Research Article
Hierarchical Network Emotional Assistance Mechanism for Emotion Cause Extraction

Yu Wang, Bingjie Wei, Shuhua Ruan, Xingshu Chen, and Haizhou Wang

1School of Cyber Science and Engineering, Sichuan University, Chengdu 610065, China
2Cyber Science Research Institute, Sichuan University, Chengdu 610065, China
3National Computer Network Emergency Response Technical Team, Beijing 100029, China

Correspondence should be addressed to Shuhua Ruan; ruanshuhua@scu.edu.cn

Received 14 December 2021; Revised 17 January 2022; Accepted 1 February 2022; Published 14 March 2022

Academic Editor: Aleksandar Jevremovic

The fine-grained task of emotion analysis, emotion cause extraction, is a current research hotspot. It aims to discover the underlying reasons behind the emotional expression in texts. Most of the existing work regards the task as an independent text clause classification problem, ignoring the relationship between the clauses and failing to use the indicative relationship between emotional sentences and emotional cause sentences. The existence of these problems greatly affects the accuracy of the task. In this work, an emotion cause extraction method based on a hierarchical network emotional assistance mechanism is proposed. This method uses a hierarchical network composed of bidirectional gated recurrent units, attention mechanism, and graph convolutional networks to capture clause context information, deep semantic information, and structural information between clause neighborhoods. At the same time, by enhancing the emotional information representation of the graph convolutional network nodes, the clause features of the text emotional keywords are introduced into the discovery of candidate cause sentences. Thus, a model of the deep neural network combined with the emotional assistance mechanism is established. Compared with the existing methods, the model established in this paper has better classification performance on the Chinese emotion cause dataset.

1. Introduction

In recent years, with the rapid development of the Internet, people can express their views and experiences and share their views and attitudes towards current events anytime and anywhere, which makes the number of available texts with emotional tendencies increase sharply, and the text emotion analysis in natural language processing has attracted more and more attention. At present, there are many research tasks in text sentiment analysis, such as text sentiment summary, stance detection, sentiment element extraction, and emotion cause extraction. Among them, the task of emotion cause extraction (ECE) was first proposed by Lee et al. [1]. The goal is to trace the source of the emotion contained in the text and find the reasons that trigger the emotional tendency. Since ECE is a more fine-grained task in text sentiment analysis, emotional reasons can provide more details about emotional stimulation and expression, and this information is often very useful for product services and public opinion supervision. For example, the company pays attention to users’ feedback on products, to better understand the actual needs of users which can improve product performance in a targeted manner. Decision-makers not only know the public’s sentimentality towards the event, but also know the reasons for this attitude, so they can provide more effective solutions to control the trend of the event.

The example of emotion cause extraction is shown in Table 1. This text contains five clauses, which are composed of three parts: event background, emotional words, and emotional reasons. The orange sentence is an emotional sentence, which contains the emotional keyword “happily.” The emotional cause of “happily” is “her son was rated as an excellent soldier” in the blue sentence, and the sentence is the emotional reason sentence. The purpose of the ECE task is to identify the cause clauses that trigger emotion according to the emotional tendency of the text.
To solve the mentioned problems, a hierarchical network emotional assistance mechanism (HNEAM) for emotion cause extraction is proposed. First, a pretraining model of Bidirectional Encoder Representation from Transformers (BERT) is used to extract the semantic feature sequence of the text, and then the semantic information within the clause is captured through the bidirectional gated recurrent units and the attention mechanism. Then, the graph convolutional network is used to capture the semantic and structural information between the clause neighborhoods. In this process, the emotional assistance mechanism is used to encode the clause where the emotional keyword is located in the text, and the emotional sentence features in the same text are used to enhance the emotional information representation of the graph convolutional network nodes. Thus, the ECE model of the deep neural network combined with the emotional assistance mechanism is established. The main contributions of this paper are summarized as follows:

1. An emotion cause extraction model based on a hierarchical network emotional assistance mechanism is proposed. The model uses the relationship between text clauses, the deep semantic information and structural information between clause neighborhoods, and the indicative relationship between emotional sentences and emotional cause sentences to help find emotional cause sentences in text.

2. Based on the close relationship between emotional reasons and emotional information, this paper proposes an emotional assistance mechanism. This mechanism enhances the expression of emotional information of graph convolutional network nodes by using the characteristics of emotional sentences in the same text. Finally, the performance of ECE is improved.

3. The validity of the proposed model is verified on the Chinese emotion cause dataset. Finally, better performance of the task is obtained in 11 compared experiments, and the F1-score value is increased from 77.43% to 79.39%.

The rest of this paper is structured as follows. Section 2 mainly discusses the related work. Section 3 introduces the HNEAM model proposed in this paper for emotion cause detection. Section 4 describes our detailed experiments and evaluates the results. Finally, Section 5 summarizes the research ideas of this paper and proposes future work.

2. Related Work

The main purpose of ECE is to automatically find the direct or indirect causes that cause emotions to be generated or changed. In this section, we will introduce the related work of emotion cause extraction and graph convolutional network.

2.1. Emotion Cause Extraction. The task of ECE has attracted attention in recent years. The research methods mainly include rule-based methods, traditional machine learning methods, and deep learning methods.

2.1.1. Emotion Cause Extraction by Rule-Based Methods. Lee et al. [1] first proposed the emotion cause extraction task in 2010 and defined it as a word-level sequence annotation problem. Firstly, they manually constructed a Chinese ECE corpus; then, they analyzed the characteristics of the corpus and manually labeled the position of emotional words, emotional reasons, and event types of the text. Through the statistical analysis of the labeled corpus, they summarized nine linguistic rules from the results and found emotional reasons based on these rules.

