_t is the reset gate, which determines whether the previous information needs to be reset. The r_t decides whether to abandon the previous activation h_t when the candidate is activated. The equations is as follows:

$$r_t = \sigma(\omega_r \cdot [h_{t-1}, x_t]) \quad (2)$$

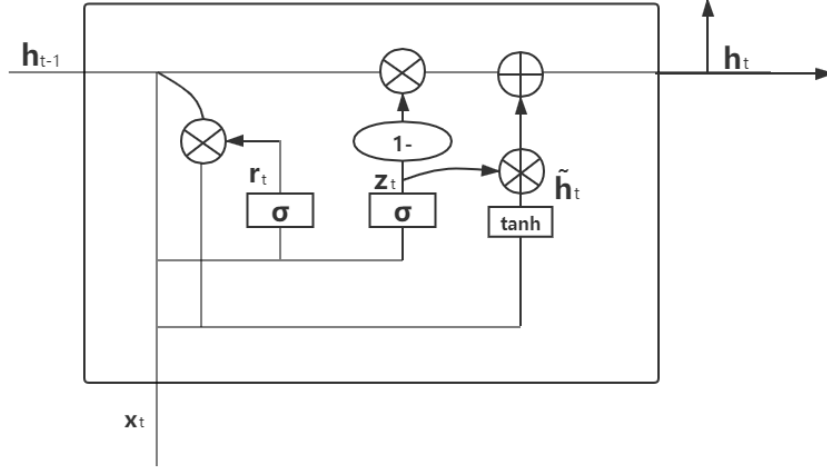


Fig. 2. The model:GRU

$$\tilde{h}_t = \tanh(\omega \cdot [r_t \cdot h_{t-1}, x_t]) \quad (3)$$

In the Equation 3, the tanh is the activation function and the h_t is the hidden layer. The calculation formula for receiving $[h_{t-1}, h_t]$ in the calculation process is as follows:

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \quad (4)$$

In the model (XLM-R combined with Bi-GRU), the XLM-R layer is the embedding layer of the Bi-GRU layer, and the Bi-GRU layer is the hidden layer. The input sequence is embedded in the model from two directions and the two-way information is stored. Last the information ($h_t^{(i)}$) is finally output. $h_t^{(i)}$ represents the bidirectional GRU information of the i-th text, and \vec{h}_t represents the forward GRU information of the i-th text. \overleftarrow{h}_t represents the reverse GRU information of the i-th text. The equations are as follows:

$$\vec{h}_t = GRU(x_t, \overrightarrow{h_{t-1}}) \quad (5)$$

$$\overleftarrow{h}_t = GRU(x_t, \overleftarrow{h_{t-1}}) \quad (6)$$

Figure 3 below shows the model diagram of XLM-R. The XLM-R model adopts the structure of the bidirectional Transformer for encoding. It can convert text information into information (vector) that can be recognized by the machine [17]. XLM-R adopts a dynamic mask and cancels the NSP (Next Sentence Prediction) task.

In Figure 3, $E_1, E_2, E_3, \dots, E_{n-1}, E_n$ represent the input (Language embeddings + Position embeddings + Token embeddings). After the bidirectional

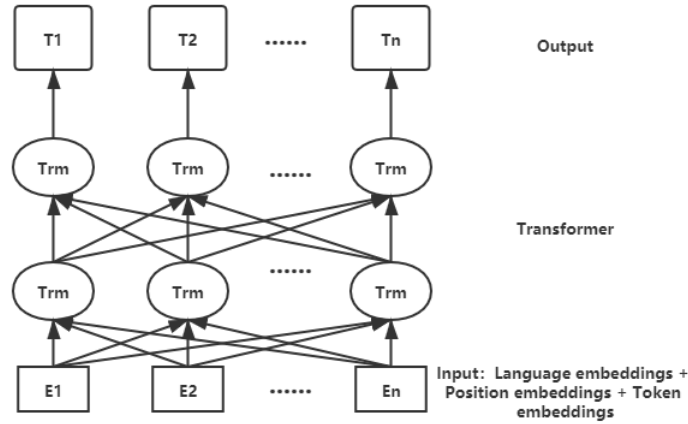


Fig. 3. The model: XLM-RoBERTa

Transformer encoder, we can get the vector of the tweet text $(T_1, T_2, \dots, T_{n-1}, T_n)$. The vector obtained through the XLM-R model training is expressed as: $W_i = (w_1^{(i)}, w_2^{(i)}, w_3^{(i)}, \dots, w_{n-1}^{(i)}, w_n^{(i)})$, where the W_i is the vector-matrix of the i -th sentence. The $W_j^{(i)}$ is the feature of each word, and the n is the maximum length of the sentence.

As shown in Figure 1, the entire XLM-R combined with Bi-GRU model is mainly divided into 5 layers. The first layer is the input layer. We input the officially provided data into the model; The second layer, we use the XLM-R model to vectorize the tweets; The third layer, we input the obtained word features into the Bi-GRU network to realize the emotion analysis of the text; The fourth layer, we use the function (Softmax) to classify the features; The fifth layer, the model output the emotion labels of the tweets (anger, disgust, fear, joy, sadness, surprise, others).

4 Experiment and Results

4.1 Data preprocessing

The data come from the official competition. We find that the data are based on events related to entertainment, disasters, politics, commemoration and global strikes that occurred in April 2019. Since the data set comes from tweets, the content of the data are not formal texts. So we need to process the data before training and testing. (1) Use "HASHTAG" to replace the tags in the data. (2) Because the username has no effect on emotion analysis and may affect the judgment of the model, a unified standard is adopted and "#@USER" is used instead of the username. (3) There are emoticons in the text. Replace all emoticons with text. (4) Uniform case.

