An Emotion Cause Detection Method Based on XLNet and Contrastive Learning

Hai Feng Zhang
School of Computer Science and
Information Engineering
Hubei University
Wuhan, China
zhf@stu.hubu.edu.cn

Cheng Zeng*
School of Computer Science and
Information Engineering
Hubei University
Wuhan, China
zc@hubu.edu.cn

Peng He
School of Computer Science and
Information Engineering
Hubei University
Wuhan, China
penghe@hubu.edu.cn

Abstract*—Emotion cause detection is a new direction in the field of emotion research and is a fine-grained analysis of emotion. However, research on emotion cause detection is still challenging due to the extreme complexity of human emotions, the difficulty of tracing emotions back to their origins, and the fact that emotion cause detection corpus annotation requires a lot of manual involvement. An emotion cause detection method incorporating contrastive learning is proposed to address this problem, which combines the autoregressive language model XLNet and a contrastive learning approach to introduce a difficult sample generation strategy and a word repetition strategy in the positive/negative example comparison pattern in the training data,and design a loss function that incorporates the classification task and the comparison learning task. Experiments on the Weibo sports game commentary dataset show better performance in terms of accuracy, macro-average F1 values, and a 2.73 percentage point improvement in accuracy compared to the baseline XLNet, demonstrating the effectiveness of the proposed method.

Keywords-emotion cause detection;contrastive learning; autoregressive anguage model;

I. Introduction

Emotion cause detection refers to an individual's cognitive process when influenced by factors such as environmental stimuli, physiological conditions and cognitive processes.[3].Analyzing the text of online media comments can dig out effective public opinion information and it is important to find the key factors in the review text that influence the user's emotional change government departments can guide public opinion and avoid major public opinion incidents [1-2].Gui et al. [4] inspired by the question-and-answer domain, used sentiment keywords as query words and their context as query text to determine whether the current clause is an emotional cause by means of question-and-answer. Li et al. [5] proposed a co-attention neural network (CANN) model based on emotional context-awareness. The method first encodes the reason candidate and sentiment clauses by a BiLSTM model, and then sends them to the convolutional layer of CNN for sentiment reason recognition. And with the widespread use of pre-trained models like BERT [6] in natural language processing, there has been a qualitative improvement in the study of emotion cause detection.The emotion cause detection research is not only dependent on the algorithms implemented but also limited by the cause-labeled corpus, and the current lack of relevant corpora has affected the depth of research in this area [7].

To address the above issues, this paper proposes an emotion cause detection method incorporating contrastive learning. The method combines a pre-trained model and a contrastive learning approach, extracts emotional text features using an autoregressive language model XLNet [8], adds a contrastive learning task during model training, and makes full use of the positive/negative example contrast patterns in labeled data to improve the classification of emotional attribution.

II. Related Technologies

A. Autoregressive language model XLNet

Fig. 1 shows the structure of the XLNnet model, the XLNet autoregressive language model is based on the core of transformer-XL framework [9]. By introducing the circular transfer mechanism and encoding relative position information, it can make full use of the textual context information and combine the information of each word context to better characterize the multi-sense of words, which reflects the superiority of the autoregressive model.In the emotion cause detection task, the XLNet sentence vector output is connected to the fully connected layer and then softmax is computed to obtain the cause category probability distribution.

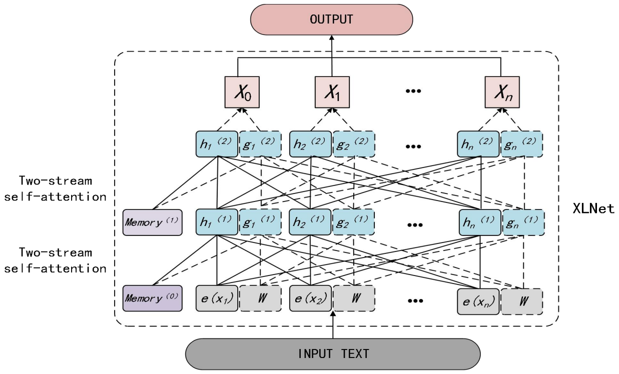

Figure 1. XLNet Model Structure.

