# Emotion-cause pair extraction based on interactive attention

**Weichun Huang**[1] · **Yixue Yang**[1] · **Xiaohui Huang**[1] · **Zhiying Peng**[1] · **Liyan Xiong**[1]

## Abstract

In recent years, a new fine-grained task has been proposed in the field of sentiment analysis, the emotion-cause pair extraction (ECPE) task, whose purpose is to extract all emotions and their causes from a document. Most of existing methods produce effective emotion-cause pairs by filtering all possible pairs. However, this types of methods ignore the relationship between emotion clauses and cause clauses when learning the representations of emotions and causes clauses. In order to solve the above problem, we propose an end-to-end framework, which uses interactive attention and its fusion mechanism to learn the relationship between emotions and causes, and then pair them. Experimental results on quasi-base corpus shows our proposed method outperform the state-of-the-art baseline.

**Keywords** Interactive attention · Emotion-cause pair extraction · Fusion mechanism

## 1 Introduction

In recent years, discovering the causes behind emotions has received more and more attention from scholars in the field of sentiment analysis. Some scholars [1] defined a clause-level emotion cause extraction (ECE) task and provided a new corpus. The ECE task is to further detect the underlying causes behind the emotions by labeling the emotions, and its goal is to judge whether each clause in the document is the cause of emotional correspondence. Analyzing the causes of emotions has a very wide range of applications,

Weichun Huang and Yixue Yang contributed equally to this work.

✉ Yixue Yang
  yang_yi_xue@163.com

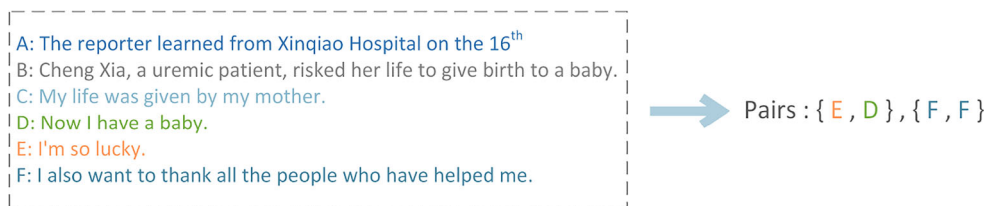  Weichun Huang
  hwc@ecjtu.edu.cn

  Xiaohui Huang
  hxh016@gmail.com

  Zhiying Peng
  pzhiying0224@163.com

  Liyan Xiong
  xly_ecjtu@163.com

[1] School of Software Department, East China Jiaotong University, Nanchang, 330013, China

such as mining user reviews on e-commerce and internet platforms, and monitoring the trend of public opinion on social platforms. In its follow-up research, the ECE task has received extensive attention. Although the ECE task is well-defined, there are still two key problems: first, the ECE task ignores the potential relationship between emotions and causes; second, the emotion needs to be annotated before the cause is extracted. This limits the application scenarios of ECE tasks to a large extent. In response to these two problems, Xia and Ding et al. [2] proposed a new task named emotion-cause pair extraction(ECPE), which aims to extract all emotion-cause pairs from a given document. As shown in Fig. 1, there are seven sentences in the document, and all clauses are used as input to the model. Clause E contains the emotion "lucky" and clause F contains the emotion "thank", corresponding to the cause clause D. ("Now I have a baby.") and the clause F ("I also want to thank all the people who have helped me."), through model training, to predict, and finally output the emotion-cause pair {E,D}, {F,F}.

For the ECPE task, Xia and Ding et al. [2] proposed a two-step framework. First, a binary classifier based on Bi-LSTM is used to extract emotion clauses and cause clauses respectively, and then emotion-cause pairs in the document are filtered through the cartesian product pairing filter. Such a two-step framework is effective, but there are still some problems. The first is that the extraction of emotions and causes in the first step of the model will directly affect the accuracy of the second step. Second, the model uses filters to filter incorrect pairs, resulting in low prediction

**Fig. 1** A sample document in the data set

A: The reporter learned from Xinqiao Hospital on the 16<sup>th</sup>
B: Cheng Xia, a uremic patient, risked her life to give birth to a baby.
C: My life was given by my mother.
D: Now I have a baby.
E: I'm so lucky.
F: I also want to thank all the people who have helped me.

Pairs : { E , D }, { F , F }

accuracy. As a result, some scholars have proposed to use an end-to-end model to solve the above problems [3–5]. With the popularity of transformer, graph convolution and other modules, some basic deep learning models are used to solve the pairing problem of emotion and causes. For example, Ding et al. [3] proposed an end-to-end model, called ECPE-2D, which is an emotion-cause pair extraction model that uses transformer to combine two-dimensional representation, interaction, and prediction. Although this type of model well integrates the hierarchical relationship between the ECPE task and the two subtasks, and fully learns the relationship between all possible emotion-cause pairs. However, this type of model also has the following two problems, which affect the accuracy and efficiency of the prediction results:
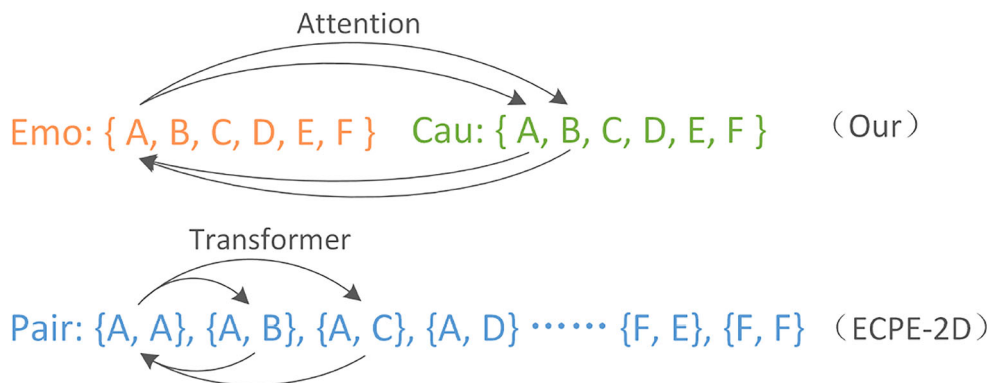
First of all, when generating the expression of emotion-cause pairs, although existing research considers the relationship between emotion-cause pairs, it ignores the relationship between emotion clauses and cause clauses. Learning and strengthening the relationship between emotion clauses and cause clauses may be more beneficial to the prediction of emotion-cause pairs. Taking the ECPE-2D model proposed by Ding et al. [3] as an example, ECPE-2D uses an end-to-end model to pair the emotion and cause expression learned by Bi-LSTM one by one, and express the emotion-cause pair as a vector. Then, using transformer to learn the relationship between each emotion-cause pair. According to Fig. 2, ECPE-2D first paired emotion clauses and cause clauses one by one to form a two-dimensional representation, that is, Pair in the Fig. 2, and then input it into transformer to learn the relationship between pairs. However, it ignores the relationship between the emotions

