Pair-Based Joint Learning with Relational Graph Convolutional Networks for Emotion-Cause Pair Extraction

Anonymous ACL submission

Abstract

Emotion-cause pair extraction (ECPE) aims to extract the emotion clauses and the corresponding cause clauses, which have recently received more attention. Previous methods sequentially encode features with a specified order, which first encode the emotion and cause features for clause extraction and then combine them for pair extraction, leading to an imbalance in inter-task feature interaction where features extracted later have no direct contact with the former. To this end, we propose a novel joint encoding network, which generates pairs and clauses features simultaneously in a joint feature learning manner to model the causal relationship from clauses. Specifically, from a multi-relational perspective, we construct a heterogeneous undirected graph and apply the Relational Graph Convolutional Network (RGCN) to capture the complex relationship between clauses and the relationship between pairs and clauses. Experimental results show that our model achieves state-of-the-art performance on the Chinese benchmark corpus.

1 Introduction

Emotion cause extraction (ECE) is a kind of emotion analysis task which is first proposed by Lee et al. (2010) and has developed for a long time. ECE extracts the cause for the input document and certain emotion labels. However, emotions in the documents need to be annotated in advance, which requires human involvement and costs lots of time (Xia and Ding, 2019; Ding et al., 2020a). Hence, Xia and Ding (2019) proposes a new task called emotion-cause pair extraction (ECPE). Given a document as the input, ECPE extracts the clauses which express emotions and their corresponding clauses which express causes (as shown in Figure 1). Intuitively, ECPE is much more challenging because the clauses classification task and the pairs matching task need to be completed simultaneously.

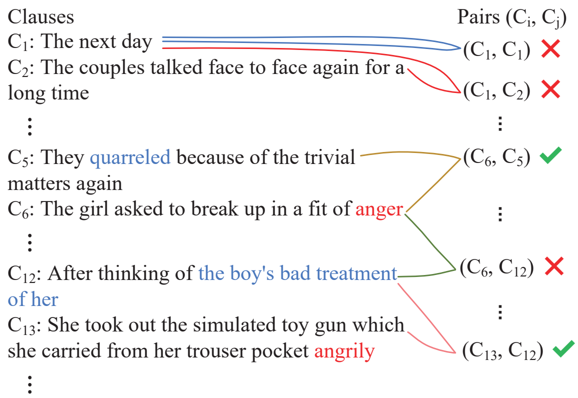

Figure 1: An example document from the ECPE corpus where c_i represents the emotion clause and c_j represents the cause clause in pair. The words in red are the keywords about emotion and the words in blue are about cause.

For ECPE, Xia and Ding (2019) first proposes a two-stage method. However, the two-stage method may cause the problem of error propagation. To solve this problem, most works use end-to-end methods (Ding et al., 2020b; Chen et al., 2020c; Singh et al., 2021). Most of them use sequential encoding, in which their task-specific features are learned sequentially in a predefined order. Specifically, following Wei et al. (2020), ECPE contains two auxiliary tasks which are emotion clause extraction (EE) and cause clause extraction (CE). Usually, the previous works first model the clauses for EE and CE and then model the pairs for ECPE.

However, the sequential encoding makes the information only flow from clauses to pairs, but can not from pairs to clauses, resulting in different amount of information exposed to pairs and clauses. Besides, the sequential encoding only considers the intra-relationship within pairs or clauses while ignoring the inter-relationship between them. Specifically, the causal relationship (Chen et al., 2020a) between emotion and cause clauses in pairs is ignored, which is a decisive factor that judgments whether emotions and causes match. For example, in Figure 1, c_6 and c_{13} both express anger, and c_{12} is cause clause. However, (c_{13}, c_{12}) is a pair but (c_6, c_{12}) is not. If we separately model the pairs and clauses, the lack of relationship information between these two clauses will make it difficult for the model to judge this situation.

To address the above issues, we propose a novel joint encoding method, which simultaneously generates pairs and clauses features in a joint feature learning manner. Specifically, we model the inter-relationship between pairs and clauses, in which a pair only interacts with the corresponding two clauses. It is conducive to learning pair representation and modeling the causal relationship from clauses and prevent interference from irrelevant information. Meanwhile, the key information about emotion and cause clauses is different. Therefore, different features should be extracted from these two clauses. Considering these complicated relationships, we construct a heterogeneous undirected graph and apply Relational Graph Convolutional Networks (RGCN) on it, which includes four kinds of nodes and four kinds of edges, utilizing different strategies to connect the nodes. Thus, it can make the information flow between emotion clauses and emotion clauses, between emotion clauses and pairs, etc., more efficient.

