POP-CEE: Position-oriented Prompt-tuning Model for Causal Emotion Entailment

Zhihan Zhou[1], **Xue Gu**[2], **Yujie Zhao**[1], **Hao Xu**[1*]
[1]College of Computer Science and Technology, Jilin University, Changchun, China
zhzhou22, zhaoyj21@mails.jlu.edu.cn, xuhao@jlu.edu.cn
[2]School of Engineering, Department of Industrial Electronics, University of Minho, Portugal
id9267@alunos.uminho.pt

Abstract

The objective of the Causal Emotion Entailment (CEE) task is to identify the causes of the target emotional utterances in a given conversation. Most existing studies have focused on a fine-tuning paradigm based on a pretrained model, e.g., the BERT model. However, there are gaps between the pretrained task and the CEE task. Although a pretrained model enhances contextual comprehension to some extent, it cannot acquire specific knowledge that is relevant to the CEE task. In addition, in a typical CEE task, there are peculiarities in the distribution of the positions with different emotion types of emotion utterances and cause utterances in conversations. Existing methods employ a fixed-size window to capture the relationship between neighboring conversations; however, these methods ignore the specific semantic associations between emotions and cause utterances. To address these issues, we propose the Position-oriented Prompt-tuning (POP-CEE) model to solve the CEE task in an end-to-end manner. Specifically, we can model the CEE task by designing prompts with multiple unified goals and by exploring the positional relationship between emotion and cause utterances using a position constraint module. Experimental results demonstrate that the proposed POP-CEE model achieves state-of-the-art performance on a benchmark dataset. Our code and data can be found at: https://github.com/Zh0uzh/POP-CEE.

1 Introduction

Social media's recent expansion has increased interest in emotion analysis in conversations (EAC) in the natural language processing (NLP) field. Existing EAC efforts primarily focus on emotion recognition in conversation (ERC), which identifies the types of emotions conveyed in utterances. However, (Poria et al., 2021) has argued that recognizing a speaker's emotion is insufficient. In other words, further reasoning is required, e.g., mining the causes behind the speaker's emotion. Thus, they proposed the RECCON task, which comprises two sub-tasks, i.e., Causal Span Extraction (CSE) at the word or phrase level and Causal Emotion Entailment (CEE) at the utterance level. In this paper, we focus on the CEE task, which attempts to predict which specific utterances in a given dialogue history contain the causes of non-neutral emotions in the target utterance.

For the CEE task, (Poria et al., 2021) proposed a baseline approach, treating the task as a natural language inference problem. However, this method only performs simple inference and does not consider the characteristics of the dialogue text. Thus, (Zhang et al., 2022), (Kong et al., 2022), and (Jiang et al., 2023) characterized the structure of the conversation and the information of the emotion and speaker in the conversation text by designing different network models, and (Li et al., 2022), (Zhao et al., 2022) incorporated the commonsense knowledge into the inference using graph models.

*Corresponding author

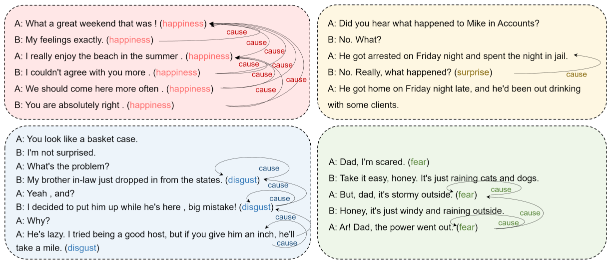

Figure 1: Position relationships between emotion and cause utterance for different emotions in a conversation scenario. Four distinct emotions are shown: **Happiness** (red), **Surprise** (yellow), **Disgust** (blue), and **Fear** (green), where the arrows point from the emotion utterance to the cause utterance. Note that *Happiness* is constantly contagious in the conversation. The conversation partner causes the *Surprise* emotion, which lasts for a short utterance. The *Disgust* and *Fear* emotions are present on only one side of the conversation.

Findings of the Association for Computational Linguistics ACL 2024, pages 4199–4210
August 11-16, 2024 ©2024 Association for Computational Linguistics

emo \ pos	0	1	2	3	4	5	6	7
happiness	56.2%	56.8%	26.3%	16.2%	13.1%	8.4%	8.4%	5.5%
surprise	42.2%	84.6%	5.2%	7.1%	1.8%	1.7%	0.6%	0.8%
sadness	76.2%	22.6%	32.2%	12.2%	19.8%	6.5%	14.6%	4.3%
anger	73.1%	38.8%	42.2%	17.1%	28.7%	13.2%	18.2%	7.1%
fear	73.4%	11.9%	32.8%	5.6%	19.1%	2.9%	6.5%	0.0%
disgust	83.9%	27.1%	31.2%	7.4%	24.3%	3.4%	6.8%	2.9%

Table 1: Relationship between emotion type and cause utterance position in the RECCON-DD dataset. emo denotes six different emotion types, and pos denotes the distance between the candidate utterance U_i and the target utterance U_t (only 0-7 are listed here), where 0 means that the candidate utterance is the target utterance itself. The data in the table indicate the proportion of candidate utterances that imply a cause utterance among the candidate utterances at a particular emotion type and a particular position.

