



# Emotion Cause Pair Extraction Based on Multitask

Yilin Li<sup>1</sup>, Dechen Gao<sup>1</sup>, Yuezhong Liu<sup>2</sup>, and Shenggen Ju<sup>1</sup>(✉)

<sup>1</sup> College of Computer Science, Sichuan University, Chengdu 610005, China  
jsg@scu.edu.cn

<sup>2</sup> Enterprise Service, Commonwealth Bank of Australia, Sydney, NSW 2000, Australia

**Abstract.** Emotion cause pair extraction is a sub-task of sentiment analysis tasks, which aims to extract all emotion clauses in a given document and the cause clauses corresponding to the emotion. At present, only the method of sharing parameters at the bottom layer is used to build the connection between tasks, and the shared word encoder is used in word encoding, the phenomenon that the model pays different attention to emotional words and cause words is ignored, and the rich interactive relationship information between the three tasks cannot be fully utilized. This paper proposes a multi-task emotion cause pair extraction model based on feature fusion. The model learns the multi-information at the word level through the feature fusion module, and uses the two tasks of emotion clause extraction and cause clause extraction as auxiliary tasks. The extracted result is transformed into label information, and the method of label embedding is used to integrate into the generation of emotion cause pair representation, thereby improving the effectiveness of the emotion cause pair extraction. The experimental results of the model in the paper on the ECPE data set prove that the Mul-ECPE model has improved in the evaluation indicator of the three tasks compared to the previous series of models.

**Keywords:** Emotion cause pair extraction · Multitask · Feature fusion · Label embedding

## 1 Introduction

The emotion analysis task is a hot research area in the field of natural language processing. As comments and opinions on the Internet play an increasingly important role in influencing people's choices or events, the reasons for people's comments become more important. The objective of the Emotion Cause Extraction [1] (ECE) task is to discover the causes behind emotions. With the development of deep learning technology, deep learning is used to solve ECE task. Inspired by the Memory networks in the Question Answering System, Gui et al. [2] designed an ECE model based on the QA system. This method ignores the context of emotional words in the text. Therefore, Li et al. [3] designed a co-attention network to make full use of the information between clauses. Furthermore, Yu et al. [4] proposed to use hierarchical network to model the information at word, phrase and clause level. Xia et al. [5] also designed a hierarchical network

based on Transformer, and combined the global prediction information with the relative position. Fan et al. [6] further introduced external emotional corpus as regular term in hierarchical network, and at the same time introduced regular term for relative position.

However, the ECE task requires marking the polarity of emotions in advance. The Emotion Cause Pair Extraction (ECPE) task is an improvement on ECE task. It does not need to mark the Emotion in advance. The task objective is to select the sentence where the emotion is and the sentence which causes the emotion. At present, the ECPE task is implemented by the deep learning model, and there are mainly two methods to solve the ECPE task: the two-step method and the end-to-end method. The two-step method is the traditional method for this task. First, all the emotion and cause clauses in the text are extracted to obtain two clause sets. A filter was then trained to pick the right emotion cause pair from the candidates.

However, it has the problem of error propagation. In addition, because the first step of the two-step method faces the emotion and cause clause extraction task, rather than the ECPE task, the effect of the model will be reduced.

In order to solve the problem, end-to-end architecture is widely used. Wu et al. [7] proposed a multi-task neural network to perform ECPE and two auxiliary tasks in a unified model. Fan et al. [8] compared the ECPE task to the directed graph generation task, and designed the transition-based ECPE model TransECPE. Wei et al. [9] proposed a more efficient model RANKCP, which used graph attention network to model document content and structure, solving the ECPE task from the perspective of ranking. In addition, in order to utilize the interactive information between different tasks, the parameter sharing method in multi-task learning is adopted to establish the relationship between the two sub-tasks and the emotion cause pair. Tang et al. [10], calculated the relationship score between emotion and cause clauses by using self-attention [11], extracted the emotion cause pair whose scores exceeded the threshold.