4.2 Experiment setting

The experimental equipments used in this experiment are as follows: Intel CORE i7 CPU (16G), Hard Disk 1T, NVIDIA RTX 3080Ti. The operating system is Windows 10. The tool of editor is the Pychrom2020. The framework of Deep Learning is PyTorch. The optimizer used is Adam, and the learning rate is $2e-5$. The specific parameters of the model are shown in Table 2.

Table 2. The Setting of Parameter.

Parameter	Value
optimizer	Adam
Max_seq_length	80
Dimension	512
Batch_size	32
dropout	0.4
epoch	70
Gradient_accumulation_steps	8

4.3 Results

The evaluation indexes for the Spanish emotion multi-classification model are F1 score, averaged recall, averaged precision and accuracy. The partial results in this competition are shown in Table 3 below.

Table 3. Evaluation results.

Team	Accuracy	Averaged precision	Averaged recall	F1 score
GSI-UPM	0.727657	0.709411	0.727657	0.717028
qu	0.449879	0.618833	0.449879	0.446947

The team GSI-UPM win the first place in the competition. However, our team qu only win the 15th place in this competition. In the competition subtask 3, we not achieve the ideal results. We think there may be three main reasons for this result: (1) The value of the hyper-parameter is unreasonable. The value of the epoch is too large and it can increase the number of iterations of the weights in the network, and cause the model to over-fit. (2) The distribution of training data is uneven. The distribution of the data amount of each label in the training data is too uneven, resulting in poor performance of the trained model. Because the linear classifier used in this model is biased to most classes, it causes the

deviation of the model. (3) The training set has only 5723 pieces of data. So the generalization ability of the model is poor.

5 Conclusion

In this paper, we mainly propose a model (XLM-R combined with Bi-GRU) for emotion detection of Spanish text (tweets). The performance of this model on the development set is good, but the performance on the test data set is not satisfactory. Therefore, I think there is still a huge room for improvement in the performance of the model. In the next work, I will reset the hyper-parameters and use the K-fold ensemble method to improve the generalization ability of the model.

Acknowledgements

The completion of the paper is attributed to many people's support. First and foremost, I want to extend my heartfelt gratitude to my supervisor, Yanhua Yang. He gives me much help and advice during the whole process of my writing. My thanks also go to the authors whose books and articles have given me inspiration in my writing of the paper. Last but not least, I would like to thank the organizers for their hard work.

References

1. Marquardt, J., Farnadi, G., Vasudevan, G., Moens, M.F., Davalos, S., Teredesai, A., De Cock, M.: Age and gender identification in social media. *Proceedings of CLEF 2014 Evaluation Labs* **1180**, 1129–1136 (2014)
2. Meina, M., Brodzinska, K., Celmer, B., Czoków, M., Patera, M., Pezacki, J., Wilk, M.: Ensemble-based classification for author profiling using various features. *Notebook Papers of CLEF* (2013)
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. Agarwal, S., Sureka, A.: Characterizing linguistic attributes for automatic classification of intent based racist/radicalized posts on tumblr micro-blogging website. *arXiv preprint arXiv:1701.04931* (2017)
5. Choe, D.E., Kim, H.C., Kim, M.H.: Sequence-based modeling of deep learning with lstm and gru networks for structural damage detection of floating offshore wind turbine blades. *Renewable Energy* **174**, 218–235 (2021)
6. Korhonen, A., Traum, D., Márquez, L.: *Proceedings of the 57th annual meeting of the association for computational linguistics*. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (2019)
7. Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., Kumar, R.: Semeval-2019 task 6: Identifying and categorizing offensive language in social media (offenseval). *arXiv preprint arXiv:1903.08983* (2019)

8. Davidov, D., Tsur, O., Rappoport, A.: Semi-supervised recognition of sarcasm in twitter and amazon. In: Proceedings of the fourteenth conference on computational natural language learning. pp. 107–116 (2010)
9. Zhang, Z., Robinson, D., Tepper, J.: Detecting hate speech on twitter using a convolution-gru based deep neural network. In: European semantic web conference. pp. 745–760. Springer (2018)
10. Poria, S., Cambria, E., Hazarika, D., Vij, P.: A deeper look into sarcastic tweets using deep convolutional neural networks. arXiv preprint arXiv:1610.08815 (2016)
11. Ghosh, A., Veale, T.: Fracking sarcasm using neural network. In: Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis. pp. 161–169 (2016)
12. Tay, Y., Tuan, L.A., Hui, S.C., Su, J.: Reasoning with sarcasm by reading in-between. arXiv preprint arXiv:1805.02856 (2018)
13. Zhang, M., Zhang, Y., Fu, G.: Tweet sarcasm detection using deep neural network. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: technical papers. pp. 2449–2460 (2016)
14. 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)
15. Plaza-del-Arco, F., Strapparava, C., Ureña-Lopez, L.A., Martin-Valdivia, M.T.: 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>
16. Golubev, A., Loukachevitch, N.: Use of bert neural network models for sentiment analysis in russian. *Automatic Documentation and Mathematical Linguistics* **55**(1), 17–25 (2021)
17. Yu, H., Ji, Y., Li, Q.: Student sentiment classification model based on gru neural network and tf-idf algorithm. *Journal of Intelligent & Fuzzy Systems* (Preprint), 1–11