These previous studies considered different dialogue characteristics. However, they ignored the relationship between emotion types and the positions of the corresponding cause clauses. When a speaker displays a particular emotion, the different types of emotion can reveal the position of the cause. For example, we consider the conversation scenarios shown in Figure 1. Here, for the happiness emotion, when Speaker A's utterance contains the happiness emotion, Speaker B's subsequent utterance is more likely to contain happiness because positive emotions are contagious among speakers. In other words, when a speaker communicates positive emotions, those around them are influenced and experience similar positive emotions. We observe that the underlying causes of positive emotions are typically the same, and the primary cause is the initial utterance of Speaker A expressing happiness. In terms of the surprise emotion, when Speaker B's utterance contains the surprise emotion, we know that Speaker B must not have known about it beforehand. The surprise emotion typically only lasts for a short period. Therefore, the cause of surprise is generally contained in the utterance nearest to the emotion utterance. Thus, surprise is typically caused by the speaker's words most intimate to the surprise emotion, frequently Speaker A. For negative emotions, e.g., disgust and fear, in a conversational situation, the goal of the conversation is to transform the negative emotion into a neutral or positive emotion. When Speaker A expresses a negative emotion, Speaker B will typically persuade Speaker A with a neutral feeling. The conversation tends to move in the direction of one person complaining and the other comforting, and the cause of the negative emotion is generally in the first utterance in which Speaker A expresses the negative emotion. In rare cases, e.g., when both speakers are arguing, Speakers A and B may represent the same emotion. The position between the emotion and the cause can facilitate effective utterance extraction in conversations. These linguistic phenomena are based on our statistics on the RECCON-DD dataset, and the relationship between emotion and position is shown in Table 1.

Thus, we propose the Position-oriented Prompt-tuning model for CEE (POP-CEE). Multiple templates are proposed to guide the pretrained model from a semantic perspective combined with constraints to filter out the model's obviously incorrect prediction results. Based on the observed positional relationship between the emotions and the causes in the target conversations, we construct multiple templates modelling the CEE tasks to direct the model's attention to the position between the emotions and causes. Then, the positional constraints module filters the obtained predictions.

We conducted extensive experiments on benchmarks, and the results demonstrate that the proposed framework outperforms all baselines on the experimental dataset. In addition, we conducted several studies and analyses to explore each module's effectiveness on the CEE task. The primary contributions of this study are summarized as follows.

- We present the POP-CEE method to solve CEE tasks in an end-to-end manner. The presented method can fully utilize the position-specific information between emotions and causes in different dialogue scenarios.
- We design four prompt templates and combine them for model inference, which can use the semantic information present in conversations to understand the relationship between emotion types and cause utterances.
- We conduct extensive experiments, demonstrating that the method proposed in this paper achieves state-of-the-art performance.

2 Related Work

Early work in sentiment analysis was applied to opinion mining tasks (Cambria et al., 2013) to recognize and extract opinions from text data. More recently, the work in sentiment analysis has been applied to a wider range of tasks, e.g., allowing for

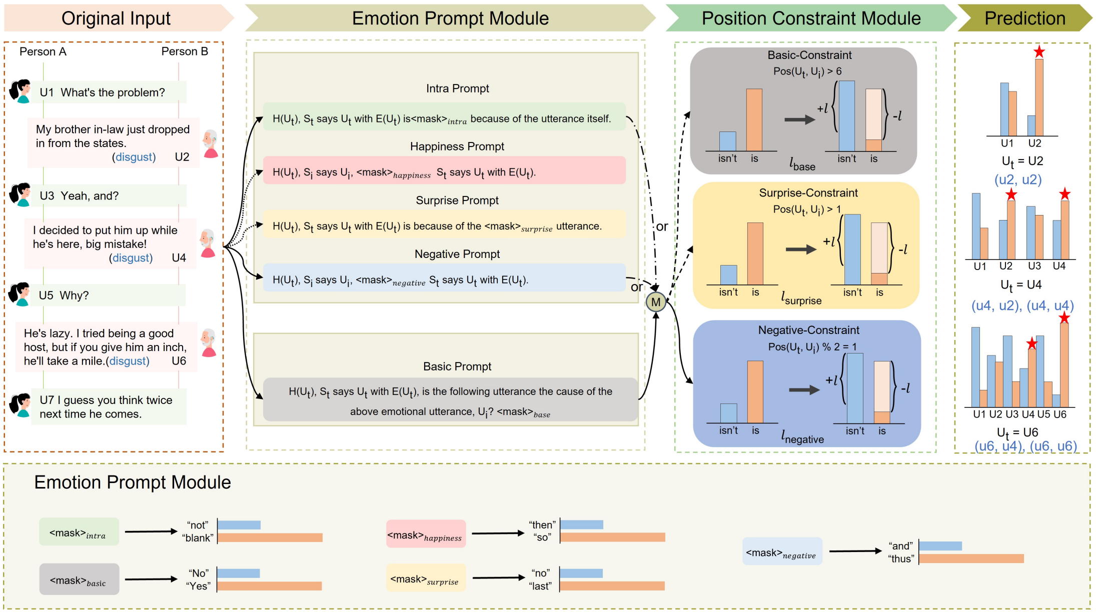

Figure 2: Framework of the proposed POP-CEE model. The structure of the model is shown at the top, and the detailed structure of the emotion prompt module is shown below. First, we input the target utterance, the candidate utterance, and their speakers for reasoning. Then, the emotion prompt module selects the corresponding prompt according to the emotion type of the target utterance. Then, the average value of the reasoning result and the reasoning result of the basic-prompt are taken to obtain the probability that the candidate utterance contains the cause of the emotion. The position constraint module then further constrains the probability according to the positional relationship between the target utterance and the candidate utterance to obtain the final prediction result. Candidate utterances containing the target utterance's cause are marked with pentagrams.

dialogueue topic segmentation (Gao et al., 2023) and speech separation (Jiang et al., 2021).