But the multi-task model cannot make full use of the rich interaction information between tasks, and the information transfer between tasks is also implicit. Label embedding has been proved to be an effective method in various fields and tasks. In the field of natural language processing, label embedding for text classification has been studied in the heterogeneous network of Tang et al. [12] and the multi-task learning environment of Zhang et al. [13]. Subsequently, Wang et al. [14] regarded the task of text classification as the problem of label and word joint embedding, and introduced an attention framework to measure the compatibility between text sequence and label embedding. Zhang et al. [15] proposed multi-task label embedding, which maps labels of each task to semantic vectors and uses a method similar to word embedding to process word sequences, so as to transform classification tasks into vector matching tasks.

The main contributions of this paper can be summarized as follows: (1) Construct the relationship between tasks by sharing parameters at the bottom level, share the results of emotion and cause extraction with the ECE task explicitly through label embedding. (2) Feature fusion is adopted to obtain feature representation with multiple relational information. (3) The model obtained optimal results on data set ECPE, which verified the validity of the model.

## 2 Related Work

Xia et al. [16] proposed an end-to-end solution to solve the error propagation problem of the two-step method, integrating the generation, pairing and final prediction of emotion cause pair representation into a joint framework. They designed a two-dimensional Transformer [17] and two variants to obtain the representation of emotion cause pair and captured the interaction between different emotions and causes in the process of generating vector representation.

Some researchers transfer the ideas of other tasks to ECPE tasks. Song et al.[18] regarded ECPE as a link prediction problem, to predict whether there is an edge from the emotion clause to the cause clause, and if there is, the two clauses constitute an emotion cause pair. Fan et al. [19] treated the ECPE task as a sequence labeling problem and proposed a multi-task sequence labeling model via label distribution refinement.

Most current graph convolution network rely on the structural information of graphs for classification. Li et al. [20] retained the structural information and feature similarity of graphs, and introduced the attention mechanism for classification. Chen et al. [21] applied a graph convolution network to the model and constructed an emotion cause pair graph to simulate three kinds of dependency relationships between candidate pairs in the local neighborhood, where each node represents a candidate emotion cause pair and the edge connecting two nodes represents the dependency relationship between two candidate pairs. Finally, the graph convolution network is used to extract these three types of edges so as to disseminate the context information in the graph.

## 3 Methodology

### 3.1 Task Description

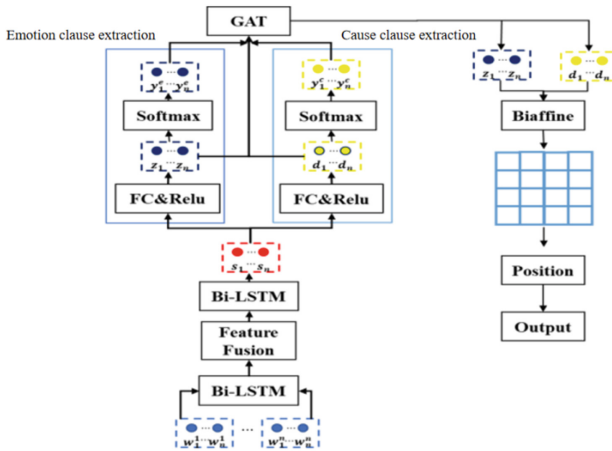
The task of ECPE in this paper is to extract emotion clauses and corresponding cause clauses from a given text. A text case in the ECPE data set is given in Table 1. The text is divided into eight clauses according to punctuation marks, the sixth clause is taken as an emotion clause, the eighth clause is the corresponding cause clause. The final goal of ECPE task is to extract emotion cause pairs composed of (C5, C6).