B. Contrastive Learning

The main idea of contrastive learning is to draw similar samples closer and push away dissimilar samples to learn a better semantic representation space from the samples.Chen et al. [10] proposed a simple framework for contrast learning for

DOI reference number: 10.18293/SEKE2022-093

visual representation (SimCLR), which has made great progress in image representation.Also in terms of text representation,Gao et al. [11] propose two ways of constructing positive and negative examples in SimCSE.Liang et al. [12] proposed an R-Drop regularization method similar to contrastive learning, which only uses the model's two dropout outputs as positive example pairs, and adds a KL-divergence loss without any structural modification, with significant improvement in a variety of NLP tasks.

The key to introduce contrastive learning in the emotion cause detection task is how to construct positive and negative example pairs and combine them with the classification task, so that the model can perform both the contrastive learning task and the classification task, design an effective contrast model, and improve the classification effect of the model by effectively combining the contrast loss function and the classification loss function to obtain a better sentence representation.

III. Approach

As shown in Fig. 2, the XLNet-CL emotion cause detection method proposed in this paper incorporates a contrastive learning approach based on the autoregressive language model XLNet.

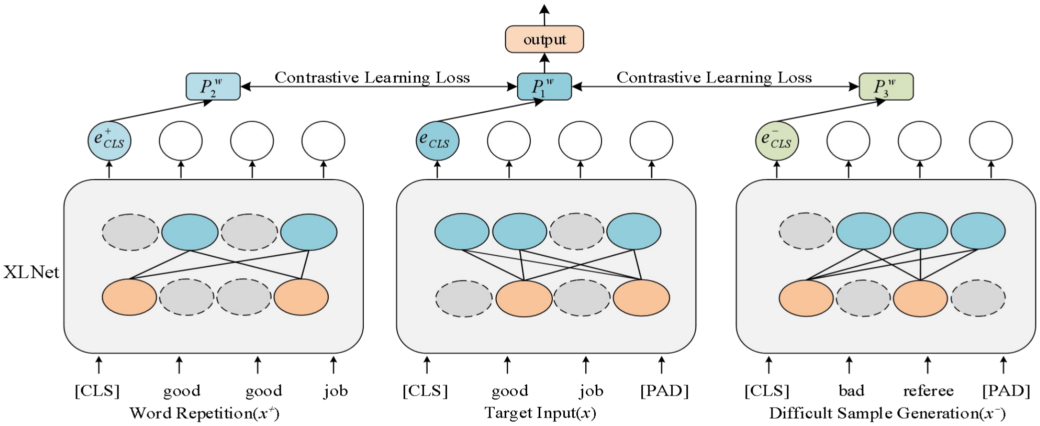

Figure 2. The framework of XLNet-CL.

The model training process consists of five main steps: data processing, comparison learning positive and negative example sample generation, dropout training, loss function fusion calculation, and difficult negative sample update. As shown in the figure, the detailed flow of each step of the fusion model in the training process is as follows.

Step 1: Data processing. Before model training, data processing is required to regularize the data text, delete the text fragments that affect the attribution of emotions, for example, the user's name of microblog comments can have an impact on the classification results, eliminate unnecessary user business card segments, and improve the normality of the data. The emotional text dataset $\{(x_i, y_i)\}_{i=1}^{N}$ is divided into training set, validation set $\{(x_i, y_i)\}_{i=1}^{D_m}$ and test set $\{(x_i, y_i)\}_{i=1}^{T_m}$.

Step 2: Contrastive learning positive and negative example sample generation. For small batches of training set data $\{(x_i, y_i)\}_{i=1}^{T_m}$, positive sample pairs are generated using the word repetition strategy for the input samples to obtain the positive sample set $\{(x_i^+, y_i^+)\}_{i=1}^{T_m}$. Initial negative samples $\{(x_i^-, y_i^-)\}_{i=1}^{T_m}$ are generated by random selection among the training set samples.

Word repetition strategy mainly solves the problem that pre-trained models may mistakenly believe that input texts of the same length have the same semantics when using only dropout strategy for positive sample generation, so word repetition is used to extend the sentence word length without changing the sentence word semantics. word repetition will randomly copy some words or phrases in a sentence. Here we take word repetition as an example, given a sequence of sentences $s = [w_1, w_2, w_3, \cdots, w_n]$, n is the length of the sequence. After the word repetition policy generates positive samples $s^+ = [w_1, \overline{w}_2, \overline{w}_2, w_3, \cdots, w_n]$, the positive sample length becomes $n+1$ and the word w_2 is repeated once in the sentence. To enable more diversity to be introduced when extending the sequence length, 10% to 30% of the words are randomly selected for word repetition, which results in a random increase of 10% to 30% in the positive sample length.

