

PromptECL: Leveraging Prompt Engineering to Unlock Emotion Classification Capabilities in LLMs

Anonymous ACL submission

Abstract

Emotion Recognition in Conversation (ERC) is a crucial task in natural language processing that aims to identify emotional states within dialogic interactions. While existing approaches typically employ conventional models like BERT, GRU, and GNN for contextual feature extraction and emotion classification, they often face limitations in interpretability, architectural complexity, and performance constraints. To address these challenges, we present PromptECL, a novel framework that harnesses few-shot prompt engineering to unlock the intrinsic emotion classification capabilities of Large Language Models (LLMs) through strategic prompt templates that activate LLMs’ generative potential for latent emotion cue extraction and semantic data augmentation, followed by fine-tuning to enhance emotional analysis proficiency. Comprehensive evaluations across four benchmark datasets demonstrate PromptECL’s superior effectiveness and generalization capacity, establishing new state-of-the-art performance with improved average weighted F1 scores on IEMOCAP, MELD, and EmoryNLP, while achieving a remarkable 7.67% micro F1 score improvement over previous best results on DailyDialog.

1 Introduction

As a critical research task in Natural Language Processing (NLP), ERC aims to identify and interpret the subjective emotional tendencies of participants by analyzing conversational texts. With the growing prevalence and deepening integration of human-computer interaction technologies and intelligent customer service systems, ERC plays a vital role in enhancing user experience and enabling intelligent service delivery.

Therefore, it is essential for the ERC task to integrate contextual dialogue information with a nuanced understanding of the underlying mechanisms driving emotional fluctuations, and to accurately

model both the speaker’s affective profile and the dynamic factors contributing to emotional shifts within conversations. To address the modeling of ERC tasks for contextual information, existing approaches have primarily focused on the following aspects, as detailed in Appendix A.2.

1) Recurrent-based methods (Lei et al., 2023b; Liu et al., 2024; Ghosal et al., 2020; Hu et al., 2023a; Majumder et al., 2019; Hu et al., 2021) leverage various types of recurrent neural networks (e.g., LSTM and GRU) to model individual affective states and global affective impacts, respectively.

2) Graph-based methods (Ghosal et al., 2019; Li et al., 2024; Liang et al., 2022; Ishiwatari et al., 2020) typically employ nodes and edges to represent the relationships between characters and dialogues in conversation modeling.

3) Transformer-based methods (Ma et al., 2024; Chudasama et al., 2022) aim to establish long-term emotional associations in conversational scenarios by either directly utilizing or enhancing the original Transformer architecture.

4) Traditional pre-trained models-based methods (Yu et al., 2024; Xue et al., 2024; Shen et al., 2021; Xue et al., 2023) utilize pre-trained models such as BERT and RoBERTa to model discourse and uncover the latent semantic information.

ERC methods face dual challenges: (1) Over-reliance on complex neural architectures for contextual modeling increases training complexity and limits generalizability (Brown et al., 2020; Ouyang et al., 2022); (2) Inherent lack of interpretability in emotional label prediction mechanisms, failing to reveal semantic rationales behind decisions (Moradi et al., 2021; Goyal et al., 2022).

While LLMs demonstrate emergent capabilities through in-context learning and chain-of-thought reasoning (Wei et al., 2022), two critical gaps persist: (1) Behavioral deviations during task execution despite prompt engineering improvements

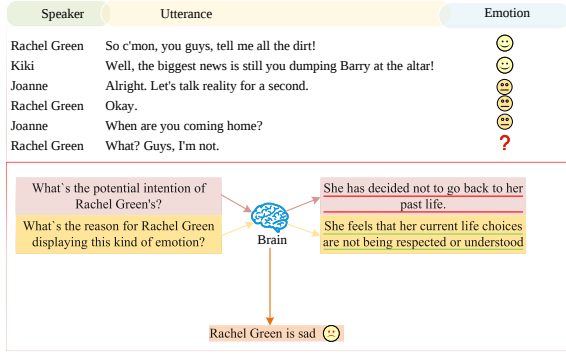


Figure 1: The process of human brain analyzing emotions in conversation texts

(Reynolds and McDonell, 2021); (2) Fundamental explainability limitations in black-box architectures, particularly problematic for safety-critical applications.

Human affective cognition (Fig. 1) uniquely integrates multi-level reasoning - simultaneously decoding linguistic content, inferring communicative intent, and hypothesizing psychological causes of emotional states. This contrasts with LLMs' predominant reliance on surface-level textual patterns, highlighting a critical divergence in contextual and causal reasoning capacities.

To address the aforementioned challenges, we propose PromptECL, a prompt engineering-based approach designed to activate the latent emotional categorization capabilities of LLMs. By leveraging carefully designed prompt, PromptECL effectively stimulates the intrinsic emotion recognition potential of LLMs during both inference and fine-tuning phases. Specifically, PromptECL first employs the LLM to perform semantic augmentation of the original training data, uncovering implicit emotional information, and subsequently fine-tunes the model using this enriched dataset to enhance its classification accuracy and interpretability.

The contributions of this work can be summarized as the follows: 1) We propose a novel prompt-word framework for conversational emotion recognition, which synthesizes and leverages conversational information through cue-word templates, enabling LLM to semantically augment the original training dataset and uncover potential emotional information. 2) To the best of our knowledge, this is the first work to leverage prompt engineering to activate the emotional categorization capabilities of an LLM, while simultaneously generating corresponding rationales for each predicted emo-

tion label, thereby enhancing the interpretability of the model's outputs. 3) Experimental results demonstrate that the proposed PromptECL method achieves new state-of-the-art performance on four widely used ERC datasets. Notably, on the Daily-Dialog dataset, it surpasses the previous state-of-the-art micro F1 score by 7.67%.

2 Methods

This section provides a comprehensive overview of the proposed PromptECL method. First, a brief overview of the ERC task is presented, followed by an exposition of the overall architecture and its constituent components.

2.1 Task Definition

Emotion recognition in dialogues is a task of identifying the emotion category corresponding to each utterance, given a piece of dialog and its participants. Specifically, let the conversation $U = \{u_1, u_2, \dots, u_N\}$ be a continuous conversation containing N sentences, where u_i denotes the content of the i -th sentence in the conversation. Each sentence u_i has a corresponding speaker $S_{u_i} \in S = \{S_{u_1}, S_{u_2}, \dots, S_{u_N}\}$, where S_{u_i} denotes the speaker of the i -th sentence.

The primary objective of this task is to accurately identify the emotional category e_i associated with each utterance u_i within a dialogue, based on a set of predefined emotion categories such as *happy*, *frustrated*, *angry*, and *surprised*. This classification is performed using only the given dialogue context as input.

2.2 Model Architecture

The model architecture is mainly composed of three modules: the emotional information mining module, the emotional perception generation module, and the emotion reasoning module, as shown in Fig. 2. In the emotional information mining module, the raw training data is input into the LLM through data-augmented prompt templates to mine potential emotional information. Next, in the emotional perception generation module, the mined emotional information is used to enhance the original training data. Simultaneously, additional potential information is extracted from within the original training data itself, such as the history of utterances, the identity of the previous speaker, and the emotional transformation of the current speaker. This enriched data is then used to

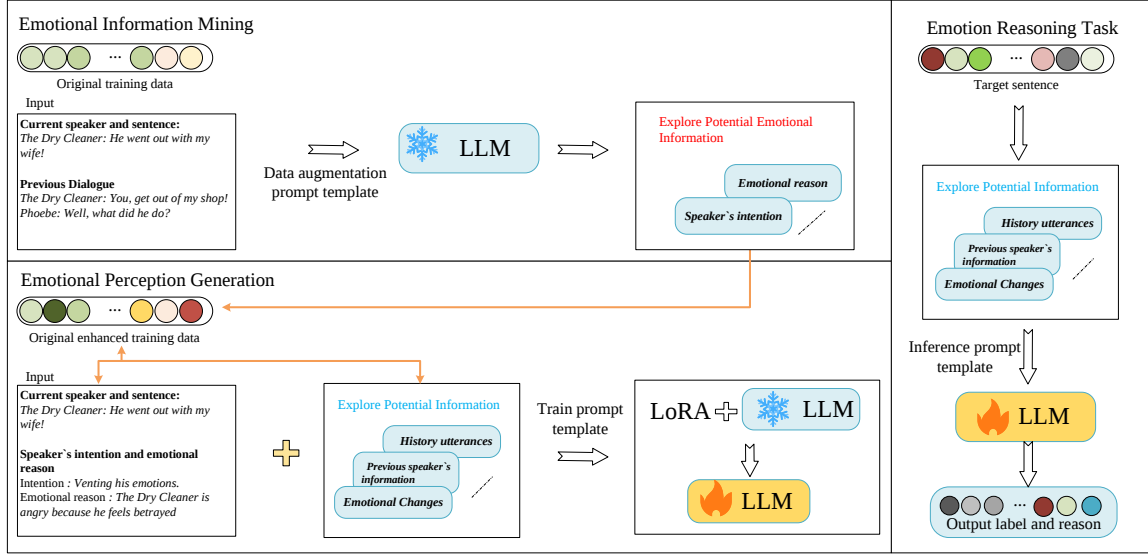


Figure 2: Overview of the PromptECL framework.

fine-tune the LLM, after being processed through training prompt templates. Finally, in the emotion reasoning module, potential information is also mined from the target utterance and input into the fine-tuned LLM via inference prompt templates to obtain both the emotional label and the emotional reasoning for the utterance.