Initially, the extraction of target emotional causes was primarily realized using rule-based methods (Lee et al., 2010; Chen et al., 2010). (Gui et al., 2016) constructed a public ECE dataset using city news as a corpus and solved the emotional causes extraction problem as a question answering task (Gui et al., 2017). Then, in conjunction with the ECE task, (Xia and Ding, 2019) proposed the emotional causes pair extraction (ECPE) task and used a two-part framework for it to process. Many end-to-end improvement methods emerged subsequently (Yuan et al., 2020; Fan et al., 2020; Wei et al., 2020; Cheng et al., 2020; Song et al., 2020; Singh et al., 2021; Zhou et al., 2022; Liu et al., 2022; Zheng et al., 2022; Hu et al., 2023). However, the ECE and ECPE datasets use news articles as the corpus, and the complexity of reasoning is lower because most expressions in news articles are more formal, the expressed emotions are typically very clear, and the information density is higher.

To further reason about emotions, (Poria et al., 2021) proposed the RECCON task, which is based on two datasets, i.e., the IEMOCAP(Busso et al., 2008) and Dailydialogue(Li et al., 2017) datasets. The RECCON task comprises two different subtasks, i.e., the CSE and CEE tasks, and most current work has focused on solving CEE, whose goal is to determine whether the target utterance in a given conversation history contains the cause of the target emotion utterance. To address this task, (Zhang et al., 2022) proposed TSAM to capture intra-speaker and inter-speaker emotional influences in the global view, (Kong et al., 2022) proposed DAM to capture the interactions between utterances and session-specific structures, (Zhao et al., 2022) proposed KBCIN to reason about emotional causes combined with commonsense knowledge, and (Jiang et al., 2023) constructed a joint framework using a Window Transformer to capture the short-time effects of information transfer between multiple utterances. In this paper, we solve the CEE task based on prompt learning.

3 Proposed Method

3.1 Task Definition

Given an utterance U_t with a non-neutral emotion, the CEE task attempts to predict which utterances in the given conversation history $H(U_t) =$

$(U_1, U_2, U_3, U_4, \ldots)$ contain the causes corresponding to U_t's emotion. Here, the inputs to the model include the target utteranceU_t, its emotion type, the conversation history $H(U_t)$, and the candidate utterance U_i. The candidate utterance U_i is classified as a **positive case** if it contains the cause corresponding to the emotion in U_t. Otherwise, it is classified as a **negative case**.

3.2 Overview

The overall framework of the proposed POP-CEE model is shown in Figure 2. This framework comprises two main modules, i.e., the emotion prompt module and the position constraint module. The emotion prompt module directs the pretrained model to prioritize the relationship between the emotion type and the position of emotion-cause utterances, thereby allowing the pretrained model to identify the cause utterance corresponding to the target emotion utterance. The position constraint module constrains the model's prediction results by considering the specific positional relationship between the emotion and the cause utterances.

3.3 Emotion Prompt Module

The emotion prompt module is divided into five elements according to the distribution of the position of emotions and causes in the given dialogue. The Intra-Prompt is applied to cases where the emotion and cause are contained in the same utterances. The Happiness-Prompt, Surprise-Prompt, and Negative-Prompt are applied to cases where the emotion is "Happiness", the emotion is "Surprise", the emotion is "Disgust", "Fear", and other negative emotions, respectively. The Basic-Prompt is applied to all cases.

Intra-Prompt is applied to determine whether the cause utterance of a given emotion utterance is the emotion utterance itself. Specifically, the intra-prompt template is designed as follows:

$T_{intra} =$ "$H(U_t)$, S_t says U_t with $E(U_t)$ is<mask> because of the utterance itself."

where U_t denotes the target emotional utterance, $H(U_t)$ is its conversation history, S_t is the speaker of U_t, and $E(U_t)$ is the emotion of U_t. The label words corresponding to <mask> are {"not", "blank"}, where "not" corresponds to a negative case and "blank" means that the <mask> is not filled with any word and corresponds to a positive case. When the candidate utterance is the target utterance itself, i.e., $U_i = U_t$, the model applies the intra template and the pretrained model gives the probability of the label word filled at <mask>. $P_M(< MASK >="not"|U_i)$ and $P_M(< MASK >="blank"|U_i)$ correspond to the probability that U_i is a negative and positive case predicted by the model, denoted as g_i^{neg} and g_i^{pos}, respectively, which are given by the following formula:

$$g_i^{neg} = g(P_M(< MASK >= ''not''|U_i)|U_i = U_t)$$

$$g_i^{pos} = g(P_M(< MASK >= ''blank''|U_i)|U_i = U_t)$$

Happiness-Prompt is applied to reason about the cause utterance corresponding to the emotion of happiness. The happiness emotion is typically transferred between two speakers, and the cause utterances it corresponds to are generally the same; thus, we designed the template for the happiness-prompt as follows:

$T_{happiness} =$ "$H(U_t)$, S_i says U_i, <mask> S_t says U_t with $E(U_t)$."