Chen et al. [2], by analyzing the corpus constructed by Lee 1, proposed a multi-tag-based ECE method to analyze the relative position of emotional keywords and emotional causes in the text. When extracting the cause features, the two methods of automatic feature extraction and manual feature extraction are combined to ensure the completeness of the extracted features.

Neviarouskay et al. [3] constructed an emotional tagging corpus specific to 22 emotions in 2013. Based on the analysis of this corpus, they used the dependency syntactic analysis and the establishment of an external emotional knowledge corpus to assist in extracting the linguistic relationship
between eight types of emotional sentences and emotional reason sentences in the text.

Different from most methods based on statistics and rules, Li et al. [4] proposed a rule method based on knowledge transfer to find emotional causes by introducing knowledge in other fields such as sociology.

The above rule-based ECE methods have good controllability and interpretability, but the construction of the ruleset requires a lot of manual participation, and the coverage of rules is not high, which cannot cover all text features.

2.1.2. Emotion Cause Extraction by Traditional Machine Learning Method. Gui et al. [5] constructed the microblog emotion cause corpus and proposed a traditional machine learning method. They extracted 25 emotion cause matching rules from the labeled corpus. Based on these rules, they extracted the features that can be used by machine learning and then used the traditional machine learning methods of support vector machine and conditional random field to extract emotion causes.

Xu et al. [6] further explored the emotional cause discovery task based on rules and knowledge transfer. Since there was no publicly available dataset, they first constructed and labeled the emotional cause dataset. Then, on this basis, an ECE method based on event extraction and a multicore support vector machine framework is proposed, and the undersampling method is used to solve the problem of imbalance of the dataset.

Gui et al. [7] released a corpus of Chinese emotional causes based on Sina news in 2016, which is now regarded as the benchmark dataset for the ECE task. The dataset was annotated according to the emotion markup language published by the World Wide Web Consortium (W3C), and the ECE task is defined as the binary classification problem of clauses to judge whether each candidate clause is an emotional cause sentence. They used a syntactic tree to represent the text and proposed an event-driven method for finding emotional causes by improving the synonym of the traditional convolution kernel.

Methods based on traditional machine learning usually need to find effective features related to the expression of emotional reasons. Although such methods have achieved better experimental results than rule-based and statistical methods, they often need empirical guidance for feature extraction and screening and are sensitive to missing data. Moreover, it is impossible to complete the modeling of the sequence relationship between clauses in the text, and it is difficult to dig into the latent semantic information of the context in the text.

2.1.3. Emotion Cause Extraction by Deep Learning Method. In recent years, with the rapid development of deep learning technology, this technology has also been applied to the extraction of emotional causes.

Gui et al. [8] regarded ECE as a question and answer task and built a question answering (QA) system on the deep network. Emotional keywords are used as query words, and context clauses are input into the system as questions. The deep relationship between emotional query words and text reason candidate clauses is learned through the network.

Through the analysis of the corpus, Ding et al. [9] found that in addition to the text content, the location information and global label information of the text are also very important for the identification of emotional causes. They proposed a model based on a neural network architecture to encode and integrate the above three elements in a unified end-to-end manner.

Yu et al. [10] proposed a clause selection framework based on a hierarchical network, which detects emotional cause sentences from word-level, phrase-level, and sentence-level coding information. The framework not only considered the semantic information within the clause, but also captured the relationship between the contexts in the text.

Xia et al. [11] proposed a joint emotion cause extraction framework called RNN-Transformer Hierarchical Network (RTHN). RTHN encoded the words in clauses by a word-level encoder based on RNN and captured the relationship between text clauses by a clause-level encoder based on the transformer to realize synchronous modeling and classification of multiple clauses in the text. At the same time, the relative position information of the text and the global label information are encoded into the transformer method.

Li et al. [12] proposed a model based on a multi-attention neural network. The model encodes the clause via bidirectional long short-term memory and captures the mutual influences between the emotional clause and candidate cause clauses through multi-attention mechanism, so as to obtain better-distributed representations of the emotional expressions and clauses.

Wu et al. [13] presents an emotion cause extraction method that incorporates rule distillation within a hierarchical attention network. The hierarchical attention network uses position encoding and residual structure to capture the latent semantic relationships within the text clauses. With the help of knowledge distillation technology, relevant linguistic rules are introduced to guide the learning of neural networks.

Hu et al. [14] proposed a graph convolutional network over the interclause dependency to fuse the semantics and structural information. The model could automatically learn and select relevant clauses useful for the task. Thus, the model could capture remote information from semantic and structural information.

Aiming at the unbalanced distribution of datasets [15], Yan et al. [16] studied the existing ECE methods and found the dependence of the model on the relative position of clauses. They proposed a novel strategy to generate adversarial examples in which the relative position information is no longer the indicative feature of cause clauses. They also proposed a novel graph-based method to enhance the semantic dependencies between a candidate clause and an emotion clause by leveraging the commonsense knowledge.

2.2. Graph Neural Network. The graph convolutional network (GCN) was first proposed by Bruna [17]. Kipf and Welling [18] first applied GCN to node classification task. Since then, GCN has been widely used in various natural
3. Methodology

In this section, we first define the task of ECE and then propose the framework of a hierarchical network emotional assistance mechanism for emotion cause extraction.