2.3 Emotional Information Mining Module

The power of LLM comes from the linguistic features and contextual relationships it learns when it is pre-trained on massive textual data. We guide the LLM by designing specific prompt template, prompting it to activate neurons related to emotions, and speculate on the speaker's intentions and potential reasons for the change in emotions through the emotional analysis results generated by the LLM.

In order to explore the implicit information of the speaker in depth, this study introduces multidimensional information related to emotion, including "speaker's intention", "the emotion label of the current sentence", and "the emotion reason for the emotion label". By fine-tuning LLM in conversational emotional analysis tasks and making full use of this emotional-related information, the aim is to warm up the model parameters for subsequent ERC tasks. Specifically, we have designed the following prompt template to guide the model in gaining clarity and understanding of the emotional characteristics to improve its ability to analyze emotions in complex conversational situations.

As shown in the data augmentation prompt template in Fig. 3 a), "Prompt" = "Instruction" + "One-

shot prompt" + "Format input/output", here "Instruction" = "You are an emotion analysis expert. The following are log records of multiple rounds of situational conversations in the format of speaker, sentence, and emotional label. Please analyze the log records and output the reasons and the current speaker's intention for emotional labels in one sentence.", The primary objective of this section is to define the role of the model as an emotional analysis expert and to provide detailed instructions on how it should interpret emotional labels within dialogue logs, including the underlying causes of these emotions and the communicative intentions of the speakers. The One-shot prompt is to manually create a concrete example of the task to help the model better understand the goal of the task and the way to execute it, so as to realize the learning of the task with fewer samples. This section reduces the model's ambiguity in task comprehension by providing a complete set of input-output pairs and showing how the model should extract emotional reasons and speaker intentions from the dialog. And "Format input/output" = "The log is as follows: Dialogue The sentences that need to be processed are as follows: u_i , Please only output the current speaker's intention and the emotional reason for this sentence, output format: Intention : t_i Emotional reason : r_i ", Where Dialogue refers to the current previous dialog record, and u_i represents the current input sentence, t_i stands for the speaker's potential intention, and r_i stands for the emotional reason.

a) Data augmentation prompt template	b) Train prompt template	c) Inference prompt template
Instruction You are an emotion analysis expert. The following are log records of multiple rounds of situational conversations in the format of speaker, sentence, and emotional label. Please analyze the log records and output the reasons and the current speaker's intention for emotional labels in one sentence.	Instruction You are an expert in dialogue emotion recognition and analysis.	Instruction You are an expert in dialogue emotion recognition and analysis.
One-shot prompt An example is as follows: The log is as follows: Speaker: "Clearly, You know, do you have like a supervisor or something?" Emotional label: neu. Speaker: "Yeah, Do you want to see my supervisor? Huh? Yeah. Do you want to see my supervisor? Fine, I'll be right back." Emotional label: ang. ... Speaker: "That would - I would appreciate that. Yeah." Emotional label: fru. The sentences that need to be processed are as follows: Speaker: "Excuse me." Emotional label: neu. Please only output the current speaker's intention and the emotional reason for this sentence, output format: Intention : xxx Emotional reason : xxx.	Previous dialogue The emotional conversation records enclosed within the '###' symbols involve multiple speakers. ###(U ₁ , U ₂ , U ₃ , ..., U _{i-1})### Input utterance <Speaker : U _i > Emotion changes considering the current speaker's emotional changes history [e ₁ , e ₂ , ..., e _{k-1}] (Historical emotional variation range window = k) Previous speaker's information the previous speaker's utterance [U _{i-1}] and emotion [e _{previous}]. Format output Based on the above information, please analyze the potential intent of the current speaker. The emotion label of the current sentence is <label : "e">. Please analyze the reason for the emotion label. Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx, xxx is the content you output	Previous dialogue The emotional conversation records enclosed within the '###' symbols involve multiple speakers. ###(U ₁ , U ₂ , U ₃ , ..., U _{i-1})### Input utterance <Speaker : U _i > Emotion changes considering the current speaker's emotional changes history [e ₁ , e ₂ , ..., e _{k-1}] (Historical emotional variation range window = k) Previous speaker's information the previous speaker's utterance [U _{i-1}] and emotion [e _{previous}]. Format output Based on the above information, please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <Neutral, Joyful, Peaceful, Powerful, Scared, Mad, Sad>, and provide the reason for the emotional label.

Figure 3: a) Data Enhancement Prompt Template. b) Train prompt template. c) Inference prompt template.

2.4 Emotional Perception Generation Module

This module involves both the prediction of current affective states and the inference of potential causes underlying affective changes. In the previous module, the model was guided to identify and explain the root causes of emotional changes. Through the analysis of the current affective state and its dynamic variations, the model can generate interpretative explanations regarding affective intentions and the underlying causes of emotional responses.

To enhance the precision of emotional analysis, a refinement strategy is applied wherein the inferred causes of emotional shifts are iteratively fine-tuned based on the model's output. In this phase, specialized prompt templates are employed to activate the LLM's intrinsic emotional analysis capabilities. As shown in the train prompt template in Fig. 3 b), "Prompt" = "Instruction" + "Previous Dialogue" + "Input Utterance" + "Emotion changes" + "Previous speaker's information" + "Formate output". Here "Instruction" = "You an expert in dialogue emotion recognition and analysis.". "Previous Dialogue" = "The emotional conversation records enclosed within the '###' symbols involve multiple speakers. ### U = {U₁, U₂, U₃, ..., U_{i-1}} ###", U denotes the historical context comprising the preceding $i - 1$ sentences of the current utterance. "Input Utterance" = "<Speaker : U_i>", U_i denotes the current sentence. "Emotion changes" = "considering the current speaker's emotional changes history [e₁, e₂, ..., e_{k-1}]", e_i stands for the user's emotion label, while k represents the

threshold of the sliding window that controls the range of emotion changes that can be seen for the current user. And "Previous speaker's information" = "the previous speaker's utterance [U_{i-1}] and emotion [e_{previous}].", is used to prompt the previous speaker's emotional message. Ultimately, "Format output" = "Based on the above information, please analyze the potential intent of the current speaker. The emotion label of the current sentence is <label : "e">. Please analyze the reason for the emotion label. Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx, xxx is the content you output" is used to control the output of the model in a canonical format.

The sliding window k plays a crucial role in modeling emotional fluctuations. The sliding window is used to control the scope of the model's attention to the historical emotional and to ensure that the model is able to synthesize the emotional states of the previous k time steps when analyzing the current emotional. Specifically, the affective history H_t of the current speaker at moment t is defined as the sequence of affective states from $t - k$ to t , as expressed in Equation 1:

$$H_t = [E_{t-k}, E_{t-k+1}, \dots, E_{t-1}] \quad (1)$$

E_t denotes the emotional state at moment t . The model takes the historical emotional information H_t as part of the input and participates in the computation along with the features S_t of the current sentence to infer the current potential emotional

information, as shown in Equation 2:

$$C_t = l_i + r_i = LLM(H_t, S_t) \quad (2)$$

Where, l_i denotes the emotional label of the current sentence, r_i denotes the reason for the emotional label of the current sentence. In this way, the model is capable of predicting current affective states with greater accuracy, taking into account historical affective changes. Based on the sliding window mechanism, we synthesize the historical emotional information H_t and the emotional features of the current sentence S_t to predict the emotional state C_t at the current moment, which is formulated as Equation 3:

$$\begin{aligned} C_t = l_i + r_i &= f(H_t, S_t) \\ &= f([E_{t-k}, E_{t-k+1}, \dots, E_{t-1}], S_t) \end{aligned} \quad (3)$$

This approach allows the model to make dynamic predictions of emotional changes within a limited context, enhancing its performance in emotional analysis tasks. In the preceding step, the model is directed to identify and elucidate the underlying causes of emotional changes. Through the analysis of the current affective state and its fluctuations, the model generates explanations regarding affective intentions and the determinants of emotional responses.

To further enhance the accuracy of emotional analysis, we employ a strategy of fine-tuning LoRA based on the inferred causes of emotional changes in the model's output. Specifically, the model generates an emotional state and the reason for its change at each prediction, and based on these results, we fine-tune LLM to perform more accurately on the emotional analysis task. The process of fine-tuning can be achieved by minimizing the cross-entropy loss function of the emotional analysis task, which calculates the error between the predicted model emotional state $\hat{C}_t$ and the true emotional state C_t . The loss function is shown in Equation 4:

$$\mathcal{L}(C_t, \hat{C}_t) = -\log p(\hat{C}_t | H_t, S_t) \quad (4)$$

Through the backpropagation algorithm, the loss function quantifies the magnitude of the error in the entire emotional prediction process, subsequently adjusting the model's internal parameters in response to this error to improve its ability to interpret emotional changes.