Step 3: Dropout training. The input mini-batch data $\{(x_i, y_i)\}_{i=1}^{T_m}$ and its positive sample pairs $\{(x_i^+, y_i^+)\}_{i=1}^{T_m}$ and negative sample pairs $\{(x_i^-, y_i^-)\}_{i=1}^{T_m}$ are trained by XLNet-dropout. The dropout training method makes the model partially deactivate the neurons during the training process, although for the same XLNet model, the output of three sub-models is actually obtained by three times XLNet forward propagation, thus obtaining the probability distribution of three outputs :P_1^w、P_2^w、P_3^w.The output e_{CLS} of the sentence vectors corresponding to the input samples are extracted by XLNet, and then the probability distribution P^w is calculated by the fully connected layer and softmax.

Step 4: Loss function fusion calculation. For the 3 probability distribution outputs P_1^w、P_2^w、P_3^w of input samples

Corresponding author: Cheng Zeng (zc@hubu.edu.cn)

x, x^+, x^-, not only the basic classification loss calculation is needed, but also the contrastive learning loss NCEloss function calculation is needed between P_1^w、P_2^w、P_3^w, so as to close the positive sample-to-vector space distance and keep away from the negative sample-to-vector space distance, so that the model can obtain a better sentence vector representation in the classification process. The contrast loss function is calculated as follows:

$$NCE = -\log \frac{e^{sim(h_i \cdot h_i^+)/\tau}}{\sum_{j=1}^{N} (e^{sim(h \cdot h_j^+)/\tau} + e^{sim(h_i \cdot h_j^-)/\tau})} \quad (1)$$

The temperature coefficient τ is a hyperparameter that can be fine-tuned to adjust the model performance, and the comparative loss is calculated by subjecting P_1^w、P_2^w, and P_3^w to the comparative loss L_{CL}:

$$L_{CL} = NCE(P_1^w, P_2^w, P_3^w), \quad (2)$$

The three probability distributions are similarly subject to the basic categorical loss function calculation:

$$L_{NLL} = -\log P_1^w(y_i \mid x_i) - \log P_2^w(y_i \mid x_i) - \log P_3^w(y_i \mid x_i), \quad (3)$$

The final model fusion loss function is calculated as follows, with the loss function fusion factor α as the hyperparameter:

$$L = L_{NLL} + \alpha \cdot L_{CL} \quad (4)$$

Step 5: Difficult negative sample generation. After the initial training, the model has a certain classification capability, and the validation set is verified on the trained model, when the dropout strategy is off and all the deep neural network neurons are in working state. The accuracy values of each category are obtained from the validation results on the model validation set. The *N* categories with poor accuracy in the training set are randomly filtered to generate negative samples, and among the total 7 emotional reason categories, only the *N* categories with the worst accuracy are selected to go into the negative sample set for comparison learning negative sample comparison. At the same time, the negative example sample pairs $\{(x_i^-, y_i^-)\}_{i=1}^{T_m}$ are updated for the next training, and *N* is also the model hyperparameter.

IV. Experiments and Analysis

A. Experimental data

The emotion cause detection dataset used in this paper is annotated by crawling on Sina Weibo sports event comment data, using manual annotation. The seven emotional reasons are labeled as: behavior itself, discourse expression, empathy, rating criteria, opinion climate, history and culture, and derivative topics. The total number of data is 11520, and the training set, validation set and test set are divided according to the ratio of 6:2:2.Some of the data are shown below:

TABLE I. DATA DISPLAY

Emotional Text	Label
All of you who have watched the whole thing know that your score shouldn't be like this, this judge is really not good, stuck with the score to give the score.	rating criteria
Great, great, superb!	empathy
He is also still our hero	derivative topics

Since this experimental task is an emotion-attributed multiclass text classification task, accuracy (Acc), macro-precision (MP), macro-recall (MR), and macro F1 value are used as the evaluation metrics for model performance.

B. Experimental details

The main hyperparameters in the contrastive learning module include the contrastive learning loss function NCEloss hyperparameter temperature coefficient τ, the negative sample pool hyperparameter *N*, and the loss function fusion factor α.After tuning the model for several times, τ is set to 0.1, *N* is set to 5, and α is set to 0.5 to achieve the best model performance. The model was trained by Adam gradient descent algorithm, with batch size of 128, maximum rounds set to 20, and initial learning rate set to 5E-5.