The main contributions are as follows:

- We propose a novel method to jointly encode the clauses and pairs for ECPE, helping the pairs learn the causal relationship between the two clauses during the encoding process.
- We propose an RGCN framework to model the complicated relationship between pairs and clauses. Different edges in the RGCN help the pairs or clauses extract more targeted information, improving the efficiency of the information flow.
- Experiments on ECPE benchmark corpus demonstrate that our model is state-of-the-art. Furthermore, some other experiments are performed to verify the effectiveness of our method.

2 Related Work

Our work is based on the emotion cause pair extraction (ECPE) task and the Relational Graph Convolutional Network (RGCN), which are developing rapidly recently.

2.1 Emotion-Cause Pair Extraction

Xia and Ding (2019) proposes ECPE task and uses a pipeline framework which first extracts the emotion and cause clauses then matches them as pairs for prediction. Due to the error propagation problem, Wei et al. (2020) proposes a unified framework which uses Graph Convolution Networks to encode the emotion and cause clauses in the same representations. However, it does not model the pairs, which makes the pairs lack contextual information. Furthermore, Ding et al. (2020a), Ding et al. (2020b) and Chen et al. (2020c) build encoders for pairs and clauses separately, which first model clauses and then concatenate them as pairs. Considering the symmetric relation between emotion clauses and cause clauses, Cheng et al. (2020) uses a local search strategy for the clauses which are predicted as emotion clauses or cause clauses.

On the other hand, Yuan et al. (2020) and Fan et al. (2021) design a novel cause-pivoted tagging scheme with a local window to predict the distance to the corresponding emotion clauses. Further, Chen et al. (2020b) uses a more fine-grained tagging scheme which combines emotion tagging and cause tagging with emotion labels separately. Finally, as another type of method, Fan et al. (2020) uses a transition-based method to solve this task.

However, these sequential encoding methods make the inter-task feature interaction unbalanced. Specifically, the features of pairs can not contact with clauses. In this paper, we will deal with this problem by joint learning network.

2.2 Relational Graph Convolutional Network

To directly model the graph-structured data, Kipf and Welling (2017) proposes the Graph Convolutional Network (GCN). However, a graph usually consists of multiple types of nodes and edges. For example, the knowledge graph has different predicates to indicate different relationships. Thus, using GCN to model this complex relationship is inappropriate. To solve this problem, Schlichtkrull et al. (2018) proposes the Relational Graph Convolutional Network (RGCN), utilizing different edges in a graph to model different relationships.

Recently, considering the powerful performance and modeling capabilities of RGCN, many works utilize it in their methods. For instance, Zhou et al. (2020) uses RGCN to encode the different relation semantics in knowledge graphs. Furthermore, Ishiwatari et al. (2020) employs RGCN to model