**Table 1.** Data example.

Text	Class
C1:When I saw the reply written to me by the leaders of ministries and commissions that the proposal was adopted	\
C2:I knew I was doing my part for the development of the country	\
C3:Bai Jinyue, an ordinary worker of Xingtai Iron and Steel Corp., Ltd. in Hebei Province	\
C4:holding the letters of thanks that the ministries and commissions of the state have given him over the years	\
C5:told Chinanew reporters excitedly	Emotion Clause
C6:that my suggestions have been adopted by the state for 27 years	Cause Clause

### 3.2 Model Description

The main architecture of GAT-ECPE, an ECPE task model proposed in this paper, is shown in Fig. 1.



**Fig. 1.** MuleCPE model architecture.

### 3.3 Input Layer

The input layer contains the word embedding and word encoding layer. Firstly, word embedding is carried out on the input text sequence, and the words which have been segmented are transformed into vectors to obtain the text sequence matrix. The  $i$ -th clause in document  $D$  can be represented by an embedding matrix  $c_i = (w_1^i, \dots, w_{|D|}^i)$ .

Bi-LSTM is used to encode words. For each clause, Bi-LSTM is used to encode the context information at the word level, as shown in the following formula.

$$\left[ h_i^1, \dots, h_i^j, \dots, h_i^{|c_i|} \right] = \left[ Bi - LSTM \left( w_i^1, \dots, w_i^j, \dots, w_i^{|c_i|} \right) \right] \quad (1)$$

where  $w_i^j$  represents the word vector of the  $j$ -th word in the  $i$ -th clause in the document, and  $h_i^j$  represents the feature representation of the clause word level.

### 3.4 Feature Fusion Layer

After passing through the input layer, the feature representation at the word level is obtained. The feature representation of each clause of the text is shown in the following formula.

$$c_i = [h_{i,1}, \dots, h_{i,n}] \quad (2)$$

Convolution kernels of different sizes are used to conduct one-dimensional convolution operations at the word level, so that the model can learn multiple features at the word level, as shown in the formula.

$$x_t^i = conv_t(c_1, \dots, c_{|c|}) \quad (3)$$

where  $conv_t$  represents the convolution operation, and  $t$  represents the size of the convolution kernel. Considering that the traditional connection cannot correctly process the language combination, the convolutional layer is densely connected:

$$x = [\oplus_t x_t] \quad (4)$$

After densely connection, the downstream layer of the convolution part can access the features generated by the upstream layer. This paper draws on the multi-scale feature attention mechanism proposed by Wang et al. [14], so that each position of the text adaptively selects features of different scales. The multi-scale feature attention mechanism consists of two steps: convolution aggregation and scale weighting (Fig. 2).

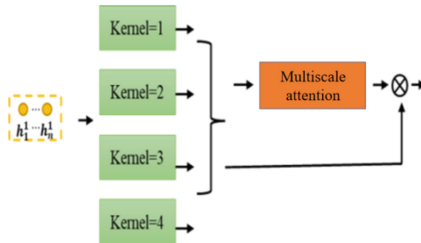


Fig. 2. Feature fusion.

First, the convolution aggregation aims to use a descriptor  $s_i^j$  to represent the features of the different blocks obtained at position  $i$ .  $K$  convolution kernels are used in each convolution block. Each eigenvector is then represented by using a scalar.

$$s_i^j = F_{ensem}(x_t^j) \quad (5)$$

where  $F_{ensem}$  represents a function that sums all  $K$  elements of the input vector. And the output scalar can be used as the description of the eigenvector.