3.1. Task Definition. ECE task is mainly to dig out the reasons behind the emotional information generated by documents through effective methods. The document and result label of the task are as follows:

Document: A passage text is composed of event background, emotional keywords, and emotional reason sentences. Since the ECE task is a clause-level task, this paper uses $d = \{c_1, c_2, \ldots, c_n\}$ to represent a text containing $n$ clauses, where $c_i$ means the $i$-th clause in the text. Each clause is composed of multiple words, so $c_i = \{w_{i1}, \ldots, w_{im}\}$ is used to indicate that the clause is composed of $m$ words.

Result label: The goal of ECE task is to predict whether a clause of the document is the cause of the affective tendency of the text. The prediction result of each clause in the text is the output of the model indicating whether the clause is an emotional reason sentence. If yes, it is represented by 1; otherwise it is represented by 0.

A specific example of the ECE task is shown in Table 2. The number of clauses in the text is 5. The position of sentence C3 is 0, which indicates the clause where the emotional keyword “happily” is located. The clauses with positions $-1$ and $-2$ indicate the clause before emotional sentence C3. The clauses with positions 1 and 2 indicate the clause after emotional sentence C3. The number indicates the relative distance from C3 sentence. The emotional cause clause is C2 sentence, so the result label value is 1, and the label of the other clauses is 0.

3.2. Hierarchical Network Emotional Assistance Mechanism. Aiming at solving the existing problems of the ECE task, this paper proposes an emotion cause extraction method based on hierarchical network emotional assistance mechanism. The overall framework of the model is shown in Figure 1. The first layer is the word-level encoder, which is composed of a BERT word embedding model, multiple bidirectional gated recurrent units (Bi-GRU) modules, and word-level interactive attention module. The second layer is the clause-level encoder. Bi-GRU is used to encode the vectors of the lower layer and learn contextual sequence information. The emotional sentences are internally learned through the Bi-GRU module and attention mechanism to complete the construction of the characteristics of the emotional assistance mechanism. The last layer is the neighborhood information encoder, which uses GCN to model the neighborhood information of clauses in the text and extract context features. Finally, the softmax layer is used for result classification.

3.3. Word-Level Encoder. The previous pretraining models such as word2vec and glove will be restricted by the one-way language model. They also limit the representation ability of the model, so that they can only obtain one-way context information. BERT uses a masked language model for pretraining and deep bidirectional transformer components to build the whole model. It finally generates deep bidirectional language representation that can fuse left and right context information. It is convenient that there are publicly available Chinese pretraining models that can be used directly. Therefore, the word embedding model in this paper selects BERT to extract the semantic features of the text.

BERT is a multilayer bidirectional converter model structure, which causes each word in the input sequence to be processed by the self-attention mechanism and obtains the new representation of the weighted sum of all word representations. Through the multilayer network structure, it can learn the representation containing more context interaction information. Therefore, when sequence $c_i = \{w_{i1}, \ldots, w_{im}\}$ is input into the BERT pretraining model, the words of each clause in the text will be mapped into 768 dimensional vector expressions. Finally, the output list of text is obtained, that is, the sequence of text semantic features.

$$c_i = \{s_{i1}, \ldots, s_{it}, \ldots, s_{im}\}. \quad (1)$$

Then, the semantic feature sequence inside the clause obtained by BERT is input into the word-level encoder composed of multiple Bi-GRUs. The GRU not only maintains the excellent effect of LSTM, but also has the advantages of relatively simple structure, less time cost, and better convergence. This paper adopts Bi-GRU neural network model. The words in each clause correspond to a Bi-GRU module. Bi-GRU is composed of forward GRU and backward GRU. This module accumulates context information for words in clauses and obtains the forward hidden state $h_t^f$ and backward hidden state $h_t^b$ of the $t$-th word in the $i$-th clause $s_i$. 
Table 2: Document and result label example of emotion cause.

| Position | Clause                                      | Label |
|----------|---------------------------------------------|-------|
| C1       | The lady lately received good news from the army | 0     |
| C2       | Her son was rated as an excellent soldier because of outstanding performance | 1     |
| C3       | The lady and her husband smiled happily looking at the good news | 0     |
| C4       | Now retired, she got back to the court      | 0     |
| C5       | Being responsible for the reception of public visits and legal questions and other work | 0     |

![Diagram of the framework of our HNEAM model](image)

Figure 1: The framework of our HNEAM model. The structure of the graph is given by a universal example; $\bar{v}_{ie}$ represents the node of the clause where the emotional keyword is located in the text, $\bar{v}_{ie-p}$ represents the first node with distance $\bar{v}_{ie}$ as $p$, and $\bar{v}_{ie-q}$ represents the last node with distance $\bar{v}_{ie}$ as $q$. In the connecting lines of node pairs, the solid lines represent the D1 edges, the dashed lines represent the D2 edges, and the alternately dotted lines represent the SL side.

\[
\begin{aligned}
\vec{h}_i &= \bar{G}_\text{RU}(s_i, \vec{h}_{i-1}), \\
\bar{h}_i &= \bar{G}_\text{RU}(s_i, \bar{h}_{i-1}).
\end{aligned}
\]

(2)

The state output of Bi-GRU is spliced by $\vec{h}_i$ and $\bar{h}_i$, that is, $h_i = [\vec{h}_i, \bar{h}_i]$. After the sequence training, the clause hidden state set $h_i = [h_i, \ldots, h_i, \ldots, h_{in}]$ is obtained.