and causes, that is, Emo and Cau in the Fig. 2. We believe that the causal relationship between the emotion clause and the cause clause in the correct emotion-cause pair is more conducive to the prediction of the emotion-cause pair. Therefore, it is more important to learn the relationship between emotion clauses and cause clauses than to learn the relationship between pairs and pairs, and the above approach fails to capture the complex relationships and correlations at different levels.

Secondly, redundant model architecture and numerous model trainable parameters are also one of the problems faced by current research. For example, ECPE-2D pairs all the clauses one by one, and then enters the Transformer for multi-level and multi-faceted training. Such a training method produces large trainable parameters, resulting in a complex network structure, time-consuming training, and a decrease in prediction accuracy, resulting in inaccurate prediction results.

In order to solve the above two problems, we propose an end-to-end model based on interactive attention, called IA-ECPE. Based on the research of Ding et al. [3] and other scholars [5], we continue to use the end-to-end network structure, and combine the resulting emotion clause set and the set of cause clauses are input to the interactive attention separately, and the degree of correlation between the emotion clause and the cause clause is learned. Then the output of the interactive attention is fused with multi-level information, and finally the prediction of the emotion-cause pair is carried out. Our model not only learns the relationship between pairs and pairs, but also considers the interaction between emotion clauses and cause clauses. At the same time, through the optimization of BiGRU

**Fig. 2** An example is used to compare the two models in the process of emotion cause pair extraction

Attention

Emo: { A, B, C, D, E, F }  Cau: { A, B, C, D, E, F }  （Our）

Transformer

Pair: {A, A}, {A, B}, {A, C}, {A, D} ······ {F, E}, {F, F} （ECPE-2D）

[6], interactive attention and other modules, the amount of model parameters has also been reduced, and the training accuracy has also been improved. The experimental results show that our method achieves the best experimental results on quasi basic corpus. Combined with the above description, the main contributions of this article are as follows:

- We designed an interactive attention module suitable for ECPE tasks, aiming to accurately capture the causal relationship between emotions and causes.
- We added different levels of fusion mechanisms to the module. At the same time, the complexity of our model architecture has been reduced, and the number of trainable parameters has also been reduced to a certain extent.
- Our model is tested on a quasi-base corpus, and the results show that our method outperform the state-of-the-art baselines.

The other parts of this paper are described as follows. In the Section 2, we introduces in detail the research and findings of relevant scholars on this task, as well as the relevant progress. In the Section 3, we mainly introduces our model. We describe in detail the interactive attention mechanism and how to use the fusion mechanism to capture different levels of information. In the Section 4, the experimental results and performance of this paper, which introduces the details and result analysis of the experiment in detail. Finally, the Section 5 summarizes the article and puts forward our prospects.

## 2 Related work

This section introduces the ECE task and the related research of ECPE task.

### 2.1 Emotion cause extraction

Lee et al. [7] first studied the task of extracting emotional causes. They designed a system based on language rules to detect cause events. In the early days, some work tried rules-based [8–10], common sense [11] and traditional machine learning [12] method to extract some causes for emotional expression. Gui et al. [1] proposed an event-driven multi-core SVM method and published a benchmark corpus. Feature-based methods [13] and neural methods [14–16] have recently been proposed. Xia et al. [17] adopted a method based on a transformer encoder enhanced by position information and an integrated global prediction embedding method to improve the prediction accuracy. Fan et al. [18] adopted an approach based on mood and position regulators to inhibit parameter learning. Hu et al. [19] used

an external emotion classification corpus to pre-train the model. In some other research fields [20], emotional causes were extracted in the context of multi-user micro blog. In addition, Kim and Klinger [21] and Bostan et al. [22] treat emotions as structured data, and study the semantic role of emotions, including trigger phrases, experiencers, goals and causes, and readers' perceptions.

