Enhancing Emotion-Cause Pair Extraction in Conversations via Center Event Detection and Reasoning

Botao Wang[1] **Keke Tang**[2,*] **Peican Zhu**[1,3,*]
[1]School of Computer Science, Northwestern Polytechnical University, China
[2]Cyberspace Institute of Advanced Technology, Guangzhou University, China
[3]School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, China
wangbotao@mail.nwpu.edu.cn, tangbohutbh@gmail.com, ericcan@nwpu.edu.cn

Abstract

Emotion-Cause Pair Extraction in Conversations (ECPEC) aims to identify emotion utterances and their corresponding cause utterances in unannotated conversations, this task that has garnered increasing attention recently. Previous methods often apply Emotion-Cause Pair Extraction (ECPE) task models, treating the entire conversation as a whole for contextual interaction. However, statistical analysis shows that the number of emotion-cause pairs in ECPEC conversation data far exceeds that in ECPE datasets, leading to interference among multiple events within a conversation and causing noise to propagate between different events. To address this issue, we propose a novel CEnter eveNT-guided framEwoRk (CENTER). This model introduces a Center Event Detection task to construct a center event-aware graph that captures the unique representations of different event regions. Additionally, mimicking human reasoning processes, we build a center event reasoning graph and use graph neural network to facilitate the flow of information between utterance pairs, thereby uncovering the relationships between emotions and their causes. Experimental results demonstrate that our approach achieves state-of-the-art performance across three benchmark datasets.

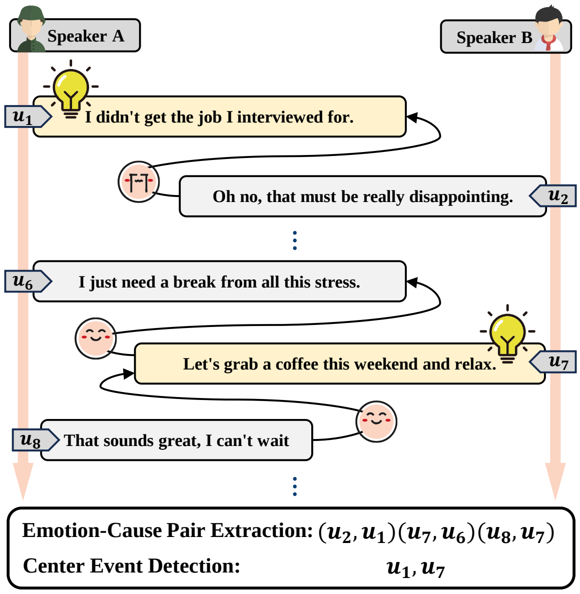

Figure 1: An example of ECPEC task, where u_i is the utterance. The emotion for each utterance is then marked. Arrow indicates the direction from the emotion to its cause, while center event is highlighted in yellow box. Based on the text content and the distribution of EC pair labels, the conversation consists of two distinct center events.

1 Introduction

With the development of news media, individuals can rapidly express and disseminate their opinions through social platforms. By analyzing the emotion polarity and intensity in conversation data (Sharma et al., 2024), it is possible to comprehend the intent of the text and uncover hidden information (Zhang et al., 2024). Given that the causes of emotions are key elements for in-depth emotion understanding, the Emotion-Cause Pair Extraction (ECPEC) task (Li et al., 2022) has garnered significant attention. This task involves extracting all potential emotion utterances and their corresponding causes from unannotated conversations, thereby gaining insights into the emotion propagation between speakers and comprehensively understanding their emotion interactions. ECPEC can be applied to various practical scenarios, such as public opinion monitoring (Zhou et al., 2024; Cheng et al., 2024), social media information detection (Hua et al., 2023), marketing (Schrama et al., 2024), and psychological interventions (Evans and Shaughnessy, 2024).

In the ECPEC task, accurately analyzing the causes of emotions is challenging due to the unique semantic properties of conversational content, such as context dependencies, complex syntactic structures, and interactivity (Hoey and Kendrick, 2017). The main approaches to addressing these chal-

[*]Corresponding authors.

Findings of the Association for Computational Linguistics: EMNLP 2024, pages 10773–10783
November 12-16, 2024

lenges can be categorized into sequential encoding and aggregated encoding: (1) **Sequential Encoding**: This method focuses on learning and enhancing context based on the sequential order of utterances in a conversation. Li et al. (2022) employ cross-attention mechanisms to learn the contextual dependencies and interactivity between utterances, effectively extracting emotion and cause utterances. Jeong and Bak (2023) incorporate speaker features as supplementary information, sequentially concatenating these with utterances to enrich the representations. (2) **Aggregated Encoding**: This approach uses connection graph to learn the syntactic structure of conversations through graph convolutional network (GCN). Li et al. (2022) use fully connected networks to represent the relationships between utterances and employ GCNs to capture the syntactic dependencies between them. Liu et al. (2023) transform the task into a machine reading comprehension problem, encoding utterances based on the positional relationships between different speakers within the conversation.

While current research address the unique semantic challenges present in conversation content, existing methods treat conversation as a single semantic unit, employing mechanisms such as graph neural networks, attention mechanisms, and recurrent neural networks for contextual enhancement. This approach can lead to the intermingling of information in lengthy conversations, resulting in counterfactual reasoning where causes appear after emotions. Inspired by embodied cognition mentioned (Zhang et al., 2021; Wilson, 2002), we observe that conversation participants may exhibit various emotions and associated causes around certain center utterances, with conversation typically exhibiting a discrete multi-event structure. Specifically, as illustrated in Figure 1, utterances u_1 and u_2 reflect speaker B's frustration over speaker A not getting the job, while u_6 and u_8 reflect speaker B's happiness about actions taken to relieve stress. If we do not differentiate between different centers of conversation content and instead encode the entire conversation as a whole for context learning, noise from unrelated content will increase the difficulty for the model to understand these scenes, making it challenging to accurately capture the intrinsic connections between emotions and their causes.