Secondly, after scale weighting, features of different scales can be weighted adaptively. The final feature representation calculation is as follows:

$$x_{atten}^i = \sum_{l=1}^L \alpha_l^i x_l^i \quad (6)$$

### 3.5 Intra-clause Information Encoding Layer

In order to further learn the semantic order information in the document, Bi-LSTM is applied. The calculation method is shown in the formula:

$$s_i = [Bi - LSTM(a_i)] \quad (7)$$

### 3.6 Label Embedding Layer

According to the Bi-affine attention mechanism, this paper takes the emotion clause as the core item and the cause clause as the dependency. The matrix composed of emotion and cause clauses are asymmetric and can sense direction. Two MLP are applied in this paper to generate the representation of emotion clauses and cause clauses respectively, as shown in the following formula:

$$z_i^e = \sigma(W^e h_i^e + b^e) \quad (8)$$

$$z_i^c = \sigma(W^c h_i^c + b^c) \quad (9)$$

where  $z_i^e$  represents the  $i$ -th emotion clause, and  $z_i^c$  represents the  $i$ -th cause clause.

The extraction of emotion and cause clause were embedded into the input of graph attention network, when processing the results extracted from the two auxiliary tasks, the results extracted from the emotion and the cause clause are first mapped to the vector representation by embedding matrix  $W$ :

$$Y_i^e = W^e y_i^e \quad (10)$$

$$Y_i^c = W^c y_i^c \quad (11)$$

where  $y$  represents the extraction result of clauses,  $Y$  is the transformed vector, and  $s$  is the context-encoded clause representation of the output of inter-clause encoding layer. After vector concatenating, they are used as the input of graph attention network to encode the representation of emotion and cause clause.

$$s_i^e = [s_i; Y_i^e] \quad (12)$$

$$s_i^c = [s_i; Y_i^c] \quad (13)$$

The concatenated vectors are transformed into emotion clause representation and cause clause representation respectively through the fully connected layer.

$$z_i = \sigma(W^e s_i^e + b^e) \quad (14)$$

$$d_i = \sigma(W^c s_i^c + b^c) \quad (15)$$

where  $z$  represents emotion clause and  $d$  represents cause clause.

### 3.7 Graph Attention Layer

Although the information between clauses has been learned through Bi-LSTM, it can only learn the single-hop relationship, while there may be a multi-hop relationship between emotion and cause. The graph attention mechanism can naturally capture the multi-hop relationship and integrate the higher-order information. In this chapter, the clause representation is updated by stacking two graph attention network layers.

The attention weight reflects correlation between clauses, which is learned by MLP. The attention between clauses is calculated by the following methods:

$$e_{ij}^{(t)} = w^{(t)T} \tanh\left(\left[W^{(t)} z_i^{(t-1)}; W^{(t)} d_i^{(t-1)}\right]\right) \quad (16)$$

$$\alpha_{ij}^{(t)} = \frac{\exp(\text{LeakyReLU}(e_{ij}^{(t)}))}{\sum_{k \in N(i)} (\exp(\text{LeakyReLU}(e_{ik}^{(t)})))} \quad (17)$$

Considering that sentiment clauses correspond to multiple cause clauses, the original neighbor aggregation method in the graph attention network is updated so that the multi-hop relationship information between two clauses can be better learned in the update. The formula is shown below:

$$z_i^t = g\left[W_1 \sum_{j \in N(i)} (\alpha_{ji}^t z_j^{t-1} + \alpha_{ij}^t d_j^{t-1}) + b_1 z_i^{t-1}\right] \quad (18)$$

$$d_i^t = g\left[W_2 \sum_{j \in N(i)} (\alpha_{ji}^t z_j^{t-1} + \alpha_{ij}^t d_j^{t-1}) + b_2 d_i^{t-1}\right] \quad (19)$$

In order to integrate these two representations into the matrix of emotion cause pairs for the final mission objective, we fold the emotion cause pairs into two parts. First, a Bi-linear like operation needs to be performed for each possible emotional cause pair. Second, Bi-affine transformations are used to deal with complex interactions, as shown in the formula:

$$M_{p,q} = \left(W^m z_p^e + b^m\right)^T z_q^c \quad (20)$$

All possible clause pairs in the original document  $d$  are regarded as candidates, assuming that the document length is  $|d|$ , then all of the possible emotion cause will

form  $|d| * |d|$  matrix,  $M_{p,q}$  means the possible correct emotion cause pairs composed of the  $p$ -th emotion clause and the  $q$ -th cause clause.