### 2.2 Emotion-cause pair extraction

In the past, emotion-cause extraction tasks all needed to mark emotion clauses. In response to this problem, Xia et al. [2] first proposed the emotion-cause pair extraction task in 2019, and proposed a two-step framework to extract emotion clauses and cause clauses respectively, and then send them to the model for training classification to filter out negative sentence pairs. Recently, some studies have focused on the ECPE task and proposed an end-to-end model, thus avoiding some cascading errors that may be caused by the two-step method proposed by Xia and Ding et al. [2]. Wei et al. [23] proposed a model named RANKCP, which uses graph neural network to propagate between clauses to learn pairwise representations and perform candidate emotion-cause pairs based on the learned pairwise representations. Sort, finally make predictions, and output the prediction results. Tang et al. [24] proposed a model called LAE-Joint-MANN-BERT, which is based on BERT for joint processing of emotion detection (ED) and ECPE tasks. Specifically, they calculated the attention value of the associative clauses in all clauses to express the relevance and importance of the associative clauses, thereby predicting the probability that each pair is an emotion-cause pair. Similarly, Song et al. [25] regarded the ECPE task as a chain network prediction problem, and proposed E2EECPE to solve the ECPE task. Specifically, the model uses the parental attention module to model the relevance and importance of clauses. Ding et al. [3] proposed the ECPE-2D model, which uses a hierarchical self-attention module to calculate the attention matrix between all clauses, and deploys a 2D converter to simulate the correlation and correlation between emotion-cause pairs. Importance. Fan et al. [18] transformed the extraction of emotion-causse pair into performing a series of actions, and modeled them to solve ECPE tasks from different perspectives. Specifically, they convert each given document into a directed graph, and convert the original data set into a sequence. On this basis, they trained a model to predict the next state of the current state. Different from the above work, Turcan et al. [26] redefines the ECPE task as a unified sequence labeling problem, and designed a unified label to replace the original label, so that their model can extract emotion-cause pairs of different emotion types and encode adjacent information to improve task performance.

# 3 Model

In this section, we propose an emotion-cause pair extraction based on interactive attention framework, called IA-ECPE. As shown in Fig. 3, IA-ECPE is a stacking framework based on interactive attention. The lower network can provide information for the upper network to optimize the results. In the embedding part of the model, since the object is a document, we use hierarchical coding, from word to sentence, and then from sentence to document. We encode words with word2vec, and the word-to-sentence process uses BiLSTM to obtain the encoding of the sentence, and then input BiGRU to encode the sentence in the document to obtain the expression of the emotion clauses and the cause clauses. Then we input the obtained clause expression into the interactive attention module to learn the relationship between emotion clauses and cause clauses, and use the fusion mechanism to strengthen the internal information of

the emotion clauses and the cause clauses, Thus, different levels of complex correlation are captured. Finally, the obtained vector and position information are fused and predicted to obtain the final prediction result. Below we will introduce the modules we designed in detail.

## 3.1 Interactive attention

As shown in Fig. 4, it is the internal overall framework of the interactive attention we designed. It can be seen from the figure that this module is mainly composed of a feedforward neural network and a weight calculation module. Specifically, according to the output of BiGRU, we can get the expression of emotion clauses $emo = \{c_1^{emo}, c_2^{emo}, \ldots\ldots, c_N^{emo}\}$ and the expression of cause clauses $cau = \{c_1^{cau}, c_2^{cau}, \ldots\ldots, c_N^{cau}\}$, where $N$ represents the number of sentences in a document. We first input the emotion clause and cause clause into a feed



**Fig. 3** The overall framework of the joint model based on interactive attention is used to solve the emotion cause pair extraction task
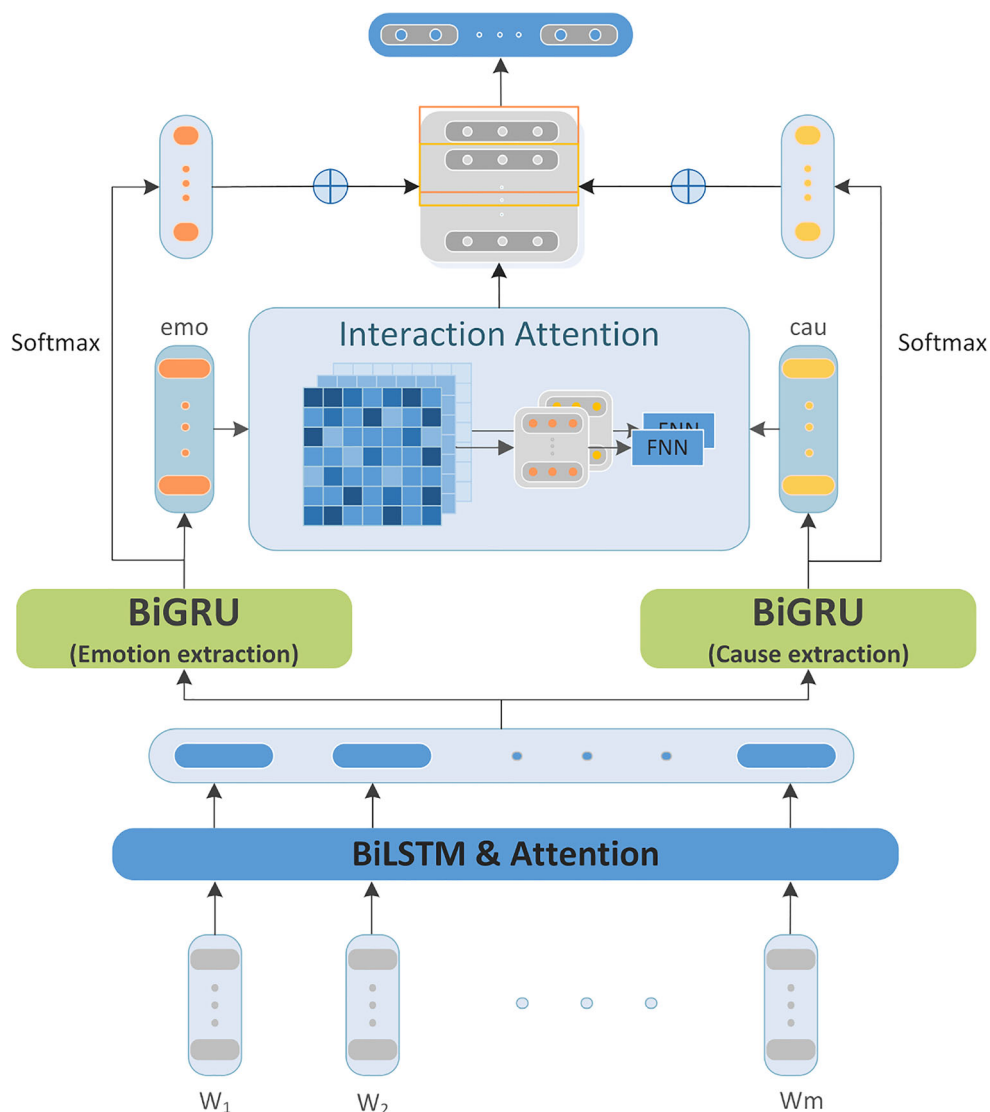
**Fig. 4** Internal structure of interactive attention
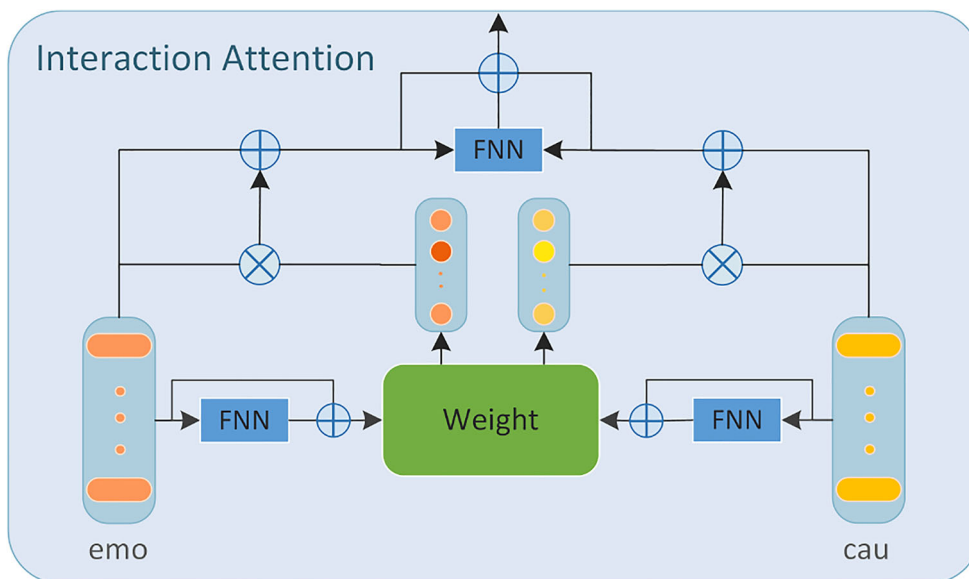