In light of the above considerations, we propose a novel **CE**nter eve**NT**-guided fram**E**wo**R**k (CENTER) for the ECPEC task. This method identifies center events within conversations and traces the causes of emotions from the perspective of human cognitive reasoning through center-aware graphs and center reasoning graphs. Specifically, we first define the concept of center events and introduce a new ECPEC auxiliary task called Center Event Detection (CED). Besides, we construct a conversation center event-aware graph from the predicted center sequence to facilitate feature aggregation within different center events. Additionally, recognizing the importance of center events in conversations, we emphasize their uniqueness in the pairing process by establishing a candidate pair center event reasoning graph, which promotes information flow between features of pairs. Finally, we employ an end-to-end multi-task framework to mitigate interference between different events and enhance the efficient flow of multi-granularity features. Experiments conduct on three ECPEC benchmark datasets demonstrate that CENTER achieves state-of-the-art performance. Our contributions are summarized as follows:

(1) We introduce an auxiliary task called conversation Center Event Detection, which helps the model recognize multiple center events within a conversation and model conversation features from the perspective of human cognition.

(2) We propose a center event-aware graph aggregation framework to model relationships between utterances, facilitating the extraction of unique features from different informational regions and enhancing the efficiency of information flow.

(3) We develop a conversation center event reasoning graph network to simulate the reason process, aiding in the propagation of center emotion-cause pair features among candidate pairs and improving reasoning capability.

2 Related Work

2.1 Emotion-Cause Pair Extraction

Since ECPEC is a variant of the Emotion-Cause Pair Extraction (ECPE) task, most methods for extracting emotion-cause (EC) pairs from conversations commonly adopt the models used in ECPE. Due to the significant role of emotions and causes in the ECPE task, prior studies predominantly treated emotion extraction (EE) and cause extraction (CE) as auxiliary tasks. Specifically, Xia and Ding (2019) introduce the ECPE task along with two auxiliary tasks. Employing a two-stage method, it initially extracts emotion and cause clauses and subsequently matches them into pairs for prediction using cartesian products. However, this approach focuses on capturing task correlations through implicit parameter sharing, which has limitations in explicitly modeling information interactions and results in lower interpretability. Fan et al. (2021) propose a multi-task sequence labeling framework, utilizing a tagging scheme based on distance encoding to simultaneously predict emotions and causes, thus enhancing the capability of model in extracting emotion-cause pairs. Yet, with the continuous refinement of encoding granularity, there exist drawbacks in simultaneously modeling specific features and interaction features. Chen et al. (2022) establish a feature-task alignment mechanism, explicitly modeling specific features and interaction features while simultaneously establishing task alignment mechanisms to achieve better label consistency. Considering the lack of modeling for pairs, Zhu et al. (2024) guide by auxiliary tasks, encode clause pair features through GNN, facilitating information propagation among candidate pairs.

These exemplary multi-task models enhance the performance of ECPE task by treating emotion extraction and cause extraction as auxiliary tasks. However, due to the differences in content between conversation and document, they often overlook the event interaction and diversity in ECPEC, failing to distinguish between events within conversations, thereby potentially leading to issues of mutual interference among multiple event information flows within conversations.

2.2 Emotion-Cause Pair Extraction in Conversations

Li et al. (2022) introduce the ECPEC task for the first time, along with a two-step framework. Initially, they encode the emotions and causes expressed in conversations using graph neural networks (GNN) and attention mechanisms. Subsequently, they employ a chunk pruning strategy to more accurately extract information features of emotion-cause pairs. Recognizing the distinct characteristics of conversations compared to ECPE research based on news articles, Jeong and Bak (2023) propose a pairing-based expert mixture model. By incorporating conversation features through gated networks, they significantly enhance the performance of emotion cause pairing. However, extracting causes without distinguishing emotions may lead to an incomplete understanding of the dependency between emotions and their causes. Therefore, Lee et al. (2023) propose a novel evaluation metric, Emotion-Cause Pair Emotion Combination Assessment, to jointly evaluate emotion-cause pairs extraction and their corresponding multiple emotion extraction. Nevertheless, prior approaches overlooked the role of speakers. Thus, An et al. (2023) construct a global view and speaker-perception extraction framework. By simulating conversations between speakers and utilizing a perception-coupled decoding module along with an emotion graph attention network to encode conversation features, they efficiently extract emotions and all associated causes in conversations.

These exemplary models for ECPEC task enhance representation capability by augmenting conversation features. However, they overlook the utilization of center events in conversations, leading to issues such as feature scarcity and unfocused attention when enhancing features in long conversations.

3 Task Definition

In the ECPEC task, the input data format consists of three hierarchical levels, i.e., "conversation-utterance-word". The input includes a series of conversations $d = \{c_1, c_2, \cdots, c_{|d|}\}$, where each is composed of n utterances $c_i = \{u_1, u_2, \cdots, u_n\}$, and each utterance consists of m words $u_j = \{w_1, w_2, \cdots, w_m\}$. The main objective of the ECPEC task is to extract all EC pairs from the conversation in d, which is described as:

$$P = \{\ldots, (u_i, u_j), \ldots\}(1 \leq i, j \leq n) \tag{1}$$

where (u_i, u_j) represents an EC pair, u_i and u_j denote the emotion and corresponding cause utterances, respectively. An emotion utterance can be associated with multiple cause ones, and the same cause may trigger various emotions. Additionally, an emotion utterance can even serve as its own cause utterance.

