Beyond Verbal Cues: Emotional Contagion Graph Network for Causal Emotion Entailment

Anonymous ACL submission

Abstract

Emotions are fundamental to conversational understanding. While significant advancements have been achieved in conversational emotion recognition and emotional response generation, recognizing the causes of eliciting emotions is less explored. Previous studies have primarily focused on identifying the causes of emotions by understanding verbal contextual utterances, overlooking that non-verbal emotional cues can elicit emotions. To address this issue, we develop an Emotional Contagion Graph Network (**ECGN**) that simulates the impact of non-verbal implicit emotions on the counterpart's emotions. To achieve this, we construct a heterogeneous graph that simulates the transmission of non-verbal emotions alongside verbal influences. By applying message passing between nodes, the constructed graph effectively models both the implicit emotional dynamics and explicit verbal interactions. We evaluate ECGN 's performance through extensive experiments on the benchmark dataset and compare it against multiple state-of-the-art models. Experimental results demonstrate the effectiveness of the proposed model.

1 Introduction

Emotions are widely present in human communication. It is crucial for humans to infer others' thoughts that are accompanied by the change of emotions. Understanding the mindset of others may involve not only understanding the contents and emotions of utterances but also digging out the potential causes of emotions. The ability of models to reason the cause of emotions is crucial in many contexts–it enhances the accuracy of responses by mining the intents, reduces the possible negative emotions for the opposite, and provides more substantive emotional support. Therefore, developing a model for recognizing the causes behind emotions is crucial for a more reliable dialogue system.

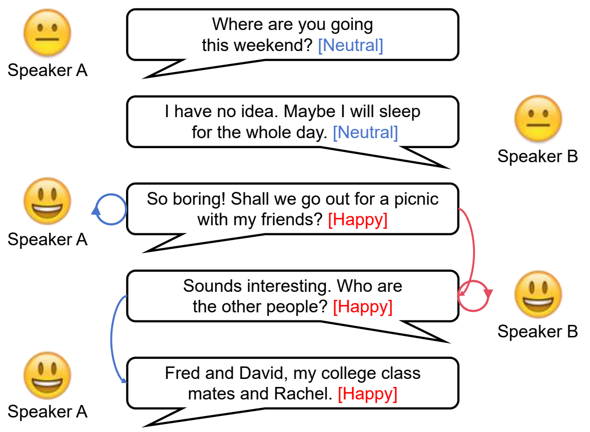

Figure 1: An example of a conversation in the RECCON-DD dataset. The arrow indicates the cause utterance for any target utterance.

Over the past few years, significant progress has been made in conversational emotion analysis. Previous studies (Hu et al., 2023; Song et al., 2022; Zhang et al., 2023a) on Emotion Recognition in Conversation (ERC) have primarily focused on labeling emotions for individual utterances, but this study often lacks recognizing the underlying emotional stimuli present in these utterances. To address this limitation, Poria et al. (Poria et al., 2021) introduce the Causal Emotion Entailment (CEE) task, which aims to determine which specific utterances stimulate a non-neutral emotional response in the target utterance. Compared to the Emotion Cause Extraction (ECE) (Lee et al., 2010; Gui et al., 2017, 2018; Fan et al., 2019) and Emotion Cause Pair Extraction (ECPE) (Xia and Ding, 2019; Hu et al., 2021b; Ding et al., 2020a; Wei et al., 2020) in discovering the cause triggers in a document, identifying conversational emotion causes is challenging because of the complex conversational structure and interactions. Many works focus on understanding verbal contextual utterances (Bosselut et al., 2019; Zhao et al., 2023a; Zhou et al., 2024a), but neglecting emotions themselves can also be the cause of emotions on the counterparts beyond verbal utterances. For example, in Figure 1, speaker B's emotion is attributed to speaker A's happiness, which makes it difficult to reason from only verbal utterances.

To address this challenge, we turn to *Emotion Contagion Theory* (Hatfield et al., 1993, 2011; Liu et al., 2024), which demonstrates a process in which a person or group influences the emotions or behavior of another person through the conscious or unconscious induction of emotion states and behavioral attitudes. This means that the emotions of counterparts can elicit emotions without any linguistic cues. Generally, emotional contagion can be either implicit (Tee, 2015; Wróbel and Imbir, 2019), which relies on mainly non-verbal communication (Schoenewolf, 1990), or explicit, which affects the emotions of counterparts by content (Kelly and Barsade, 2001).

Inspired by the emotional contagion process, we propose a novel Emotional Contagion Graph Network (ECGN) to identify emotion causes, which simulates the emotional contagion process through both explicit and implicit emotional pathways. Explicit emotional contagion is modeled through the interactions of verbal utterances called verbal cues, while implicit emotional contagion is captured through the dynamics of non-verbal emotional labels called non-verbal cues. ECGN consists of several key steps. First, ECGN extracts both non-verbal and verbal cues from the conversational context and constructs a heterogeneous conversational graph. This graph captures two types of interactions: implicit emotional contagion from non-verbal emotional labels, and explicit emotional contagion from verbal utterances. Moreover, ECGN effectively transmits the dynamics within and between non-verbal and verbal cues through relational graph neural networks. Finally, a classifier predicts the emotion cause based on the integrated information.

To evaluate the proposed ECGN, we conduct extensive experiments on the RECCON-DD dataset. Results consistently demonstrate that ECGN effectively promotes the detection of causal utterances from the target utterance.