In order to verify the effectiveness of the proposed method of fusing contrastive learning for emotion cause detection, several deep learning text classification methods are selected for experimental comparison, among which the text classification methods based on pre-trained models are BERT, RoBERTa, baseline XLNet, XLNet-RCNN [16], and the text classification methods combined with Word2Vec word vectors are TextCNN, BiLSTM-Att [15]. and some other fusion ratio learning methods XLNet-Rdrop, XLNet-SimCSE.The experiments in this paper all use the Chinese pre-training models BERT, RoBERTa, and XLNet proposed by Cui [13] et al.XLNet-Rdrop: Integration of XLNet and R-Drop methods for experimental comparison of emotion cause detection.XLNet-SimCSE: Fusing XLNet and SimCSE, the SimCSE approach is used in the contrastive learning mode, using dropout output as positive example pairs and other categories of the same batch training data as negative example pairs, with the same loss function calculation as in this paper.

V. Analysis of experimental results

The experimental results are shown in the table for the comparative model experiments on the Weibo sports event commentary dataset. As can be seen in the model comparison data experimental table, the text classification method using pre-trained model is significantly more effective than the text classification method TextCNN, BiLSTM-Att combined with Word2Vec experiments.BERT, RoBERTa and XLNet are not much different in experimental effects, RoBERTa has slightly higher classification effect due to longer training time, larger batch size and more training data on the basis of BERT.All the XLNet-based fusion models have improved over the baseline XLNet in terms of experimental results, especially the XLNet-CL method proposed in this paper has improved 2.73 percentage points in accuracy and 16.47 percentage points in F1 value.Compared with the other two methods of fusing XLNet with contrastive learning, XLNet-Rdrop and XLNet-SimCSE, it is obvious that the XLNet-CL method proposed in this paper works better.

Corresponding author: Cheng Zeng (zc@hubu.edu.cn)

TABLE II. EXPERIMENTAL RESULTS

Model	*Acc*	*MP*	*MR*	*F1*
TextCNN	0.7550	0.5173	0.5203	0.5174
BiLSTM-Att	0.7432	0.5252	0.5160	0.5197
BERT	0.7763	0.5766	0.5711	0.5718
RoBERTa	0.7822	0.6782	0.5997	0.6196
XLNet	0.7763	0.5745	0.5555	0.5619
XLNet-RCNN	0.7893	0.6996	0.6940	0.6890
XLNet-Rdrop	0.7846	0.6977	0.6906	0.6931
XLNet-SimCSE	0.7917	0.7311	0.7164	0.7199
XLNet-CL	**0.8036**	**0.7424**	**0.7159**	**0.7266**

VI. HYPERPARAMETER INFLUENCE

The main hyperparameters of the emotion cause detection model proposed in this paper include: the temperature coefficient $\tau$of the contrastive learning loss function, the fusion factor α of the loss function, and the difficult negative sample hyperparameter N. As shown in Fig. 3, the experimental comparison after fine-tuning the three hyperparameters shows that the model performs best when the temperature coefficient is set to 0.1, the fusion factor $\alpha$is set to 0.5, and the negative sample hyperparameter N is set to 5.

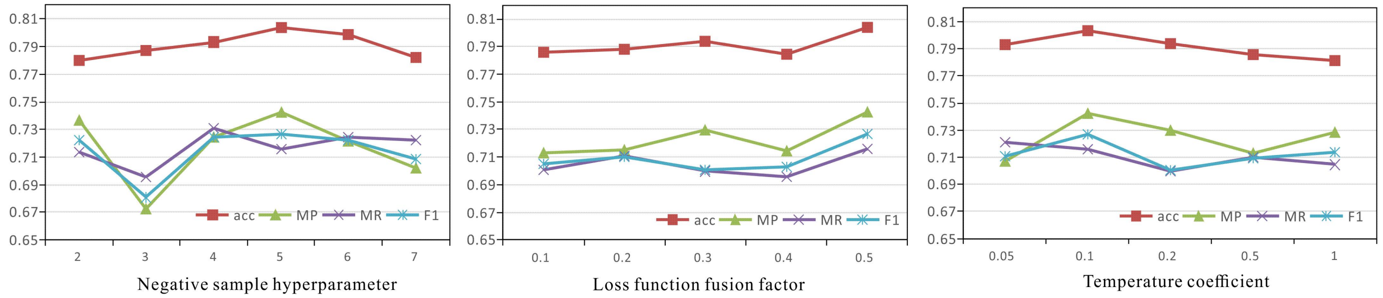

Figure 3. Hyperparameter Comparison.