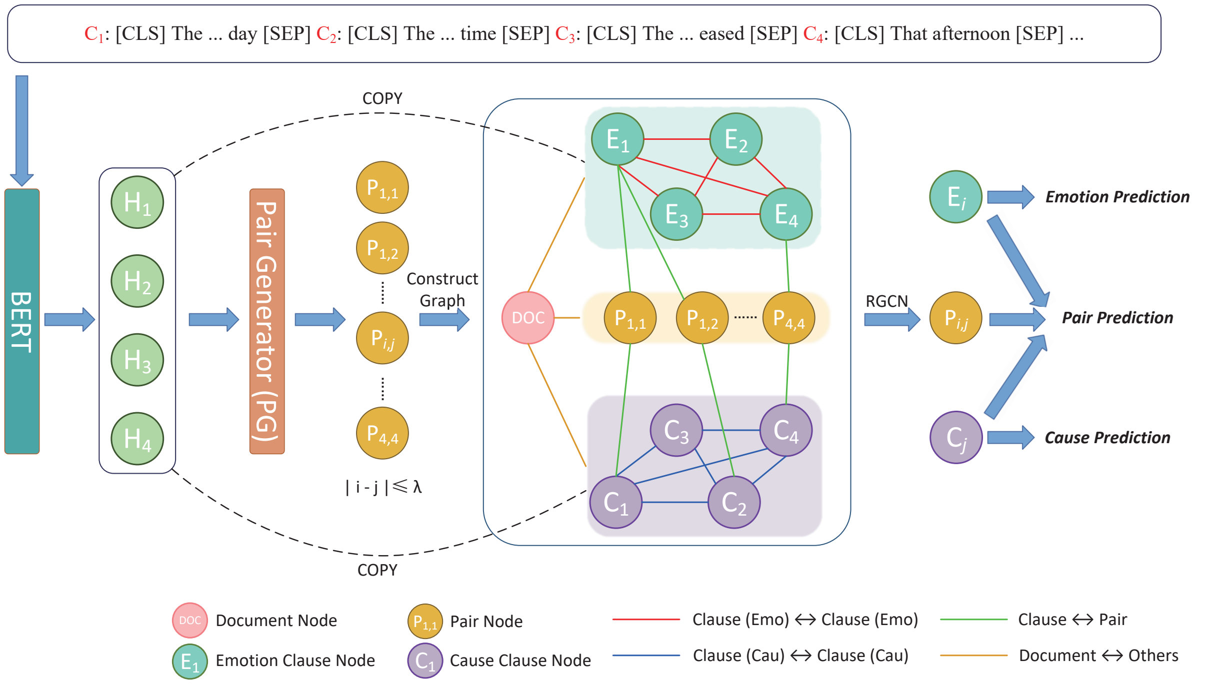

Figure 2: The overall architecture of the proposed model. First, the clauses are input into the pre-trained BERT to get the representations. Then, the representations of the pair are obtained by the pair generator. Next, we construct a heterogeneous undirected graph with the emotion clause nodes, cause clause nodes, pair nodes, and document node. Finally, after applying RGCN, we use the last layer's representations of the node for prediction.

the different relationships between speakers and time in conversation. Finally, Zeng et al. (2020) proposes an RGCN-based method to model the intra-entity edge, inter-entity edge, and document edge in the document-level relation extraction task.

3 Task Definition

Given a document $D = (c_1, c_2, \ldots, c_N)$ of N clauses and the i-th clauses $c_i = (w_1^i, w_2^i, \ldots, w_M^i)$ of M words, ECPE task aims to extract all the emotion-cause pairs in D:

$$P = \{\ldots, (c_i, c_j), \ldots\} \quad (1 \leq i, j \leq N) \tag{1}$$

where c_i and c_j represent the emotion clause and corresponding cause clause in pairs.

Meanwhile, ECPE has two auxiliary tasks which are emotion clauses extraction and cause clauses extraction. A clause c_i is emotion clause if any pair (c_i, c_j) is established, which can be defined as follow:

$$y_i^{emo} = \begin{cases} 1, & if \quad \exists c_j \in D, (c_i, c_j) \in P \\ 0, & otherwise \end{cases} \tag{2}$$

where $y_i^{emo} = 1$ means c_i is the emotion clause. The extraction of cause clauses is the same as emotion clauses.

4 Approach

In this section, we mainly describe our method, which encodes the pairs and clauses simultaneously and models the causal relationship from clauses in Relational Graph Convolutional Network (RGCN). The overall structure of our model is shown in Figure 2

4.1 Pair Generator

Following Wei et al. (2020), given a document $D = (c_1, c_2, \ldots, c_N)$ consisting of N clauses, we feed D into pre-trained BERT (Devlin et al., 2019). Specifically, we add a token [CLS] at the beginning and a token [SEP] at the end for each clause and concatenate all the clauses together as input. Finally, we use the representation of token [CLS] as the representation of the corresponding clause. Hence, the document with N clauses can be represented as:

$$H = \{h_1, h_2, \ldots, h_N\} \tag{3}$$

where $h_i \in \mathbb{R}^d$ and d is the hidden size of BERT.

To obtain the representations of pairs, we apply the Pair Generator (PG). Specifically, we concatenate the corresponding two clauses and project them with a learnable relative position embedding:

$$p_{ij} = W_p[h_i, h_j] + b_p + r_{i-j} \tag{4}$$

where $p_{ij} \in \mathbb{R}^d$ represents the pair consisting of c_i as an emotion clause and c_j as a cause clause, $W_p \in \mathbb{R}^{d\times 2d}$ and $b_p \in \mathbb{R}^d$ are learnable parameters, $r_{i-j} \in \mathbb{R}^d$ is the relative position embedding, and $[,]$ denotes the concatenating operation. In addition, following Wei et al. (2020), we set a hyperparameter λ as the local window ($|i-j| \leq \lambda$) to limit the number of pairs.

4.2 Pair-based Joint Encoder

To balance the interaction of pairs and clauses and capture the causal relationship in pairs, we construct a heterogeneous undirected graph. It can deal with the complex relationship between pairs and clauses as well as the relationship between clauses efficiently.

The graph has four kinds of nodes: emotion clause nodes, cause clause nodes, pair nodes, and document node. Intuitively, the emotion information and cause information in a clause are contained in different words. Hence, we separately use two kinds of nodes to represent the emotion clause and the cause clause. Meanwhile, we add a document node to the graph, which can provide some global information (e.g., topics) for the other nodes and interact with others like a pivot.

Moreover, there are mainly four kinds of inter-node edges in our graph:

- **Clause(Emotion)-Clause(Emotion) Edge:** All emotion clause nodes are fully connected, using this edge. It can help the emotion clause nodes interact with others to get the contextual information.