2 Related Work

Causal Emotion Entailment Poria et al.(Poria et al., 2021) introduced the RECCON task to identify the causes of a speaker's emotions in conversations. Based on the granularity of causes, it is divided into the CEE task (utterance-level causes) and the CSE task (phrase-level causes). Their approach concatenates potential causal utterances with the target utterance but overlooks conversational interactions. To improve this, recent works focus on contextual understanding. For instance, $\text{MuTEC}_{\text{CEE}}$(Bhat and Modi, 2023) employs multi-task learning to model conversational context, KEC (Li et al., 2022) and KBCIN (Zhao et al., 2023a) incorporate commonsense knowledge via directed acyclic graphs, PAGE (Gu et al., 2023) leverages positional relationships, TSAM (Zhang et al., 2022) integrates attention for intra- and inter-speaker influences, and recent works (Huang et al., 2024; Zhou et al., 2024b) explores reasoning with Large Language Models (LLMs). The above works take emotion as auxiliary information accompanied by the utterances and pay attention to the verbal information, but neglect the effects of non-verbal emotional dynamics themselves. ECGN recognizes and bridges this gap.

Emotion Recognition in Conversations. Emotion Recognition in Conversations (ERC) is a highly relevant task to CEE, which involves identifying emotion categories for the target utterance. ERC needs to predict unknown emotions in the conversation, differentiating from CEE which emotion is already known. Most of the present works adopt graph-based and sequence-based methods. The former (Ghosal et al., 2019; Ishiwatari et al., 2020; Hu et al., 2021c; Shen et al., 2021; Zhang et al., 2023a) builds a graph to handle interactions between utterances and speakers.

Another group of works exploits transformers and recurrent models to learn the interactions between utterances (Majumder et al., 2019; Hu et al., 2021a; Liu et al., 2022). Commensense Knowledge is explored by KET (Zhong et al., 2019). Contrastive learning methods are also prevailing for ERC (Lewis et al., 2019; Song et al., 2022; Yu et al., 2024) which separates utterances from representation space. The above approaches use encoders to extract utterance representations. Several recent works explore LLMs (Lei et al., 2023; Zhang et al., 2023b; Wu et al., 2024b) for ERC tasks. Unlike ERC methods that only rely on contextual utterances for prediction, ECGN introduces contextual emotional interactions to enhance cause predictions.

Emotion Cause (Pair) Extraction. Emotion cause extraction (ECE) aims to identify the causes or stimuli that trigger the emotions in each sentence in a long document, which was first proposed by (Lee et al., 2010). Early studies are devoted to designing rule-based methods (Chen et al., 2010; Neviarouskaya and Aono, 2013). Recent works propose various deep networks to tackle this task (Cheng et al., 2017; Zheng et al., 2022).

The ECE task has been researched for nearly a decade, but its reliance on additional emotion annotations limits its applicability in real-world scenarios. To this end, Emotion-Cause Pair Extraction (ECPE) (Xia and Ding, 2019) is proposed to extract all pairs of emotions and corresponding causes in a document without emotion annotation. They propose a two-step framework to perform ECPE. In the following work, ECPE-2D (Ding et al., 2020a) utilizes a 2D Transformer to model clause pairs. Sequence-labeling scheme is also constructed for ECPE (Yuan et al., 2020; Cheng et al., 2021; Wu et al., 2023). Recent works have started to explore the strong reasoning and understanding abilities of LLMs for ECPE (Wu et al., 2024a; Gu et al., 2024). Unlike these two tasks, which predict emotional causes in documents, ECGN focuses on capturing the complex emotional interactions between interlocutors in real-life conversation scenarios.

3 Methodology

3.1 Problem Definition

We start by formulating the CEE task. Consider a conversation as a sequence of utterances with speakers and emotions as $\boldsymbol{C} = \{(u_1, e_1, s_1), (u_2, e_2, s_2), \dots, (u_T, s_T, e_T)\}$, where u_t is the utterance at the timestamp t in the conversation, $e_t \in \{$*happy*, *angry*, *sad*, *disgusted*, *fearful*, *surprised*, *neutral*$\}$ is the corresponding emotion label, and s_t is the speaker identity of u_t. The goal of CEE is to identify the set of utterances $\{u_i\}(i \le t)$ which are the emotion causes of u_t in the conversation history if u_t is a non-neutral utterance.

3.2 Model Overview

Figure 2 shows the pipeline of ECGN. It consists of several key components designed to simulate explicit and implicit emotional contagion.

The first component is to encode utterances and emotions, using a language model to extract textual representations while generating emotion representations with emotion labels.

The second component is the construction of the emotion contagion graph with the extracted representations. The emotion contagion graph contains the explicit and implicit ones. The explicit emotion contagion graph simulates the triggering of emotions by language content in the conversational context. The implicit emotion contagion graph simulates the influence of non-verbal cues on emotions in the conversational context, which are represented by emotion encodings. In this graph, vertices represent utterances or emotions. Interactions within emotion nodes pass unconscious contagion silently, dynamics between emotion and utterances or utterances themselves actively trigger emotions. To learn the transition process, we employ relational graph neural networks and graph transformers to integrate such interactive relationships, which allows ECGN to capture causes in terms of contents and emotions.

The last component combines both the learned emotional and utterance information together to construct cause representations which are used to distinguish the causal and non-causal utterances.

3.3 Context Encoding with Emotions

Given an utterance history $\boldsymbol{U} = \{u_1, u_2, \dots, u_T\}$ and emotion history $\boldsymbol{E} = \{e_1, e_2, \dots, e_T\}$, where T is the number of utterances contained in a conversation, we use a language model to extract verbal utterance representations. More specifically, we add special tokens such as [CLS] and [SEP] which serve as markers to indicate the beginning and end of each utterance. To facilitate verbal utterance representations with emotion semantics, we construct a prompt:

$$X_i(s_i, u_i, e_i) = s_i\ e_i\ says:\ u_i, \tag{1}$$

where $X(\cdot,\cdot,\cdot)$ transforms each utterance into an implicit emotion-rich form. For instance, an utterance can be organized as *John happily says: I'm so glad I bought this watch!* Finally, we concatenate all the prompts in a conversation and feed them into a pretrained language model:

$$\boldsymbol{H}^t = \mathrm{PLM}([\mathrm{CLS}]X_1[\mathrm{SEP}]\dots[\mathrm{CLS}]X_T[\mathrm{SEP}]), \tag{2}$$

Where the conversational textual representations $\boldsymbol{H}^t = \mathrm{Concat}(h_1^t, h_2^t, \dots, h_T^t) \in \mathbb{R}^{T\times d}$ is the concatenation of all last hidden states at the [CLS] token's position, d is the dimension of hidden states.