In order to highlight the importance of the key information in the clause to the discovery of text reasons, a word-level attention mechanism is introduced to assign different probability weights to different word vectors. The input of attention layer is the output vector of Bi-GRU layer activation processing.

\[
u_{it} = \tanh(w_{it}^{(n)}h_{it} + b_{it}),
\]

(3)

\[
\alpha_{it} = \frac{\exp(u_{it}u_{iw})}{\sum\exp(u_{it}u_{iw})},
\]

(4)

\[
v_{it} = \sum_{i=1}^{t} \alpha_{it} h_{it}.
\]

(5)

In (3)-(5), $h_{it}$ is the output vector of the upper layer, $b_{it}$ is the bias coefficient, $w_{it}^{(n)}$ is the weight coefficient, and $u_{it}$ is the randomly initialized attention matrix. Through the attention mechanism, different weight probabilities are allocated and the product sum of each hidden layer state is accumulated. Finally, $v_{it}$ is the semantic expression of the clause.

3.4. Clause-Level Encoder. The word-level encoder can only capture the semantic information within the clause, so the clause-level encoder is used to learn the context information of the text and extract the features of the clause where the emotional keyword is located.

3.4.1. Contextual Information Extraction. In order to better learn the context information and the deep-seated feature extraction of the text, the Bi-GRU neural network is adopted to make the output of the current time related to the state of the previous time and the state of the next time. The forward hidden state $\vec{h}_i$ and backward hidden state $\bar{h}_i$ of the $i$-th clause in text $d$ are obtained through Bi-GRU neural network.

\[
\begin{aligned}
\vec{h}_i &= \bar{G}_\text{RU}(d, \vec{h}_{i-1}), \\
\bar{h}_i &= \bar{G}_\text{RU}(d, \bar{h}_{i+1}).
\end{aligned}
\]

(6)

The state output of Bi-GRU is spliced by $\vec{h}_i$ and $\bar{h}_i$, that is, $v_i = [\vec{h}_i, \bar{h}_i]$. After the sequence training, the clause hidden state set $v_i = [v_1, \ldots, v_i, \ldots, v_n]$ is obtained.
3.4.2. Emotional Feature Extraction. The clause in which the emotional keyword is located in the text has an indicative relationship to the emotional cause sentence; that is, the candidate cause sentence does not exist independently of the emotional sentence. As shown in Figure 2, the clause where the emotional keyword is located has a certain guiding role and unique directivity for the discovery of the cause clause. Therefore, this paper proposes an emotional assistance mechanism to mine this relationship.

In the ECE task, according to the rules established by the dataset, each text has only one emotion keyword and at least one reason sentence. Therefore, when constructing the emotion assistance mechanism, it is not necessary to consider the fusion of multiple emotion key sentences in a single text.

To better explore the characteristics of the clause where the emotional keyword is located, after obtaining the expression of the clause \( v_{IE} \) where the emotional keyword is located, the clause is first copied \( n \) times. Then, the vector is sent to the Bi-GRU network and the attention mechanism. This makes it better to learn the emotional features in the clause where the emotional keyword is located and obtain the final emotional feature vector \( E \).

\[
    h_{IE} = \text{BiGRU}\left[ v_{IE1}, \ldots, v_{IEq}, \ldots, v_{IE\text{En}} \right],
\]

\[
    u_{IE} = \tanh(\omega_{IE} h_{IE} + b_{ue}),
\]

\[
    \alpha_{IE} = \frac{\exp(u_{IE}u_{eIE})}{\sum_{i=1}^{n} \exp(u_{IE}u_{iE})},
\]

\[
    E = \sum_{i=1}^{n} \alpha_{IE} h_{IE}.
\]

In (8) and (9), \( b_{ue} \) is the bias coefficient, \( \omega_{IE} \) is the weight coefficient, and \( u_{eIE} \) is the randomly initialized attention matrix. Through the attention mechanism, different weight probabilities are allocated, and the product sum of each hidden layer state is accumulated. Finally, \( E \) is the emotional feature vector.

3.5. Emotional Assistance Implement of GCN. The clause-level encoder of the HNEAM model only learns the context order information of the clause. In order to enable the model to learn more information provided by the text, not limited to the sequential learning of the context, this paper uses GCN to model the neighborhood information of the text.

3.5.1. Node Emotional Information Enhancement. The node \( V_i \) of graph convolution neural network is composed of the feature vector \( \tilde{V}_i \) of each clause in the text and the emotion auxiliary vector \( E \) of the text. The purpose is to enhance the emotional information representation of graph convolution neural network nodes.

\[
    \tilde{V}_i = \tilde{V}_i + E.
\]

3.5.2. Neighborhood Information Coding. According to the relevant information of the dataset provided by Gui et al. [7] (Table 3), it can be seen that 85.5% of the corpus’s cause sentences are located in 1 relative position of the emotional sentence, and 95.43% of the corpus’s cause sentences are located in 2 relative positions of the emotional sentence. Therefore, the positional structure information between the emotional cause clauses and the clause where the emotional keyword is located is the key information in the task of ECE. In order to make full use of the relative position information, the construction of the graph edges in this paper refers to Chen et al. [22]. Three types of edges are defined to represent the influence of the current node on the neighbor nodes, namely, the SL, D1, and D2 edges:

(1) SL edge is self-looping edge of node.

(2) D1 edge indicates the edge with relative position of 1 between two clause nodes.