- **Clause(Cause)-Clause(Cause) Edge:** All cause clause nodes are fully connected. Similarly, the edge is conducive to the learning of cause clause nodes.

- **Clause-Pair Edge:** All pair nodes are connected to their corresponding emotion clause nodes and cause clause nodes with this edge. The edge can help these three types of nodes transmit causal relationship between emotion and cause to each other.

- **Document-Others Edge:** The document node is connected to all other nodes with this edge, transmitting the global information in document to others.

Besides, each type of node has a kind of self-loop edge, which can help each node to keep its feature in the process of interaction.

Next, the Relational Graph Convolutional Network (RGCN) (Schlichtkrull et al., 2018) is applied on our heterogeneous undirected graph to aggregate the features from neighbors. First, we use the representation of clause to initialize each emotion and cause clause node:

$$H_E^{(0)} = H,\ H_C^{(0)} = H \tag{5}$$

where $H_E^{(0)}$ is the representation of emotion clause nodes and $H_C^{(0)}$ is the representation of cause clause nodes. Then, we use the representations of pairs to initialize the pair nodes:

$$H_P^{(0)} = \{p_{11}, p_{12}, \ldots, p_{NN}\} \tag{6}$$

In addition, we use the average pooling of the representations of clause to initialize the document node:

$$H_D^{(0)} = Avgpool(H) \in \mathbb{R}^d \tag{7}$$

After that, we apply the RGCN on our graph. Given a node u, it is defined as:

$$s_u^{(l)} = W_s^{(l)} h_u^{(l)} + b_s^{(l)} \tag{8}$$

$$t_u^{(l+1)} = s_u^{(l)} + \sum_{r\in\mathcal{R}} \sum_{v\in\mathcal{N}_r(u)} \frac{1}{|\mathcal{N}_r(u)|} W_r^{(l)} h_v^{(l)} + b_r^{(l)} \tag{9}$$

$$h_u^{(l+1)} = ReLU\left(t_u^{(l+1)}\right) \tag{10}$$

where l is the l-th layer of RGCN, $\mathcal{R}$ are different types of edges, $W_s^{(l)} \in \mathbb{R}^{d\times d}$, $b_s^{(l)} \in \mathbb{R}^d$, $W_r^{(l)} \in \mathbb{R}^{d\times d}$ and $b_r^{(l)} \in \mathbb{R}^d$ are learnable parameters, $\mathcal{N}_r(u)$ is the neighbours for node u connected with the edge of type r, and $ReLU$ is the ReLU activation function.

Finally, we select the last layer as the final representation of all nodes after convolutional operation of θ layers:

$$E = H_E^{(\theta)},\ C = H_C^{(\theta)},\ P = H_P^{(\theta)} \tag{11}$$

4.3 Classification

After getting all the representations of nodes, we use a simple MLP to obtain the prediction of emotion-cause pairs:

$$\hat{y}_{ij}^p = \sigma\left(MLP\left([P_{ij}, E_i, C_j]\right)\right) \tag{12}$$

where MLP includes two full-connected layers and a ReLU activation function between them, σ is the sigmoid activation function.

Correspondingly, the binary cross entropy loss is utilized as loss of ECPE:

$$\mathcal{L}_p = -\sum_{i}^{N}\sum_{j}^{N} y_{ij}^{p}\log(\hat{y}_{ij}^{p}) \tag{13}$$

where y_{ij}^{p} is the ground truth label.

Following the settings in (Wei et al., 2020), we set two auxiliary tasks which are emotion clauses extraction and cause clauses extraction in order to make the clause nodes learn the key contextual information about emotion or cause in the clauses. We compute the probability as follows:

$$\hat{y}_i^e = \sigma\left(W_e E_i + b_e\right) \tag{14}$$

$$\hat{y}_j^c = \sigma\left(W_c C_j + b_c\right) \tag{15}$$

where $\hat{y}_i^e$ and $\hat{y}_j^c$ are the probability of emotion and cause clauses separately, σ is the sigmoid activation function, $W_e \in \mathbb{R}^{1\times d}$, $W_c \in \mathbb{R}^{1\times d}$, $b_e \in \mathbb{R}$ and $b_c \in \mathbb{R}$ are learnable parameters.

Similarly, they have the corresponding loss:

$$\mathcal{L}_e = -\sum_{i}^{N} y_i^e \log(\hat{y}_i^e) \tag{16}$$

$$\mathcal{L}_c = -\sum_{j}^{N} y_j^c \log(\hat{y}_j^c) \tag{17}$$

where y_i^e and y_j^c are the ground truth labels.