The sigmoid function is further used to activate the emotion cause pair matrix:

$$\tilde{M}_{p,q} = g(M_{i,j}) \quad (21)$$

The location information between emotion and cause clause is also an important factor influencing the accuracy, emotion clause is close to cause clause, so the weight of the setting is bigger in the closer position, as shown the following formula.

$$A_{p,q} = \frac{|D|-|p-q|+\epsilon}{|D|+\epsilon} \quad (22)$$

where  $\epsilon$  is the smooth term, integrates position matrix with the emotion cause pair:

$$\hat{M}_{p,q} = \tilde{M}_{p,q} \odot A_{p,q} \quad (23)$$

### 3.8 Extraction Results

In the E2EECP task proposed by Song et al. [18], the emotion and cause clause extraction is regarded as two independent tasks. In this paper, the previous task is improved by adding the result of extraction of emotion and cause clauses into the representation of subsequent emotion and cause clauses.

The formulas of emotion and cause extraction of clauses are as follows:

First, the feature representation of the clause used for extraction is obtained:

$$z_i^{ae} = \sigma(W^{ae}h_i^{ae} + b^{ae}) \quad (24)$$

$$z_i^{ac} = \sigma(W^{ac}h_i^{ac} + b^{ac}) \quad (25)$$

Input to softmax layer after passing through two fully connected layers:

$$y^{ae} = \text{softmax}(\hat{W}^{ae}z_i^{ae} + \hat{b}^{ae}) \quad (26)$$

$$y^{ac} = \text{softmax}(\hat{W}^{ac}z_i^{ac} + \hat{b}^{ac}) \quad (27)$$

The method of obtaining the final result of emotion extraction and cause extraction is similar to the formula of auxiliary task, but the parameters are different:

$$\tilde{z}_i^e = \sigma(\tilde{W}^e h_i^e + \tilde{b}^e) \quad (28)$$

$$\tilde{z}_i^c = \sigma(\tilde{W}^c h_i^c + \tilde{b}^c) \quad (29)$$

where  $\tilde{W}^e \in R^{d_z \times 2d_h}$ ,  $\tilde{b}^e \in R^{d_z}$  and  $\tilde{W}^c \in R^{d_z \times 2d_h}$ ,  $\tilde{b}^c \in R^{d_z}$ .



### 3.9 Prediction Layer

Compare  $\hat{M}$  with the set threshold  $\eta$  to judge whether it is for emotion cause pair:

$$\begin{cases} \hat{Y}_{p,q} = 1 \left( \hat{M}_{p,q} > \eta \right) \\ \hat{Y}_{p,q} = 0 \left( \hat{M}_{p,q} \leq \eta \right) \end{cases} \quad (30)$$

The whole model structure can be trained by standard gradient descent. The loss is the combination of cross entropy and L2 regularization, and consists of two parts:

The first part is the loss function of task for emotion causes:

$$L_{pair} = - \sum_{p,q} Y_{p,q} \log(\hat{M}_{p,q}) - \sum_{p,q} (1 - Y_{p,q}) \log(1 - \hat{M}_{p,q}) \quad (31)$$

The second part is the loss of emotion extraction and cause extraction for clauses:

$$L_{class} = - \sum_i \left[ \sum_k y_k^e \log(\hat{y}_k^e) + \sum_k y_k^c \log(\hat{y}_k^c) \right] \quad (32)$$

$y_k^e, y_k^c, Y_{p,q}$  represent the extraction of emotion clauses, the extraction of cause clauses, and the correct judgment result of ECPE.

If it is an emotion clause then  $y^e = 1$ , otherwise  $y^e = 0$ . Same rule for  $y^c, Y_{p,q}, y^e$ .