forward neural network with a non-linear layer Relu to update the vector expression of emotion and cause. The vector representations $e_i$ and $c_i$ corresponding to emotion clauses and cause clauses are obtained by feedforward neural network, and then this module use the fusion mechanism to fuse the output of BiGRU with its features to obtain the vector representations $E_i$ and $C_i$ that are finally input to the attention module. The specific formula is expressed as:

$$e_i = F_1\left(W^e c_i^{emo} + b^e\right), \tag{1}$$

$$c_i = F_2\left(W^c c_i^{cau} + b^c\right), \tag{2}$$

$$E_i = f_a\left(c_i^{emo}, e_i\right), \tag{3}$$

$$C_i = f_a\left(c_i^{cau}, c_i\right), \tag{4}$$

where $e_i$ represents the update expression of the i-th emotion clause in the document, $c_i$ represents the update expression of the i-th cause clause in the document, $W^e$ and $W^c$ are the coefficient matrices of emotion and cause linear transformation, respectively, $b^e$ and $b^c$ are bias terms, $F$ represents the function representation corresponding to the feedforward neural network, and $f_a$ represents the corresponding addition function between the vectors.

The obtained $E_i$ and $C_i$ are jointly calculated to obtain the attention weight matrix. The rows and columns of the matrix represent the relevance and importance of the emotions and the causes clause. Specifically, the rows of the matrix represent the degree of relevance of each emotion clause to all the cause clauses in the document, and the columns of the matrix represent the degree of relevance of each cause clause to all the emotion clauses in the document. After that, the attention weight matrix is

standardized. As shown in Fig. 5, the attention matrix is normalized in different directions, and the specific formula is as follows:

$$a_{ij} = e_i^T c_i, \tag{5}$$

$$\alpha_{ij} = \frac{\exp\left(a_{ij}\right)}{\sum_m^N \exp\left(a_{im}\right)} \quad j \in [0, N], \tag{6}$$

$$\beta_{ij} = \frac{\exp\left(a_{ij}\right)}{\sum_m^N \exp\left(a_{mj}\right)} \quad i \in [0, N], \tag{7}$$

where $\alpha_{ij}$ and $\beta_{ij}$ represent the correlation weight coefficients corresponding to the emotions and causes respectively, and $a_{ij}$ is the correlation degree of the i-th row and the j-th column of the interactive attention matrix.

## 3.2 Fusion mechanism

Through the previous calculations, we found that the vector expression of clauses may lose some key information during the training process, so we added a fusion mechanism to the interactive attention framework to strengthen the correlation between clauses and capture the complex dependence between different levels. The specific operation is as follows. Through the above processing, we get the corresponding attention weight, that is, correlation coefficient $\alpha_i$ of emotion clause corresponding to cause clause, and the correlation coefficient $\beta_j$ of the cause clause corresponding to the emotion clause. The attention weight is respectively fused with the updated emotion clause vector expression $E_i$ and the cause clause vector expression $C_i$, and the correlation degree between the clauses is fused with the original information to obtain the vector expression of the attention mechanism $Emo^{com}$ and $Cau^{com}$. As shown