2.5 Emotion Reasoning Module

In the emotion reasoning module, we utilize a fine-tuned LLM for emotional classification. To enhance the accuracy of emotional classification, the model incorporates contextual information from multiple rounds of dialogue as input. It also employs specifically designed inference prompt to guide the prediction process. These prompts are crafted to align with the emotional flow of the conversation and accurately capture shifts in sentiment. By leveraging conversational history, the model is better equipped to extract key information and assign appropriate emotional labels to the current speaker.

The design of the inference prompt is shown in Fig. 3 c), for ERC task, each input consists of four parts: instructions, historical content, label statement, and demonstration retrieval. The components "Prompt" = "Instruction" + "Previous dialogue" + "Input utterance" + "Emotion changes" + "Previous speaker's information" + "Format output" are consistent with the train prompt template used during training. "Format output" = "Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <Neutral, Joyful, Peaceful, Powerful, Scared, Mad, Sad>, and provide the reason for the emotional label."

In the inference process, we employ a greedy search strategy for LLM to select the optimal emotional labels. Specifically, the model outputs a probability distribution over the possible labels. Let $P(Y | x)$ denote the conditional probability distribution of the set of emotional labels Y , where $Y \subset R^n$ represents the set of n possible label categories, given the input dialogue context x . For each input context x , the objective of the greedy search is to identify the label $\hat{y}$ that maximizes this probability, as formally expressed below:

$$\hat{y} = \arg \max_{y \in Y} P(y | x), \quad Y \subset R^n \quad (5)$$

During inference, the model computes the conditional probability for each potential emotional label and deterministically selects the label corresponding to the maximum probability as the output $\hat{y}$. This strategy ensures an efficient and direct inference process, facilitating rapid and accurate emotion classification within the dialogue.

3 Experiments

3.1 Implementation Details

This study employs four widely used benchmark datasets for conversational emotion recognition research: IEMOCAP, MELD, EmoryNLP, and DailyDialog. A comprehensive description of the datasets can be found in Appendix A.1.

For each dataset, we report the Weighted-average F1 score (Weighted-F1) to evaluate the model performance. Specifically, for the DailyDialog dataset, we also report the Micro-average F1 score (Micro-F1) to evaluate the model performance. Subsequently, we show the average performance compared to previous work on these datasets.

In the emotional information mining module, we use the Qwen2.5-72B-instruct model as the emotional information enhancement model. In the emotional perception generation module, the Qwen2.5-7B-instruct model is used as the base model, and the fine-tuning method is chosen to be the Lora method with Float16 data type. The sliding window of the speaker’s emotion range is between {2, 4}, and the window is dynamically adjusted according to each task to optimize the model’s effect.

3.2 Comparison Methods

In order to make a comprehensive assessment, we compare our approach to the following baselines:

1) Sequence-based models: EmotionIC (Liu et al., 2024), SACL (Hu et al., 2023a) and DialogueCRN (Hu et al., 2021).

2) Graph-based models: S+PAGE (Liang et al., 2022) and GraphCFC (Li et al., 2024).

3) Transformer-based models: COSMIC (Ghosal et al., 2020), BERT-ERC (Qin et al., 2023), EACL (Yu et al., 2024), AccWR (Xue et al., 2023), InstructERC (Lei et al., 2023a), BiosERC (Xue et al., 2024).

3.3 Experimental Results and Analysis

According to the experimental procedures and evaluation indexes of the relevant datasets, this section analyzes the experiments on IEMOCAP, MELD, EmoryNLP and DailyDialog datasets respectively. Since the dataset division and experimental procedures of the experiments in this paper are fully in accordance with the specification of each dataset for the experiments, the experimental results summarized in the table are exclusively drawn from

those reported in the corresponding research papers.

The specific experimental results are shown in Table 1, which demonstrates in detail the F1 score performance of different models on several benchmark datasets, including the sequence-based model, the graph-based model, and the Transformer-based model. Our method, **PromptECL**, shows significant advantages on multiple datasets, and especially achieves the current SOTA performance on the IEMOCAP, EmoryNLP, and DailyDialog datasets, which fully demonstrates the effectiveness of the PromptECL framework.

Specifically, PromptECL’s Weighted-F1 metric on the IEMOCAP dataset is 0.72% higher than the current SOTA method SDT model, on the EmoryNLP dataset, the Weighted-F1 metric outperforms the current SOTA method BiosERC model by 0.36%, and on the DailyDialog dataset, the Micro-F1 metric improves 7.67% compared to the current SOTA method S+PAGE model, which is far more than the current SOTA methods. PromptECL fails to achieve the optimal performance on the MELD dataset, but its performance is still close to the best, mainly due to two reasons: firstly, the MELD dataset is a multimodal dataset, which, besides the textual information, also contains visual and speech, etc., in addition to textual information, while only a single textual modality is used in our experiments; second, compared to the other three datasets, the MELD dataset is more limited in terms of contextual information, which puts some constraints on our feature extraction and understanding of emotion expression. Meanwhile, the average performance of our method on the three datasets of IEMOCAP, EmoryNLP, and DailyDialog outperforms the current SOTA method, InstructERC model, by 0.36%, which indicates that PromptECL is highly adaptable and generalizable in terms of capturing emotional expressions and contextual information.

3.4 Ablation studies

To validate the effectiveness of each module of the PromptECL framework, we systematically analyze it through four sets of experiments. The contribution of each component to the emotional classification performance is quantitatively assessed by gradually removing the emotional information mining, potential information mining and emotional perception generation modules and comparing them with the baseline model. The experimental results

Method	IEMOCAP Weighted-F1	MELD Weighted-F1	EmoryNLP Weighted-F1	Average Weighted-F1	DailyDialog Micro-F1
<i>Sequence-based models</i>					
EmotionIC	69.61	66.32	40.25	58.72	60.13
SACL	69.22	66.45	39.65	58.44	–
DialogueCRN	66.20	58.39	–	–	–
<i>Graph-based models</i>					
S+PAGE	68.72	63.32	39.14	57.06	64.07
GraphCFC	68.91	58.85	–	–	–
<i>Transformers-based models</i>					
COSMIC	65.30	65.21	38.11	56.20	58.48
BERT-ERC	71.70	67.11	39.84	59.55	61.42
EACL	70.41	67.12	40.24	59.25	–
ACCWR	67.65	64.58	39.33	57.16	59.22
SDT	74.08	66.60	–	–	–
InstructERC	71.39	69.15	41.37	60.64	–
BiosERC	69.02	68.72	41.44	59.72	–
PromptECL	74.80	66.21	41.80	60.94	71.74

Table 1: Comparison with the state-of-the-art methods on four datasets (%).

Method	IEMOCAP Weighted-F1	MELD Weighted-F1	EmoryNLP Weighted-F1	DailyDialog Micro-F1
PromptECL	74.80	66.21	41.80	71.74
- w/o EIM	76.34 (↑1.54)	65.09 (↓1.12)	37.85 (↓3.95)	65.40 (↓6.34)
- w/o PIE	46.34 (↓28.46)	61.05 (↓5.16)	36.06 (↓5.74)	61.38 (↓10.36)
- w/o EPG	61.22 (↓13.58)	53.57 (↓12.64)	39.57 (↓2.23)	62.79 (↓8.95)
Base Model	37.12 (↓37.68)	58.96 (↓7.25)	32.80 (↓9.00)	60.21 (↓11.53)

Table 2: Results(%) of eliminating different modules. "- w/o EIM" means elimination of Emotional Information Mining. "- w/o PIE" means elimination of Potential Information Exploring. "- w/o EPG" means elimination of Emotional Perception Generation.

confirm that working with working with multiple moframeworksignificantly improve the abilitability to perceive emotionslog scenarios.

1) Elimination of Emotional Information Mining (EIM):The ablation study eliminates the emotional information mining module and directly fine-tunes the LLM using original training data with latent emotion features. As shown in Table 2, this leads to performance drops of 1.00% (Weighted-F1) on MELD, 3.95% on EmoryNLP, and 6.34% (Micro-F1) on DailyDialog, demonstrating the importance of emotion feature mining. However, IEMOCAP shows a 1.54% Weighted-F1 improvement, attributed to its strong reliance on multi-modal contextual features from long dialogues and implicit emotional expressions, where modeling dialogue history becomes more critical than explicit emotion mining.

2) Elimination of Potential Information Exploring (PIE):This ablation study eliminates the Potential Information Exploring module during both training and inference. The LLM is fine-tuned using only latent emotion features extracted by the emotion mining module from raw data, while eliminating latent information mining and related prompt templates in inference. Results in Table 2 show significant performance degradation: 5.16% (MELD), 5.74% (EmoryNLP) in Weighted-F1, and 10.36% (DailyDialog) in Micro-F1. Notably, IEMOCAP suffers a 28.46% Weighted-F1 drop, revealing its extreme dependence on contextual information modeling facilitated by the removed module. This confirms the critical role of latent information mining in enhancing LLMs’ contextual emotion understanding.

3) Elimination of Emotional Perception Generation (EPG):This experiment disables the LLM’s intrinsic emotional perception while retaining prompt templates for information exploration during inference. Results in Table 2 reveal substantial performance degradation: 13.58% (IEMOCAP), 12.6% (MELD), and 2.23% (EmoryNLP) drops in Weighted-F1, along with an 8.95% Micro-F1 decrease on DailyDialog. The drastic decline on IEMOCAP and MELD highlights their dependence on activated emotion perception, particularly for modeling complex contextual interactions. This

confirms that emotion perception fine-tuning is essential for unlocking LLMs’ latent emotional understanding capabilities.