4.4 Training Object

We train our model by jointly optimize the three sub-tasks using the AdamW optimizer (Loshchilov and Hutter, 2018). The total training object is defined as follow:

$$\mathcal{L} = \alpha\mathcal{L}_p + \beta\mathcal{L}_e + \gamma\mathcal{L}_c \tag{18}$$

where α, β and γ are hyperparameters.

5 Experiments

Extensive experiments are conducted to verify the effectiveness of the proposed model.

5.1 Experimental Setup

Dataset and Evaluation Metrics

We use the Chinese benchmark dataset released by Xia and Ding (2019), which is pre-processed from the dataset released by Gui et al. (2016) for the emotion cause extraction (ECE) task. Table 1 shows the detail of the dataset. Following Xia and Ding (2019), we use the 10-fold cross-validation as the data split strategy and the precision P, recall R and F-score $F1$ as evaluation metrics on three tasks: emotion-cause pair extraction, emotion clause extraction and cause clause extraction.

Item	Quantity
Document with one pair	1,746
Document with two pairs	177
Document with three or more pairs	22
Avg. # of clauses per document	14.77
Max. # of clauses per document	73
Total # of documents	1,945

Table 1: The detail of the Chinese corpus.

Comparative Approaches

We compare our model with the following methods, which use the pre-trained BERT as encoder:

- **ECPE-2D** (Ding et al., 2020a): This method uses the 2D representation to construct a pairs matrix and utilizes the 2D transformer module to interact with other pairs for prediction.
- **TransECPE** (Fan et al., 2020): It is a transition-based method which transforms the task into a procedure of parsing-like directed graph construction.
- **RankCP** (Wei et al., 2020): This method tackles emotion-cause pair extraction from a ranking perspective, which ranks pairs in a document and proposes a one-step neural approach to extract.
- **PairGCN** (Chen et al., 2020c): This method constructs a graph using the pair nodes and a Pair Graph Convolutional Network to model the dependency relations among candidate pairs.
- **ECPE-MLL** (Ding et al., 2020a): It is the current state-of-the-art method, which employs two joint frameworks, including the emotion-pivot cause extraction and cause-pivoted emotion extraction with sliding window strategy.
- **MTST-ECPE** (Fan et al., 2021): This method uses a multi-task sequence tagging framework with refining the tag distribution.

Approach	Emotion-Cause Pair Extraction			Emotion Clause Extraction			Cause Clause Extraction		
	P	R	$F1$	P	R	$F1$	P	R	$F1$
ECPE-2D	72.92	65.44	68.89	86.27	**92.21**	89.10	73.36	69.34	71.23
TransECPE	73.74	63.07	67.99	87.16	82.44	84.74	75.62	64.71	69.74
RankCP	71.19	**76.30**	73.60	**91.23**	89.99	**90.57**	74.61	77.88	76.15
PairGCN	76.92	67.91	72.02	88.57	79.58	83.75	79.07	68.28	73.75
ECPE-MLL†	77.00	72.35	74.52	86.08	91.91	88.86	73.82	**79.12**	76.30
MTST-ECPE⋄	75.78	70.51	72.91	85.83	80.94	83.21	77.64	72.36	74.77
Ours	**77.97**	72.95	**75.30***	90.89	87.64	89.19	**79.74**	74.79	**77.12**

Table 2: The results comparison with baselines on the ECPE corpus for emotion-cause pair extraction and the two sub-tasks: emotion clause extraction and cause clause extraction. The best performance is in **bold** and the second best performance is underlined. Result with † is previous state-of-the-art method. Approach with ⋄ is based on our implementation. * denotes $p < 0.05$ for a two-tailed t-test against the RankCP.

Approach	Emotion-Cause Pair Extraction			Emotion Clause Extraction			Cause Clause Extraction		
	P	R	$F1$	P	R	$F1$	P	R	$F1$
Ours	**77.97**	**72.95**	**75.30**	90.89	87.64	**89.19**	**79.74**	74.79	**77.12**
- w/o Clause Edge	76.52	71.56	73.90	89.72	86.39	87.98	78.63	73.62	75.99
- w/o Pair Node	75.57	72.87	74.13	90.51	**87.72**	89.07	79.33	75.12	**77.12**
- w/o PG	77.57	71.87	74.55	**91.08**	76.89	88.9	79.71	73.99	76.68
- w/o Pair Node & PG	72.82	72.45	72.55	89.20	87.63	88.38	77.86	**75.28**	76.48
- w/o Doc. Node	76.95	71.50	74.08	89.76	85.99	87.8	79.15	73.71	76.29

Table 3: The results of ablation study on the benchmark corpus for emotion-cause pair extraction and the two sub-tasks. The best performance is in **bold** and the second best performance is underlined.