**Fig. 5** Calculation principle of attention weight matrix
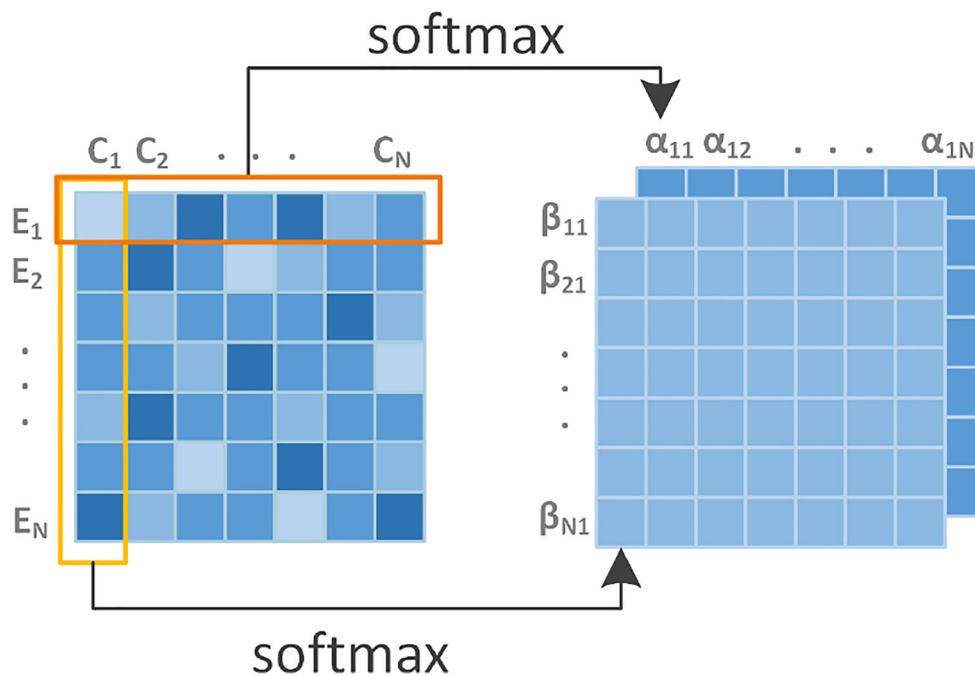


in Fig. 4, because we consider all possible pairs, we fuse the features of all emotion clauses and cause clauses. Then, emotion clauses and cause clauses are paired one by one to connect, and get the output vector $Y_{ij}$ of the final interactive attention. The specific formula is as follows:

$$Emo_j^{com} = f_a \left( c_j^{emo}, \alpha_{ij} C_j \right), \tag{8}$$

$$Cau_i^{com} = f_a \left( c_i^{cau}, \beta_{ij} E_i \right), \tag{9}$$

$$Y_{ij} = \left[ f_a \left( Emo_i^{com}, F_2 \left( Emo_i^{com} \right) \right), f_a \left( Cau_j^{com}, F_2 \left( Cau_j^{com} \right) \right) \right], \tag{10}$$

where, $Emo_i^{com}$ and $Cau_j^{com}$ represent the vectors of emotion and cause clauses are obtained by fusing the original information and attention information, and $f_a$ represents the corresponding addition function between the vectors, $F_2$ represents the feedforward neural network function with Relu.

### 3.3 Joint expression

Our model is a stacking framework. The results of the subtasks in the lower layer of our task may affect the upper task. Therefore, we input the emotion and cause sentence expression output by BiGRU into softmax for prediction, and the predicted value $y^{emo}$ and $y^{cau}$ of the emotion and cause clauses is obtained. Each emotion clause and cause clause are paired one by one, and the output $Y$ of the interactive attention is fused with the predicted values $y^{emo}$,

$y^{cau}$ and location information $P$ to obtain the representation $U$ of the emotion-cause pair,

$$U_{ij} = Concat \left( Y_{ij}, y_i^{emo}, y_j^{cau}, P_{ij} \right), \tag{11}$$

where $W_1$ and $W_2$ are weight matrices, $b_1$ and $b_2$ are bias terms, $P_{ij}$ is the relative position information between the i-th emotion clause and the j-th cause clause, and $U_{ij}$ represents the joint expression of the i-th emotion clause and the j-th cause clause.

### 3.4 Predictions of affective causes

Through the above calculation, we have obtained all possible emotion-cause pairs for the expression $U$. Based on common sense, we found that causes generally appear around emotion. Therefore, we set up a sliding window. For each emotion-cause pair, only the cause clauses within a certain range around it are used for prediction. The setting of sliding window not only reduces the prediction range, but also reduces the trainable parameters of the model to a certain extent. Then we predict the obtained $U_w$ with softmax. The loss functions for emotion clause classification, cause clause classification and emotion-cause pair classification are as follows:

$$L^{emo} = - \sum_{i=1}^{|N|} Y_i^e \cdot \log \left( y_i^{emo} \right), \tag{12}$$

$$L^{cau} = - \sum_{i=1}^{|N|} Y_i^c \cdot \log \left( y_i^{cau} \right), \tag{13}$$

$$L^{pair} = -\sum_{i=1}^{|N|} \sum_{j=1}^{|M|} u_{i,j}^{pair} \cdot log \left( U_{ij} \right), \tag{14}$$

where $Y_i^e$, $Y_i^c$, $u_{i,j}^{pair}$ are the ground truth distribution of emotion, cause and emotion-cause pair, respectively. Combine the three to get the final loss function $L$:

$$L = L^{emo} + L^{cau} + L^{pair} + \lambda \|\theta\|^2, \tag{15}$$

where $\lambda$ is the weight of the $l_2$ regularization term. $\theta$ represents all model parameters.

## 4 Experiment

In this section, we will introduce the details of our experimental research and evaluate the feasibility and effectiveness of our model.