4 Proposed Model

In this section, we describe our approach in detail. As in Figure 2, our CENTER consists of four main components: (1) **Conversation Encoder**, which encodes the input text and learns contextual features; (2) **Center Event-Aware (CEA) Graph**, which constructs a graph comprising center event nodes and non-center ones for utterance-level feature enhancement; (3) **Center Event Reasoning (CER) Graph**, which is constructed by center pairs and non-center pairs for reasoning between emotion and cause; (4) **Conversation Emotion-Cause Extraction**, which integrates multi-dimensional feature representations and uses a multi-task framework for final classification.

4.1 Conversation Encoder

For the input conversation $c_i = \{u_1, u_2, \cdots, u_n\}$, combined with the representational capacity of pre-trained models, we encode utterance representations by learning contextual features. Specifically, we first insert $[CLS]$ and $[SEP]$ tokens at the beginning and end of each utterance, respectively. Then, we construct token sequences for each utterance by including all words and the two special tokens $\{[CLS], w_1, w_2, \cdots, w_m, [SEP]\}$. We obtain the utterance representation v_i using BERT, the process is provided as:

$$v_i = BERT(token, mask, type) \tag{2}$$

where $mask$ represents the sequence for the self-attention operation, and $type$ denotes the sentence position encoding. Finally, by allowing each word to interact with its surrounding words, the $[CLS]$ token at the end of each utterance serves as the feature for the entire utterance, resulting in the conversation sequence, i.e., $\{v_1, v_2, \cdots, v_n\}$.

4.2 Center Event-Aware Graph

Annotation of Center Events We utilize objective annotation rules to annotate center events within public dataset, aiming to investigate their impact on dialogues. Specifically, we adopt the following rules: (1) center events serve as focal points of the conversation; (2) center events should be directly related to EC pairs; (3) the number of center events should be minimized while accurately and comprehensively expressing the main content of the conversation; (4) the granularity of center events is at the utterance level. Following these rules, we conduct a statistical analysis based on EC pair labels, defining center event utterances as those emotion utterances containing more than α causes and those cause utterances that appear α or more times.

Center Event Detection Given significant role of center events in conversations, it is crucial to accurately identify multiple center events and convey center information within utterance representations. Specifically, we propose an auxiliary task for center event detection. This task involves feeding the conversation sequence representations v_i into a fully connected layer. Using an activation function, we regard the most likely utterances as center events. The process is:

$$\hat{y}_i^{ce} = argmax\left(linear\left(W_{ce}v_i + b_{ce}\right)\right) \tag{3}$$

where $\hat{y}_i^{ce}$ represents the predicted label indicating whether utterance i is a center event or not. $linear()$ denotes a linear layer, and $argmax()$ denotes the function that selects the index of the maximum value. W_{ce} and b_{ce} are trainable parameters. This produces the label sequence $\{\hat{y}_1^{ce}, \hat{y}_2^{ce}, \cdots, \hat{y}_n^{ce}\}$ of the CED task.

Center Event-Aware Graph Construction Given the crucial role of center events in conversation and the tendency for surrounding utterances to revolve around these events, we construct a center event-aware graph to facilitate the transmission of key information among utterances. Specifically, we treat each utterance as a node. We suppose that utterances being predicted as center events and their surrounding contexts describe the same event. The specific connection method is provided as:

$$M_{ce_{i,j}} = \begin{cases} 1 & \hat{y}_i^{ce} = 1 \; and \; |i-j| \leq \beta \\ 0 & otherwise \end{cases} \tag{4}$$

where β is the influence range of a center event. The dimension of matrix M_{ce} for CEA graph is n^2.

Utterance Feature Aggregation To facilitate the transmission of key information within the conversation, we consider the conversation sequence, i.e., $\{v_1, v_2, \cdots, v_n\}$, as nodes and use the matrix M_{ce} to establish edges. We utilize a GNN to effectively aggregate the features of the utterances:

$$r_i^t = ReLU\left(\sum_{j=N_i} M_{ce_{i,j}} W_u r_j^{t-1}\right) \tag{5}$$

where t represents the current layer, $ReLU()$ is the activation function, r_j^{t-1} indicates the utterance features from the $(t-1)$-th layer of the GNN, N_i represents all neighbors of the i-th utterance, and W_u represents trainable parameters. Finally, we obtain the utterance feature vector enhanced by centrality, i.e., $\{r_1, r_2, \cdots, r_n\}$.

4.3 Center Event Reasoning Graph

Center Event Reasoning Graph Construction To explore the causal relationships between emotion and cause utterances, we establish a center