Two auxiliary tasks are also set, which also need to calculate the loss function:

$$L_{aux} = - \sum_i \left[ \sum_k y_k^e \log(y_k^{ae}) + \sum_k y_k^c \log(y_k^{ce}) \right] \quad (33)$$

The final training objective is composed of the above objective function, and coefficients are added to control its influence, as shown in the formula:

$$L = L_{pair} + L_{class} + \beta L_{aux} + \lambda \|\theta\|_2 \quad (34)$$

$\beta$  is used to adjust the influence of auxiliary tasks on the overall model,  $\theta$  represents all parameters that is optimizable, and  $\lambda$  represents the parameters of L2.

The detailed approaches are stated in Algorithm 1.

**Algorithm 1:** MulGAT-ECPE**Input:** text embedding vector  $w$ **Output:** Emotion Cause Pair  $\{E_p, C_q\}$ for  $i = 0 \rightarrow \text{epoch}$  do  *#Step 1: feature fusion and encode for input embedding*   $h = \text{Bi-LSTM}(w)$    $a = F_{\text{fusion}}(h)$  //multi-scale convolution operation is used for feature fusion   $s = \text{Bi-LSTM}(a)$  //Further study the semantic order information in the text  *#Step 2: emotion clause and cause clause extraction*   $y_e, y_c = \text{MLP}(s)$    $z, d = \text{GAT}(y_e, y_c)$  //Capture multi-hop relationships and integrate higher-order information  *#Step 3: calculate emotion cause pairs matrix*   $M = \text{Biaffine}(z, d)$   *#Step 4: predict*

$$\hat{Y} \leftarrow \begin{cases} 1 & \text{if } M_{p,q} > \eta \\ 0 & \text{if } M_{p,q} \leq \eta \end{cases}$$

  return  $E_p, C_q$ 

end for

## 4 Experiment

### 4.1 Experiment Setup

#### Dataset

The dataset was constructed by Xia et al. [16], each document contains only one emotion and its corresponding one or more causes. The average document length is 14.77, and the maximum length is 73. Table 2 shows the proportion of documents with different emotion cause pairs.

**Table 2.** Dataset information

The Document Contains The Number of Emotion Cause Pairs	Quantity	Proportion
One Pair	1746	89.77%
Two Pairs	177	9.10%
Three or More Pairs	22	1.13%

In order to better meet the task for ECPE, the documents with the same text content are integrated into one document. According to statistics, there are 1,945 paragraphs in total, including 2,167 emotion cause pairs, and the number of documents which only have one pair of emotion cause pair accounts for 89.77% of the total documents.

#### Experimental Parameters

In this paper, the experimental conditions are GTX2080Ti, PyTorch framework is used,

Word2Vec trained on Weibo is used as the word vector, the number of filters is set to 50. The specific experimental parameters are shown in Table 3.

**Table 3.** Dataset information.

Parameters	Value
Batch Size	16
Dimension of Word Vector	200
Dimension of Bi-LSTM Hidden State	300
Dimension of MLP	100
Layer of Graph Attention Network	2
Multi Head Attention	3
Smooth Item $\epsilon$	1
L2 Regularized Coefficient $\lambda$	$10^{-5}$
Threshold $\eta$	0.3
Layer of Convolution	4
Kernel Size	(1,2,3,4)
K	128

## 4.2 Experimental Results and Analysis

Table 4 shows a specific document, and the final analysis result is shown in Fig. 3. A lighter color indicates a higher score for the candidate emotion cause pair.

**Table 4.** Case analysis.

Text	Class
C1:Relying on parents' support and the money saved by working in the school	\
C2:they invested more than 500,000 yuan in total	\
C3:and started everything from scratch	\
C4: to set up a chicken farm	\
C5:In 3 months, the sales have reached more than 500,000 yuan	\
C6:With such a good beginning	Cause Clause
C7:they are both proud and confident about the prospect	Emotion Clause

Figure 3 shows that all the output result values are small, this is because after sigmoid, the output result of the double affine attention layer is multiplied by the weight matrix, resulting in a lower score.