Implementation Details

We implement our model based on Transformers[1] (Wolf et al., 2020), and use the default parameters in BERT, setting the hidden size d to 768. Besides, the hyperparameters λ and θ are set to 3 and 1, separately. And the α, β and γ are all set to 1. We train our model through AdamW optimizer and the learning rate is 2e-5. Finally, we set the mini-batch to 4 and the training epoch to 25. The experiments are run on the PyTorch-1.9.0 platform and Ubuntu 18.04 using the Intel(R) Core(TM) i7-8700K CPU, 64GB RAM and NVIDIA GeForce RTX 2080 Ti 11GB GPU.

5.2 Experimental Results

Table 2 shows the results on the emotion-cause pair extraction (ECPE) task and two sub-tasks: emotion clause extraction (EE) and cause clause extraction (CE). Our model shows a clear advantage over previous works. Specifically, our model obtains 0.78% and 1.70% $F1$ improvements on ECPE compared with the previous best methods ECPE-MLL and RankCP, separately. We argue that the pair-based joint encoding plays an important role in it, making the interaction bidirectional and balancing the information obtained by pairs and clauses. Moreover, we get competitive improvement (0.82% on $F1$) on CE and slight improvement (0.33% on $F1$) on EE, which can help improve the performance on ECPE with the consideration of causal relationship.

Although our model is not the best in EE, our model can balance the EE and CE. Specifically, RankCP gets a huge improvement on EE (1.38% on $F1$) to our model, but achieves poor performance on CE, leading to the sharply dropped in ECPE. Similarly, MTST-ECPE and ECPE-2D encounter the imbalance problem compared with our model, in which MTST-ECPE performs well on CE and ECPE-2D performs well on EE. We argue that the balance is benefit from modeling two types of clauses efficiently. Meanwhile, our model achieves better results than PairGCN, which also uses the Graph Neural Network. We believe our new strategy to construct the graph mainly leads to this improvement.

5.3 Ablation Study

Ablation studies are conducted to verify the effectiveness of the Pair Generator (PG) and different relationship edges in our graph. Table 3 shows the results of the ablation studies.

w/o Clause Edge We use a single type of edge to replace the Clause(Emotion)-Clause(Emotion) Edge and the Clause(Cause)-Clause(Cause) Edge,

[1] https://github.com/huggingface/transformers

which means the model does not distinguish the emotion clause nodes and the cause clause nodes. Without these two edges, the performance of our model sharply drops both on EE and CE, further leading to the drop on ECPE, which indicates that the contextual information in emotion clauses is different from those in cause clauses. Therefore, using the same representations for prediction on EE and CE will blur their features and lead to a drop in results.

w/o Pair Node We remove the pair nodes and separately model the emotion and cause clauses using the Relational Graph Convolutional Network (RGCN). The pairs from PG are utilized to replace the pairs after RGCN, and they are concatenated with the clauses after RGCN for prediction. In this way, the two types of clauses can not interact, and the pairs can not learn the causal relationship. Although the performance on EE and CE is similar with the complete model, the $F1$ of ECPE is sharply dropped by 1.17%, which means the problem we describe in Section 1 appears. Without the causal relationship, the model may combine two unrelated emotion clause and cause clause into a pair.

w/o PG On the other hand, we remove the PG and use another relative position embedding to replace the representations of the pair, which means the pairs having the same relative position will have the same initial representations in the RGCN and do not contain any clause information. Without the PG, the performance slightly drops on ECPE and is similar on EE and CE compared with the complete model. Although lacking the clause information, the pairs can learn the clause features and the causal relationship by the Clause-Pair Edge. We argue that the causal relationship is crucial to the modeling of pairs. Therefore, compared with the model without pair nodes learning the causal relationship, the model without PG achieves better performance on ECPE.

w/o Pair Node & PG Moreover, we remove the pair nodes and PG together, similar to previous works which only encode the clauses for prediction. The $F1$ on ECPE is sharply dropped by 2.75%, which is caused by the ignorance of pair modeling and the causal relationship in pairs of clauses.

w/o Doc. Node Finally, we remove the document node in the RGCN. The drop in performance mainly occurs in ECPE and CE. We believe that the global information of documents (e.g., topics) is beneficial for the ECPE. We will explore it in detail in Section 5.4.

#Clauses per Doc.	% in Corpus	Approach	P	R	$F1$
< 14	45.71	Ours	**78.94**	74.03	**76.41**
		- w/o Doc. N.	78.19	73.01	75.51
		RankCP	69.82	**77.49**	73.46
≥ 14	54.29	Ours	**76.74**	71.93	**74.26**
		- w/o Doc. N.	75.60	70.21	72.81
		RankCP	72.00	**75.28**	73.60
≥ 20	16.25	Ours	**73.93**	67.70	**70.67**
		- w/o Doc. N.	70.68	64.33	67.35
		RankCP	67.37	**71.91**	69.57

Table 4: The result of ECPE for documents with different numbers of clauses.