VII. CONCLUSION

In this paper, we propose an emotion cause detection method based on XLNet and contrastive learning. The results of several experimental comparisons show the effectiveness of the proposed fusion model on the emotion cause detection task in this paper. The disadvantage of the model in this paper is the large number of parameters during model training, which is not suitable for long text sentiment analysis tasks. The difficult sample generation strategy is although simple negative sample generation for difficult category data can alleviate the data imbalance problem, it still has not fully explored the difficult samples in depth. In the next work, we will optimize the difficult sample generation strategy and try to conduct experiments on multi-granularity sentiment analysis tasks to find more suitable task scenarios for this model.

ACKNOWLEDGMENT

This work is supported by the National Key R&D Program of China (No. 2018YFB1003801), the National Natural Science Foundation of China (Nos. 61832014, 61902114, 61977021), and the Science and Technology Innovation Program of Hubei Province under Grant (No. 2018ACA133, 2019ACA144), and the Open Foundation of Hubei Key Laboratory of Applied Mathematics (No. HBAM201901).

REFERENCES

[1] Ravi K, Ravi V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications[J]. Knowledge-based systems, 2015, 89: 14-46.

[2] MU Yongli;LI Yang;WANG Suge, Emotion cause detection Based on Ensembled Convolution Neural Networks[J], Journal of Chinese Information Processing,2018,32(02):120-128.

[3] ZHANG Haitao， ZHANG Xinrui， ZHOU Honglei， SUN Tong, Key Factors and Influencing Mechanisms of User Emotion Evolution in Public Health Emergencies[J],Information Science,2020,38(07):9-14.

[4] Gui L, Hu J, He Y, et al. A question answering approach to emotion cause extraction[J]. arXiv preprint arXiv:1708.05482, 2017.

[5] Li X， Song K， Feng S , et al. A Co-Attention Neural Network Model for Emotion Cause Analysis with Emotional Context Awareness[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018.

[6] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [EB/OL]. [2019-08-17]. https://arxiv.org/pdf/1810.04805.pdf.

[7] Deriu J， Lucchi A， Luca V D , et al. Leveraging Large Amounts of Weakly Supervised Data for Multi-Language Sentiment Classification[J]. International World Wide Web Conferences Steering Committee, 2017.

[8] YANG Z, DAI Z, YANG Y, et al. XLNet: Generalized autoregressive pretraining for language understanding[EB/OL].Neural Information Processing Systems.Canada,2019: 5754-5764.https://arxiv.org/abs/1904.09482

[9] DAI Z, YANG Z, YANG Y, et al. Transformer-XL: Attentive Language Models beyond a Fixed-Length Context[C]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Italy, 2019: 2978-2988.

[10] Chen X, Fan H, Girshick R, et al. Improved baselines with momentum contrastive learning[J]. arXiv preprint arXiv:2003.04297, 2020.

[11] Gao T, Yao X, Chen D. Simcse: Simple contrastive learning of sentence embeddings[J]. arXiv preprint arXiv:2104.08821, 2021.

[12] Liang X， Wu L， Li J , et al. R-Drop: Regularized Dropout for Neural Networks[J].arXiv preprint arXiv:2106.14448,2021.

[13] Cui Y， Che W， Liu T , et al. Revisiting Pre-Trained Models for Chinese Natural Language Processing[J]. 2020.

[14] KIM Y. Convolutional neural networks for sentence classification [C]// Proceedings of the 2014 Conference of Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2014: 1746-1751.

[15] Zhou Peng,Shi Wei,Tian Jun,et al.Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 54th Annual Meeting of the ACL.Stroudsburg, PA: Association for Computational Linguistics,2016:207-212.

[16] PAN Lie,ZENG Cheng,et al.Text sentiment analysis method combining generalized autoregressive pre-training language model and recurrent convolutional neural network[J/OL].Journal of Computer Applications.2021.

Corresponding author: Cheng Zeng (zc@hubu.edu.cn)