4) Elimination of PromptECL: This study evaluates PromptECL by disabling the entire framework and using the base LLM directly for emotion classification. Results in Table 2 show catastrophic performance degradation: 37.68% (IEMOCAP), 9% (EmoryNLP), and 7.25% (MELD) drops in Weighted-F1, with an 11.53% Micro-F1 decrease on DailyDialog. The extreme 37.68% plunge on IEMOCAP particularly underscores its critical reliance on contextual modeling mechanisms within PromptECL, which are essential for decoding complex multimodal emotional cues in long dialogues. This comprehensively validates the framework’s effectiveness in enhancing LLMs’ emotional comprehension through structured prompt engineering and latent feature exploration.

3.5 Experiments on Fine-Grained Emotional Classification

This experiment aims to systematically evaluate the performance of the proposed PromptECL model on fine-grained emotion classification tasks, and to assess its effectiveness and robustness in multi-class emotion recognition by benchmarking it against mainstream models (e.g., EmotionIC, DialogueCRN, SACL) on the IEMOCAP, MELD, and EmoryNLP datasets. PromptECL demonstrates strong overall performance in fine-grained emotion classification tasks and is further validated for effectiveness and robustness in multi-category emotion recognition scenarios.

As shown in Table 3, PromptECL achieved an average accuracy of 73.85% on the IEMOCAP dataset, outperforming all baseline models. It exhibited particularly strong performance on dynamic emotion categories such as “Excited” with 80.92% accuracy and “Frustrated” with 71.22%, demonstrating its ability to capture complex emotional dynamics in dialogue. Additionally, it maintained high accuracy on more common emotions, achieving 83.6% on “Sad” and 73.63% on “Neutral”, further validating its balanced performance across different emotional categories.

As shown in Table 4, PromptECL achieved an average accuracy of 50.08% on the MELD dataset, which is slightly lower than SACL’s 50.76%, yet demonstrates superior performance on low-frequency emotion categories. Specifically, it substantially outperformed baseline models on emo-

Method	Ang.	Exc.	Fru.	Hap.	Neu.	Sad	Ave.
EmotionIC	60.49	72.94	65.78	64.02	69.72	80.94	70.48
DialogueCRN	62.09	67.33	64.22	54.28	69.57	81.34	66.47
SACL	64.09	69.70	65.02	56.91	70.00	84.78	68.42
PromptECL	69.67	80.92	71.22	64.08	73.63	83.6	73.85

Table 3: Weighted-F1 results (%) for fine-grained emotion classification on IEMOCAP.

Method	Ang.	Dis.	Fea.	Joy.	Neu.	Sad	Sur.	Ave.
EmotionIC	54.05	23.91	17.91	64.52	79.83	42.19	60.32	48.96
DialogueCRN	52.53	32.07	18.26	64.56	79.72	39.30	57.62	49.15
SACL	52.35	31.47	26.23	64.98	80.17	41.34	58.77	50.76
PromptECL	53.47	34.43	22.79	59.25	78.73	43.03	58.86	50.08

Table 4: Weighted-F1 results (%) for fine-grained emotion classification on MELD.

Method	Joy.	Mad.	Neu.	Pea.	Pow.	Sad	sca.	Ave.
EmotionIC	-	-	-	-	-	-	-	-
DialogueCRN	54.42	36.44	53.83	10.18	4.55	25.74	37.49	31.81
SACL	54.78	37.68	55.42	11.66	5.43	25.83	37.11	32.56
PromptECL	52.07	45.07	47.31	21.48	28.07	30.11	40.25	33.77

Table 5: Weighted-F1 results (%) for fine-grained emotion classification on EmoryNLP.

tions such as “Disgust” with 34.43% accuracy and “Sadness” with 43.03%.

As shown in Table 5, PromptECL achieved an average accuracy of 33.77% on the EmoryNLP dataset, outperforming all baseline models. It demonstrated notable advantages on long-tailed emotion categories such as Powerful at 28.07%, and on ambiguous categories such as Mad at 45.07%, highlighting its effectiveness in handling sparse and challenging emotional data.

These results confirm the effectiveness of PromptECL in fine-grained emotion classification, particularly in dynamic and low-frequency categories. By incorporating a prompt strategy, it successfully captures complex emotional states in conversational contexts and exhibits strong generalization across datasets.

4 Conclusions

This study proposes the PromptECL method, which effectively activates the latent capabilities of LLMs in emotion classification tasks through a few-shot prompt template learning strategy. Although performance on the MELD dataset slightly trails current SoTA systems, ablation studies and fine-grained emotion classification experiments confirm the pivotal role of prompt template learning mechanisms in enhancing emotional discriminability. PromptECL provides a novel and effective solution for ERC tasks, with strong scalability and adaptability, and is expected to be promoted and applied in more practical applications in the future.

5 Limitations

Although the PromptECL method has achieved satisfactory experimental results, certain limitations remain. First, the performance on some datasets (e.g., MELD) still has room for improvement. Future work can further enhance the method’s effectiveness by employing more refined prompt engineering techniques or by utilizing more efficient LLMs. Second, while the analysis of emotional change causes offers a preliminary step toward emotional understanding, conducting a more systematic and comprehensive interpretable analysis of the emotion categorization decision-making process remains an important direction for future research. PromptECL presents a novel and effective approach to emotion categorization, demonstrating strong scalability and adaptability. It holds great potential for broader adoption and application in real-world scenarios.

References

- Tom B. Brown and 1 others. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901.
- Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeanette N Chang, Sungbok Lee, and Shrikanth S Narayanan. 2008. Iemocap: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42:335–359.
- Vishal Chudasama, Purbayan Kar, Ashish Gudmalwar, Nirmesh Shah, Pankaj Wasnik, and Naoyuki Onoe. 2022. M2FNet: Multi-modal fusion network for emotion recognition in conversation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 4651–4660.
- Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. Cosmic: COmmonSense knowledge for eMotion identification in conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2470–2481.
- Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. DialogueGCN: A graph convolutional neural network for emotion recognition in conversation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 154–164.
- Tanya Goyal, Junyi Jessy Li, and Greg Durrett. 2022. News summarization and evaluation in the era of GPT-3. *Computing Research Repository*, arXiv:2209.12356. Version 2.

- Dou Hu, Yinan Bao, Lingwei Wei, Wei Zhou, and Songlin Hu. 2023a. Supervised adversarial contrastive learning for emotion recognition in conversations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10835–10852.
- Dou Hu, Lingwei Wei, and Xiaoyong Huai. 2021. Dialoguecrn: Contextual reasoning networks for emotion recognition in conversations. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7042–7052.
- Hanxu Hu, Hongyuan Lu, Huajian Zhang, Yun-Ze Song, Wai Lam, and Yue Zhang. 2023b. Chain-of-symbol prompting elicits planning in large language models. *Computing Research Repository*, arXiv:2305.10276. Version 7.
- 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*, pages 7360–7370.
- Shanglin Lei, Guanting Dong, Xiaoping Wang, Keheng Wang, Runqi Qiao, and Sirui Wang. 2023a. InstructERC: Reforming emotion recognition in conversation with multi-task retrieval-augmented large language models. *Computing Research Repository*, arXiv:2309.11911. Version 6.
- Shanglin Lei, Xiaoping Wang, Guanting Dong, Jiang Li, and Yingjian Liu. 2023b. Watch the speakers: A hybrid continuous attribution network for emotion recognition in conversation with emotion disentanglement. In *Proceedings of the 35th IEEE International Conference on Tools with Artificial Intelligence (IC-TAI)*, pages 881–888.
- Jiang Li, Xiaoping Wang, Guoqing Lv, and Zhigang Zeng. 2024. Graphcfc: A directed graph based cross-modal feature complementation approach for multimodal conversational emotion recognition. *IEEE Transactions on Multimedia*, 26(1):77–89.
- Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. *Computing Research Repository*, arXiv:1710.03957. Version 1.
- Chen Liang, Jing Xu, Yangkun Lin, Chong Yang, and Yongliang Wang. 2022. S+PAGE: A speaker and position-aware graph neural network model for emotion recognition in conversation. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*, pages 148–157.
- Yingjian Liu, Jiang Li, Xiaoping Wang, and 1 others. 2024. EmotionIC: Emotional inertia and contagion-driven dependency modeling for emotion recognition

