

JoPR: Joint Emotion Perception and Reasoning for Conversational Emotion Recognition

Yumeng Fu¹, Weitao Huang¹, Junjie Wu²,
Hao Teng¹, Meishan Zhang³, Bingquan Liu^{1*},

¹School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China,

²School of Computer Science and Technology, Soochow University, Suzhou, China,

³School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China,

24b303004@stu.hit.edu.cn, huangweitao@insun.hit.edu.cn, 20224027010@stu.suda.edu.cn

25s103237@stu.hit.edu.cn, mason.zms@gmail.com, liubq@hit.edu.cn

Abstract

Emotion Recognition in Conversation (ERC), the task of identifying the emotion of each utterance in a conversation, is crucial for human-machine interaction. Existing LLM-based ERC methods focus on standard prompting and slow thinking for emotion analysis. However, they suffer from the lack of human-like emotion reasoning and discrimination between similar emotions, thus limiting accurate emotion predictions. To this end, we present JoPR, jointing perception-curriculum learning and emotional reasoning for conversational emotion recognition. Specifically, we devise a multi-dimension curriculum with long CoT fine-tuning to clone human-like emotion reasoning. We further design an emotion-specific reward function in a novel reinforcement learning framework, thereby enhancing the discernment between similar emotions. We conduct extensive experiments on three widely used benchmark datasets, and the results demonstrate that our JoPR achieves consistent and significant improvements over baselines.

1 Introduction

Emotion recognition in conversation (ERC) aims to identify the emotion of each utterance within a conversation, which is a crucial task in the community of natural language processing (NLP). The task has received increasing attention due to its potential applications, such as mental health assessment (Talati, 2023), personalized content recommendation (Song et al., 2024), socially adaptive robots (Ackermann et al., 2025), and so on.

With the rapid development of large language models (LLMs), researchers have exhibited the perception and reasoning capabilities of large models in the realm of ERC. For example, InstructERC (Lei et al., 2023) leverages prompt tuning to incorporate speaker identity and conversational context.

*Corresponding author

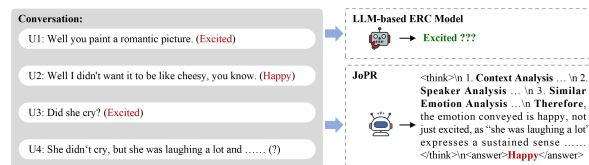


Figure 1: Comparison between previous ERC models and our method. (a) Existing ERC methods achieve emotion perception. (b) Our method performs deep thought process over context, speaker, and the distinctions between similar emotions, to facilitate accurate emotion reasoning and perception.

BiosERC (Xue et al., 2024) enhances emotional understanding by generating speaker biographical knowledge. LaERC-S (Fu et al., 2025) dynamically integrates commonsense knowledge. Despite their impressive performance, they still struggle to accurately provide the boundary between similar emotion labels.

Recent advancements in language reasoning models (LRMs), such as o1 (Jaech et al., 2024) and DeepSeek R1 (Guo et al., 2025), have attracted an enormous amount of attention from both classification and reasoning. Their success stems primarily from slow thinking (De Neys, 2023). Advanced methods leverage reinforcement learning (RL) with reward functions to guide models for slow thinking, thus yielding impressive performance in complex reasoning tasks (Chen et al., 2025; Xiang et al., 2024; Zhu et al., 2025; Tang et al., 2025). However, such attempt for the task of ERC remain under-explored. Therefore, utilizing reinforcement learning with emotion-specific reward function is potential for accurate emotion predictions.

In this paper, we introduce JoPR, a novel framework that joints perception-curriculum learning and emotional reasoning for conversational emotion recognition. JoPR combines curriculum learning (CL) with RL to alleviate the confusion between similar emotion labels in a human-like reasoning manner. First, we use supervised fine-tuning (SFT)

to inject emotion-specific structured reasoning and long chain-of-thought (CoT) capability into the model. After then, leveraging direct preference optimization (DPO) to achieve the consistency between emotion reasoning and answers. Finally, we employ group-relative policy optimization (GRPO) with emotion-level reward to guide the model in freely exploring unseen scenarios. In addition, we propose a multi-dimensional difficulty measurer to guide an easy-to-hard perception curriculum learning process, thereby emulating the human cognitive mechanism of stable and progressive learning.