5.4 The Effect of Document Node

To verify the effect of document node, some extensive experiments are conducted to explore the impact of document node in different lengths of a document, according to the average number of clauses per document 14.77 and the median 14. Besides, we consider the extreme case in which a document contains 20 clauses or more.

As shown in Table 4, the improvement in long documents on ECPE (≥ 14, 1.14%, 1.72%, 1.45% on P, R and $F1$, separately) is much more than in short documents (< 14, 0.75%, 1.02%, 0.90% on P, R and $F1$, separately) with the help of document node. When the document is long, there will be many emotion clause nodes and cause clause nodes in a graph. Hence, each emotion and cause clause node can hardly learn the effective contextual information for the competitive fully connected graph. In this situation, the document node can filter the invalid information and integrate them into global information, then transmits them to other nodes through the Document-Others Edge. Furthermore, in the extremely long documents which contain 20 clauses or more, the improvement is even more obvious (3.25%, 3.37% and 3.32% on P, R and $F1$ separately), which shows the advantage in long documents. Finally, our model completely surpasses RankCP on P and $F1$ in long documents with the help of the document node.

5.5 Case Study

We present the case studies with three examples selected from the benchmark corpus to demonstrate the effectiveness of our considering the causal relationship in our model. The ground truths and the predicted results and RankCP are shown in Table 5. We choose the RankCP to compare with our model

ID	Examples	Predicted Pairs		Ground Truths
		Ours	RankCP	
1	... He learned information from a friend by accident(c_{10}) Someone finally found his mother in Xinzheng City, Henan Province(c_{11}) He was so happy(c_{12}) But he was also worried about the difficulty of setting up his mother's registered residence(c_{13}) ...	$[c_{12}, c_{11}]$ $[c_{13}, c_{13}]$	$[c_{12}, c_{11}]$ $[c_{12}, c_{13}]$ $[c_{13}, c_{13}]$	$[c_{12}, c_{11}]$ $[c_{13}, c_{13}]$
2	... She and her family are very healthy(c_2) So they can continue to donate blood to contribute to the society(c_3) She used to worry about limiting the age of blood donation to 55 years old(c_4) She is used to donating blood now(c_5) If she can't continue donating blood because of her age, she will be very disappointed(c_6) ...	$[c_4, c_4]$ $[c_6, c_6]$	$[c_4, c_4]$ $[c_4, c_6]$ $[c_6, c_6]$	$[c_4, c_4]$ $[c_6, c_6]$
3	... He gave all the money to the girl who he love(c_{11}) However, something happened(c_{12}) He found out that she had already concealed him from getting married and having a daughter(c_{13}) And she squandered all the money(c_{14}) What made people wired is(c_{15}) He didn't complain about her(c_{16}) ...	$[c_{15}, c_{16}]$	$[c_{15}, c_{13}]$ $[c_{15}, c_{14}]$ $[c_{15}, c_{16}]$	$[c_{15}, c_{16}]$

Table 5: Examples predicted by our model and RankCP. The words in red are the emotion keywords, and the words in blue are the cause keywords. The pairs in green are the correct prediction, and the pairs in red are incorrect.

because it is more representative.

For the first example, although RankCP extracts all the ground truths, it extracts another incorrect pair (c_{12}, c_{13}). The emotion clause c_{12} expresses happy and the cause clause c_{13} expresses concern about the difficulty in registered residence. Obviously, c_{12} is not the reason to cause c_{13}. By considering the causal relationship, our model avoids this situation.

Next, for the second example, RankCP encountered the same problem as the first example. Further, the emotion clause c_4 expresses worry and c_6 expresses disappointment, which are both negative emotions. Moreover, the cause clause c_4 describes the same thing with cause clause c_6. We think that it is more difficult for RankCP to judge this situation. Nevertheless, our model successfully deals with this situation.

Finally, for the last example, RankCP and our model both extract the correct emotion clause. However, RankCP predicts another two wrong pairs. Although the c_{13} and c_{14} contain something that makes the man feel disappointed, there is no corresponding emotion clause in the text. We believe that our model catches the causal relationship between the emotion feeling wired and the cause that the man did not complain to avoid making the incorrect prediction.

5.6 Hyperparameters Discussion

As shown in Figure 3, we examine the effects of different values of θ on ECPE. We can observe that the performance tends to drop with the increasing of the layers of RGCN. We believe that the multi-hop of RGCN causes this problem. Specifically, when the θ is more than one, the features of the emotion node can be transmitted to the cause nodes, which disturbs the prediction of cause clause extraction and further leads to the drop of performance on ECPE by affecting the modeling of causal relationship, vice versa. Besides, more layers indicates more learnable parameters, which will result in over-fitting.