3.4 Emotion Encoding

To leverage the non-verbal cues, we generate emotional representations at each time step with emo-

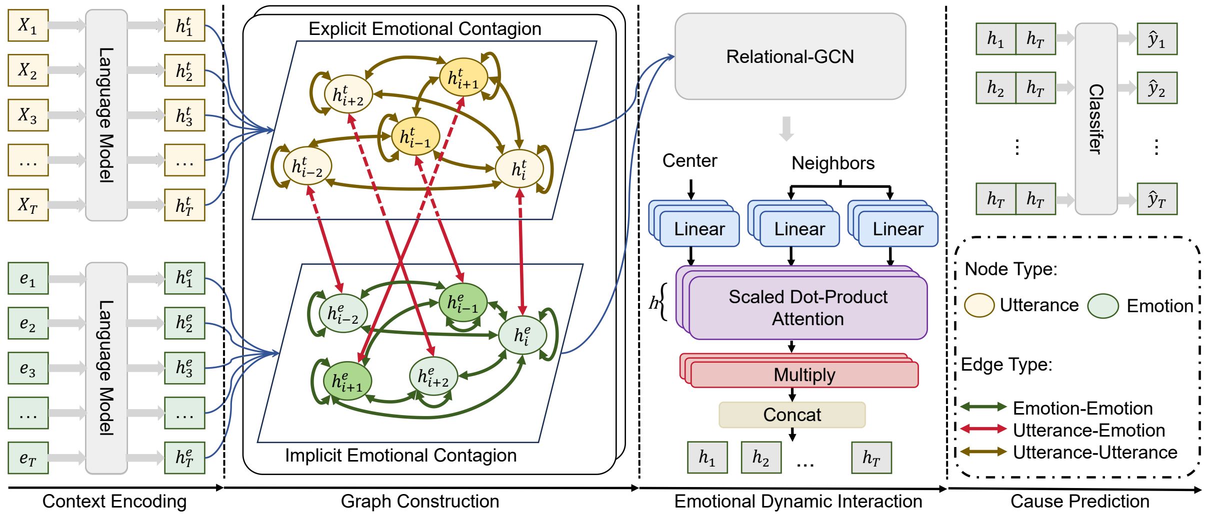

Figure 2: Overview of our proposed method. The structure of the model is shown at the bottom. First, we input the utterances and emotions into the language model to obtain the encodings of them. Then we construct a heterogenous graph modeling the complex interaction relations, including the simulated implicit and explicit emotional contagion. Having the heterogeneous graph, we build up a graph-learning model for learning dynamics between different node features. Different relations indicate distinct information passing needed. Finally, a cause prediction module is employed to identify the causes of emotions within the conversation.

tion labels. Given a candidate set of emotion labels $\boldsymbol{S} = \{e_1, e_2, \ldots, e_{|S|}\}$, each emotion e_i can be represented as an embedding vector:

$$g_i = \mathrm{PLM}(e_i), \tag{3}$$

Where $g_i \in \mathbb{R}^d$, and then we concatenate emotion representations as a lookup table $\boldsymbol{P} = \mathrm{Concat}(g_1, g_2, \ldots, g_{|S|}) \in \mathbb{R}^{|S| \times d}$. These emotion representations are initialized with the original pre-trained language model. Given an emotion history $\boldsymbol{E} = \{e_1, e_2, \ldots, e_T\}$, we generate the representations for e_t through:

$$h_t^e = \mathrm{Lookup}(\boldsymbol{P}, e_t), \tag{4}$$

Where $h_t^e \in \mathbb{R}^d$ is the representation of e_t, concatenating them can get the conversation emotional representation $\boldsymbol{H^e} = \mathrm{Concat}(h_1^e, h_2^e, \ldots, h_T^e) \in \mathbb{R}^{T \times d}$.

3.5 Emotion Contagion Graph Construction

To mimic both explicit and implicit emotional contagion processes, we construct a heterogeneous graph for each conversation history. We denotes a graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$, with vertices $v_i \in \mathcal{V}$, edges $\epsilon_k \in \mathcal{E}$, $r_{ij} \in \mathcal{R}$ is the relation type between v_i and v_j.

Our graph $\mathcal{G}$ contains two types of nodes:

Utterance node: We consider ith utterance in the conversation as a node $v_i^t \in \mathcal{V}^t$, whose representations are initialized with its utterance-level features $h_i^{t,(1)} = h_i^t$ for any time step i.

Emotion node: We treat each emotion in the conversation as a node $v_i^e \in \mathcal{V}^e$ and initialize the representations with $h_i^{e,(1)} = h_i^e$.

Then the set of nodes can be represented as:

$$\mathcal{V} = \mathcal{V}^t \cup \mathcal{V}^e, \tag{5}$$

where utterance node $\mathcal{V}^t = \{u_i\}$, emotion node $\mathcal{V}^e = \{e_i\}$ and $i \in [1, T]$.

Our graph $\mathcal{G}$ contains three types of edges:

Emotion-Emotion edge: To simulate the non-verbal implicit emotion contagion, we connect the current utterance i with a past context window size of p and a future context window size of f. We believe that the adjacent utterances of utterance i have the most significant impact. For the sake of message passing between utternaces, each utterance vertex has an edge with the timestamp i utterance of the past: $v_{i-p}^e, v_{i-p+1}^e, \ldots, v_{i-1}^e$, the future utterances: $v_{i+1}^e, v_{i+2}^e, \ldots, v_{i+f}^e$ and v_i^e itself. These edges are denoted as $\mathcal{E}^{uu} = \{(e_i, e_j), (e_j, e_i)\}$, where $\max(0, i-p) \le j \le \min(i+f, T)$, and $i \in [1, T]$. $\mathcal{E}^{ee}$ enables the non-verbal emotional information to transmit intra- and inter- speakers.