where U_i denotes the candidate utterance, and S_i represents the speaker of the candidate utterance. The label words corresponding to <mask> are {"then", "so"}, where "then" indicates that U_i and U_t have a succession relation, which means a negative case. Furthermore, "so" ndicates that U_i and U_t have a causal relation, which refers to a positive case. When the candidate utterance is not the target utterance itself and emotion type $E(U_t) = happiness$, the model applies the happiness template. The pretrained model gives the probability of the label words filled at <mask> $P_M(< MASK >="then"|U_i)$ and $P_M(< MASK >="so"|U_i)$, correspond to the probability that U_i is a negative case and a positive case, respectively, as predicted by the model, which are given by the following formulas:

$$g_i^{neg} = g(P_M(< MASK >= ''then''|U_i)|E(U_t))$$

$$g_i^{pos} = g(P_M(< MASK >= ''so''|U_i)|E(U_t))$$

Surprise-Prompt is applied to reason about the cause utterance corresponding to the emotion of surprise. Note that the duration of the surprise emotion in a dialogue scene is generally short; thus, the distance between the emotion utterance and the cause utterance will be very small. Here, we only consider the previous utterance of the target utterance. The designed surprise-prompt template is expressed as follows:

$T_{surprise}$ = "$H(U_t)$, S_t says U_t with $E(U_t)$ is because of the <mask> utterance."

The label words filled at <mask> are {"no", "last"}, where "no" and "last" correspond to a negative case and a positive case, respectively.

Negative-Prompt is applied to reason about the cause of the emotion utterances containing negative emotions, e.g., anger, disgust, fear, and sadness, which are typically not transmitted between speakers. Therefore, its cause utterance is typically a conversation before the U_t's speaker S_t. The template for the negative-prompt is the same as that for the happiness-prompt, i.e., $T_{negative} = T_{happiness}$, with the difference being $< mask >= \{"and","thus"\}$, where "and" and "thus" correspond to a negative case and a the positive case, respectively.

Basic-Prompt is a generic template for all situations:

T_{basic} = "$H(U_t)$, S_t says U_t with $E(U_t)$, is the following utterance the cause of the above emotional utterance, U_i? <mask>."

The corresponding label words at <mask> are {"No", "Yes"}, where "No" and "Yes" correspond to a negative case and a positive case, respectively. Basic-Prompt balances the four templates mentioned above and mitigates their overfitting. Therefore, for all U_i, the model will apply Basic-prompt to calculate the probability of their negative and positive cases, which are denoted as g^{neg}_{basic} and g^{pos}_{basic}, respectively, and the specific formulas are shown below:

$$g^{neg}_{basic} = g(P_M(< MASK >= ''No''|U_i))$$

$$g^{pos}_{basic} = g(P_M(< MASK >= ''Yes''|U_i))$$

The model then computes the mean of g^{neg}_i and g^{pos}_i, and the mean of g^{neg}_{basic} and g^{pos}_{basic}, respectively, as follows:

$$p^{neg}_i = mean(g^{neg}_i, g^{neg}_{basic})$$

$$p^{pos}_i = mean(g^{pos}_i, g^{pos}_{basic})$$

where p^{neg}_i denotes the probability of a negative case, i.e., the probability that U_i and U_t are uncorrelated, and p^{pos}_i denotes the probability of a positive case, i.e., the probability that U_i and U_t are correlated, and finally p^{neg}_i and p^{pos}_i are formed into a binary group:

$$p^{prompt}_i = (p^{neg}_i, p^{pos}_i)$$

Note that p^{prompt}_i is the final output of the prompt module.

3.4 Position Constraint Module

The constraint module filters out irrelevant candidate utterances based on the positional relationship. In the proposed model, we consider all conversations before the target emotional utterance U_t as the local context of that utterance, for which we design a total of three constraint strategies, i.e., the basic-constraint, surprise-constraint, and negative-constraint strategies. In addition, we set a value of l for each constraint. When the candidate utterance U_i and target utterance U_t satisfy the corresponding constraint relation C, the constraint module subtracts l from the probability of the positive case and adds l to the probability of the negative case. These three constraint strategies are described in detail in the following.

Basic-Constraint is a universal constraint for all emotion utterances. As shown in Table 1, there is a position bias in the RECCON-DD dataset. For the majority of pairs, the distance between its emotion utterances and the corresponding cause utterances is less than 7. This means that no cause utterance can be found when the distance is greater than 6. The condition of the basic-constraint is designed as follows:

$$c_{basic} : pos(U_t, U_i) > k_{basic}$$

where $pos(U_t, U_i)$ denotes the distance between the target utterance and the candidate utterance, and k_{basic} is set to 6.

Surprise-Constraint is applied to utterances with the surprise emotion. As shown in Table 1, the distance between the emotion utterance and the cause utterance for surprise is less than 2, which means that the cause utterance can only be the target utterance or the preceding utterance. The surprise-constraint is expressed as follows.

$$c_{surprise} : pos(U_t, U_i) > k_{surprise}$$

Here, $k_{surprise}$ is set to 1.

Negative-Constraint is applied to all negative emotion types, e.g., anger, disgust, fear, and sadness. Negative emotions are typically not transmitted between two speakers; thus, the causal utterance of a negative emotion is generally a previous utterance by the speaker of the target utterance. Thus, the negative-constraint is given as follows.

$$c_{negative} : pos(U_t, U_i)\%2 = k_{negative}$$

Here, $k_{negative}$ is set to 1.