(3) D2 edge indicates the edge with relative position of 2 between two clause nodes.

Through the construction of three edges, the neighborhood information of the clause is allowed to be transmitted through three edges to better extract the context information. Different weight matrices are used for context propagation according to different edges. In this paper, two-layer GCN is used to capture the neighborhood information of nodes through two graph transformations.

\[
    g^1_i = \sigma \left( \frac{1}{z} \sum_{k \in D1} v_k W^1_{D1} + \frac{1}{z} \sum_{n \in D2} v_n W^1_{D2} + \frac{1}{z} v_i W^1_{SL} \right),
\]

\[
    g^2_i = \sigma \left( \sum_{k \in D1} g^1_k W^2_{D1} + \sum_{n \in D2} g^1_n W^2_{D2} + g^1_i W^2_{SL} \right).
\]

In the first transformation formula (12), \( W^1_{D1}, W^1_{D2}, \) and \( W^1_{SL} \) represent the weight matrix between the node and the three edges; \( z \) represents the node degree of the node; and \( \sigma \) represents the RELU activation function. The second transformation formula (13) uses the first output to obtain the representation \( g^2_i \) of node \( V_i \). \( W^2_{D1}, W^2_{D2}, \) and \( W^2_{SL} \) are the weight matrix updated after the first transformation.

3.6. Output Layer. The essence of ECE task is classification. Therefore, this paper directly passes the representation of clauses through classifier and softmax layer to obtain the final output probability distribution \( O_i \). In (14), \( W \) is the weight matrix and \( b \) is the offset vector. Finally, the output is calculated by linear transformation.

\[
    O_i = \text{softmax}(W \cdot g^2_i + b).
\]

3.7. Loss Function. The model proposed in this paper uses the backpropagation method to train the model end-to-end and uses the cross entropy loss function as the function of the optimization objective. The loss function is defined as follows.
In (15), $y_i$ is the real label of the clause and $\tilde{y}_i$ is the predicted value of the clause. In this paper, the stochastic gradient descent method is used to optimize the objective function, and the dropout technique is used to alleviate the problem of overfitting.

\[ \text{Loss} = -\sum_{i=1}^{\text{num}} y_i \cdot \log(\tilde{y}_i). \]  

(15)

4. Experiments and Analysis

In this section, we first introduce the basic information of the dataset used in the experiment, then describe the evaluation indicators and experimental settings of the experiment, carry out the experiment, and finally analyze the results of the model proposed in this paper.

4.1. Dataset

We select the Chinese emotion cause dataset based on Sina news [7]. This dataset is the largest Chinese dataset for the ECE task. The corpus has 2105 texts and 11799 clauses, including 2167 emotional cause sentences. Each text contains an emotional sentence and one or more emotional cause sentences. The details of the dataset are shown in Table 4.

| Number | Instances | Clauses | Emotion causes | Doc with 1 cause | Doc with 2 causes | Doc with 3 causes |
|---|---|---|---|---|---|---|
| Number | 2105 | 11799 | 2167 | 2046 | 56 | 3 |

4.2. Evaluation Metrics

We divided the dataset into training data and test data in a ratio of 9:1, respectively. In order to obtain statistically reliable results, the experiment was repeated 10 times, and the average results were taken to evaluate the performance of the model. The experiment uses precision, recall, and F1-score to verify the method proposed in this paper.

In this article, positive samples represent reason sentences in the text, and negative samples represent non-cause sentences in the text. The confusion matrix is shown in Table 5, where TP represents the number of samples with the correct prediction of causal sentences, TN represents the number of samples with non-cause sentences predicted as non-cause sentences, FP represents the number of samples with the wrong prediction of causal sentences, and FN represents the number of samples with non-cause sentences predicted as causal sentences.

The calculation method of each indicator is as follows.

\[ \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \]

\[ \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \]

\[ F_1 - \text{score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \]  

(16)

4.3. Training Details

We use a Chinese pretraining model based on BERT, where the number of hidden layers is set to 768 layers, the number of network layers is 12 and the word embedding dimension is 768. Other parameters of our model are listed in Table 6.

The numbers of the hidden BI-GRUs and the hidden GCN units are the optimal values obtained by fine-tuning the model based on the experimental settings of Chen [22]. Each training batch contains 32 texts. If the batch size is too small, the training time of the model increases and the gradient oscillation is serious; if the batch size is too large, the final convergence accuracy fails to reach the optimal value. With the increase of the learning rate, the loss of the model will first decrease and then increase. Therefore, the optimal learning rate is selected from the region with the least loss. Through the experiment, it is found that the model in this paper has the best performance when the learning rate is set at 0.001.
4.4. Result Analysis. In this section, we first verify the superiority of the HNEAM model proposed in this paper through comparative experiments, then show the influence of the optimization strategy inside the HNEAM model on the ECE task, and analyze the results.

4.4.1. Comparison with Existing Methods. To verify the effectiveness of the HNEAM model, based on the dataset described in 4.1, the HNEAM model is compared with a variety of benchmark methods. The specific conditions and parameters of the benchmark method in the experiment are described as follows:

| Method | Parameter | Value |
|--------|-----------|-------|
| Hidden BI-GRUs | 300 |
| Hidden GCN units | 300 |
| Batch size | 32 |
| Learning rate | 0.001 |

The learnable parameters (including all weight matrices and bias vectors) of the model are randomly initialized by a uniform distribution of $U(-0.01, 0.01)$. Moreover, the model is based on the Adam optimizer training.