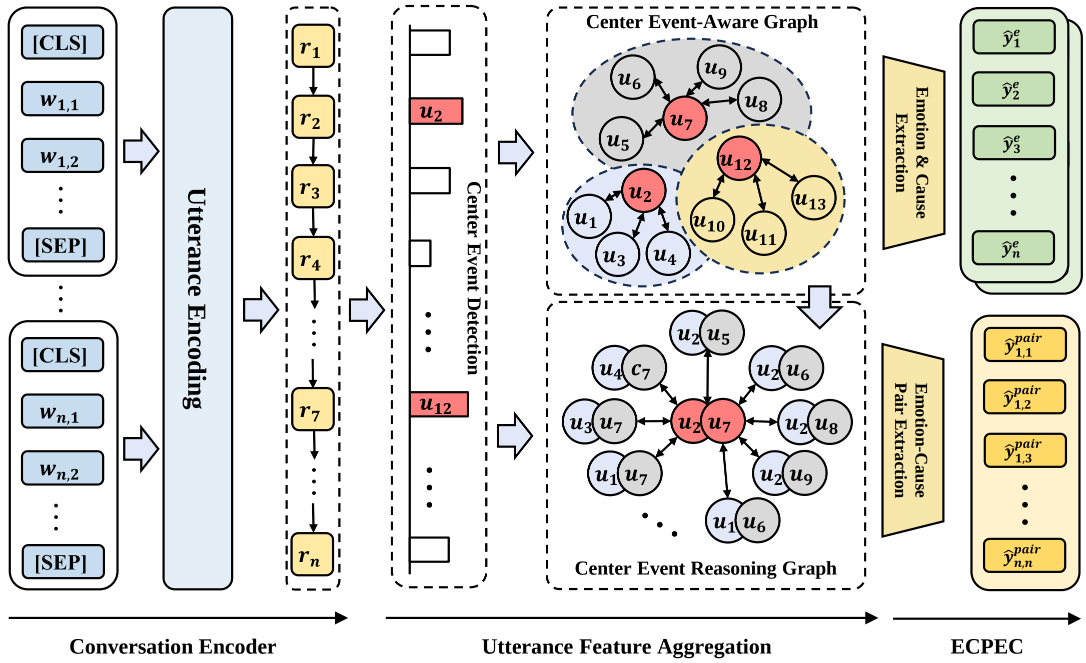

Figure 2: Overall architecture of CENTER. First, the conversation text is input into a pre-trained BERT model for encoding. Then, utilizing the CED task, we construct the CEA graph and CER graph to aggregate multi-granularity features. Finally, the centrality enhanced representations are used for multi-task prediction.

event reasoning graph that also leverages the pivotal role of center events in conversation and augments the reasoning capacity between emotion and cause. Specifically, we initially pair the n utterances. Each candidate pair of utterances serves as a node, and based on the predicted labels of center events, we define the pairs between center events as center utterance pair. Subsequently, we establish connections between center utterance pair and candidate pair as:

$$M_{pair_{(i,j),(z,w)}} = \begin{cases} 1 & \hat{y}_i^{ce} = 1 \; and \; \hat{y}_j^{ce} = 1 \\ & |i-z| \leq \delta \; and \; |j-w| \leq \delta \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

S where δ is the connection range of center pairs. The dimension of the M_{pair} for the CER graph is n^4.

Center Event Reasoning Graph Aggregation After concatenating the center-enhanced utterance vectors with the utterance distance encoding, we obtain the feature vector for each pair of utterances, i.e., $p_{i,j} = [r_i; r_j; d_{i,j}]$. We treat each utterance pair as a node and utilize the matrix of the CER graph M_{pair} to establish edges. This enables the utilization of GNN to aggregate the feature vectors of utterance pairs. The specific calculation is outlined as:

$$x_{i,j}^l = ReLU\left(\sum_{z,w=N_{i,j}} M_{pair_{(i,j),(z,w)}} W_p x_{i,j}^{l-1}\right) \tag{7}$$

where l denotes the current layer. When it is the first layer, the input is the feature vector of the utterance pair $\{p_{1,1}, p_{1,2}, \cdots, p_{n,n}\}$, W_p represents the trainable parameters, $x_{i,j}^{l-1}$ denotes the feature of the utterance pair corresponding to the $(l-1)$-th layer of the GNN. $N_{i,j}$ represents all neighbors of the utterance pair $p_{i,j}$. Ultimately, through the multi-layer aggregation of GNN, the feature vectors of the enhanced utterance pairs $\{x_{1,1}, x_{1,2}, \cdots, x_{n,n}\}$ can be obtained.

4.4 Conversation Emotion-Cause Extraction

Emotion Extraction and Cause Extraction Following Li et al. (2022), we treat EE and CE as auxiliary tasks to facilitate the learning of utterances within context. We feed the feature vectors of the enhanced utterances $\{r_1, r_2, \cdots, r_n\}$ into two separate activation functions to predict whether the utterances are emotion or cause, i.e.,

$$\hat{y}_i^e = softmax\left(W_e r_i + b_e\right) \tag{8}$$

$$\hat{y}_j^c = softmax\left(W_c r_j + b_c\right) \tag{9}$$

where $\hat{y}_i^e$ and $\hat{y}_j^c$ represent the predicted labels of emotion and cause utterance respectively, W_e, b_e, W_c, and b_c are trainable parameters.

Binary cross entropy loss is utilized for both EE and CE tasks which are provided as:

$$L_e = \sum_n^i -\left(y_i^e \log \hat{y}_i^e\right) \tag{10}$$

$$L_c = \sum_{n}^{j} - \left(y_j^c \log \hat{y}_j^c\right) \tag{11}$$

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

Emotion-Cause Pair Extraction After obtaining the center enhanced utterance pair features, we try to predict the EC pairs:

$$\hat{y}_{i,j}^{pair} = softmax\left(W_p x_{i,j} + b_p\right) \tag{12}$$

where $\hat{y}_{i,j}^{pair}$ is the predicted label of EC pairs, W_p and b_p are trainable parameters.

We employ the same loss calculation method:

$$L_{pair} = \sum_{n}^{i}\sum_{n}^{j} - \left(y_{i,j}^{pair} \log \hat{y}_{i,j}^{pair}\right) \tag{13}$$

where $y_{i,j}^{pair}$ is the truth label.