We conduct experiments on three widely benchmark datasets, i.e., IEMOCAP, EmoryNLP and MELD. The results show that our method leads to significant improvements compared to baseline model, achieving increases of 3.61% on IEMOCAP (Busso et al., 2008), 2.70% on MELD (Poria et al., 2018) dataset and 3.12% on EmoryNLP (Zahiri and Choi, 2018). We further conduct analyses to better understand the efficacy of our method and uncover several noteworthy findings. We find that this emotion perception curriculum learning approach for measuring sample difficulty is not only well-suited for our own model but also yields improvements in existing LLM-based ERC models, such as InstructERC and LaERC-S. This demonstrates the generalizability of our method. Furthermore, the thought process behind this grading reward function in reinforcement learning significantly enhances the ability to distinguish similar emotions when analyzing the sources of those emotions.

In summary, our contributions are three-fold:

- We propose JoPR, a coarse-to-fine learning framework (SFT-DPO-GRPO), which simultaneously introduces joint emotion perception and human reasoning for conversational emotion recognition.
- We progressively enhance human emotion reasoning capabilities by a GRPO approach with emotion-level reward and propose a multi-dimensional difficulty measurer to guide easy-to-hard perception curriculum learning.
- Extensive experiments conducted on three benchmark datasets IEMOCAP, MELD, and EmoryNLP demonstrate that our method significantly surpasses the previous methods.

2 Related Work

2.1 Emotion Recognition in Conversation

Previous approaches to ERC can be broadly classified into three categories: sequence-based methods, GNN-based methods, and LLM-based methods. **Sequence-based ERC models** (Majumder et al., 2019; Ghosal et al., 2020; Hu et al., 2023) capture emotional dynamics by modeling the temporal dependencies between utterances using architectures like RNNs or Transformers, often incorporating speaker-aware context to enhance emotion recognition. **GNN-based ERC models** (Ghosal et al., 2019; Hu et al., 2021; Shen et al., 2021b; Feng and Fan, 2025) construct dialogue graphs to capture inter-utterance and inter-speaker dependencies, leveraging graph neural networks to integrate contextual and relational information for accurate emotion recognition. **LLM-based ERC models** (Lei et al., 2023; Fu et al., 2025; Xue et al., 2024; Zhang et al., 2023b) leverage instruction tuning to integrate context, speaker, and knowledge, enabling LLMs to generate emotion.

Unlike previous work, ERC paradigms that directly map utterances to emotion labels, our reasoning model is designed to simulate the cognitive process of emotion inference, thus offering enhanced emotion perception and reasoning.

2.2 Curriculum Learning

Curriculum Learning (CL) (Wang et al., 2021) is a training strategy for machine learning models that mimics the human learning process, gradually progressing from easy to hard data. This strategy enhances the model’s comprehension capability. Curriculum Learning (CL) offers significant advantages across machine translation (Platanios et al., 2019), reinforcement learning (Narvekar et al., 2020), supervised learning (Wang et al., 2021).

In ERC task, both HCL (Yang et al., 2022) and SPCL (Song et al., 2022) attempt to incorporate curriculum learning, but they each assess sample difficulty from a single perspective. HCL is trained by increasing the number of samples with each epoch, leading to slow convergence. SPCL uses a fixed number of samples per round and focusing solely on easy samples initially and difficult samples later, which results in instability and significant training bias when there is a large disparity between difficult and easy sample distributions. To address these issues, we propose an approach based on a multi-dimension and dynamic curriculum.