The constraint module filters out utterances with very low relevance based on the emotion type of the target utterance and the distance between the target and candidate utterances, which is expressed as follows:

$$f_c(p_i^{prompt}) = P((p_i^{neg} + l, p_i^{pos} - l)|c)$$
$$,(l \in L, c \in C)$$

where $C = \{c_{basic}, c_{surprise}, c_{negative}\}$ and $L = \{l_{basic}, l_{surprise}, l_{negative}\}$. The result is then mapped to the probabilities of positive and negative cases using the softmax between 0 and 1 to obtain the final prediction result $\hat{y}_i$.

$$\hat{y}_i = softmax(f_c(p_i^{prompt}))$$

3.5 Cause Prediction and Optimization

The final prediction of the model for the CEE task is implemented based on the prediction results of the constraint module and optimization of the model parameters.

Cause Prediction We obtain the model's final classification result by comparing the magnitude of the probability of the positive and negative cases in $\hat{y}_i$.

$$P_{CEE}(U_i) = \begin{cases} uncorrelated, & \arg\max \hat{y}_i = 0 \\ correlated, & \arg\max \hat{y}_i = 1 \end{cases}$$

Here, $\hat{y}_i$ denotes the final prediction result of the constraint module, $\hat{y}_i$ is a binary group, the position of 0 denotes the prediction probability of the negative case, and the position of 1 denotes the prediction probability of the positive case. $P_{CEE}(U_i)$ is the final prediction result of the model, **uncorrelated** means that the candidate's utterance is unrelated to the target utterance (a non-causal utterance), and **correlated** means that the candidate utterance is related to the target utterance (a causal utterance).

Optimization The proposed model is optimized using an end-to-end optimization process. Based on the probability of the positive and negative cases predicted by the model, cross-entropy is employed to calculate the loss and optimize the model parameters.

$$loss = \sum_{i \in d} -(y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i))$$

Here, y_i indicates whether the candidate utterance is the cause for the target utterance.

4 Experiments

4.1 Experimental Datasets

We conducted experiments on the RECCON's benchmark dataset containing two subsets, i.e., RECCON-DD and RECCON-IE. RECCON-DD was constructed based on the Dailydialogue dialogue dataset. The Dailydialogue dataset comes from various websites used by English language learners to practice dialogue in their daily lives, in which each dialogue contains the following sentiment categories: {Anger, Disgust, Fear, Happy, Neutral, Sad, and Surprise}. RECCON-IE was constructed based on the IEMOCAP dialogue dataset. The IEMOCAP is a dataset of two-person conversations labeled in English across 16 conversational scenarios with six emotion categories: {Anger, Excited, Frustrated, Happy, Neutral and Sad}. Both RECCON-DD and RECCON-IE were constructed by randomly selecting conversations and annotating them with information about the causes of each utterance's sentiment.

4.2 Implementation

We used RoBERTa-base as the pretrained model to facilitate a fair experimental comparison. We set the batch size to 8 during the training process, and the learning rate was set to $1e - 5$. Here, we used the AdamW optimizer for model training, and cross-entropy was used as the loss function. The model was trained over 12 epochs, and we saved the best results for each generation of the model. Finally, the model with the highest macro-averaged F1-score was used to evaluate the test set. We also used the PyTorch framework for the experimental models, and all experiments were executed on a computer system running the Ubuntu OS with NVIDIA GeForce RTX 3090 Ti (24 GB memory) GPUs.

Method	Pos.F1(%)	Neg.F1(%)	Macro F1(%)
RoBERTa-base(Poria et al., 2021)	64.28	88.74	76.51
RoBERTa-large(Poria et al., 2021)	66.23	87.89	77.06
ECPE 2D(Ding et al., 2020a)	55.50	94.96	75.23
ECPE-MLL(Ding et al., 2020b)	48.48	94.68	71.59
RankCP(Wei et al., 2020)	33.00	97.30	65.15
KAG(Yan et al., 2021)	55.52	94.49	75.02
Adapted(Turcan et al., 2021)	62.47	95.67	79.07
MuTEC(Bhat and Modi, 2022)	69.20	85.90	77.55
TSAM(base)(Zhang et al., 2022)	68.50	89.75	79.17
TSAM(large)(Zhang et al., 2022)	70.00	90.48	80.24
KBCIN(Zhao et al., 2022)	68.59	89.65	79.12
Window transformer(Jiang et al., 2023)	63.10	**97.96**	80.53
POP-CEE	**70.71***	90.48	**80.60**

Table 2: Experimental results obtained on RECCON-DD dataset. Bold indicates the best experimental results, and "*" indicates that the proposed POP-CEE model achieved a statistically significant improvement compared to the best baseline method.

Method	Pos.F1(%)	Neg.F1(%)	Macro F1(%)
RoBERTa-base(Poria et al., 2021)	28.02	95.67	61.85
RoBERTa-large(Poria et al., 2021)	40.83	95.68	68.26
ECPE 2D(Ding et al., 2020a)	28.67	**97.39**	63.03
ECPE-MLL(Ding et al., 2020b)	20.23	93.55	57.65
RankCP(Wei et al., 2020)	15.12	92.24	54.75
POP-CEE	**43.24***	96.98	**70.11**

Table 3: Experimental results obtained on RECCON-IE dataset. Bold indicates the best experimental results, and "*" indicates that the proposed POP-CEE model achieved a statistically significant improvement compared to the best baseline method.