Utterance-Utterance edge: Verbal communications may elicit emotions, we connect utterance nodes to construct explicit emotion contagion graph to capture the conscious emotions as $\mathcal{E}^{ee} = \{(u_i, u_j), (u_j, u_i)\}$, which allows utterances themselves to cause the emotions.

Utterance-Emotion edge: To further establish the interactions between emotions and utterances, we connect utterance node i with its emotion node to model the interaction within a speaker. The edges can be represented as $\mathcal{E}^{ue} = \{(u_i, e_i), (e_i, u_i)\}$, which connects the mutual effect of emotion and utterance within a speaker. Besides, multi-hop message passing enables such an effect to spread across speakers.

Then the set of edges can be represented as:

$$\mathcal{E} = \mathcal{E}^{uu} \cup \mathcal{E}^{ue} \cup \mathcal{E}^{ee}, \tag{6}$$

where $\mathcal{E}$ includes non-verbal implicit emotional dynamics $\mathcal{E}^{ee}$, and verbal explicit emotional dynamics $\mathcal{E}^{ue}$ and $\mathcal{E}^{uu}$.

3.6 Emotional Dynamic Interaction

To effectively pass the information between nodes and learn the dynamics, we utilize R-GCN (Schlichtkrull et al., 2018), which can integrate different relationships between vertices and learn the node representations:

$$h_i^{*,(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_i^r}\frac{1}{|\mathcal{N}_i^r|}W_r^l h_j^{*,(l)} + W_0^l h_i^{*,(l)}\right), \tag{7}$$

Where $\mathcal{N}_i^r$ is the set of neighboring nodes of node i under the relationship r, $h_i^{*,l}$ is the representations for node i which is either emotional or textual node after layer $l \in [1, L]$, $W_r \in \mathbb{R}^{d_1\times d_2}$ and $W_0 \in \mathbb{R}^{d_1\times d_2}$ are learnable parameters to transform the neighborhood information within relationship r. R-GCN layers not only transmit the emotional dynamics within emotional nodes and utterance nodes but also capture the interactions between emotions and utterances. Then the node representations are mapped to a shared representation space. To step further, We exploit GraphTransformer (Shi et al., 2020) to learn rich utterance representations. More specifically, the representations can be calculated as follows:

$$h_i^{*,(l+1)} = W_1 h_i^{*,(l)} + \sum_{j\in\mathcal{N}_i}\alpha_{i,j}W_2 h_j^{*,(l)}, \tag{8}$$

$$\alpha_{i,j} = \text{Softmax}\left(\frac{(W_3 h_i^{*,(l)})(W_4 h_j^{*,(l)})}{\sqrt{d}}\right), \tag{9}$$

where the $\alpha_{i,j}$ is the attention coefficient and d is the hidden size. The final utterance representation is then obtained by concatenating the emotional and utterance node representations at layer L:

$$h_i = \text{Concat}(h_i^{t,(L)}, h_i^{e,(L)}), \tag{10}$$

3.7 Cause Prediction

To predict the cause of the target utterance, we obtain the cause representation c_t by concatenating the utterance representations between the target utterance T and historical utterance i:

$$c_i = \text{ReLU}(W_5[h_i; h_t] + b_1), \tag{11}$$

$$\hat{y}_i = \text{Sigmoid}(W_6 c_i + b_2), \tag{12}$$

Where $\hat{y}_i$ is the probability of utterance i is the cause of emotion in the target utterance. $W_5 \in \mathbb{R}^{d_2\times d_3}$, $W_6 \in \mathbb{R}^{d_3\times 1}$ $b_1 \in \mathbb{R}^{d_3}$, $b_2 \in \mathbb{R}$ are learnable parameters. CrossEntropy loss function is adopted for optimization.

4 Experimental settings

4.1 Dataset and evaluation metrics

Dataset. We conduct experiments on a benchmark dataset RECCON-DD (Poria et al., 2021), which is built upon the DailyDialog dataset (Li et al., 2017). The detail of the RECCON-DD dataset is shown in Table 2. The data samples used for the experiments were constructed by pairing each non-neutral emotional utterance with its historical utterances, including itself, one by one. If a historical utterance was found to be the cause of an emotional utterance, the utterance pair was labeled as positive; otherwise, the pair was labeled as negative. Besides, we analyze the distribution of cause pairs in the conversations, as shown in Figure 3, about 80 % of emotion causes are located within two-time steps before the target utterances, indicating the high impact of neighbor emotions and utterances.

Metrics. Following previous work (Poria et al., 2021), we adopt the macro-averaged F1 score for evaluating performance. Also, the F1 score for positive and negative samples is reported.

4.2 Baselines

For a comprehensive evaluation, We compare our method with the following baselines:

(1) ECE and ECPE methods: **KAG** (Yan et al., 2021) that alleviates positional bias problem and improves the semantic dependencies using CSK; **Adapted** (Turcan et al., 2021) jointly detecting emotion and emotion cause enhanced by CSK; **ECPE-2D** (Ding et al., 2020a) uses the 2D representation to simulate emotion-cause pairs interactions with a 2D transformer; **ECPE-MLL** (Ding

Methods	Negative F1 (%)	Positive F1 (%)	Macro F1 (%)
ECE Methods			
KAG (Yan et al., 2021)	86.35	58.18	72.26
Adapted (Turcan et al., 2021)	88.18	64.53	76.36
ECPE Methods			
ECPE-2D♣ (Ding et al., 2020a)	94.96	55.50	75.23
ECPE-MLL♣ (Ding et al., 2020b)	94.68	48.48	71.59
RankCP♣ (Wei et al., 2020)	97.30	33.00	65.15
HCL-ECPE (Hu et al., 2024)	88.52	66.47	76.93
CEE Methods			
ChatGPT 0-shot† (Zhao et al., 2023b)	85.25	51.33	68.29
ChatGPT 1-shot† (Zhao et al., 2023b)	82.10	52.84	67.47
MuTEC$_{\text{CEE}}$ (Bhat and Modi, 2023)	83.46	61.62	72.54
PAGE (Gu et al., 2023)	89.42	65.20	77.02
KEC † (Li et al., 2022)	88.85	66.55	77.70
KBCIN † (Zhao et al., 2023a)	89.65	68.59	79.12
TSAM (Zhang et al., 2022)	<u>89.75</u>	68.59	79.17
DAM (Kong et al., 2023)	89.35	<u>69.32</u>	<u>79.34</u>
ECGN(ours)	**90.57**$^{*}_{\pm 0.19}$	**69.78**$_{\pm 0.54}$	**80.17**$^{*}_{\pm 0.24}$