Training Based on CED labels described in 4.2, we compute the loss as:

$$L_{ce} = \sum_{n}^{i} - \left(y_i^{ce} \log \hat{y}_i^{ce}\right) \tag{14}$$

where y_i^{ce} is the truth label.

We jointly optimize these four subtasks to train CENTER and overall training loss is defined as:

$$L = \gamma_e L_e + \gamma_c L_c + \gamma_{ce} L_{ce} + \gamma_{pair} L_{pair} \tag{15}$$

where γ_e, γ_c, γ_{ce}, and γ_{pair} are hyper-parameters.

5 Experiments

5.1 Dataset and Evaluation Metrics

We utilize ConvECPE dataset released by Li et al. (2022). Furthermore, we also consider RECCON dataset (Poria et al., 2021). This dataset comprises DailyDialog (Li et al., 2017) (named ECPE-D-DD) and IEMOCAP (Busso et al., 2008) (named ECPE-D-IE). In these datasets, utterances being labeled as happy, sad, angry, excited, or frustrated are considered as emotion utterances, while neutral utterances are categorized as non-emotion ones.

We split the datasets into 80/10/10 for training, validation, and testing, respectively. Given the small scale of ECPE-D-IE, we use ECPE-D-IE exclusively as test data. Furthermore, Precision (P), recall (R), and F1-score ($F1$) are adopted as evaluation metrics.

5.2 Implementation Details

Our model is implemented based on the Transformers framework, utilizing the default parameters of bert-base-cased, with a hidden size being set to 768. Additionally, the hyperparameters β, δ, and α are set to 3, 2, and 2 respectively, while γ_e, γ_c, γ_{ce}, and γ_{pair} are all assigned to 1. We train CENTER using Adam optimizer with a learning rate of 0.001. The training epoch is set to 40. Experiments are conducted on the PyTorch-1.12.0 platform and Ubuntu 20.04.1, using Intel(R) Xeon(R) Silver 4310 CPU and NVIDIA GeForce RTX 3090 24GB GPU.

5.3 Baselines

In this manuscript, we considered the following baselines for comprehensive performance comparison. The corresponding details are provided as:

ECPE-2D (Ding et al., 2020a): This introduces a 2D representation scheme to depict emotion-cause pairs. It employs a unified end-to-end framework for modeling, interaction, and prediction.

ECPE-MLL (Ding et al., 2020b): This approach incorporates a sliding window-based multi-label learning framework. It extracts emotion-cause pairs by designating either emotion clauses or cause clauses as the focal point.

RANKCP (Wei et al., 2020): This model presents a ranking-based end-to-end extraction method. It simulates interactions between speakers through a graph attention mechanism and uses kernel-based relative position embeddings for effective ranking.

RECCON (Poria et al., 2021): Utilizing a pre-trained RoBERTa model with a classification layer, this method models and extracts emotion-cause pairs. For a fair comparison, we configure the language model to BERT.

ECPE-MM-R (Zhou et al., 2022): This approach transforms the ECPE task into a document-level machine reading comprehension (MRC) task. It introduces a multi-round MRC framework to simulate complex relationships between emotions and causes.

MRC (Liu et al., 2023): This model implements multi-turn MRC for ECPEC. It adopts a position-aware graph convolutional network framework to model conversation and utterance features.

PRG-MoE (Jeong and Bak, 2023): This method introduces a pairwise relationship-guided mixture-of-experts model. It utilizes speaker information to enhance utterance features and employs a gated network to model the causality of emotions from the perspective of cause classification.

Joint (Li et al., 2022): This method employs a two-step model, utilizing GCN or Cross Attention (Xatt) to generate utterance-level representations. Subsequently, it prunes emotion-cause pairings using a window-limiting approach. Specifically, Joint-GW denotes the use of GCN and the window-limiting method, while Joint-XW denotes the use of Xatt and the window-limiting method.

5.4 Main Results

Table 1 and Table 2 present the experimental results of CENTER and baseline methods on there datasets. On the ECPE-D-DD dataset, CENTER shows no significant performance difference in addressing the EE task compared to other models. This is because the introduction of the center event auxiliary task diminishes the distinction between emotion and cause. However, despite similar performance in EE task, our primary focus is on extracting EC pairs, where CENTER demonstrates a clear advantage over previous approaches. Furthermore, the ECPE-D-IE and ConvECPE datasets contain approximately eight times more EC pairs than that of ECPE-D-DD, leading to significant interference between events and subsequently poorer model performance. Compared to other models, CENTER effectively balances various metrics and exhibits robust performance.

Specifically, we believe that center events play a crucial role in the CENTER model, facilitating feature interactions within their respective regions and mitigating the propagation of noise from irrelevant events. Notably, compared to ECPE-MLL, CENTER exhibits similar performance in addressing the EE task but shows significant improvement in handling the ECPE task. This suggests that the introduction of center events in CENTER helps distinguish between different events and accurately match more EC pairs. Compared to MRC, CENTER exhibits a certain degree of performance decline in the EE task. We attribute this to MRC's prioritization of emotion in machine reading comprehension, enhancing emotion representation through multiple rounds of querying. However, by utilizing center event detection and reasoning methods to extract emotion-cause relationships, CENTER achieves significant improvements in the ECPE task. Similarly, PRG-MoE achieves several second-best results across the datasets, but its emphasis on cause classification and overall conversation enhancement leads to suboptimal EC pair extraction performance. Compared to the Joint model, CENTER demonstrates significant improvements in both EE and ECPE tasks. We attribute this to the fact that the Joint model employs a two-step extraction process and uses window-based pruning, leading to error accumulation. In contrast, our model leverages the concept of center events to effectively model utterance and utterance pair representations. This end-to-end approach not only enhances EE performance but also significantly improves the performance of handling ECPE task.