Table 6: The experiment settings of our models.

According to the performance comparison of the ECE methods in Table 7, the HNEAM model proposed in this paper obtains the best performance. Firstly, in the rule-based methods, the accuracy of RB is high, but the recall rate is low. The CB method has a high recall rate, but the accuracy rate is very low. The difference between accuracy and recall obtained by traditional machine learning methods is relatively stable and higher than those of rule-based methods. It can be seen that the methods based on traditional machine learning are better than the methods based on rules and external knowledge introduction. This is because, through complex feature engineering, effective features related to emotional reasons can be extracted to obtain better performance. However, the HNEAM model proposed in this paper is superior to all traditional machine learning methods. Finally, we compare the method with the methods based on the neural network proposed by predecessors. Compared with the rule-based and the traditional machine learning methods, the performance of the hierarchical network and attention-based models such as RTHN and MANN has been significantly improved. That is, the methods based on the neural network can not only obtain the semantic features of the text sequence, but also capture the relationship between clauses in the text. These methods ignore the remote dependence between clauses in the text, resulting in the loss of information. FSS-GCN and KAG methods use the graph neural network model to effectively capture the clause information in the text, which improves the performance of the model. Nevertheless, the HNEAM captures the characteristics of text structure information and the characteristics closely related to emotional reasons and emotional information, so it achieves better performance.

In summary, the accuracy of the HNEAM model proposed in this paper is significantly higher than that of other existing methods. It proves that the HNEAM model reduces the information loss caused by the long-distance dependence of clauses due to the learning of the semantic and structural information of the text and can effectively complete the task of ECE. The recall rate is also higher than that of the previous optimal RTHN model, and the value increased by 0.0075. This shows that in the case of unbalanced dataset categories, the model uses GCN and assigns weights to different nodes, which can effectively increase the probability that the clause is predicted correctly. Due to the increase of $P$ and $R$, F1 is also significantly higher than that of the previous method. F1 is 0.0196 higher than KAG model, which further proves the superiority of HNEAM model.

(vi) FSS-GCN: Hu et al. [14] proposed an emotion cause extraction method based on GCN, which could automatically learn and select relevant clauses useful for the task.

(vii) KAG: Yan et al. [16] proposed a novel graph-based method to explicitly model the emotion triggering paths by leveraging the commonsense knowledge to enhance the semantic dependencies between a candidate clause and an emotion clause.

(i) RB: Lee et al. [1] proposed an emotion cause extraction method based on rule matching.

(ii) CB: Russo et al. [23] proposed an emotion cause extraction method based on common sense matching.

(i) RB + CM + ML: support vector machine classifier is used based on CB method and RB method [2].

(ii) Multikernel: Gui et al. [7] proposed a method on multicore support vector machine classifiers.

(i) PAE-DGL: Ding et al. [9] proposed a prediction model that integrates relative position information and dynamic global labels.

(ii) HCS: Yu et al. [10] proposed an emotional cause extraction method based on CNN-RNN hierarchical network.

(iii) RTHN: Xia et al. [11] proposed a hierarchical emotion cause extraction method based on RNN and transformer.

(iv) MANN: the model proposed by Li et al. [12] uses a multi-attention mechanism to capture the interaction between emotional clauses and candidate causal clauses to obtain better clause representation.

(v) RD-HAN: Wu et al. [13] proposed a method of finding emotional cause combined with rule distillation.
4.4.2. Ablation Study. In order to evaluate the effectiveness of the model proposed in this paper, ablation experiments are performed on parts of the structure of the model HNEAM. The ablation methods are as follows.

(i) HNEAM-noEmo: in this method, the HNEAM model lacks emotional assistance mechanism.

(ii) HNEAM-noGCN: in this method, the HNEAM model lacks neighborhood information encoder.

It can be seen from Table 8 that after the HNEAM model removes the emotional assistance mechanism, the values of the three evaluation indicators are reduced. The R and F1 are reduced by 0.0136 and 0.0092, respectively. This proves that the emotional assistance mechanism can fully explore the guiding effect of emotional sentences on emotional reason sentences, increase the proportion of correct reason sentences in the text, and increase the stability of the model. When the domain information encoder was removed from the HNEAM model, the values of the three evaluation indicators all dropped significantly. This is because the removal of the neighborhood information encoder not only removes the GCN module, but also removes the emotional assistance mechanism, which exists in the nodes of the GCN. At the same time, this also illustrates the importance of the deep semantic information and structural information between clause neighborhoods in the task of ECE.

4.4.3. Impact of Word Embedding. In order to verify the superiority of the semantic sequence obtained by BERT model, different word embedding methods are introduced to compare. The random initialization method initializes the word vector randomly for each word in the text according to a certain probability distribution. Word2vec is a popular word embedding method. The experimental results are shown in Figure 3. The effect of the HNEAM model is better than the above two traditional word embedding methods, indicating the effectiveness of BERT’s Chinese pretraining model in fusing the semantic information of the text.