Table 1: Performance of baselines and ours on the RECCON-DD dataset. †and ♣ denotes the results obtained from (Zhao et al., 2023b) and (Poria et al., 2021).* represents our method is significant statistically (p-value < 0.05).

RECCON-DD	Train	Dev	Test
Positive Pairs	7026	328	1767
Negative Pairs	20558	838	5296
Number of Dialogues	834	47	225

Table 2: Statistics of the RECCON-DD dataset.

et al., 2020b) extends ECPE-2D by incorporating multi-label learning to extract emotion cause. **RankCP** (Wei et al., 2020) emphasizes inter-clauses modeling with a ranking perspective for ECPE; **HCL-ECPE** (Hu et al., 2024) introduces hierarchical contrastive learning for ECPE.

(2) CEE methods: **KEC** (Li et al., 2022) injects commonsense knowledge for a directed acyclic graph; **KBCIN** (Zhao et al., 2023a) leverage event-centered commonsense knowledge (Bosselut et al., 2019) to capture the inter-utterance relationships; **PAGE** (Gu et al., 2023): A position-aware graph-based model distinguishes different speakers for causal entailment. **MuTEC$_{\text{CEE}}$** (Bhat and Modi, 2023) exploits multi-task learning for extracting conversational emotions, emotion causes, and entailment. **TSAM** (Zhang et al., 2022) proposes a two-stream attention model to separately model the emotions and speakers. **In-Context-Learning** (Zhao et al., 2023b): uses ChatGPT (GPT-3.5-turbo-0301) with few-shot demonstrations to test the CEE performance.

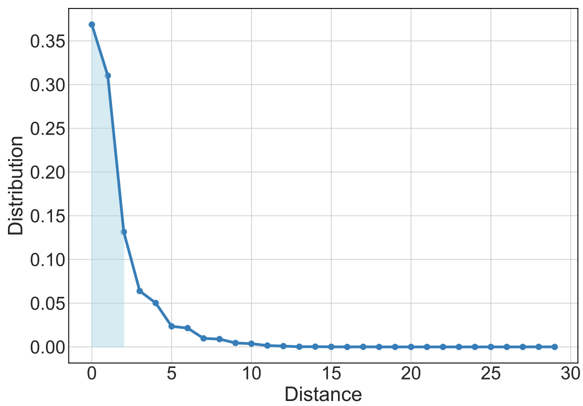

Figure 3: Distribution of the distance between positive pairs. The distance denotes the temporal difference between the causal and target utterances. The blue part indicates the portion that distance is less than 2.

4.3 Implementation Details

We conduct all the experiments based on Roberta-base (Liu et al., 2019) for a fair comparison. All experiments are run on a single A100 80GB GPU under Pytorch (Paszke et al., 2019) and Torch Geometric (Fey and Lenssen, 2019) framework for five repetitions.

5 Results and analysis

Table 1 shows the performance comparison of ECGN with state-of-the-art methods. It is observed that ECE and ECPE methods perform worse than CEE methods. For example, Adapted (Turcan

Layers	Neg. F1	Pos. F1	Macro F1
1	90.22	68.27	79.24
2	**90.57**	**69.78**	**80.17**
3	90.38	68.67	79.52
4	90.11	68.70	79.40
5	90.43	68.15	79.29
6	90.46	69.02	79.74

Table 3: The performance of using a different number of R-GCN layers under the window size 2.

et al., 2021) serves as the best method among ECE and ECPE methods that can achieve 76.36% macro F1 score, which performs mediocrely among CEE methods. ECGN surpasses 3.81%, indicating the effectiveness of our design for CEE. ECPE and ECE models fail to leverage available emotion labels of utterances and model utterance and emotion interactions in conversation structure, leading to their worse performance.

Compared to CEE methods, we outperform the second-best model TSAM by 1% overall and KBCIN which incorporates external knowledge. While the second-best baseline DAM incorporates discourse parsing to enhance long-distance cause classification. However, as shown in Figure 3, most causal relations occur in the local context, highlighting the effectiveness of our emotional contagion simulation in the local context for improving the overall performance. In addition, ECGN has an overwhelming performance advantage over ChatGPT, the possible reason is that ChatGPT is not well aligned to the complex data annotation for CEE. The experimental results are significantly better than the baselines under the t-test, which validates the robustness of ECGN.

5.1 Ablation Study

We conducted a series of ablation studies on ECGN. The results, as depicted in Table 4, highlight the criticality of each element in our approach. Removing the graph structure and concatenating the emotional representations with utterance representations, as well as removing the implicit emotional graph part harm the performance to a large extent. This result demonstrates the effectiveness of ECGN in dealing with emotional causes happening in the local time with the mutual influence of emotions. In addition, removing e_i in X_i decreases Macro F1 0.3%, indicating the importance of the influence of emotional state for utterance features.