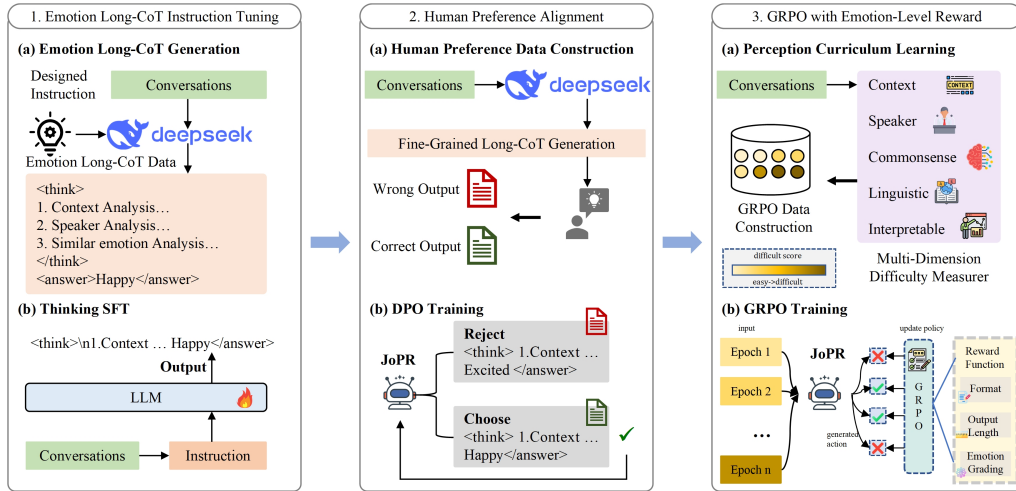


Figure 2: The overview framework of JoPR model.

2.3 Reinforcement Learning

The emergence of DeepSeek R1 (Guo et al., 2025), featuring its groundbreaking Reinforcement Learning with Verifiable Reward (RLVR), has brought reinforcement learning (RL) to the forefront of LLM research. RLVR’s rule-based reward design enables efficient, reliable, and controllable model optimization. Building on this paradigm, recent applications such as SQL-R1 (Ma et al., 2025), VideoChat-R1 (Li et al., 2025), and R1-Searcher (Song et al., 2025) have achieved notable progress in reasoning, multimodal perception, and autonomous information retrieval through tailored RL or GRPO-based training strategies. However, they suffer from the lack of human-like reasoning. And slow thinking for the task of ERC remain under-explored.

3 Method

Our method primarily consists of three parts as show in Figure 2: Emotion long chain-of-thought instruction tuning, Human-preference alignment via DPO, GRPO with emotion-level reward.

3.1 Definition

Emotion Recognition in Conversation (ERC) is the task of identifying the emotional state of a speaker within a conversational exchange. Formally, given a conversation represented as a sequence of utterance-speaker pairs, $\{(u_1, s_1), (u_2, s_2), \dots, (u_n, s_n)\}$, where u_t is the utterance and s_t is the speaker at turn t , and a predefined set of emotion labels \mathcal{E} (e.g., {excited, frustrated, sad, neutral, angry, happy}), the goal is to predict the emotion label $e_t \in \mathcal{E}$ for each current utterance u_t .

3.2 Emotion Long CoT Instruction Tuning

To inject emotion domain-specific reasoning knowledge and facilitate learning the thinking format, an instruction tuning phase. We distilled high-quality structured thought processes from the DeepSeek-R1-Distill-Qwen-32B (Team, 2024) model (prompts refer to the Appendix A.1). We decompose reasoning into three steps, including historical context, speaker information, and similar emotion decision making. The desired format includes “<think> chain of thought </think><answer> emotion label</answer>”. This thought process is then used as the training target to fine-tune large language model, thereby imbuing it with the ability to think.

3.3 Human-preference Alignment via DPO

To align the model with human reasoning, we fine-tune JoPR using Direct Preference Optimization (DPO) (Rafailov et al., 2023) on a curated set of preference pairs. Each data pair consists of a model-generated wrong output y^- and a human-preferred correct output y^+ , where y^- is the reasoning trace generated *without* the ground-truth emotion label, whereas y^+ is generated *with* it. Specially, we extract instances of model-prone thinking errors from the emotion CoT model, which mainly originate from hallucinated, inaccurate, and biased reasoning. Then, we use a large language model to annotate the correct reasoning processes, which are subsequently reviewed and corrected by five human experts as human-validated preferred outputs. Consistent with the DPO objective (Rafailov et al., 2023), the model is trained to increase the likelihood of y^+ directly, eliminating the need for

an explicit reward model. The corresponding loss function is formulated as:

$$\mathcal{L}(\theta) = -\mathbb{E}_{(x, y^+, y^-)} \left[\log \sigma \left(\beta \left(\log \frac{\pi_{\theta}(y^+|x)}{\pi_{\theta}(y^-|x)} - \log \frac{\pi_{\text{ref}}(y^+|x)}{\pi_{\text{ref}}(y^-|x)} \right) \right) \right] \quad (1)$$

Here, the term inside the logarithm represents the log-ratio of likelihoods between the current policy π_{θ} and a fixed reference policy π_{ref} . The loss drives the model to assign higher likelihood to the human-preferred response y^+ than to the less-preferred response y^- , thereby aligning the model’s output with human preferences.