4.3 Baselines and Evaluation Metrics

The proposed model was compared with the following baseline models. RoBERTa-base/RoBERTa-large(Poria et al., 2021), ECPE 2D(Ding et al., 2020a), ECPE-MLL(Ding et al., 2020b), RankCP(Wei et al., 2020), KAG(Yan et al., 2021), Adapted(Turcan et al., 2021), MuTEC(Bhat and Modi, 2022), TSAM(Zhang et al., 2022), TSAM(Zhang et al., 2022), KBCIN(Zhao et al., 2022), Window Transformer(Jiang et al., 2023).

Based on previous work(Jiang et al., 2023; Zhao et al., 2022; Zhang et al., 2022), the Macro-F1-score was used to evaluate the experimental results. Here, the F1-score was calculated separately for the positive and negative cases (labeled Pos F1 and Neg F1, respectively).

4.4 Main Results

Results on REECCON-DD: Table 2 shows the experimental results of REECCON-DD. Here, the results for all baseline models were obtained from the corresponding literature (Zhao et al., 2022; Jiang et al., 2023). As can be seen, the proposed method outperformed the compared baseline methods, with a 0.07% improvement in the Macro-F1 score compared to the previous best performance obtained by the Windows Transformer model. In addition, the Pos.F1 score obtained by the proposed method was improved by 0.71% compared to the TSAM (large) model, which confirms the effectiveness of the proposed method on the CEE task.

w/o prompt	Pos.F1(%)	Neg.F1(%)	MacroF1(%)
POP-CEE	**70.71**	**90.48**	**80.60**
w/o intra	68.37	89.84	79.11
w/o happiness	70.27	90.24	80.25
w/o surprise	67.41	89.88	78.64
w/o negative	69.10	89.50	79.30
w/o basic	69.14	90.24	79.69

Table 4: Effect of different prompts in the prompt module on the performance of the proposed POP-CEE model.

The experimental results of REECCON-DD demonstrate that the proposed POP-CEE model transforms the CEE task into a cloze task that is the same as the pretrained model's task, and the entire model architecture only uses the Roberta model without additional neural network layers, thereby maximizing the pretrained model's ability to understand semantic information. In addition, compared to the RANKCP, KAG, and TSAM models using graph networks, the proposed method employs the constraint module to direct the pretrained model to pay more attention to utterances that are closer to the target utterance because the possibility of sentiment relevance to the target utterance decreases considerably as the distance to the target utterance increases. Finally, compared with the Window Transformer model, the proposed method considers the relationship between the emotion type and the position of the utterance because different emotion types typically imply the position of the cause utterance.

Results on REECCON-IE: Table 3 shows the experimental results of REECCON-IE. The POP-CEE model outperformed the compared baseline methods, with a 1.85% improvement in the Macro-F1 score compared to the best performance obtained by the RoBERTa-large model. In addition, the Pos.F1 score obtained by the POP-CEE model was improved by 2.41% compared to the RoBERTa-large model. The experimental results of REECCON-IE demonstrate that the POP-CEE model is able to maintain a high performance in datasets with multiple conversational scenarios, and shows a good generalization.

4.5 Ablation Study

To analyze the role of each module in the model, we conducted two sets of ablation experiments to investigate the effects of the emotion prompt module and position constraint module.

4.5.1 Emotion Prompt module

The emotion prompt module includes five distinct prompts. Ablation experiments were conducted on different prompts to determine whether each prompt achieves the pretrained model's reasoning for the cause utterances. Table 4 shows the experimental results.

As shown in Table 4, the lack of any of the prompts yields a reduction in model performance, which demonstrates that all five prompts play critical roles in the model learning process, thereby verifying the effectiveness of the prompts in the proposed model. Specifically, eliminating the surprise-prompt resulted in a considerable decrease in model performance because the position features of the cause utterance corresponding to the surprise emotion are the most obvious. Thus, the pretrained model achieved the best prediction with the surprise prompt. In addition, the happiness-prompt was found to have a relatively small effect on the performance of the model compared to the other prompts because happiness is the most common emotion; thus, the relationship between the happiness emotion and the position of the corresponding cause utterance is more complex and difficult to reason, and the absence of the negative-prompt caused the model's accuracy to drop for negative cases. The lack of the intra-prompt made the model less accurate in reasoning for positive cases. As a generalized prompt, the basic-prompt reduced the inference error of the other prompts and is an indispensable component of the prompt module.

4.5.2 Position Constraint Module

Three constraints are designed in the constraint module. To confirm the validity of these constraints, we conducted ablation experiments on the three types of rules, i.e., $c_{basic}, c_{surprise}, c_{negative}$, their corresponding l-values, and the k_{basic}-values.

w/o constraint	Pos.F1(%)	Neg.F1(%)	MacroF1(%)
POP-CEE	**70.71**	**90.48**	**80.60**
w/o surprise	68.18	89.91	79.04
w/o negative	69.97	89.96	79.81
w/o basic	68.16	90.03	79.09

Table 5: Effect of different constraints in the constraint module on the performance of the proposed POP-CEE models.