Emotion	Graph	Neg. F1	Pos. F1	Macro F1
✓	✓	**90.57**	**69.78**	**80.17**
✓	✗	89.98	68.48	79.23
✗	✓	89.77	68.45	79.11
✗	✗	89.38	68.29	79.04

Table 4: Effects of different components on the performa

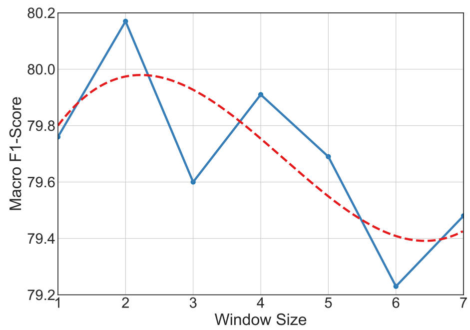

Figure 4: Test results under different window sizes on the RECCON-DD dataset. The red dashed line depicts the trend generated by polynomial fitting.

5.2 Effect of the Number of R-GCN Layers

We conducted an investigation into the influence of the number of layers in the Relational-GCN architecture. The findings, as presented in Table 3, indicate that incorporating more global information with the deeper graph networks introduces confused context since most causes are adjacent to the target utterances. Besides, using deep graph neural networks resulted in performance degradation due to over-smoothing, as reported in previous studies (Kipf and Welling, 2016). Experiments show that employing two layers of Relational-GCN proved to be a balanced approach. The decrease in performance when the number of layers is excessive possibly owing to redundant information.

5.3 Effect of Contextual Window Size

We also report the performance under a large range of window sizes. In Figure 4, the trend of performance has a trend that first increases with the growth of window size. The increase in window size should have a larger perception field to aggregate more information, however, the intrinsic property of a conversation decides that a non-neutral emotion is more likely to be triggered by the neighbor's utterances and distant utterances may introduce irrelevant information (Ding et al., 2019).

Turn	Speaker	Utterance	Emotion	w/IEC	w/o IEC	Label
1	A	Hey George, how is your chicken?	Neutral	-	-	-
2	B	My chicken tastes all right, but it is pretty dry. How is your fish?	Neutral	-	-	-
3	A	My fish is pretty dry too.	Sad	[3]	[3]	[3]
4	B	It's almost as if this food has been sitting a little too long. It doesn't seem fresh.	Sad	[3, 4]	[4]	[3, 4]

Table 5: Case study of a conversation instance shows that non-verbal implicit emotional dynamics (IEC) enables the model to rectify incorrect cause predictions.

5.4 Can Implicit Emotional Dynamics Identify Causes?

To verify how much the causes depend on the implicit emotion dynamics, we remove utterance nodes and only retain emotion nodes to determine the causes. As reported in Figure 5, the performance works almost equal to ChatGPT 1-shot and slightly better than RankCP on the Macro F1 score. It indicates that implicit non-verbal emotional dynamics play a critical role in causal emotion entailment, as demonstrated by its performance even in the absence of explicit utterance-level information.

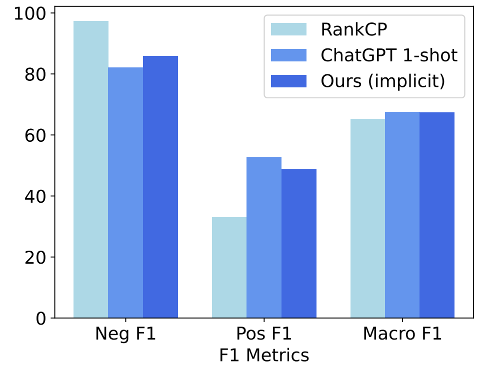

Figure 5: Performance comparison with baselines using only non-verbal implicit emotional dynamics.

6 Case Study

We exhibit a case study in Table 5. In this case, the speaker S_A first feels sad when finding his chicken tastes dry, which elicits a sad emotion. Subsequently, speaker S_B turns his emotion from neutral to sad not only because speaker S_B's tastes are dry, but also is influenced by S_A. As shown in Table 5, removing the implicit emotion contagion network enables the model to understand only utterance semantic, which overlooks speaker S_A's sadness as an emotion trigger of S_B, leading to mistaken classification. Besides, by analyzing our predicted emotion causes, we find that the following aspects mainly cause prediction errors: First, the causal relationship happens when the distance between the target utterance and the cause utterance is large. This type of error presents a challenge to trace back to distant previous dialog history. The second category is sudden emotional change, which confuses the model about causal relations. Solving these two kinds of errors needs a more fine-grained reasoning process to understand the mental state, e.g. Theory of Mind (ToM) (Ma et al., 2023; Jin et al., 2024; Strachan et al., 2024) because conversational context is simple, which is unable to provide sufficient information to accurately identify those causes, and external memory to retrieve relevant information from long conversation histroy (Zhong et al., 2024; Maharana et al., 2024) also have potential capturing distant information.

7 Conclusion and Future Work

In this paper, we introduce the Emotional Contagion Graph Network (ECGN) as an innovative model that improves causal emotion entailment in conversations by simulating the impact of non-verbal implicit emotions on the counterpart's emotions. By drawing inspiration from the Emotional Contagion Theory, the model constructs a heterogeneous conversational graph to capture explicit and implicit emotional dynamics between speakers, simulating the influence between emotions themselves and interactions with utterances in a conversation to determine the emotion causes. Extensive experimentation on the RECCON-DD dataset demonstrates superior performance of ECGN over state-of-the-art baselines. Ablation studies and evaluations further validate the robustness and effectiveness of the approach, as well as the importance of implicit emotional dynamics in the conversation for causal emotion entailment.

8 Limitations

Our method primarily considers the conversation-level emotional and utterance information transmission but does not consider mental activities among different speakers. We believe future work can benefit from a cognitive science perspective to enable more complex mental state reasoning for CEE task. Besides, multimodal information such as video and audio can better help recognize non-verbal emotional contagion such as facial expressions, body language, posture, tone of voice, and other non-verbal signals. Compared to the information transmission between emotion labels and utterance representations, it can better reflect emotional contagion in real-life scenarios. This leaves a large room for future benchmarks and the development of new methods.

9 Ethical Consideration

This study focuses on causal emotion entailment in conversations, which involves processing and analyzing emotional expressions. Ethical concerns include ensuring data privacy and avoiding unintended biases in the model. The dataset used is publicly available, and we adhere to ethical guidelines for handling conversational data. While our model aims to improve emotional understanding, it should not be used for manipulative or deceptive purposes.