3.4 GRPO with Emotion-Level Reward

This section consists of two parts: Perception Curriculum Learning and GRPO Training. By first employing curriculum learning to progressively acquire knowledge from easy to hard, followed by GRPO to reinforce exploration with hierarchical rewards, the model can consolidate its foundation while more efficiently learning fine-grained emotion reasoning. The approach gradually shifts from easy to difficult samples to ensure stable learning and prevent divergence caused by drastic distribution changes during training.

(a) Perception Curriculum Learning In emotion recognition in conversation, we define samples with easily confusable emotions as difficult samples. To better differentiate these similar samples, we adopted a **Perception Curriculum Learning (PCL)** process, progressing from easy to hard. This process primarily includes two parts: multi-dimension difficulty measurer and a dynamic curriculum strategy.

Multi-dimension Difficulty Measurer. Inspired by the Mixture-of-Experts (MoE) paradigm (Masoudnia and Ebrahimpour, 2014), we develop multi-expert models through LLM fine-tuning to serve as a sample difficulty measurer (prompts refer to the Appendix A.2). As shown in Table 1, we evaluate sample difficulty scores based on five distinct dimensions: context (situational details influencing emotion), speaker (individual characteristics and background), commonsense (incorporating knowledge to understand emotions), linguistic features (discourse structural information), and interpretability. The five dimensions are derived from psychological perspectives on emotion assessment. (Lazarus, 1991). Specifically, the difficulty score d for a given sample is formalized as follows:

$$d = 1 - \frac{1}{|E|} \sum_{j \in E} c_j \quad (2)$$

Here E is the set of experts, and c_j denotes the confidence score of the j -th expert, defined as the softmax probability assigned to the true class. A higher d indicates experts are less confident to classifying the sample into the true emotion.

Dynamic Curriculum Strategy. After scoring the difficulty of samples across the entire training set, we employ the Efraimidis–Spirakis Weighted Random Sampling Algorithm (Efraimidis and Spirakis, 2008) to dynamically select samples in training epoch. This enables a stable and adaptive sampling mechanism to construct a series of subsets that gradually transition from easy to hard.

First, we normalize the difficulty scores using the following transformation to mitigate the impact of extreme values and ensure stable sample selection.

$$\tilde{d}_i = \frac{d_i - \mu_d}{\sigma_d} \quad (3)$$

where d_i is the raw difficulty score of sample i , and μ_d, σ_d denote the mean and standard deviation of the difficulty scores. We set the sample’s weight w using the following formula:

$$w_i^{(t)} = \text{sigmoid} \left(-\tilde{d}_i \cdot \left(1 - \frac{t}{T}\right)^{\gamma} \right) \quad (4)$$

where \tilde{d}_i denotes the difficulty score of sample i . T is the total number of training steps, and γ controls the rate of transition. $\left(1 - \frac{t}{T}\right)^{\gamma}$ gradually decays from 1 to 0 as training progresses, modulating the influence of difficulty. Consequently, easier samples are more likely to be selected early in training, while harder samples are increasingly incorporated in later stages, enabling a smooth and adaptive curriculum learning effect. Specifically, for each sample i , we compute a random key:

$$k_i^{(t)} = r_i^{1/w_i^{(t)}} \quad (5)$$

where $r_i \in (0, 1]$ is uniformly sampled, and $w_i^{(t)}$ is computed as in Equation (4). We then select the m samples with the highest key values to form the training data. This ensures that the probability of selecting a sample is proportional to its current weight, enabling a smooth and adaptive curriculum learning process throughout training. The curriculum learning is illustrated in line 1–8 of Algorithm 1.