Constraint Ablation Table 5 shows the results of the constraint ablation experiments. As can be seen, all three constraints play an essential role in the model's inference process. We found that the surprise-constraint has the most significant influence on the model's performance for the same reason as the surprise-prompt, i.e., the positional features of the cause utterance corresponding to surprise are the most significant. The basic-constraint has a more pronounced influence on the model's inference effect in the positive case, and the negative-constraint has a relatively small impact on the performance of the model, which may be caused by the fact that the target utterance of the negative emotion had a relatively small proportion in the experimental dataset.

l-value Ablation The results of the ablation experiments for each l-value of the constraint are shown in Table 6. As shown, the model achieved the best performance when $l_{basic} = 5, l_{surprise} = 3$, and $l_{negative} = 3$. We found that l_{basic} had a

constraint	l	Pos.F1(%)	Neg.F1(%)	MacroF1(%)
basic	1e1	68.45	90.16	79.30
	1e2	69.10	89.98	79.54
	1e3	68.80	90.15	79.47
	1e4	69.97	90.23	80.10
	1e5	**70.71**	**90.48**	**80.60**
	1e6	70.80	90.15	80.47
	1e7	70.70	89.84	80.27
surprise	1e1	69.94	89.83	79.88
	1e2	70.64	90.16	80.40
	1e3	**70.71**	**90.48**	**80.60**
	1e4	69.46	89.83	79.65
	1e5	70.13	89.76	79.94
	1e6	68.71	89.68	79.20
	1e7	67.42	89.83	78.63
negative	1e1	68.14	90.20	79.17
	1e2	69.11	90.24	79.68
	1e3	**70.71**	**90.48**	**80.60**
	1e4	69.95	90.06	80.01
	1e5	69.46	90.20	79.83
	1e6	70.75	90.19	80.47
	1e7	69.35	89.85	79.60

Table 6: Effect of l-values of each constraint in the constraint module on the performance of the proposed POP-CEE model (values in bold indicate the best results).

constraint	k_{basic}	Pos.F1(%)	Neg.F1(%)	MacroF1(%)
basic	0	66.15	90.01	78.08
	1	66.07	90.15	78.11
	2	69.08	90.20	79.64
	3	69.65	89.92	79.79
	4	70.40	89.77	80.09
	5	70.48	90.24	80.36
	6	**70.71**	**90.48**	**80.60**
	7	70.73	90.22	80.47
	8	70.77	89.92	80.34

Table 7: Effect of k_{basic} in basic-constraint on the performance of the proposed POP-CEE model (values in bold indicate the best results).

relatively small impact on the performance of the model because it has the most relaxed constraints. In addition, $l_{surprise}$ exhibited a significant impact on the accuracy of the negative cases because the limitations on the surprise-constraint are the most stringent, which may make the model overly dependent on the surprise-constraint in the reasoning process for the target utterances of the surprise type, which in turn produces a bias in the semantic understanding of the negative cases of the surprise type. Thus, we recognized many negative cases as positive cases. The effect of $l_{negative}$ on the model performance was less regular, which could be due to the fact that negative emotion involves more types of emotion, and the negative-constraint is more complex.

k_{basic}-value Ablation The results of thek_{basic} ablation experiments under the basic-constraint are shown in Table 7. As can be seen, when k_{basic} is smaller, there is a great impact on the model performance, especially during the inference process for positive cases, the value of Macro-F1 at $k_{basic} = 0$ is 2.52% lower than that at $k_{basic} = 6$, which is because the correlation between candidate utterances and target utterances is higher when the distance between candidate utterances and target utterances is closer. In addition, a large k_{basic} value yields a smaller effect on the model's performance, thereby providing relatively stable results. Note that the difference between the Macro-F1 value at $k_{basic} = 6$ and $k_{basic} = 8$ is only 0.26%.

5 Conclusion

In this study, we have applied prompt learning to the CEE task and designed a new prompt learning–based paradigm to strengthen the pretrained model's knowledge of the positional relationship between an emotion and its cause utterances to guide the model to recognize the corresponding emotional cause utterance effectively. Experimental results demonstrate that the proposed model outperforms several state-of-the-art models, and additional ablation experiments and analyses verify the effectiveness of each component implemented in the proposed model.

6 Limitations

In this paper, we propose the Position-oriented Prompt-tuning (POP-CEE) model to solve the CEE task in an end-to-end manner. However, prompts and constraints are somewhat empirically designed, and an automated approach is necessary. In addition, our current approach is based on RoBERTa-base, and we have not yet verified its effectiveness on generative large language models. Future research will focus on addressing these limitations and exploring the potential of generative models.

7 Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant No.62077027 and Grant No.72091315, the Department of Science and Technology of Jilin Province, China under Grant No.20230201086GX and Industry University Research Innovation Fund of the Ministry of Education under Grant No.2022XF017.

References

Ashwani Bhat and Ashutosh Modi. 2022. Multi-task learning framework for extracting emotion cause span and entailment in conversations.

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*, page 335–359.

Erik Cambria, Björn Schuller, Yunqing Xia, and Catherine Havasi. 2013. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent systems*, 28(2):15–21.

Ying Chen, SophiaYatMei Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion cause detection with linguistic constructions. *International Conference on Computational Linguistics,International Conference on Computational Linguistics*.

Zifeng Cheng, Zhiwei Jiang, Yafeng Yin, Hua Yu, and Qing Gu. 2020. A symmetric local search network for emotion-cause pair extraction. In *Proceedings of the 28th International Conference on Computational Linguistics*.

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*.

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)*.

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

Haoyu Gao, Rui Wang, Ting-En Lin, Yuchuan Wu, Min Yang, Fei Huang, and Yongbin Li. 2023. Unsupervised dialogue topic segmentation with topic-aware contrastive learning. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2481–2485.

Lin Gui, Jiannan Hu, Yulan He, Ruifeng Xu, Lu Qin, and Jiachen Du. 2017. A question answering approach for emotion cause extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.

Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. 2016. Event-driven emotion cause extraction with corpus construction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*.

Guimin Hu, Yi Zhao, and Guangming Lu. 2023. Emotion prediction oriented method with multiple supervisions for emotion-cause pair extraction.