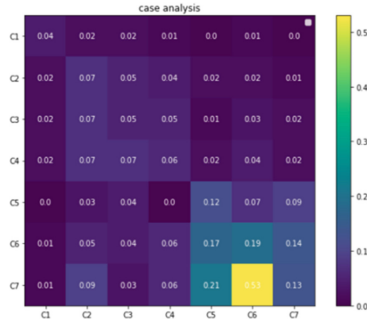


Fig. 3. Result analysis.

As shown in Fig. 4, the ablation experiment was set up to better analyze the effect of each module proposed in this paper.

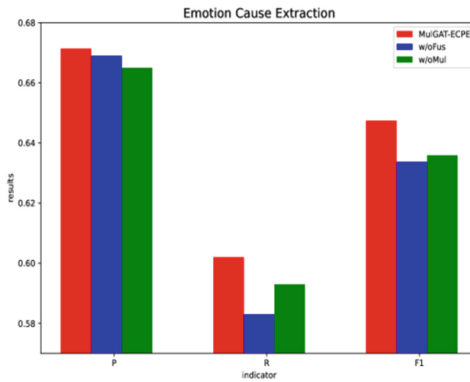


Fig. 4. Comparison of emotion cause pair extraction results.

According to the ablation experiment, when label embedding and feature fusion were added to the model in this paper, the three evaluation indexes of ECPE task were greatly improved. This is because label embedding makes up for the shortcoming of the shared feature parameters. In addition, with feature fusion, the indicators also increased, indicating that the multi-scale features obtained by this method can help the computer better understand emotion words and cause words.

In order to better verify the validity, the Mul-ECPE model is compared with other models in Table 4. Table shows that the proposed model is superior to the latest model in the F1 values of three tasks. The F1 value of emotion ECPE task was 0.8% higher than that of the latest model (Table 5).

**Table 5.** Experiment Results.

Model	Class	Emotion clause extraction			Cause clause extraction			Emotion cause clause extraction		
		P	R	F1	P	R	F1	P	R	F1
Inter-CE(2019)	Two Step	0.8458	0.8035	0.8263	0.6838	0.5754	0.6231	0.6780	0.5254	0.5896
Inter-EC(2019)	Two Step	0.8406	0.8097	0.8242	0.6989	0.5991	0.6426	0.6691	0.5503	0.6013
MAM(2021)	Two Step	0.8486	0.8123	0.8319	0.7035	0.6093	0.6516	0.6893	0.5604	0.6165
DQAN(2021)	Two Step	-	-	-	-	-	-	0.6733	0.6040	0.6362
RHNSC(2020)	End to End	-	-	-	-	-	-	0.6956	0.5871	0.6357
E2EECP(2020)	End to End	0.8595	0.7915	0.8238	0.7062	0.6030	0.6503	0.6478	0.6105	0.6280
Hier-BiLSTM(2020)	End to End	0.6816	0.6629	0.7480	0.7227	0.5532	0.6248	0.6925	0.5371	0.6030
PairGCN(2020)	End to End	0.8587	0.7208	0.7829	0.7283	0.5953	0.6541	0.6999	0.5779	0.6321
MTNECP(2020)	Multi-Task	0.8662	0.8393	0.8520	0.7400	0.6378	0.6744	0.6828	0.5894	0.6321
LAE-Joint(2020)	Multi-Task	0.8810	0.7810	0.8260	-	-	-	0.6880	0.5960	0.6400
TDGC(2020)	Multi-Task	0.8080	0.8406	0.8256	0.6742	0.6534	0.6636	0.6515	0.6354	0.6394
MulGAT-ECPE	Text	0.8643	0.8421	0.8617	0.7310	0.6015	0.6814	0.6714	0.6070	0.6484

## 5 Conclusion

The Mul-ECPE model proposed in this paper is used for ECPE. This model makes full use of the effect of feature fusion, multi-scale information is allocated to the word vector representation so that the model can better understand the emotion word and cause word. The addition of label embedding can make full use of the rich interactive relationship information among the three tasks. Experimental results also show that the proposed model achieves good results on ECPE data sets. Existing methods remain at the clause level, so future work can be improved from the perspective of granularity level, so as to extract emotion cause pairs at a fine-grained level.