728	in conversation. <i>Science China Information Sciences</i> ,	783
729	67:182103.	784
730	Jieyi Long. 2023. Large language model guided	785
731	tree-of-thought . <i>Computing Research Repository</i> ,	786
732	arXiv:2305.08291.	787
733	Hui Ma, Jian Wang, Hongfei Lin, Bo Zhang, Yijia	788
734	Zhang, and Bo Xu. 2024. A transformer-based model	789
735	with self-distillation for multimodal emotion recog-	790
736	nition in conversations . <i>IEEE Transactions on Multi-</i>	791
737	<i>media</i> , 26:776–788.	792
738	Navonil Majumder, Soujanya Poria, Devamanyu Haz-	793
739	arika, Rada Mihalcea, Alexander Gelbukh, and Erik	794
740	Cambria. 2019. DialogueRNN: An attentive RNN for	795
741	emotion detection in conversations. In <i>Proceedings</i>	
742	<i>of the Thirty-Third AAAI Conference on Artificial</i>	
743	<i>Intelligence</i> , pages 6818–6825.	
744	Milad Moradi, Kathrin Blagec, Florian Haberl, and	
745	Matthias Samwald. 2021. GPT-3 models are poor	
746	few-shot learners in the biomedical domain . <i>Comput-</i>	
747	<i>ing Research Repository</i> , arXiv:2109.02555. Version	
748	2.	
749	Long Ouyang and 1 others. 2022. Training language	
750	models to follow instructions with human feedback.	
751	In <i>Advances in Neural Information Processing Sys-</i>	
752	<i>tems</i> , volume 36, pages 27730–27744.	
753	Soujanya Poria, Navonil Majumder, Rada Mihalcea,	
754	and Eduard Hovy. 2019. Emotion recognition in con-	
755	versation: Research challenges, datasets, and recent	
756	advances. <i>IEEE access</i> , 7:100943–100953.	
757	Xiangyu Qin, Zhiyu Wu, Tingting Zhang, Yanran Li,	
758	Jian Luan, Bin Wang, Li Wang, and Jinshi Cui. 2023.	
759	Bert-erc: fine-tuning bert is enough for emotion	
760	recognition in conversation. In <i>Proceedings of the</i>	
761	<i>Thirty-Seventh AAAI Conference on Artificial Intel-</i>	
762	<i>ligence and Thirty-Fifth Conference on Innovative</i>	
763	<i>Applications of Artificial Intelligence and Thirteenth</i>	
764	<i>Symposium on Educational Advances in Artificial</i>	
765	<i>Intelligence</i> , pages 13492–13500.	
766	Laria Reynolds and Kyle McDonell. 2021. Prompt	
767	programming for large language models: Beyond	
768	the few-shot paradigm. In <i>Extended Abstracts of the</i>	
769	<i>CHI Conference on Human Factors in Computing</i>	
770	<i>Systems</i> , pages 1–7.	
771	Weizhou Shen, Siyue Wu, Yunyi Yang, and Xiaojun	
772	Quan. 2021. Directed acyclic graph network for	
773	conversational emotion recognition. In <i>Proceedings</i>	
774	<i>of the 59th Annual Meeting of the Association for</i>	
775	<i>Computational Linguistics and the 11th International</i>	
776	<i>Joint Conference on Natural Language Processing</i> ,	
777	pages 1551–1560.	
778	Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le,	
779	Ed Chi, Sharan Narang, Aakanksha Chowdhery, and	
780	Denny Zhou. 2022. Self-consistency improves chain	
781	of thought reasoning in language models . <i>Computing</i>	
782	<i>Research Repository</i> , arXiv:2203.11171. Version 4.	
	Zilong Wang, Hao Zhang, Chun-Liang Li, Julian Martin	
	Eisenschlos, Vincent Perot, Zifeng Wang, Lesly Mi-	
	culicich, Yasuhisa Fujii, Jingbo Shang, Chen-Yu Lee,	
	and Tomas Pfister. 2024. Chain-of-table: Evolving	
	tables in the reasoning chain for table understanding .	
	<i>Computing Research Repository</i> , arXiv:2401.04398.	
	Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten	
	Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le,	
	and Denny Zhou. 2022. Chain-of-thought prompt-	
	ing elicits reasoning in large language models. In	
	<i>Proceedings of the 36th International Conference</i>	
	<i>on Neural Information Processing Systems</i> , pages	
	24824–24837.	
	Jiawei Xue, Minh Phuc Nguyen, Blake Matheny, and	
	Linh Mai Nguyen. 2024. BiosERC: Integrating bi-	
	ography speakers supported by LLMs for ERC tasks.	
	In <i>Proceedings of the International Conference on</i>	
	<i>Artificial Neural Networks (ICANN)</i> , pages 277–292.	
	Jieying Xue, Minh Phuong Nguyen, Blake Matheny,	
	and Minh Le Nguyen. 2023. Accumulating word	
	representations in multi-level context integration for	
	ERC task. In <i>Proceedings of the 15th International</i>	
	<i>Conference on Knowledge and Systems Engineering</i> ,	
	pages 1–6.	
	Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran,	
	Thomas L. Griffiths, Yuan Cao, and Karthik	
	Narasimhan. 2023. Tree of thoughts: deliberate	
	problem solving with large language models. In	
	<i>Proceedings of the 37th International Conference</i>	
	<i>on Neural Information Processing Systems</i> , pages	
	11809–11822.	
	Yao Yao, Zuchao Li, and Hai Zhao. 2024. Got: Effec-	
	tive graph-of-thought reasoning in language models.	
	In <i>Findings of the Association for Computational</i>	
	<i>Linguistics: NAACL 2024</i> , pages 2901–2921.	
	Fangxu Yu, Junjie Guo, Zhen Wu, and Xinyu Dai. 2024.	
	Emotion-anchored contrastive learning framework	
	for emotion recognition in conversation. In <i>Find-</i>	
	<i>ings of the Association for Computational Linguis-</i>	
	<i>tics: NAACL 2024</i> , pages 4521–4534.	
	Sayyed M Zahiri and Jinho D Choi. 2018. Emotion	
	detection on tv show transcripts with sequence-based	
	convolutional neural networks. In <i>AAAI Workshops</i> ,	
	volume 18, pages 44–52.	
	Zhuosheng Zhang, Aston Zhang, Mu Li, and Alex	
	Smola. 2022. Automatic chain of thought prompt-	
	ing in large language models . <i>Computing Research</i>	
	<i>Repository</i> , arXiv:2210.03493. Version 1.	
	Yucheng Zhou, Xiubo Geng, Tao Shen, Chongyang Tao,	
	Guodong Long, Jian-Guang Lou, and Jianbing Shen.	
	2023. Thread of thought unraveling chaotic contexts .	
	<i>Computing Research Repository</i> , arXiv:2311.08734.	
	Version 1.	

A Appendix

In this appendix, Section A.1 describes related work on conversational emotion recognition and prompt engineering, Section A.2 presents experimental case studies and discussions, and Section A.3 show cases selected examples from the experimental datasets.

A.1 Datasets

(1)**IEMOCAP**(Busso et al., 2008): The IEMOCAP dataset, constructed in 2008, is a multimodal corpus featuring ten speakers engaged exclusively in dyadic conversations. Although the original dataset includes annotations for eleven emotional categories, only six primary emotion labels are consistently utilized in practice.

(2)**MELD**(Poria et al., 2019): Introduced in 2019, MELD is a multimodal (visual, acoustic, and textual) emotion recognition dataset derived from the television series Friends. It contains over 1,400 dialogues comprising approximately 13,000 utterances. The dataset provides two levels of emotional annotation: coarse-grained emotion categories and fine-grained sentiment labels.

(3)**EmoryNLP**(Zahiri and Choi, 2018): Similar to MELD, EmoryNLP is also sourced from Friends, consisting of 97 episodes with 897 scenes and 12,606 utterances.

(4)**DailyDialog**(Li et al., 2017): As the largest dataset in our evaluation, DailyDialog comprises everyday conversations annotated with seven different emotional categories. Consistent with prior research protocols, we exclude the NEUTRAL-labeled instances during performance evaluation.

Detailed statistical characteristics of these four datasets are presented in Table 6.

Dataset	Dialogue			Utterances			CLS
	train	dev	test	train	dev	test	
IEMOCAP	108	12	31	5163	647	1623	6
MELD	1039	114	280	9989	1109	2610	7
EmoryNLP	659	89	79	7551	954	984	7
DailyDialog	11118	1000	1000	87823	7912	7836	7

Table 6: Statistical information on four datasets.

A.2 Related Work

A.2.1 Emotion Recognition in Conversation

ERC aims to analyze the emotional state of each utterance in a conversation, which is a key task in building human-computer interaction systems with empathy. ERC models can be classified into

key categories based on baseline architectures that model contextual dialogue information through diverse technical implementations.

Recurrent-based methods: DialogueRNN (Majumder et al., 2019) explicitly models interactions between speakers by dynamically tracking the state of each participant in a conversation through the attention mechanism and RNN. The method exploits the temporal modelling capability of the RNN at the feature extraction layer to capture the local dependencies of the conversation. COSMIC (Ghosal et al., 2020) enhances the context modeling capabilities of GRU by integrating commonsense knowledge, such as mental states and causal reasoning, to mitigate issues related to emotion transfer and category confusion. SACL-LSTM (Hu et al., 2023a) introduces an adversarial contrastive learning framework based on LSTM, which employs supervised adversarial training by generating perturbed samples and optimizing the feature distribution through a label-aware contrastive loss. Its recurrent module (LSTM) enhances the model’s robustness to contextual perturbations while preserving fine-grained emotional features. HCAN (Lei et al., 2023b) proposes an architecture that integrates recurrent and attentional mechanisms to model affective continuity and captures intra- and inter-speaker emotional dynamics through affective attribution coding. Additionally, a cognitive loss function is introduced to mitigate overfitting in speaker modeling and enhance cross-scene generalization. EmotionIC (Liu et al., 2024) further innovates in the recursive framework by proposing the Dialogue Gating Recursive Unit (DiaGRU), which combines Identity Masked Multihead Attention (IMMHA) to model local (speaker-perceived) and global (identity-perceived) context dependencies, respectively. Its recursive module dynamically fuses temporal information through the gating mechanism, which significantly improves the modeling of complex emotional inertia and contagion effects. DialogueCRN (Hu et al., 2021) integrates contextual cues by employing an LSTM network to acquire situational-level and speaker-level context, and retaining and updating dynamic memories.