Dazhi Jiang, Zhihui He, Yingqing Lin, Yifei Chen, and Linyan Xu. 2021. An improved unsupervised single-channel speech separation algorithm for processing speech sensor signals. *Wireless Communications and Mobile Computing*, 2021:1–13.

Dazhi Jiang, Hao Liu, Geng Tu, and Runguo Wei. 2023. Window transformer for dialogue document: a joint framework for causal emotion entailment. *International Journal of Machine Learning and Cybernetics*, pages 1–11.

Dexin Kong, Nan Yu, Yun Yuan, Guohong Fu, and Chen Gong. 2022. Discourse-aware emotion cause extraction in conversations.

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.

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.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. *International Joint Conference on Natural Language Processing,International Joint Conference on Natural Language Processing*.

Junlong Liu, Xichen Shang, and Qianli Ma. 2022. Pair-based joint encoding with relational graph convolutional networks for emotion-cause pair extraction.

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.

Aaditya Singh, Shreeshail Hingane, Saim Wani, and Ashutosh Modi. 2021. An end-to-end network for emotion-cause pair extraction. *arXiv: Computation and Language,arXiv: Computation and Language*.

Haolin Song, Chen Zhang, Qiuchi Li, and Dawei Song. 2020. End-to-end emotion-cause pair extraction via learning to link. *Cornell University - arXiv,Cornell University - arXiv*.

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: Computation and Language,arXiv: Computation and Language*.

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*.

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*.

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

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)*.

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*.

Weixiang Zhao, Yanyan Zhao, Zhuojun Li, and Bing Qin. 2022. Knowledge-bridged causal interaction network for causal emotion entailment.

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.

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.

A Case Study

To further validate the POP-CEE's effectiveness, we selected several representative cases to analyze the model's prediction results.

Each utterance is enumerated according to its order, and if the utterance contains a non-neutral emotion, the emotion, the ground truths of cause utterances and predictions are listed at the end separately.

Case One: Here is a conversation where the boss praises his employee.

1. A: You have been doing a great job this year and I am very satisfied with your work. (happiness)**(Ground truths: 1)(Predictions: None)**

2. B: I am very happy to know my work could be recognized by you and our company. (happiness)**(Ground truths: 1, 2)(Predictions: 1, 2)**

3. A: Based on your contribution, I would like to give you a pay raise. Your monthly salary will be increased by 800 Yuan. This increase includes an 80% married increase based on your achievements and increase responsibilities and then the additional 20% increase were reflecting the changes in cost of living. (happiness)**(Ground truths:1, 3)(Predictions: 1, 3)**

4. B: I really appreciate it. Thank you. (happiness)**(Ground truths:1, 2, 3)(Predictions: 3, 4)**

In Case One, POP-CEE successfully predicted the emotional reasons for the second and third utterances, i.e., (1, 2) and (1, 3), which demonstrates that POP-CEE captures the characteristics of the happiness conversation, i.e., the emotion and the reason are transmitted between the two speakers. However, for the first utterance POP-CEE did not predict correctly, this is probably because there is no prior knowledge for the first utterance and it is difficult for the model to predict the emotional reason. For the fourth utterance, the model predicted only partially correct answers, this because the first utterance and the second utterance are relatively far from the fourth utterance, and the model tends to think that they are not relevant to the fourth utterance.

Case Two: Here is a conversation between two speakers discussing an athlete.

1. A: What happened? Why didn't he win?

2. B: Didn't you hear? He was disqualified.

3. A: How did that happen? He's so talented! I thought he had a great chance of winning a gold medal! (surprise)**(Ground truths: 2)(Predictions: 2)**

4. B: If he didn't have any drug problems, he would have won.

5. A: What? What kind of drugs was he using? (surprise)**(Ground truths: 4)(Predictions: 4)**

6. B: He was taking steroids to make him stronger and faster.

7. A: I thought that all athletes were supposed to be regularly tested for drugs.

8. B: They are. The only reason they didn't disqualify him until after the race is because the results from the text only came back afterwards.

In Case Two, POP-CEE successfully predicted the causes of all surprise utterances, which demonstrates that POP-CEE can capture the cause characteristics of surprise utterances well.

Case Three: Here is a conversation between two drivers.

1. A: Hey, look out!
2. B: What happened?
3. A: You've just scratched my car. Oh, God, a paint was scratched off. (anger)**(Ground truths: 3)(Predictions: 3)**
4. B: Where? my car? (surprise)**(Ground truths: 3)(Predictions: 3)**
5. A: No, mine! (anger)**(Ground truths: 3, 5)(Predictions: 3)**
6. B: Thank goodness!
7. A: I've just had it repainted.
8. B: That's terrible. (sadness)**(Ground truths: 3, 7)(Predictions: 3, 7)**
9. A: I am sorry to say this, sir, but you should've been more careful.
10. B: I apologize for that. But the space is too small. (sadness)**(Ground truths: 3, 7, 10)(Predictions: 3, 10)**
11. A: What about the damage to my car? What are you gonna do about that?
12. B: Can we solve this later? I am calling the insurance company.
13. A: OK. I gotta call mine too.

In Case Three, POP-CEE can perfectly predict the cause of emotion in the third, fourth fifth and eighth utterances, and part of the cause in the tenth utterance. This demonstrates that POP-CEE can capture the cause characteristic of negative emotional utterances, i.e., negative emotions and causes are transmitted in one speaker's utterances.