5.5 Ablation Studies

To validate the effectiveness of the CEA and CER graphs, we conduct corresponding ablation studies. The results are presented in Table 3.

Effect of CER We replace the utterance pair features aggregated under the guidance of the CER graph with direct concatenation and matching of utterance features. This diminishes the model's reasoning capability, leading to the second-worst performance in EE and ECPE. This demonstrates that the aggregation method based on center utterance pairs enhances the flow of key information and the ability to trace emotions in conversations.

Effect of CEA We replace the features aggregated under the guidance of CEA graph with embeddings. Without this graph, our model's performance in EE task significantly decreases, leading to a further decline in ECPE task. This indicates that the aggregation method based on center events is crucial for effective conversation modeling.

Effect of CEA & CER We remove both graphs and GNN aggregation method, retaining only the CED as an auxiliary task. The lack of critical information flow in utterance-level and candidate pair-level features results in the most significant performance decline in the EE and ECPE tasks. However, the performance still surpasses some baselines, indicating the effectiveness of addressing the CED task.

5.6 In-Depth Analyses

Effect of Center Event Distance To evaluate the impact of the center events' influence distance on the ECPEC task, we conduct additional experiments under varying interaction distances. As in Figure 3, the distance β from the center represents the distance between center events and surrounding nodes in the CEA graph, as well as the distance δ between center pairs and surrounding pairs in the CER graph. Specifically, we first fix $\beta = 3$ and vary δ. The results indicate that the optimal performance is achieved at $\delta = 2$. Increasing δ beyond this point introduces excessive interaction

Dataset	ECPE-D-DD						ECPE-D-IE					
Model	ECPE			EE			ECPE			EE		
	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)
ECPE-2D	48.34	49.34	47.37	74.44	71.91	**77.16**	19.42	45.17	12.37	**60.36**	78.16	**49.16**
ECPE-MLL	46.86	50.71	43.56	71.23	68.97	73.64	4.13	43.41	2.17	12.66	83.82	6.85
RANKCP	17.77	58.09	10.49	31.87	75.85	20.17	2.09	**65.91**	1.10	6.26	**91.75**	3.24
RECCON	39.68	49.31	33.19	-	-	-	7.92	46.52	4.33	-	-	-
ECPE-MM-R	50.22	49.36	51.32	74.31	75.41	73.48	13.91	34.30	8.82	33.23	76.23	21.43
MRC	52.47	52.19	52.86	**75.49**	**76.26**	74.96	20.96	59.59	16.08	39.54	90.32	29.49
PRG-MoE	57.26	58.95	55.67	73.86	71.76	73.86	28.90	51.95	20.02	57.29	85.58	43.06
CENTER	**62.50**	**64.94**	**60.24**	71.97	74.39	72.77	**52.94**	43.55	**67.50**	49.22	69.81	46.59

Table 1: The performance of CENTER and baseline models in ECPE and EE tasks. All models are trained exclusively on ECPE-D-DD dataset. To validate the model's generalization capability, we also test it using the ECPE-D-IE dataset.

Dataset	ConvECPE					
Model	ECPE			EE		
	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)
RANKCP	1.44	45.16	0.73	4.09	**83.87**	2.1
Joint-GW	46.43	38.16	**59.27**	66.10	58.28	76.34
Joint-XW	44.67	37.12	56.09	66.10	58.28	76.34
PRG-MoE	49.17	-	-	72.35	-	-
CENTER	**50.55**	**45.73**	56.50	**79.04**	78.67	**80.04**

Table 2: Experimental results on ConvECPE, which contains more EC pairs and longer conversations.

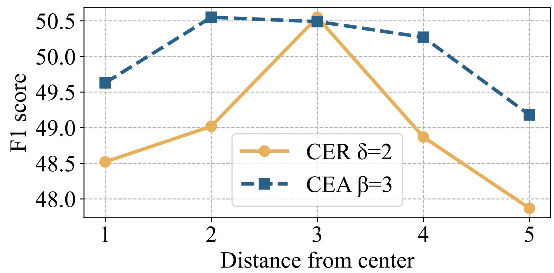

Figure 3: Influence of the center event distance.

noise, leading to a decline in performance. Next, we fix $\delta = 2$ and vary β. The results show that performance consistently improves with the increase of β, peaking at $\beta = 3$. Beyond this, the performance declines due to the over-reasoning of event influence. These findings suggest that the optimal performance is achieved at $\beta = 3$ and $\delta = 2$. Deviations from these values result in suboptimal performance.

Effect of Center Event Annotation To verify the importance of center event annotation, we vary α to observe their effect on the ECPE task. As in Figure 4, setting $\alpha = 1$ results in a significant decline in performance, which is attributed to the high overlap between center events and emotion cause annotations. As α increases, performance improves because most irrelevant events are filtered out. However, when α continues to increase, performance gradually declines. We attribute this to the uneven label distribution incurred by too few center events, particularly when α exceeds 6, leading to an average of less than one center event per conversation. The results indicate that setting $\alpha = 2$ allows for an optimal number of center events to be annotated, achieving peak performance.