**Acknowledgments.** This work was supported by the Key projects of the National Natural Science Foundation of China (No.62137001).

## References

1. Lee, S.Y.M., et al.: Emotion cause events: corpus construction and analysis. In: International Conference on Language Resources & Evaluation. DBLP (2012)
2. Gui, L., et al.: A Question Answering Approach for Emotion Cause Extraction (2017)
3. Li, X., et al.: A co-attention neural network model for emotion cause analysis with emotional context awareness. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (2018)
4. Yu, X., et al.: Multiple level hierarchical network-based clause selection for emotion cause extraction. IEEE Access (2019)
5. Xia, R., Zhang, M., Ding, Z.: RTHN: A RNN-Transformer Hierarchical Network for Emotion Cause Extraction (2019)

6. Fan, C., et al.: A knowledge regularized hierarchical approach for emotion cause analysis. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (2019)
7. Wu, S., et al.: A multi-task learning neural network for emotion-cause pair extraction. ECAI IOS Press **2020**, 2212–2219 (2020)
8. Fan, C., et al.: Transition-based directed graph construction for emotion-cause pair extraction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (2020)
9. Wei, P., Zhao, J., Mao, W.: Effective inter-clause modeling for end-to-end emotion-cause pair extraction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (2020)
10. Tang, H., Ji, D., Zhou, Q.: Joint multi-level attentional model for emotion detection and emotion-cause pair extraction. Neurocomputing **409**, 329–340 (2020)
11. Vaswani, A., et al.: Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 6000–6010 (2017)
12. Tang, J., Qu, M., Mei, Q.: Pte: predictive text embedding through large-scale heterogeneous text networks. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1165–1174 (2015)
13. Zhang, Y., et al. Deconvolutional paragraph representation learning. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 4172–4182 (2017)
14. Wang, G., et al.: Joint embedding of words and labels for text classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 2321–2331 (2018)
15. Zhang, H., et al.: Multi-task label embedding for text classification. Proceedings of the Conference on Empirical Methods in Natural Language Processing. **2018**, 4545–4553 (2018)
16. Xia, R., Ding, Z.: Emotion-cause pair extraction: a new task to emotion analysis in texts. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 1003–1012 (2019)
17. Ding, Z., Xia, R., Yu, J.: ECPE-2D: emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 3161–3170 (2020)
18. Song, H., et al.: End-to-end emotion-cause pair extraction via learning to link. arXiv preprint [arXiv:2002.10710](https://arxiv.org/abs/2002.10710) (2020)
19. Fan, C., et al.: Multi-task sequence tagging for emotion-cause pair extraction via tag distribution refinement. IEEE/ACM Transactions on Audio, Speech, and Language Processing **29**, 2339–2350 (2021)
20. Li, C., Zhai, R., Zuo, F., Yu, J., Zhang, L.: Mixed Multi-channel Graph Convolution Network on Complex Relation Graph. In: Xing, C., Fu, X., Zhang, Y., Zhang, G., Borjigin, C. (eds.) WISA 2021. LNCS, vol. 12999, pp. 497–504. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-87571-8\\_43](https://doi.org/10.1007/978-3-030-87571-8_43)
21. Chen, X., Li, Q., Wang, J.: Conditional causal relationships between emotions and causes in texts. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) **2020**, 3111–3121 (2020)