### 4.1 Dataset, metrics and experimental settings

We evaluate the model using two publicly available ECPE datasets [2, 5]. Among them, the Chinese dataset constructed by Gui et al. [1], and contains 1945 documents from news websites, and the English dataset contains 2843 documents from English novels. We divided the data set into a training set and a testing set, where 90% of the data is the training set, and the remaining 10% is the testing set. In order to further verify the robustness of the model statistics, we perform 10-fold cross-validation and report the average results of these experiments. We used the same data partitioning as Xia and Ding [2]. We use precision, recall and F1 score as evaluation indicators. Among the three indexes, F1, the main evaluation index, is the comprehensive evaluation of precision and recall rate. The calculation is as follows:

$$P = \frac{\sum correct_{ECP}}{\sum proposed_{ECP}}, \tag{16}$$

$$R = \frac{\sum correct_{ECP}}{\sum annotated_{ECP}}, \tag{17}$$

$$F1 = \frac{2 \times P \times R}{P + R}, \tag{18}$$

where $correct_{ECP}$ represents the number of emotion-cause pairs marked and predicted, $proposed_{ECP}$ represents the number of emotion-cause pairs predicted by the model, and $annotated_{ECP}$ represents the total number of emotion-cause pairs marked in the data set.

Regarding the experimental parameter settings, the details are as follows. We use word2vec [27] to pre-train the word vectors in the Chinese Weibo corpus. We set the pre-training dimension to 200 dimensions. For the BiLSTM used in the model, the number of hidden layers of BiGRU is set to 100. The hidden state in the attention mechanism is set to 30. In the feed-forward neural network, the dimension of the middle layer is 30, and the dropout is set to 0.9. The initial dimension of the relative position vector is 50. The sliding window size is 3. In the training process, we set the learning rate to 0.005 and the batch size to 32. For the optimization process, we use the Stochastic Gradient Descent (SGD) algorithm and Adam optimizer. $\lambda$ is set to $1e-5$.

### 4.2 Baseline models

We compared our model with the following baseline model:

- Indep: This method is the first model proposed by Xia and Ding et al. [2], which treats emotion extraction and cause extraction as two independent tasks. Emotions and causes are extracted by multi task learning, and then emotions and causes are paired and filtered by filter.
- Inter-CE: An interactive multi-task learning method that uses the prediction of cause extraction to improve emotion extraction. The rest of the model is the same as Indep.

**Table 1** The performance of our model and baseline model on ECPE task and two subtasks based on accuracy, recall and measurement index F1

| Models | Emotion Ext | | | Cause Ext | | | Emotion-Cause Pair Ext | | |
|---|---|---|---|---|---|---|---|---|---|
| | P | R | F1 | P | R | F1 | P | R | F1 |
| Indep | 83.75 | 80.71 | 82.10 | 69.02 | 56.73 | 62.05 | 68.32 | 50.82 | 59.18 |
| Inter-CE | 84.94 | 81.22 | 83.00 | 68.09 | 56.34 | 61.51 | 69.02 | 51.35 | 59.01 |
| Inter-EC | 83.64 | 81.07 | 82.30 | 70.41 | 60.83 | 65.07 | 67.21 | 57.05 | 61.28 |
| E2EECPE | **85.95** | 79.15 | 82.38 | 70.62 | 60.30 | 65.03 | 64.78 | **61.05** | 62.80 |
| ECPE-2D | 84.63 | **81.95** | 83.19 | 72.17 | 62.66 | 67.01 | **71.31** | 57.86 | 63.65 |
| | (±0.13) | (±0.09) | (±0.09) | (±0.58) | (±0.27) | (±0.36) | (±0.75) | (±0.31) | (±0.37) |
| IA-ECPE | 85.24 | 81.38 | **83.23** | **72.53** | **63.22** | **67.43** | 69.80 | 60.56 | **64.78** |
| | (±0.11) | (±0.08) | (±0.06) | (±0.61) | (±0.26) | (±0.33) | (±0.62) | (±0.32) | (±0.39) |

The best performances are marked in bold

**Table 2** The parameters of our model and ECPE-2D

| Method | Trainable Parameters |
|---|---|
| IA-ECPE(ours) | 978134 |
| ECPE-2D | 1060116 |

- Inter-EC: This is another interactive multi-task learning method that uses the prediction of emotion extraction to improve the cause extraction. The rest of the model is the same as Indep.
- ECPE-2D: Proposed by Ding et al. [3], they propose to realize the interaction of all emotion-cause pairs in 2D form, and use the self-attention mechanism to calculate the attention matrix of emotion-cause pairs. Here we choose the Inter-EC model with better effect.
- E2EECPE: An end-to-end model proposed by Song et al. [25]. This model is a multi-task learning linkage framework that uses a biaffine attention to mine the relationship between any two clauses.

### 4.3 Overall performance

The overall effect of our model is shown in Table 1. It can be seen that in the three tasks of our model, the performance of indicator F1 is the best, which fully proves the effectiveness of our model.

First of all, our model is significantly better than the three models of Indep, Inter-CE and Inter-EC, regardless of the results on the two subtasks or on the ECPE task. Compared with Inter-EC, which performs best among them, our model improves F1 by about 5.28% on the task of Emotion-cause pair extraction.

Secondly, we compared the model with ECPE-2D, the model with the best result among all baseline models. Our model improved in three tasks, among which the F1 value on ECPE task increased more significantly than ECPE-2D. This fully proves the effectiveness of the interactive attention module. This further confirms our guess that the relationship between emotion and cause may be more

conducive to pairing than the relationship between pair and pair. At the same time, our model not only uses a simpler architecture to achieve similar or even better results than ECPE-2D, but also has better performance in terms of parameters, as shown in Table 2. Compared with ECPE-2D, the parameter quantity is reduced by 7.73%.