References

Ashwani Bhat and Ashutosh Modi. 2023. Multi-task learning framework for extracting emotion cause span and entailment in conversations. In *Transfer Learning for Natural Language Processing Workshop*, pages 33–51. PMLR.

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. Comet: Commonsense transformers for automatic knowledge graph construction. *arXiv preprint arXiv:1906.05317*.

Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion cause detection with linguistic constructions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 179–187.

Xiyao Cheng, Ying Chen, Bixiao Cheng, Shoushan Li, and Guodong Zhou. 2017. An emotion cause corpus for chinese microblogs with multiple-user structures. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 17(1):1–19.

Zifeng Cheng, Zhiwei Jiang, Yafeng Yin, Na Li, and Qing Gu. 2021. A unified target-oriented sequence-to-sequence model for emotion-cause pair extraction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:2779–2791.

Zixiang Ding, Huihui He, Mengran Zhang, and Rui Xia. 2019. 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*, volume 33, pages 6343–6350.

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.

Chuang Fan, Hongyu Yan, Jiachen Du, Lin Gui, Lidong Bing, Min Yang, Ruifeng Xu, and Ruibin Mao. 2019. A knowledge regularized hierarchical approach for emotion cause analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5618–5628. Association for Computational Linguistics.

Matthias Fey and Jan Eric Lenssen. 2019. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*.

Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. Dialoguegcn: A graph convolutional neural network for emotion recognition in conversation. *arXiv preprint arXiv:1908.11540*.

Xiaojie Gu, Renze Lou, Lin Sun, and Shangxin Li. 2023. Page: A position-aware graph-based model for emotion cause entailment in conversation. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

Xue Gu, Zhihan Zhou, Ziyao Meng, Jian Li, Tiago Gomes, Adriano Tavares, and Hao Xu. 2024. Emoprompt-ecpe: Emotion knowledge-aware prompt-tuning for emotion-cause pair extraction. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5678–5688.

Lin Gui, Jiannan Hu, Yulan He, Ruifeng Xu, Qin Lu, and Jiachen Du. 2017. A question answering approach to emotion cause extraction. *arXiv preprint arXiv:1708.05482*.

Lin Gui, Ruifeng Xu, Dongyin Wu, Qin Lu, and Yu Zhou. 2018. Event-driven emotion cause extraction with corpus construction. In *Social Media Content Analysis: Natural Language Processing and Beyond*, pages 145–160. World Scientific.

Elaine Hatfield, John T Cacioppo, and Richard L Rapson. 1993. Emotional contagion. *Current directions in psychological science*, 2(3):96–100.

Elaine Hatfield, Richard L Rapson, and Yen-Chi L Le. 2011. Emotional contagion and empathy. *The social neuroscience of empathy*, 19.

Dou Hu, Yinan Bao, Lingwei Wei, Wei Zhou, and Songlin Hu. 2023. Supervised adversarial contrastive learning for emotion recognition in conversations. *arXiv preprint arXiv:2306.01505*.

Dou Hu, Lingwei Wei, and Xiaoyong Huai. 2021a. Dialoguecrn: Contextual reasoning networks for emotion recognition in conversations. *arXiv preprint arXiv:2106.01978*.

Guimin Hu, Guangming Lu, and Yi Zhao. 2021b. Bidirectional hierarchical attention networks based on document-level context for emotion cause extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 558–568.

Guimin Hu, Yi Zhao, and Guangming Lu. 2024. Improving representation with hierarchical contrastive learning for emotion-cause pair extraction. *IEEE Transactions on Affective Computing*.

Jingwen Hu, Yuchen Liu, Jinming Zhao, and Qin Jin. 2021c. Mmgcn: Multimodal fusion via deep graph convolution network for emotion recognition in conversation. *arXiv preprint arXiv:2107.06779*.

Zhaopei Huang, Jinming Zhao, and Qin Jin. 2024. Ecrchain: Advancing generative language models to better emotion-cause reasoners through reasoning chains. *arXiv preprint arXiv:2405.10860*.

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.

Chuanyang Jin, Yutong Wu, Jing Cao, Jiannan Xiang, Yen-Ling Kuo, Zhiting Hu, Tomer Ullman, Antonio Torralba, Joshua B Tenenbaum, and Tianmin Shu. 2024. Mmtom-qa: Multimodal theory of mind question answering. *arXiv preprint arXiv:2401.08743*.

Janice R Kelly and Sigal G Barsade. 2001. Mood and emotions in small groups and work teams. *Organizational behavior and human decision processes*, 86(1):99–130.

Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Dexin Kong, Nan Yu, Yun Yuan, Xin Shi, Chen Gong, and Guohong Fu. 2023. Discourse-aware causal emotion entailment. In *International Conference on Neural Information Processing*, pages 389–401. Springer.

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.

Shanglin Lei, Guanting Dong, Xiaoping Wang, Keheng Wang, and Sirui Wang. 2023. Instructerc: Reforming emotion recognition in conversation with a retrieval multi-task llms framework. *arXiv preprint arXiv:2309.11911*.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Jiangnan Li, Fandong Meng, Zheng Lin, Rui Liu, Peng Fu, Yanan Cao, Weiping Wang, and Jie Zhou. 2022. Neutral utterances are also causes: Enhancing conversational causal emotion entailment with social commonsense knowledge. *arXiv preprint arXiv:2205.00759*.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. *arXiv preprint arXiv:1710.03957*.

Yingjian Liu, Jiang Li, Xiaoping Wang, and Zhigang Zeng. 2024. Emotionic: emotional inertia and contagion-driven dependency modeling for emotion recognition in conversation. *Science China Information Sciences*, 67(8):1–17.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Yuchen Liu, Jinming Zhao, Jingwen Hu, Ruichen Li, and Qin Jin. 2022. Dialogueein: Emotion interaction network for dialogue affective analysis. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 684–693.