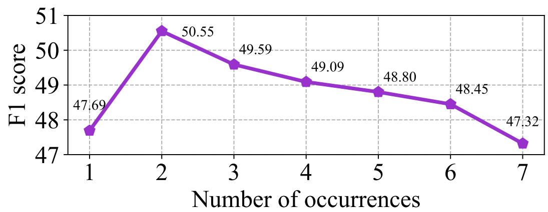

Figure 4: Results of the center event annotation experiment for different numbers of center event occurrences.

Effect of Center Event Detection To evaluate the effectiveness of the CED task, we conduct experiments through our CENTER on different datasets. CENTER demonstrates a clear advantage across all three datasets, with $F1$, P, and R metrics all exceeding 60%.

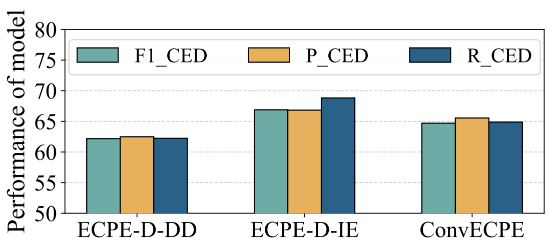

Figure 5: Performance of addressing CED task across different datasets.

In ECPE-D-DD, which has a sufficient number of conversations but an average length of only 10, the number of center events per conversation is relatively low, resulting in poorer model performance. Conversely, ECPE-D-IE, with an average conversation length of 41, contains an adequate number of center events, leading to the best extraction perfor-

Dataset	ConvECPE						RECCON					
Model	ECPE			EE			ECPE			EE		
	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)
w/o CEA & CER	48.54	44.61	53.21	66.88	70.28	76.34	56.59	48.60	**67.72**	69.06	69.63	69.08
w/o CER	49.64	44.99	55.35	67.10	68.78	76.16	56.88	52.95	61.45	69.52	69.59	69.50
w/o CEA	49.98	**46.06**	54.62	67.67	70.09	76.28	59.44	57.68	61.31	70.18	71.01	70.59
full model	**50.55**	45.73	**56.50**	**79.04**	**78.67**	**80.04**	**62.50**	**64.94**	60.24	**71.97**	**74.39**	**72.77**

Table 3: Ablation studies of CEA and CER.

mance by the model. In ConvECPE, the average conversation length is 49, but the higher number of conversations yields suboptimal performance in the CED task. These findings indicate that our model excels at extracting center events in longer conversations that contain multiple EC pairs.

Case Study We analyze examples from benchmark corpus to verify the effectiveness of CENTER in predicting center events. As in Table 4, u_2 and u_8 are marked as center events. Our model successfully identifies all EC pairs with a high confidence level of over 80%. In contrast, Joint-GW only captures (u_2, u_2) with a high confidence.

To reveal the causal reasoning capability, we visualize the results of test data. In Figure 6(a), our CENTER focuses on the beginning and end of long conversation, thereby identifying the center events accurately. In Figure 6(b), our model successfully captures emotions in both head and tail parts, whereas Joint-GW model, due to the lackage of guidance from center events, concentrates attention mainly on the tail part. Notably, for ECPE task, Joint-GW model, as in Figure 6(d), treats the conversation as a whole and enhances the utterances globally. It causes the pairing of emotions with largely irrelevant contexts. According to human language habits, emotions usually arise after their corresponding causes, so EC pairs above the diagonal are considered as incorrect extractions. We attribute this phenomenon to the interference of multiple events, which affects the model's reasoning direction and scope. Our CENTER, as in Figure 6(c), enhances utterance centrality within events, mitigating the issue of divergent pairing ranges. Through correct guidance from center pairings, most of the extracted EC pairs are distributed below the diagonal, alleviating the problem of unclear reasoning direction.

6 Conclusion

In this paper, we propose a novel **CE**nter eve**NT**-guided fram**E**wo**R**k (CENTER) for emotion-cause pair extraction in conversation. This framework detects center events within conversations and constructs a center event-aware network and a center event inference network. By enhancing the centrality of utterance and utterance pair representations, it mitigates the interference between different events and effectively captures the relationships between emotions and their corresponding causes. We evaluate our approach using the ECPE-D-DD, ECPE-D-IE, and ConvECPE datasets, demonstrating that CENTER outperforms other models in the ECPEC task.

...I want to know what to do to stop it from happening again (u_2: frustrated). This happens every two weeks (u_3)? Yeah. That my service just goes out (u_4: frustrated). You've lost your connection (u_5). ... And I'll reset the IP address. Direct connect the modem to my computer, and then reset everything (u_8: frustrated)...
CENTER: $(u_2, u_2, 98.1\%)(u_4, u_2, 84.3\%)(u_8, u_8, 81.5\%)$
Joint-GW: $(u_2, u_2, 92.1\%)(u_4, u_2, 59.8\%)(u_8, u_8, 58.6\%)$
Truth: $(u_2, u_2)(u_4, u_2)(u_8, u_8)$

Table 4: EC pair predictions by CENTER and Joint-GW. Red utterances indicate center events, while blue words denote annotated emotion utterances.