4.4.4. Discussion on \( n_c \) for Emotional Assistance Mechanism. In Section 3.4.2, it is stated that after obtaining the expression of the clause \( v_{IE} \) where the emotional keyword is located, the clause is first copied \( n_c \) times, and then the final emotional feature vector \( E \) is obtained through the emotional assistance mechanism. The selection of \( n_c \) is also the value with the best comprehensive effect of the evaluation index after experimental comparison. In this paper, the \( n_c = 1/0.5n/1.5n/2n/3n \) numerical values are selected in turn for experimental comparison, and the results are shown in Figure 4. It can be seen from the figure that with the increase of \( n_c \), the neural network learns more information. However, the learning content is limited, so the accuracy rate increases first and then gradually stabilizes with the increase of \( n_c \). The recall rate begins to decrease when the value of \( n_c \) increases to a certain value, indicating that when \( n_c \) reaches the value of \( n \), the proportion of the reason sentences in the corpus that are predicted correctly is the largest. Therefore, by combining the two evaluation indexes, when \( n_c = n \), the addition of emotional assistance mechanism makes the model achieve the best performance.

4.4.5. Different Emotional Assistance Mechanism. The emotional assistance mechanism in the HNEAM model proposed in this paper mainly uses the Bi-GRU network combined with the attention mechanism. This mechanism can pay attention to the key information of clauses where emotional keywords are located and conduct sufficient learning. In order to verify the effectiveness of this method, the direct assistance method of emotion and the method of bidirectional gated recurrent unit are used to explore the method.

(i) HNEAM-Di: in this method, after the text is learned by the clause-level encoder, the feature vector \( v_{IE} \) of the clause where the emotional keyword is located is directly spliced to the node of the constructed graph neural network.

| Methods       | P    | R    | F1   |
|---------------|------|------|------|
| RB            | 0.6712 | 0.5247 | 0.5890 |
| CB            | 0.2672 | 0.7130 | 0.3887 |
| RB + CB + ML  | 0.5921 | 0.5307 | 0.5597 |
| Multikernel   | 0.6588 | 0.6927 | 0.6752 |
| PAE-DGL       | 0.7619 | 0.6908 | 0.7242 |
| HCS           | 0.7388 | 0.7154 | 0.7269 |
| RTHN          | 0.7697 | 0.7662 | 0.7677 |
| MANN          | 0.7843 | 0.7587 | 0.7706 |
| RD-HAN        | 0.7706 | 0.7203 | 0.7446 |
| FSS-GCN       | 0.7861 | 0.7572 | 0.7714 |
| KAG           | 0.7912 | 0.7581 | 0.7743 |
| HNEAM         | 0.8164 | 0.7737 | 0.7939 |

Table 8: Comparison of ablation experiments.

| Methods       | P    | R    | F1   |
|---------------|------|------|------|
| HNEAM         | 0.8164 | 0.7737 | 0.7939 |
| HNEAM-noEmo   | 0.8154 | 0.7585 | 0.7848 |
| HNEAM-noGCN   | 0.7438 | 0.6669 | 0.7024 |

Figure 3: Performance comparison of different word embedding methods.
(ii) HNEAM-Bi: in this method, after the text is learned by the clause-level encoder, the feature vector of the clause where the emotional keyword is located is copied $n$ times to align with the total number of sentences in the text. Furthermore, only the Bi-GRU network is used to preserve the important features of the clause.

From the analysis of the results in Table 9, it can be seen that the values of the direct assistance method of emotion to the graph nodes is lower than those of the method of learning the emotional sentence features again. It shows that only learning simple context information cannot effectively extract the characteristics of the clause where the emotional keyword is located. The $P$ and $F1$ of the experimental results using Bi-GRU network combined with attention mechanism are higher than those using Bi-GRU network only, which shows that Bi-GRU network combined with attention mechanism can more effectively mine the important features of clauses when learning the features between the same clauses. This proves the effectiveness of the emotional assistance mechanism method proposed in this paper.

5. Conclusion and Future Work

A method based on hierarchical network emotional assistance mechanism for ECE tasks is proposed. By constructing a hierarchical network to encode the deep semantic information within the text clauses and between the clauses neighborhoods, the relationship between multiple clauses in the text can be effectively learned. In addition, in order to solve the problem of not using the indicating relationship of the emotional sentence to the emotional cause sentence, the emotional assistance mechanism is proposed. This mechanism uses the clause features of emotional keywords to enhance the emotional representation of graph convolutional neural network nodes, thereby assisting the discovery of emotional reason sentences. The experimental results on Chinese emotion cause dataset show that the method in this paper achieves the best performance known in accuracy, recall, and $F1$.

However, during the research process, it was also found that the Chinese annotation corpus for the ECE task only contains news texts, so the transferability of the model proposed in this paper cannot be verified in other texts (such as social network dataset). At the same time, due to the lack of labeled corpus, the model is easy to overfit, which also has an impact on the design and training of the model. In the follow-up work, we will consider constructing a corpus with topical information and dig deeper into the important emotional cause of topic events. Aiming at solving the problems of feature selection and hyperparameter optimization involved in emotion cause extraction, swarm intelligence approaches will be considered [24–28].

Data Availability

The data supporting this paper were obtained from previously reported studies and datasets, which are explained in Section 4.1.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 61802271, 61802270, and U19A2081), the Joint Research Fund of China Ministry of Education and China Mobile Company (no. CM20200409), and the Science and Engineering Connotation Development Project of Sichuan University (no. 2020SCUNG129).