To verify the robustness of our model, we apply our model to the English dataset to observe its performance. According to Table 3, on the emotion extraction task, IA-ECPE is only slightly higher than ECPE-2D in recall rate. However, on the task of cause extraction, the P, R, and F1 values of the IA-ECPE model are all better than the baseline model, and the comprehensive index F1 is 2.62% higher than the optimal baseline. This fully demonstrates that our model can effectively capture the underlying causes in the document. On the emotion-cause pair extraction task, the IA-ECPE model also achieved better results. We speculate that this may be because our model improves the accuracy of cause clause extraction, thereby improving the accuracy of emotion-cause pair extraction.

### 4.4 Ablation study

To prove the effectiveness of various modules of our model, we conducted ablation experiments on the model, as shown in Table 4. First of all, "-IA" means that the model removes the interactive attention we designed, and uses the output of BiGRU to directly perform joint expressions and then make predictions. According to the Table 4, the scores of F1 decreased in different degrees in the three tasks, which fully indicates that removing the interactive attention will lead to a decrease in prediction accuracy. At the same time, compared to ECPE-2D, even without interactive attention, our ECPE task F1 has achieved better results when the subtask does not have a good result of ECPE-2D. As a result, this shows that our framework has certain advantages in the pairing of emotion-cause pairs.

Secondly, we conducted an ablation experiment on the fusion mechanism, "+IA-M", which means that the interactive attention has been added but the fusion

**Table 3** Experimental performance of emotion-cause pair extraction, emotion extraction and cause extraction on English datasets

| Models | Emotion Ext | | | Cause Ext | | | Emotion-Cause Pair Ext | | |
|---|---|---|---|---|---|---|---|---|---|
| | P | R | F1 | P | R | F1 | P | R | F1 |
| Inter-EC | 67.41 | 71.60 | 69.40 | 60.39 | 47.34 | 53.01 | 46.94 | 41.02 | 43.67 |
| E2EECPE | 65.19 | 66.09 | 65.64 | 54.95 | 46.01 | 50.08 | 50.02 | 37.16 | 42.63 |
| ECPE-2D | **74.32** | 69.78 | **71.92** | 62.75 | 52.83 | 57.30 | 59.86 | 43.00 | 49.92 |
| IA-ECPE | 73.98 | **69.85** | 71.80 | **63.87** | **54.55** | **58.80** | **60.14** | **43.03** | **50.05** |
| | (±0.10) | (±0.06) | (±0.04) | (±0.10) | (±0.07) | (±0.05) | (±0.27) | (±0.11) | (±0.11) |

We omit Indep and Inter-CE due to their poor performances

The boldface is used to mark the maximum value of each parameter

**Table 4** Ablation study of our method

| Models | Emotion Ext | | | Cause Ext | | | Emotion-Cause Pair Ext | | |
|---|---|---|---|---|---|---|---|---|---|
| | P | R | F1 | P | R | F1 | P | R | F1 |
| –IA | 84.78 | 81.62 | 83.14 | 71.35 | 62.41 | 66.49 | 68.35 | 60.23 | 63.91 |
| +IA-M | 85.10 | **81.76** | **83.32** | 72.23 | 61.87 | 66.52 | 69.03 | 60.10 | 64.18 |
| IA-ECPE | **85.24** | 81.38 | 83.23 | **72.53** | **63.22** | **67.43** | **69.80** | **60.56** | **64.78** |

The boldface is used to mark the maximum value of each parameter

mechanism in the interactive attention has been deleted. Compared with the former, the F1 of the three tasks has increased, and the P has also increased significantly, which fully demonstrates the effectiveness of interactive attention. By looking at R, it can be seen that IA-ECPE has improved scores on both subtasks and the ECPE task compared to the +IA-M model. We guess that IA-ECPE is beneficial to improve the recall rate, and it is precisely because of the improvement of the index R that our F1 on the three tasks is improved. These results fully demonstrate the effectiveness of our proposed IA-ECPE framework.

# 5 Conclusion

In response to the new tasks in the field of sentiment analysis in recent years, the emotion-cause pair extraction task (ECPE), we propose an end-to-end model based on interactive attention. In our model, the emotion clause set and the cause clause set are input into the interactive attention respectively to learn the relationship between the emotions and the causes, get new clauses representation, and then conduct fusion pairing. This approach not only learns the relationship between pairs and pairs, but also considers the interaction between affective clauses and cause clauses. At the same time, through BiGRU, interactive attention reduces the amount of model parameters, reduces training time and improves training accuracy. The experimental results show that our method shows good experimental results on the quasi-base data set. Moreover, the ablation experiment further proved the effectiveness of our method. In the future, we plan to study the conditional random field on the ECPE task and apply it to our model, and use BERT for word pre-training in our framework. We think this may be more conducive to improving the prediction accuracy.

# References

1. Gui L, Wu D, Xu R, Lu Q, Zhou Y (2016) Event-driven emotion cause extraction with corpus construction. In: Proceedings of the 2016 conference on empirical methods in natural language processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016, pp 1639–1649
2. Xia R, Ding Z (2019) 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, ACL 2019, pp 1003–1012
3. Ding Z, Xia R, Yu J (2020) 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, ACL 2020, pp 3161–3170
4. Chen X, Li Q, Wang J (2020) A unified sequence labeling model for emotion cause pair extraction. In: Proceedings of the 28th international conference on computational linguistics, COLING 2020, pp 208–218
5. Singh A, Hingane S, Wani S, Modi A (2021) An end-to-end network for emotion-cause pair extraction. Association for Computational Linguistics:84–91
6. Wang Z, Yang B (2020) Attention-based bidirectional long short-term memory networks for relation classification using knowledge distillation from bert. In: 2020 IEEE Intl conf on dependable, pp 562–568
7. Lee SYM, Chen Y, Huang C-R (2010) A text-driven rule-based system for emotion cause detection. In: Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, pp 45–53
8. Gao K, Xu H, Wang J (2015) Emotion cause detection for chinese micro-blogs based on ecocc model. In: Advances in knowledge discovery and data mining-19th Pacific-Asia conference, PAKDD 2015, pp 3–14
9. Chen Y, Lee SY-M, Li S, Huang C-R (2010) Emotion cause detection with linguistic constructions. In: 23rd international conference on computational linguistics, proceedings of the conference, COLING 2010, pp 179–187
10. Neviarouskaya A, Aono M (2013) Extracting causes of emotions from text. In: Sixth international joint conference on natural language processing, IJCNLP 2013, pp 932–936
11. Russo I, Caselli T, Rubino F, Boldrini E, Martínez-barco P (2011) Emocause: an easy-adaptable approach to extract emotion cause contexts. In: Proceedings of the 2nd workshop on computational