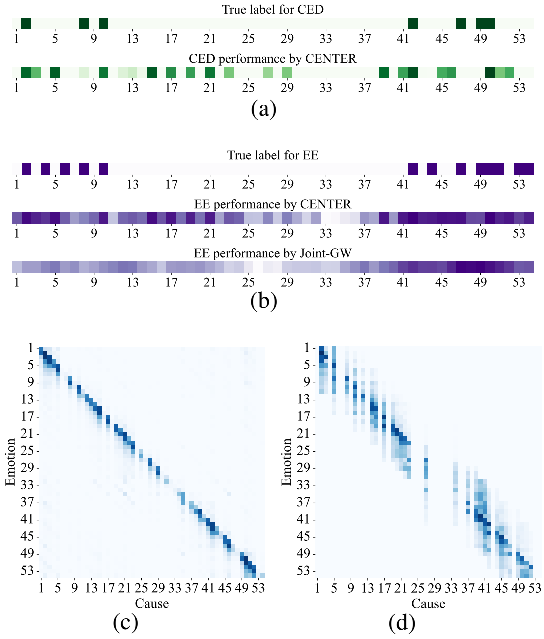

Figure 6: (a) True and predicted labels for CED task; (b) True labels, predicted emotions by CENTER and Joint-GW in EE task. Probability of correctly extracted EC pairs by CENTER (c) and Joint-GW (d). Darker color indicates higher confidence.

Limitations

Following previous studies, we develop CENTER to extract emotion-cause pairs from unannotated textual data. However, in pratice, emotions and their causes can originate from human visual, auditory, and social information. The multimodal ECPEC problem often presents additional challenges, which will be a focus of our future research.

Acknowledgements

The authors would like to thank the anonymous reviewers for their insightful comments. This work was supported in part by the National Natural Science Foundation of China (Grant nos. 62472117, 62102105, 62073263), the Fundamental Research Funds for the Central Universities (Grant no. D5000230112), the Open Research Subject of State Key Laboratory of Intelligent Game (Grant nos. ZBKF-24-02).

References

Jiaming An, Zixiang Ding, Ke Li, and Rui Xia. 2023. Global-view and speaker-aware emotion cause extraction in conversations. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. 2008. Iemocap: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42:335–359.

Shunjie Chen, Xiaochuan Shi, Jingye Li, Shengqiong Wu, Hao Fei, Fei Li, and Donghong Ji. 2022. Joint alignment of multi-task feature and label spaces for emotion cause pair extraction. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6955–6965.

Le Cheng, Peican Zhu, Chao Gao, Zhen Wang, and Xuelong Li. 2024. A heuristic framework for sources detection in social networks via graph convolutional networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.

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.

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.

Spencer C Evans and Shannon Shaughnessy. 2024. Emotion regulation as central to psychopathology across childhood and adolescence: a commentary on nobakht et al.(2023). *Journal of child psychology and psychiatry*, 65(3):354–357.

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.

Elliott M Hoey and Kobin H Kendrick. 2017. Conversation analysis. *Research methods in psycholinguistics and the neurobiology of language: A practical guide*, pages 151–173.

Jiaheng Hua, Xiaodong Cui, Xianghua Li, Keke Tang, and Peican Zhu. 2023. Multimodal fake news detection through data augmentation-based contrastive learning. *Applied Soft Computing*, 136:110125.

DongJin Jeong and Jinyeong Bak. 2023. Conversational emotion-cause pair extraction with guided mixture of experts. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3288–3298.

Jaehyeok Lee, DongJin Jeong, and JinYeong Bak. 2023. Enhancing emotionâ\euro "cause pair extraction in conversation with contextual information. *Authorea Preprints*.

Wei Li, Yang Li, Vlad Pandelea, Mengshi Ge, Luyao Zhu, and Erik Cambria. 2022. Ecpec: Emotion-cause pair extraction in conversations. *IEEE Transactions on Affective Computing*.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 986–995.

Chen Liu, Changyong Niu, Jinge Xie, Yuxiang Jia, and Hongying Zan. 2023. Emotion-cause pair extraction in conversations based on multi-turn mrc with position-aware gcn. In *2023 International Conference on Asian Language Processing (IALP)*, pages 25–30. IEEE.

Soujanya Poria, Navonil Majumder, Devamanyu Hazarika, Deepanway Ghosal, Rishabh Bhardwaj, Samson Yu Bai Jian, Pengfei Hong, Romila Ghosh, Abhinaba Roy, Niyati Chhaya, et al. 2021. Recognizing emotion cause in conversations. *Cognitive Computation*, 13:1317–1332.

Reini Schrama, Dorte Sindbjerg Martinsen, and Ellen Mastenbroek. 2024. European administrative networks during times of crisis: Exploring the temporal development of the internal market network solvit. *Regulation & Governance*.

Shivam Sharma, S Ramaneswaran, Md Shad Akhtar, and Tanmoy Chakraborty. 2024. Emotion-aware multimodal fusion for meme emotion detection. *IEEE Transactions on Affective Computing*.

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.

Margaret Wilson. 2002. Six views of embodied cognition. *Psychonomic bulletin & review*, 9:625–636.

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.

Xiaoheng Zhang, Weigang Cui, Bin Hu, and Yang Li. 2024. A multi-level alignment and cross-modal unified semantic graph refinement network for conversational emotion recognition. *IEEE Transactions on Affective Computing*.

Xiyang Zhang, Muhao Chen, and Jonathan May. 2021. Salience-aware event chain modeling for narrative understanding. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1418–1428.

Changzhi Zhou, Dandan Song, Jing Xu, and Zhijing Wu. 2022. A multi-turn machine reading comprehension framework with rethink mechanism for emotion-cause pair extraction. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6726–6735.

Zhipeng Zhou, Xingnan Zhou, Yudi Chen, and Haonan Qi. 2024. Evolution of online public opinions on major accidents: Implications for post-accident response based on social media network. *Expert Systems with Applications*, 235:121307.

Peican Zhu, Botao Wang, Keke Tang, Haifeng Zhang, Xiaodong Cui, and Zhen Wang. 2024. A knowledge-guided graph attention network for emotion-cause pair extraction. *Knowledge-Based Systems*, 286:111342.