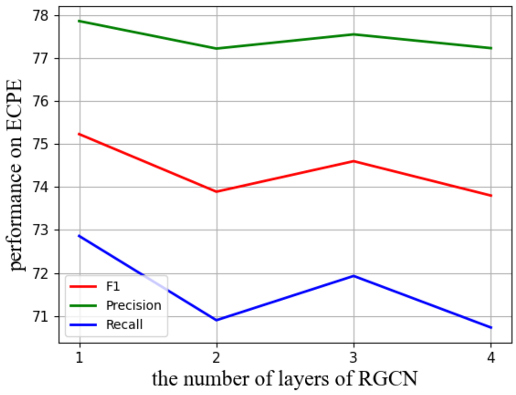

Figure 3: The influence of different θ on ECPE.

6 Conclusion

In this paper, we propose a novel joint encoding network which generates the pairs and clauses feature simultaneously to model the causal relationship in pairs, which can balance the inter-task feature interaction compared with sequential encoding, and model the causal relationship from clauses. Moreover, from a multi-relational perspective, we propose a Relational Graph Convolutional Network (RGCN) framework to capture the relationship between pairs and clauses, including four types of node and four types of edge. The experiments on the Chinese benchmark corpus show that our model achieves state-of-the-art performance, and the extensive experiments demonstrate the effectiveness of our proposed modules.

References

Xinhong Chen, Qing Li, and Jianping Wang. 2020a. Conditional causal relationships between emotions and causes in texts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3111–3121, Online. Association for Computational Linguistics.

Xinhong Chen, Qing Li, and Jianping Wang. 2020b. A unified sequence labeling model for emotion cause pair extraction. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 208–218, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Ying Chen, Wenjun Hou, Shoushan Li, Caicong Wu, and Xiaoqiang Zhang. 2020c. End-to-end emotion-cause pair extraction with graph convolutional network. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 198–207, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Zifeng Cheng, Zhiwei Jiang, Yafeng Yin, Hua Yu, and Qing Gu. 2020. A symmetric local search network for emotion-cause pair extraction. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 139–149, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Zixiang Ding, Rui Xia, and Jianfei Yu. 2020a. 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*, pages 3161–3170, Online. Association for Computational Linguistics.

Zixiang Ding, Rui Xia, and Jianfei Yu. 2020b. End-to-end emotion-cause pair extraction based on sliding window multi-label learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3574–3583, Online. Association for Computational Linguistics.

Chuang Fan, Chaofa Yuan, Jiachen Du, Lin Gui, Min Yang, and Ruifeng Xu. 2020. Transition-based directed graph construction for emotion-cause pair extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3707–3717, Online. Association for Computational Linguistics.

Chuang Fan, Chaofa Yuan, Lin Gui, Yue Zhang, and Ruifeng Xu. 2021. Multi-task sequence tagging for emotion-cause pair extraction via tag distribution refinement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:2339–2350.

Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. 2016. Event-driven emotion cause extraction with corpus construction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1639–1649, Austin, Texas. Association for Computational Linguistics.

Taichi Ishiwatari, Yuki Yasuda, Taro Miyazaki, and Jun Goto. 2020. Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7360–7370, Online. Association for Computational Linguistics.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*.

Sophia Yat Mei Lee, Ying Chen, and Chu-Ren Huang. 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*, pages 45–53, Los Angeles, CA. Association for Computational Linguistics.

Ilya Loshchilov and Frank Hutter. 2018. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer.

Aaditya Singh, Shreeshail Hingane, Saim Wani, and Ashutosh Modi. 2021. An end-to-end network for emotion-cause pair extraction. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 84–91, Online. Association for Computational Linguistics.

Penghui Wei, Jiahao Zhao, and Wenji Mao. 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*, pages 3171–3181, Online. Association for Computational Linguistics.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin

Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Rui Xia and Zixiang Ding. 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*, pages 1003–1012, Florence, Italy. Association for Computational Linguistics.

Chaofa Yuan, Chuang Fan, Jianzhu Bao, and Ruifeng Xu. 2020. Emotion-cause pair extraction as sequence labeling based on a novel tagging scheme. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3568–3573, Online. Association for Computational Linguistics.

Shuang Zeng, Runxin Xu, Baobao Chang, and Lei Li. 2020. Double graph based reasoning for document-level relation extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1630–1640, Online. Association for Computational Linguistics.

Kun Zhou, Wayne Xin Zhao, Shuqing Bian, Yuanhang Zhou, Ji-Rong Wen, and Jingsong Yu. 2020. Improving conversational recommender systems via knowledge graph based semantic fusion. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining*, KDD '20, page 1006–1014, New York, NY, USA. Association for Computing Machinery.