Graph-based methods: DialogueGCN (Ghosal et al., 2019) leverages interlocutor self and inter-speaker dependencies to recognize emotions by modeling conversational context with graph convolutional neural networks. It addresses the context propagation limitations of RNN-based methods by utilizing graph structures to capture complex re-

lational dynamics within conversations. S+PAGE (Liang et al., 2022) enhances edge representation in emotion recognition by incorporating relative positional encoding into edge weights and speaker-dependent encoding into edge types, resulting in a more effective aggregation algorithm for ERC. In addition, we propose the Dual-Stream Dialog Transformer to extract contextual features related to both the target speaker and other interlocutors for each utterance, effectively capturing speaker- and location-aware dialog structure information. GraphCFC (Li et al., 2024) introduces a cross-modal feature complementation module based on directed graphs to address the heterogeneity gap in multimodal fusion, employing multiple subspace extractors and pairwise cross-modal complementation strategies. It extracts various edge types from the constructed graphs for encoding and designs a GNN structure combining GAT and MLP, providing a unified framework for multimodal learning that effectively models context and interaction information. RGAT (Ishiwatari et al., 2020) introduces relational location encoding to equip relational graph attention networks with sequential information that reflects the structure of the relational graph. This enables the model to capture both speaker-dependent and sequential information, effectively addressing the limitation of graph neural networks in dialog emotion recognition, which typically neglect sequential dependencies.

Transformer-based methods:The transformer-based model with self-distillation (SDT) (Ma et al., 2024) utilizes both intramodal and intermodal transformer to capture the interaction information between intramodal and intermodal modes. By designing a hierarchical gated fusion strategy, different inter-modal weights are dynamically learned to achieve multi-modal information fusion. In order to learn more expressive modal representations, a self-distillation mechanism is introduced to transfer the hard label and soft label knowledge of the model to each modality to assist training.

The Multi-modal Fusion Network (M2FNet) (Chudasama et al., 2022) extracts emotionally relevant features from visual, audio, and textual modalities, and employs a fusion mechanism based on multiple heads of attention to combine the emotionally rich latent representations of each modality. Furthermore, a novel feature extractor is introduced to process both audio and visual modalities, which is trained using an adaptive marginal triad loss function to maximize the relevance of the ex-

tracted features to emotion.

Traditional PLM-based methods:The EACL framework (Yu et al., 2024) constructs emotion-anchored contrastive learning via SimCSE-Roberta-Large. The method innovatively uses label encoding as an anchor to guide representation learning, designs an auxiliary loss function to enhance the differentiation ability of similar emotions (e.g., excitement vs. happiness), and adjusts the anchor classification effect through an adapter. The DAG-ERC model (Shen et al., 2021) encodes the dialog structure as a directed acyclic graph. The synergistic advantage of graph neural networks and recurrent neural networks is achieved by adopting RoBERTa-Large as a feature extractor, obtaining discourse representations through pooled embedding of CLS tokens, and combining with a DAG network to simultaneously capture the information flow between the long-range conversational context and the neighboring context. (Xue et al., 2023) proposed a hierarchical fine-tuning strategy based on RoBERTa-large, wherein the embedding layer and the first 10 layers of the Transformer were frozen, and only the last two layers were fine-tuned. This approach enhanced contextual integration through a cumulative word vector representation, coupled with a transentropic loss function category weighting strategy, which effectively mitigated the issue of the uneven distribution of emotional labels.

Generative PLM-based methods:The BiosERC framework (Xue et al., 2024) utilizes a LLM to extract background knowledge such as the speaker’s personality traits and injects them into the model as complementary information to assist in classifying the emotional labels of each sentence. The Instructerc framework (Lei et al., 2023a) designs a retrieval template module to integrate multi-granularity dialog supervision information. This approach introduces two auxiliary tasks of speaker identification and emotional prediction to help LLMs model conversational role relationships and future emotional tendencies.

A.2.2 Prompt Engineering

Prompt engineering is a pivotal technique in natural language processing that facilitates downstream task adaptation of pre-trained language models without updating their parameters. Instead of traditional fine-tuning, it leverages carefully designed contextual cues to activate the model’s latent knowledge.

COT (Wang et al., 2022) guides models to simulate the human logical decomposition process by explicitly presenting examples of multistep reasoning paths, such as step-by-step solutions to mathematical problems. Auto-CoT (Zhang et al., 2022) reduces the cost of manual annotation by automatically generating diverse reasoning chains from “step-by-step” instructions and constructing a robust demonstration set via multi-path sampling, significantly improving few-shot learning performance. Self-Consistency decoding (Wang et al., 2022) enhances the reliability of complex reasoning in prompt-based models by generating multiple diverse reasoning paths through stochastic decoding and selecting the most consistent answer via majority voting.

Chain-of-Symbols (CoS) (Hu et al., 2023b) utilizes formal symbolic representations to reduce semantic ambiguity in spatial reasoning tasks. This approach replaces natural language with formal symbols, significantly enhancing the model’s interpretability in reasoning tasks. Tree-of-Thoughts (ToT) (Yao et al., 2023; Long, 2023) addresses nonlinear reasoning by dynamically evaluating intermediate states through tree search algorithms, enabling systematic path exploration and backtracking. This method has improved the model’s capability in multi-step, complex reasoning tasks. Graph-of-Thoughts (GoT) (Yao et al., 2024) extends the tree-based approach by incorporating graph structures, supporting the interaction and aggregation of multi-branch reasoning, and overcoming the limitations of traditional linear reasoning. Zhou and Geng (Zhou et al., 2023) develop the ThoT (Thought of Thought) framework, which adopts a two-phase contextual partitioning strategy: fragmented semantic parsing of chaotic information, followed by incremental refinement to generate the final response. Wang and Zhang (Wang et al., 2024) proposed the Chain-of-Table approach for structured data, which realizes the visualization and verifiability of table reasoning by dynamically generating SQL-like operations, and effectively solves the problem of the gap between symbolic logic and natural language representations.

A.3 Case Study and Discussion

An in-depth analysis of four cross-domain dialogue cases reveals that the LLM guided by a specialized prompt template exhibits notable advantages

in emotion analysis. By incorporating intent recognition and requiring the model to infer the underlying causes of emotions alongside label prediction, the approach enables a human-like reasoning process. It allows the model to progressively trace the emotional logic from surface semantics to deeper contextual cues, and from isolated emotional states to complex emotional chains, ultimately producing highly credible conclusions. This approach improves the depth of the model’s emotional analysis and significantly enhances its interpretability.

The prompt-based approach enables the model to analyze the logic of emotion generation from multiple dimensions by jointly predicting emotion labels and their underlying causes. For example, in the IEMOCAP dataset, the model labels the utterance “Yeah” as “excited” and further explains the emotion’s origin: the label “excited” reflects heightened enthusiasm and positive affect, indicating the speaker’s joy or eagerness regarding the topic. This form of analysis aligns closely with human reasoning. The model integrates multiple cues—such as the topic of “marriage,” laughter, and repeated exclamations like “Oh my gosh”—to infer the speaker’s excitement. By requiring the model to generate the cause of the emotion, this approach prompts it to extract salient contextual signals and construct a causal chain of emotion generation, thereby enhancing the interpretability of the analysis.

Secondly, intention recognition is incorporated into the emotion perception stage of LLMs by explicitly embedding it in the prompt cues. This allows the model to better infer the speaker’s underlying motivations and emotional states. For instance, in the EmoryNLP dataset, the utterance “That only took me an hour” is labeled as Sad through the reasoning chain: “young divorce → social pressure → self-deprecating defense.” This simulates the human cognitive mechanism of intention recognition: the model first identifies the key feature of a “divorce at age 26,” links it to societal expectations around the timing of marriage, and then constructs a hypothesis based on psychological defense mechanisms. Secondly, intention recognition is integrated into the emotion perception stage of LLMs by embedding it explicitly within prompt cues, enabling deeper inference of the speaker’s underlying motivations and emotional states. For example, in the EmoryNLP dataset, the utterance “That only took me an hour” is labeled as Sad through the reasoning chain: “young divorce → social pres-

sure → self-deprecating defense.” This mirrors the human cognitive process of intention recognition: the model identifies the key factor—divorce at age 26—relates it to societal expectations regarding marriage timelines, and infers a psychological defense mechanism.