approaches to subjectivity and sentiment analysis, WASSA@ACL 2011, pp 153–160

12. Ghazi D, Inkpen D, Szpakowicz S (2015) Detecting emotion stimuli in emotion-bearing sentences. In: Computational linguistics and intelligent text processing - 16th international conference, CICLing 2015, pp 152–165

13. Xu R, Hu J, Lu Q, Wu D, Gui L (2017) An ensemble approach for emotion cause detection with event extraction and multi-kernel svms. Tsinghua Sci Technol 22:646–659

14. Gui L, Hu J, He Y, Xu R, Lu Q, Du J (2017) A question answering approach for emotion cause extraction. In: Proceedings of the 2017 conference on empirical methods in natural language processing, EMNLP 2017, pp 1593–1602

15. Li X, Song K, Feng S, Wang D, Zhang Y (2018) A co-attention neural network model for emotion cause analysis with emotional context awareness. In: Proceedings of the 2017 conference on empirical methods in natural language processing, EMNLP 2018, pp 4752–4757

16. Yu X, Rong W, Zhang Z, Ouyang Y, Xiong Z (2019) Multiple level hierarchical network-based clause selection for emotion cause extraction. IEEE Access 7:9071–9079

17. Xia R, Zhang M, Ding Z (2019) Rthn: a rnn-transformer hierarchical network for emotion cause extraction. In: Proceedings of the twenty-eighth international joint conference on artificial intelligence, IJCAI 2019, pp 5285–5291

18. Fan C, Yuan C, Du J, Gui L, Yang M, Xu R (2020) Transition-based directed graph construction for emotion-cause pair extraction. In: Proceedings of the 58th annual meeting of the association for computational linguistics, ACL 2020, pp 3707–3717

19. Hu J, Shi S, Huang H (2019) Combining external sentiment knowledge for emotion cause detection. In: Natural language processing and Chinese computing - 8th CCF international conference, NLPCC 2019

20. Cheng X, Chen Y, Cheng B, Li S, Zhou G (2017) An emotion cause corpus for chinese microblogs with multiple-user structures. ACM Trans Asian Low-Res Lang Inf Proc TALLIP 2017 17:1–19

21. Kim E, Klinger R (2018) Who feels what and why? annotation of a literature corpus with semantic roles of emotions. In: Proceedings of the 27th international conference on computational linguistics, COLING 2018, pp 1345–1359

22. Bostan LAM, Kim E, Klinger R (2020) Goodnewseveryone: a corpus of news headlines annotated with emotions, semantic roles, and reader perception. In: Proceedings of The 12th language resources and evaluation conference, LREC 2020

23. Wei P, Zhao J, Mao W (2020) 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, ACL 2020, pp 3171–3181

24. Tang H, Ji D, Zhou Q (2020) Joint multi-level attentional model for emotion detection and emotion-cause pair extraction. Neurocomputing 409:329–340

25. Song H, Zhang C, Li Q, Song D (2020) End-to-end emotion-cause pair extraction via learning to link. arXiv:2002.10710

26. Turcan E, Wang S, Anubhai R, Bhattacharjee K, Al-Onaizan Y, Muresan S (2021) Multi-task learning and adapted knowledge models for emotion-cause extraction. arXiv:2106.09790

27. Mikolov T, Corrado G, Kai C, Dean J (2013) Efficient estimation of word representations in vector space. In: Proceedings of the international conference on learning representations, ICLR 2013

**Weichun Huang**, a professor and a master tutor, he graduated from Jiangxi Normal University with a bachelor's degree and graduated from Huazhong University of Science and Technology with a master's degree. Presided over and participated in the national 863 project, 16 provincial and municipal scientific research projects, 1 provincial education reform project, and several horizontal projects; won 4 provincial and municipal project awards; published 25 papers in domestic and foreign academic journals and conferences, Editor-in-chief of 1 textbook. The main research direction: computer application, data mining.

**Yixue Yang** graduated from the School of Information Technology of Jingdezhen Ceramic University, majoring in digital media technology, and is currently a postgraduate of the School of Software, East China Jiaotong University. The main research direction is natural language processing.

**Xiaohui Huang** received the BEng degree and the master's degree from Jiangxi Normal University, Nanchang, P.R.China, in 2005 and 2008, respectively, and the PhD degree in the Harbin Institute of Technology, P.R.China, in 2014. Since Dec 2014, he has been at the School of Information Engineering Department, East China Jiaotong University, P.R. China, where he is currently an associate professor of Computer Science. His research interests are in the areas of machine learning, deep learning, and clustering algorithm.

**Zhiying Peng** graduated from Jiangxi University of Science and Technology with a major in Information and Computing Science. He is currently a postgraduate in Computer Science and Technology at the School of Information Engineering, East China Jiaotong University. The main research direction is traffic flow prediction based on deep learning.

**Liyan Xiong**, female, professor and postgraduate tutor of the School of Information Engineering. Main research directions: software engineering, database systems, decision support systems, data mining. Young and middle-aged backbone teachers in Jiangxi Province, head of computer application technology disciplines, review experts from the National Natural Science Foundation of China and the Ministry of Science and Technology, and science and technology review experts from Hubei, Fujian, Shandong and other provinces.