Ziqiao Ma, Jacob Sansom, Run Peng, and Joyce Chai. 2023. Towards a holistic landscape of situated theory of mind in large language models. *arXiv preprint arXiv:2310.19619*.

Adyasha Maharana, Dong-Ho Lee, Sergey Tulyakov, Mohit Bansal, Francesco Barbieri, and Yuwei Fang. 2024. Evaluating very long-term conversational memory of llm agents. *arXiv preprint arXiv:2402.17753*.

Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. 2019. Dialoguernn: An attentive rnn for emotion detection in conversations. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 6818–6825.

Alena Neviarouskaya and Masaki Aono. 2013. Extracting causes of emotions from text. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 932–936.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

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.

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 *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15*, pages 593–607. Springer.

Gerald Schoenewolf. 1990. Emotional contagion: Behavioral induction in individuals and groups. *Modern Psychoanalysis*.

Weizhou Shen, Siyue Wu, Yunyi Yang, and Xiaojun Quan. 2021. Directed acyclic graph network for conversational emotion recognition. *arXiv preprint arXiv:2105.12907*.

Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjin Wang, and Yu Sun. 2020. Masked label prediction: Unified message passing model for semi-supervised classification. *arXiv preprint arXiv:2009.03509*.

Xiaohui Song, Longtao Huang, Hui Xue, and Songlin Hu. 2022. Supervised prototypical contrastive learning for emotion recognition in conversation. *arXiv preprint arXiv:2210.08713*.

James WA Strachan, Dalila Albergo, Giulia Borghini, Oriana Pansardi, Eugenio Scaliti, Saurabh Gupta, Krati Saxena, Alessandro Rufo, Stefano Panzeri, Guido Manzi, et al. 2024. Testing theory of mind in large language models and humans. *Nature Human Behaviour*, pages 1–11.

Eugene YJ Tee. 2015. The emotional link: Leadership and the role of implicit and explicit emotional contagion processes across multiple organizational levels. *The Leadership Quarterly*, 26(4):654–670.

Elsbeth Turcan, Shuai Wang, Rishita Anubhai, Kasturi Bhattacharjee, Yaser Al-Onaizan, and Smaranda Muresan. 2021. Multi-task learning and adapted knowledge models for emotion-cause extraction. *arXiv preprint arXiv:2106.09790*.

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.

Monika Wróbel and Kamil K Imbir. 2019. Broadening the perspective on emotional contagion and emotional mimicry: The correction hypothesis. *Perspectives on Psychological Science*, 14(3):437–451.

Jialiang Wu, Yi Shen, Ziheng Zhang, and Longjun Cai. 2024a. Enhancing large language model with decomposed reasoning for emotion cause pair extraction. *arXiv preprint arXiv:2401.17716*.

Zehui Wu, Ziwei Gong, Lin Ai, Pengyuan Shi, Kaan Donbekci, and Julia Hirschberg. 2024b. Beyond silent letters: Amplifying llms in emotion recognition with vocal nuances. *arXiv preprint arXiv:2407.21315*.

Zhen Wu, Xinyu Dai, and Rui Xia. 2023. Pairwise tagging framework for end-to-end emotion-cause pair extraction. *Frontiers of Computer Science*, 17(2):172314.

Rui Xia and Zixiang Ding. 2019. Emotion-cause pair extraction: A new task to emotion analysis in texts. *arXiv preprint arXiv:1906.01267*.

Hanqi Yan, Lin Gui, Gabriele Pergola, and Yulan He. 2021. Position bias mitigation: A knowledge-aware graph model for emotion cause extraction. *arXiv preprint arXiv:2106.03518*.

Fangxu Yu, Junjie Guo, Zhen Wu, and Xinyu Dai. 2024. Emotion-anchored contrastive learning framework for emotion recognition in conversation. *arXiv preprint arXiv:2403.20289*.

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.

Duzhen Zhang, Feilong Chen, and Xiuyi Chen. 2023a. Dualgats: Dual graph attention networks for emotion recognition in conversations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7395–7408.

Duzhen Zhang, Zhen Yang, Fandong Meng, Xiuyi Chen, and Jie Zhou. 2022. Tsam: A two-stream attention model for causal emotion entailment. *arXiv preprint arXiv:2203.00819*.

Yazhou Zhang, Mengyao Wang, Prayag Tiwari, Qiuchi Li, Benyou Wang, and Jing Qin. 2023b. Dialoguellm: Context and emotion knowledge-tuned llama models for emotion recognition in conversations. *arXiv preprint arXiv:2310.11374*.

Weixiang Zhao, Yanyan Zhao, Zhuojun Li, and Bing Qin. 2023a. Knowledge-bridged causal interaction network for causal emotion entailment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 14020–14028.

Weixiang Zhao, Yanyan Zhao, Xin Lu, Shilong Wang, Yanpeng Tong, and Bing Qin. 2023b. Is chatgpt equipped with emotional dialogue capabilities? *arXiv preprint arXiv:2304.09582*.

Xiaopeng Zheng, Zhiyue Liu, Zizhen Zhang, Zhaoyang Wang, and Jiahai Wang. 2022. Ueca-prompt: Universal prompt for emotion cause analysis. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 7031–7041.

Peixiang Zhong, Di Wang, and Chunyan Miao. 2019. Knowledge-enriched transformer for emotion detection in textual conversations. *arXiv preprint arXiv:1909.10681*.

Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. 2024. Memorybank: Enhancing large language models with long-term memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19724–19731.

Changzhi Zhou, Dandan Song, Zhijing Wu, Linmei Hu, Jing Xu, and Yanru Zhou. 2024a. A multi-level multi-task progressive framework based on relational graph convolutional networks for causal emotion entailment. *Knowledge-Based Systems*, page 112108.

Zhihan Zhou, Xue Gu, Yujie Zhao, and Hao Xu. 2024b. Pop-cee: Position-oriented prompt-tuning model for causal emotion entailment. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 4199–4210.