In addition, the approach enables the model to track emotional dynamics throughout a dialogue by monitoring changes in emotional states. For instance, in the DailyDialog dataset, the speaker’s emotional trajectory follows the pattern anger → neutral → anger. The model labels the utterance “This is highway robbery. I’ve got a good mind to call the police!” as anger, and further explains that the speaker feels cheated and frustrated by unreasonable charges, considering legal action as a response to perceived injustice. Beyond identifying basic emotional triggers, the model also captures a deeper sense of institutional anger. This capability to model emotion evolution demonstrates the effectiveness of prompt cues in retaining emotional memory, thereby enhancing both the depth and interpretability of emotion analysis.

Furthermore, the approach leverages the model’s inherent capacity for metaphorical interpretation, allowing it to handle unconventional emotional expressions in complex contexts. For example, in the MELD dataset, the model labels the utterance “Bumpy?” as surprise, and further explains: “Ross is surprised by Jade’s description of the experience as ‘bumpy,’ which appears unusual or unexpected.” This analysis illustrates the model’s ability to comprehend metaphorical language. It not only identifies the semantic shift of “bumpy” in an otherwise neutral context, but also incorporates Jade’s earlier reference to “awkwardness” to infer Ross’s surprise.

Our research demonstrates that with carefully structured prompt templates, LLMs can transcend shallow feature extraction and achieve multi-level alignment with human emotional cognition. This provides a methodological foundation for the development of affective computing systems rooted in social common sense, and simultaneously advances the cognitive evaluation of interpretable AI.

A.4 Presentation of Samples

This appendix shows partial examples of conversation datasets for fine-tuned LLMs, covering four multi-domain conversation datasets, IEMOCAP (as shown in Table 7), MELD (as shown in Table 8), EmoryNLP (as shown in Table 9), and DailyDia-

log (as shown in Table 10). In each table, section **Q** consists of bolded fixed prompt templates with italicized contextual fills to form the model input, and section **A** corresponds to the model’s generated responses. These examples visualize the model’s response patterns in diverse dialog contexts. The selected examples take into account the representativeness of dialogue types and the diversity of question-answer quality, and focus on presenting the implementation of contextual linkage and command control in the prompting project, which provides a traceable empirical basis for the analysis of model behaviors in the main text.

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Ses05F_impro03_F: Okay, so big news. Ses05F_impro03_M: What? Ses05F_impro03_M: Come on, what? Ses05F_impro03_F: I'm getting married. Ses05F_impro03_M: Shut up. Ses05F_impro03_F: Yeah. Ses05F_impro03_M: No way. Ses05F_impro03_F: Yeah. Ses05F_impro03_M: He asked you? Ses05F_impro03_F: Uh-huh. Ses05F_impro03_M: No way, when? when,when, When did it happen? Ses05F_impro03_F: Just a couple days ago. Ses05F_impro03_M: Oh my gosh. Ses05F_impro03_F: [LAUGHTER] Ses05F_impro03_M: And you actually said yes. I can't believe it. I never thought you would get married. Ses05F_impro03_F: I know me neither. Ses05F_impro03_M: Oh my gosh. Ses05F_impro03_F: Right, I thought I would be way older too. I thought I would be I be at least thirty something. Ses05F_impro03_M: Oh absolutely. Ses05F_impro03_M: Well, we always said we'd be in our thirtieth when we got married. And we were going to get married to each other if we didn't by then###symbols involve multiple speakers. Based on the above historical dialogue sentences, Please analyze the emotional label and reasons for the current sentence <Ses05F_impro03_F: Yeah.>, considering the current speaker's emotional changes history [unknown->excited->excited], and the previous speaker's utterance [Well, we always said we'd be in our thirtieth when we got married. And we were going to get married to each other if we didn't by then] and emotion [excited]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <neutral,frustrated,angry,sad, happy,excited>, and provide the reason for the emotional label.Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output.

A: label:excited emotional reason:The excited (exc) label indicates a high level of enthusiasm and positive emotion, reflecting the speaker's joy or excitement about the topic.

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Ses05F_impro04_F: Brian, I need help. Ses05F_impro04_M: Babe, I don't know what to tell you. Don't give up. Ses05F_impro04_F: I don't-I'm just-I'm thinking maybe I should move back home or something. I don't I don't know what to do. I can't - I cant keep living the way I'm living. Ses05F_impro04_M: Well, of course not, but what are you going to do if you move back home? Are you going to be able to find a job there? Ses05F_impro04_F: I don't know. I mean, what am I supposed to do? I've It's not for lack of effort. I've been trying.###symbols involve multiple speakers. Based on the above historical dialogue sentences, Please analyze the emotional label and reasons for the current sentence <Ses05F_impro04_M: I wish I had some answers for you, babe. I mean-I don't know what to tell you.>, considering the current speaker's emotional changes history[unknown->neutral->neutral], and the previous speaker's utterance [I don't know. I mean, what am I supposed to do? I've It's not for lack of effort. I've been trying.] and emotion [frustrated]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <neutral,frustrated,angry,sad, happy,excited>, and provide the reason for the emotional label.Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output.

A: label:neutral emotional reason:The neutral (neu) label indicates a tone of empathy and support without expressing strong emotions, reflecting a sense of helplessness in finding a solution.

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Ses05F_impro04_F: Brian, I need help. Ses05F_impro04_M: Babe, I don't know what to tell you. Don't give up. Ses05F_impro04_F: I don't-I'm just-I'm thinking maybe I should move back home or something. I don't I don't know what to do. I can't - I cant keep living the way I'm living. Ses05F_impro04_M: Well, of course not, but what are you going to do if you move back home? Are you going to be able to find a job there? Ses05F_impro04_F: I don't know. I mean, what am I supposed to do? I've It's not for lack of effort. I've been trying. Ses05F_impro04_M: I wish I had some answers for you, babe. I mean-I don't know what to tell you.###symbols involve multiple speakers. Based on the above historical dialogue sentences, Please analyze the emotional label and reasons for the current sentence <Ses05F_impro04_F: How did you get a job? What am I - What am I supposed to do?>, considering the current speaker's emotional changes history [sad->unknown->frustrated], and the previous speaker's utterance [I wish I had some answers for you, babe. I mean-I don't know what to tell you.] and emotion [frustrated]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <neutral,frustrated,angry,sad, happy,excited>, and provide the reason for the emotional label.Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output.

A: label:frustrated emotional reason:The frustration (fru) label indicates a sense of desperation and exasperation, reflecting the speaker's struggle and seek for a way out of their current situation.

Table 7: Example of selecting a full Q&A pair from IEMOCAP

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Jade: Oh, Bob, he was nothing compared to you. I had to bite my lip to keep from screaming your name. Chandler: Well, that makes me feel so good. Jade: It was just so awkward and bumpy.###symbols involve multiple speakers. Based on the above historical dialogue sentences, Please analyze the emotional label and reasons for the current sentence <Ross: Bumpy?>, considering the current speaker’s emotional changes history [unknown->unknown->unknown], and the previous speaker’s utterance [It was just so awkward and bumpy.] and emotion [neutral]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <neutral,surprise,fear,sadness,joy,disgust,anger>, and provide the reason for the emotional label.Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output.
A: label:neutral emotional reason:Ross is feeling neutral, likely relieved that the challenge has been completed, but not overly excited or emotional.

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Jade: Oh, Bob, he was nothing compared to you. I had to bite my lip to keep from screaming your name. Chandler: Well, that makes me feel so good. Jade: It was just so awkward and bumpy.###symbols involve multiple speakers. Based on the above historical dialogue sentences, Please analyze the emotional label and reasons for the current sentence <Ross: Bumpy?>, considering the current speaker’s emotional changes history [unknown->unknown->unknown], and the previous speaker’s utterance [It was just so awkward and bumpy.] and emotion [neutral]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <neutral,surprise,fear,sadness,joy,disgust,anger>, and provide the reason for the emotional label.Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output.
A: label:surprise emotional reason:Ross is surprised by Jade’s description of the experience as "bumpy," which seems unusual or unexpected.

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Phoebe: Yeah! Sure! Yep! Oh, y’know what? If I heard a shot right now, I’d throw my body on you. Gary: Oh yeah? Well maybe you and I should take a walk through a bad neighborhood. Phoebe: Okay! Phoebe: Bye! Joey: Bye!###symbols involve multiple speakers. Based on the above historical dialogue sentences, Please analyze the emotional label and reasons for the current sentence <Joey: Cut it out Ross! I hate to have to save your life and kick your ass in the same day!>, considering the current speaker’s emotional changes history [unknown->unknown->neutral], and the previous speaker’s utterance [Bye!] and emotion [neutral]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <neutral,surprise,fear,sadness,joy,disgust,anger>, and provide the reason for the emotional label.Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output.
A: label:anger emotional reason:Joey is feeling angry due to Ross’s repeated actions that require him to intervene, both literally and metaphorically.

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Ross: Hi! Rachel and Bonnie: Hi! Bonnie: Rachel was just helping me out. My head got all sunburned. Ross: Awww.###symbols involve multiple speakers. Based on the above historical dialogue sentences, Please analyze the emotional label and reasons for the current sentence <Bonnie: Thanks a million.>, considering the current speaker’s emotional changes history [unknown->unknown->neutral], and the previous speaker’s utterance [Awww.] and emotion [joy]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <neutral,surprise,fear,sadness,joy,disgust,anger>, and provide the reason for the emotional label.Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output.
A: label:anger emotional reason:Joey is feeling angry due to Ross’s repeated actions that require him to intervene, both literally and metaphorically.