References

[1] Y. M. S. Lee, Y. Chen, and C.-R. Huang, “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, CA, USA, June 2010.
[2] Y. Chen, Y. M. S. Lee, S. Li, and C.-R. Huang, “Emotion cause detection with linguistic constructions,” in Proceedings of the 23rd International Conference on Computational Linguistics, pp. 179–187, Beijing, China, August 2010.
[3] A. Neviarouskaya and M. Aono, “Extracting causes of emotions from text,” in Proceedings of the Sixth International Joint Conference on Natural Language Processing, pp. 932–936, Nagoya, Japan, October 2013.
[4] W. Li and H. Xu, “Text-based emotion classification using emotion cause extraction,” Expert Systems with Applications, vol. 41, no. 4, pp. 1742–1749, 2014.
[5] G. Lin, Y. Li, X. Ruieng, L. Bin, L. Qin, and Z. Yu, “Emotion cause detection with linguistic construction in Chinese weibo
text,” in Natural Language Processing and Chinese Computing, pp. 457–464, Springer, Berlin, Germany, 2014.

[6] X. Ruifeng, H. Jianan, L. Qin, W. Dongyn, and G. Lin, “An ensemble approach for emotion cause detection with event,” Extraction and Multikernel Smrs, vol. 22, no. 6, pp. 646–659, 2017.

[7] L. Gui, D. Wu, R. Xu, Q. Lu, and Y. Zhou, “Event-driven emotion cause extraction with corpus construction,” in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 1639–1649, TX, USA, November 2016.

[8] L. Gui, J. Hu, Y. He, R. Xu, Q. Lu, and J. Du, “A question answering approach to emotion cause extraction,” in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1593–1602, Copenhagen, Denmark, September 2017.

[9] Z. Ding, H. He, M. Zhang, and R. Xia, “From independent prediction to reordered prediction: integrating relative position and global label information to emotion cause identification,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 6343–6350, HI, USA, January 2019.

[10] X. Yu, W. Rong, Z. Zhang, Y. Ouyang, and Z. Xiong, “Multiple level hierarchical network-based clause selection for emotion cause extraction,” IEEE Access, vol. 7, pp. 9071–9079, 2019.

[11] R. Xia, M. Zhang, and Z. Ding, “RTHN: a rnn-transformer hierarchical network for emotion cause extraction,” in Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, pp. 5285–5291, Macao, China, August 2019.

[12] X. Li, S. Feng, D. Wang, and Y. Zhang, “Context-aware emotion cause analysis with multi-attention-based neural network,” Knowledge-Based Systems, vol. 174, pp. 205–218, 2019.

[13] W. Jipeng, B. Jianzhu, L. Gongqiang, and X. Ruifeng, “Emotion cause extraction using rule distillation,” Journal of Tsinghua University, vol. 60, no. 5, pp. 422–429, 2020.

[14] G. Hu, G. Lu, and Y. Zhao, “FSS-GCN: a graph convolutional networks with fusion of semantic and structure for emotion cause analysis,” Knowledge-Based Systems, vol. 212, no. 1, 2021.

[15] J. Ding and M. Kejriwal, “An Experimental Study of the Effects of Position Bias on Emotion Cause Extraction,” 2020.

[16] H. Yan, L. Gui, G. Pergola, and Y. He, “Position bias mitigation: a knowledge-aware graph model for emotion cause extraction,” in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, vol. 1, pp. 3364–3375, Bangkok, Thailand, August 2021.

[17] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, “Spectral Networks and Locally Connected Networks on Graphs. International Conference on Learning Representations,” 2014, https://arxiv.org/abs/1312.6203.

[18] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in Proceedings of the 5th International Conference on Learning Representations, Tou- lon, France, April 2017.

[19] D. Marcheggiani and I. Titov, “Encoding sentences with graph convolutional networks for semantic role labeling,” in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, September 2017.

[20] C. Zhang, Q. Li, and D. Song, “Aspect-based sentiment classification with aspect-specific graph convolutional networks,” in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 4567–4577, Hong Kong, China, November 2019.

[21] D. Ghosal, N. Majumder, S. Poria, N. Chhaya, and F. A. Gelbukh, “Dialoguegcn: a graph convolutional neural network for emotion recognition in conversation,” in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 154–164, Hong Kong, China, November 2019.

[22] Y. Chen, W. Hou, S. Li, C. Wu, and X. Zhang, “End-to-end emotion-cause pair extraction with graph convolutional network,” in Proceedings of the 28th International Conference on Computational Linguistics, pp. 198–207, Barcelona, Spain, September 2020.

[23] I. Russo, T. Caselli, F. Rubino, E. Boldrini, and P. Martinez-Barco, “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 2.011), pp. 153–160, Portland, Oregon, June 2011.

[24] S. Yildirim, Y. Kaya, and F. Killiç, “A modified feature selection method based on metaheuristic algorithms for speech emotion recognition,” Applied Acoustics, vol. 173, 2021.

[25] N. Bacanin, T. Bezdan, K. Venkatachalam et al., “Artificial Neural Networks Hidden Unit and Weight Connection Optimization by Quasi-Reflection-Based Learning Artificial Bee colony Algorithm,” IEEE Access, vol. 9, Article ID 169155, 2021.

[26] M. Zivkovic, N. Bacanin, K. Venkatachalam et al., “Covid-19 Cases Prediction by Using Hybrid Machine Learning and Beetle Antennae Search Approach,” Sustainable Cities And Society, vol. 66, 2021.

[27] Z. Zhang, “Speech feature selection and emotion recognition based on weighted binary cuckoo search,” Alexandria Engi- neering Journal, vol. 60, no. 1, pp. 1499–1507, 2021.

[28] S. Nematzadeh, F. Kiani, M. T. Afshar, and N. Aydin, “Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases,” Computational Biology and Chemistry, vol. 97, Article ID 107619, 2022.