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Joey: That part was perfect for me! I can’t believe I didn’t get it! Ross: I’m sorry, man. Ross: Hey, y’know what you should do? Ross: You should make something happen for yourself. Ross: Y’know, like-like write a play. Ross: Write a movie! Ross: Huh? Ross: I mean, what about those###symbols involve multiple speakers. Based on the above historical dialogue sentences, Please analyze the emotional label and reasons for the current sentence <Joey: Come on Ross be realistic, y’know? If I did write something, what are the chances I could get those guys to star in it?>, considering the current speaker’s emotional changes history [unknown->unknown->anger], and the previous speaker’s utterance [I mean, what about those] and emotion [neutral]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <neutral,surprise,fear,sadness,joy,disgust,anger>, and provide the reason for the emotional label.Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output.
A: label:sadness emotional reason:Joey is feeling sad due to the realism of his situation and the improbability of his idea being realized.

Table 8: Example of selecting a full Q&A pair from MELD

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Ross Geller: I'm divorced! I'm only 26 and I'm divorced! Joey Tribbiani: Shut up! Chandler Bing: You must stop!###symbols involve multiple speakers. Based on the above historical dialogue sentences, please analyze the potential intent of the current speaker in the sentence <Ross Geller: That only took me an hour.>, considering the current speaker's emotional changes history [none->none->Sad], and the previous speaker's utterance [You must stop!] and emotion [Mad]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <Neutral, Joyful, Peaceful, Powerful, Scared, Mad, Sad>, and provide the reason for the emotional label.Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output.

A: label:Sad emotion reason:Ross's previous expression of being divorced at a young age and the context of Chandler's angry reaction ("You must stop!") suggest that Ross is feeling down about his situation. The statement "That only took me an hour" can be interpreted as a sarcastic or self-deprecating remark about how quickly his marriage ended, reflecting his sad and frustrated emotional state.

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Joey Tribbiani: Of course it was a line! Monica Geller: Why?! Why? Why, why would anybody do something like that? Ross Geller: I assume we're looking for an answer more sophisticated than 'to get you into bed'. Monica Geller: I hate men! I hate men! Phoebe Buffay: Oh no, don't hate, you don't want to put that out into the universe. Monica Geller: Is it me? Is it like I have some sort of beacon that only dogs and men with severe emotional problems can hear? Phoebe Buffay: All right, c'mere, gimme your feet. Monica Geller: I just thought he was nice, y'know? Joey Tribbiani: I can't believe you didn't know it was a line! Rachel Green: Guess what? Ross Geller: You got a job? Rachel Green: Are you kidding? I'm trained for nothing! I was laughed out of twelve interviews today. Chandler Bing: And yet you're surprisingly upbeat. Rachel Green: You would be too if you found John and David boots on sale, fifty percent off! Chandler Bing: Oh, how well you know me... Rachel Green: They're my new 'I don't need a job, I don't need my parents, I've got great boots' boots! Monica Geller: How'd you pay for them? Rachel Green: Uh, credit card. Monica Geller: And who pays for that?###symbols involve multiple speakers. Based on the above historical dialogue sentences, please analyze the potential intent of the current speaker in the sentence <Rachel Green: Um... my... father.>, considering the current speaker's emotional changes history [Joyful->Joyful->Neutral], and the previous speaker's utterance [And who pays for that?] and emotion [Neutral]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <Neutral, Joyful, Peaceful, Powerful, Scared, Mad, Sad>, and provide the reason for the emotional label.Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output.

A: label:Neutral emotion reason:Rachel's shift from a joyful state to a more neutral one is evident as she is confronted with the reality of her financial dependence, which contrasts with her newfound confidence from the boot purchase. Her response is somewhat defensive and reluctant, indicating an awareness of the potential criticism or judgment from Monica, but not necessarily feeling sad or joyful about it.

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Monica Geller: Hey! Rachel Green: Umm, when were you gonna tell me that you're going out with Chip Matthews? Monica Geller: Now? Is it okay if I go out with Chip Matthews? Rachel Green: Nooo! It's not okay! I can't believe you would want to after what he did to me! Monica Geller: What, that little thing at the prom? Rachel Green: Monica! I couldn't find him for two hours! He was having sex with Amy Welch! Monica Geller: Come on, that was back in high school! How could that still bother you? Rachel Green: I mean why, of all people would you want to go out with Chip?! Monica Geller: Look, you and I went to different high schools...###symbols involve multiple speakers. Based on the above historical dialogue sentences, please analyze the potential intent of the current speaker in the sentence <Rachel Green: Okay, that doesn't help me, because we went to the same high school.>, considering the current speaker's emotional changes history [Mad->Mad->Mad], and the previous speaker's utterance [Look, you and I went to different high schools...] and emotion [Neutral]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <Neutral, Joyful, Peaceful, Powerful, Scared, Mad, Sad>, and provide the reason for the emotional label.Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output.

A: label:Mad emotion reason:Rachel is already upset and angry about the situation, and Monica's statement about attending different high schools seems to invalidate Rachel's past experiences and emotions, further aggravating her frustration.

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Rachel Green: Is that the heartbeat?###symbols involve multiple speakers. Based on the above historical dialogue sentences, please analyze the potential intent of the current speaker in the sentence <The Doctor: That's it.>, considering the current speaker's emotional changes history [none->none->none], and the previous speaker's utterance [Is that the heartbeat?] and emotion [Joyful]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <Neutral, Joyful, Peaceful, Powerful, Scared, Mad, Sad>, and provide the reason for the emotional label.Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output.

A: label:Joyful emotion reason:The Doctor is likely responding to Rachel's joyful question with a confirmation that matches her positive emotional state. The context suggests a shared moment of excitement or discovery, which is mirrored in the Doctor's response, indicating a joyful emotion.

Table 9: Example of selecting a full Q&A pair from EmoryNLP

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Speaker 1: Believe it or not , tea is the most popular beverage in the world after water . Speaker 2: Well , people from Asia to Europe all enjoy tea . Speaker 1: Right . And China is the homeland of tea . Speaker 2: Yes , Chinese people love drinking tea so much . Some even claim they can't live without tea . Speaker 1: Do you know there are several catagories of Chinese tea ? Speaker 2: Yes , I believe there are green teas , black teas and scented teas . Any Others ? Speaker 1: Well , have you ever heard of Oulong tea and compressed tea ? Speaker 2: Oh , yeah . Oulong tea is good for one's health . isn't it ? Speaker 1: You surely know a lot about Chinese tea . Speaker 2: Sure , I like drinking tea at teahouses . Speaker 1: Oh , so do I . Speaker 2: Why don't we go for one now ?###symbols involve multiple speakers. Based on the above historical dialogue sentences, Please analyze the emotional label and reasons for the current sentence <Speaker 1: Great . We can chat while enjoying a cup there .>, considering the current speaker's emotional changes history [neutral->neutral->neutral]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <neutral,disgust,happiness,sadness,anger,surprise,fear>, and provide the reason for the emotional label. Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output. A: label:happiness emotion reason:Expresses happiness and looking forward to the enjoyable experience of drinking tea and chatting together.

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Speaker 1: Hi , Dan . What's new ? Speaker 2: Where have you been ? I've been trying to get in touch with you for two days . Speaker 1: I went to Salt Lake City on business . What's up ?###symbols involve multiple speakers. Based on the above historical dialogue sentences, Please analyze the emotional label and reasons for the current sentence <Speaker 2: I got fired .>, considering the current speaker's emotional changes history [unknown->unknown->neutral]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <neutral,disgust,happiness,sadness,anger,surprise,fear>, and provide the reason for the emotional label. Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output. A: label:sadness emotion reason:Expresses sadness and disappointment over being laid off.

Q: You are an expert in dialogue emotion recognition and analysis. The emotional conversation records enclosed within the ###Speaker 1: You guys are charging me \$ 10 for a movie that I never ordered or saw . Speaker 2: Let's see , sir . According to your file , you watched ' Titanic ' Monday evening . Speaker 1: Well , the wrong information is in my file . I was at a concert Monday night . Speaker 2: Well , your word overrules the file , sir . One moment , please . Speaker 1: I knew you'd see it my way . Speaker 2: Sir , I deleted the \$ 10 , but I had to add a \$ 2 service charge to your bill . Speaker 1: Am I in the Twilight Zone ? You're charging me for a movie I never saw ? Speaker 2: Please don't blame me , sir . Blame the computer programmer .###symbols involve multiple speakers. Based on the above historical dialogue sentences, Please analyze the emotional label and reasons for the current sentence <Speaker 1: This is highway robbery . I've got a good mind to call the police !>, considering the current speaker's emotional changes history [anger->neutral->anger]. Please analyze the potential intention of the current speaker, choose the emotional label of the sentence, which can only be selected from <neutral,disgust,happiness,sadness,anger,surprise,fear>, and provide the reason for the emotional label. Only allow you to output the following format content, the output format is: label: xxx emotion reason: xxx , xxx is the content you output. A: label:anger emotion reason:The speaker is angry and feels cheated and frustrated by the charges, considering legal action to address the injustice.

Table 10: Example of selecting a full Q&A pair from DailyDailog