(b) GRPO Training

Group Relative Policy Optimization. Given an input dialogue, GRPO samples G reasoning

Dimension	The conceptual meaning in emotion recognition	The corresponding domain in psychology
Context	Emotions are context-dependent. For example, the expression ‘fine’ may convey anger, helplessness, or calmness in different situations.	Pragmatics
Speaker	Speaker identity and intent	Social Cognition
Commonsense	Reasoning about the speaker’s emotional responses using commonsense knowledge	Psychological Emotion Reasoning
Linguistic Feature Interpretable	Discourse structural information conveys emotional cues Explainability of Emotion Perception	Psycholinguistics Metacognition

Table 1: Sample Difficulty Evaluation System

Algorithm 1 Perception Curriculum Learning for ERC

Inputs: D_{train} : training data with L samples, R : total epochs, M : context encoder, \mathcal{E} : emotion label set, EXPERTS = $\{E_1, E_2, \dots, E_5\}$

Output: M_{policy} : optimized model

```

1: for each sample  $x_i \in D_{\text{train}}$  do
2:    $d_i \leftarrow \sum_{j=1}^5 [E_j(x_i) \neq y_i]$ 
3: end for
4:  $\tilde{d} \leftarrow \text{Normalize}(d)$ 
5: for epoch  $k = 1$  to  $R$  do
6:    $T \leftarrow k$ 
7:    $w^{(t)} = \text{Sigmoid}(-\tilde{d} \cdot (1 - \frac{t}{T})^\gamma)$ 
8:   Sample subset  $D_k \leftarrow \text{Sampling}_{E-S}(D, w)$ 
9:   for each batch  $B \in D_k$  do
10:     $p_b \leftarrow \text{FormatPrompt}(B)$ 
11:     $o_{\text{policy}} \leftarrow \text{Generate}(p_b, M_{\text{policy}})$ 
12:     $r_b \leftarrow \text{LooseRewardFunction}(o_{\text{policy}})$ 
13:     $l_{\text{grpo}} \leftarrow \text{GRPOLossFunction}(o_{\text{policy}}, r_b, M_{\text{ref}})$ 
14:    Backward( $l_{\text{grpo}}$ )
15:    Update model  $M_{\text{policy}}$ 
16:   end for
17: end for
18: return  $M_{\text{policy}}$ 

```

processes $\{o_i\}_{i=1}^G$ from $\pi_{\theta_{\text{old}}}$ and computes their rewards $\{r_i\}$. The relative quality is normalized as:

$$A_i = \frac{r_i - \text{mean}(\{r_j\})}{\text{std}(\{r_j\})}. \quad (6)$$

To constrain deviation from π_{ref} , GRPO maximizes:

$$\max_{\pi_{\theta}} \mathbb{E}_{o \sim \pi_{\text{old}}} \left[\sum_{i=1}^G \frac{\pi_{\theta}(o_i)}{\pi_{\text{old}}(o_i)} A_i \right] - \beta D_{KL}(\pi_{\theta} || \pi_{\text{ref}}) \quad (7)$$

where β controls the regularization strength.

Emotion-level rewards.

The Reward Function R includes format, output length, and emotion grading reward, defined as:

$$R = R_{\text{format}} + R_{\text{len}} + R_{\text{grading}} \quad (8)$$

(1) **Format Reward.** We use a format reward (R_{format}) to enforce structured outputs verified by regex.

$$R_{\text{format}} = \begin{cases} 1, & \text{if format is correct} \\ 0, & \text{if format is incorrect} \end{cases} \quad (9)$$

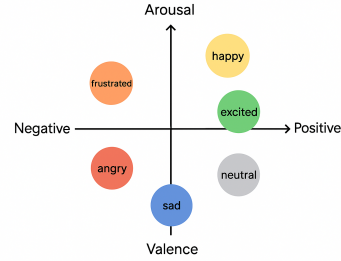


Figure 3: Valence-Arousal Diagram of Emotional Distribution.

(2) **Output Length Reward.** To limit overly long or irrelevant responses, we apply a length-based penalty, where $r_i = \text{len}(o_i)$ represents the word count of the i -th response.

$$R_{\text{len}}(o_i) = \begin{cases} 1 & \text{if } r_i < 200 \\ \frac{200}{r_i} & \text{if } r_i \geq 200 \end{cases} \quad (10)$$

(3) **Emotion Grading Reward.** As shown in Figure 3, the hierarchical reward mechanism is inspired by the Valence-Arousal model in psychology (Russell, 1980), a classical two-dimensional framework, also known as the Circumplex Model of Affect for representing emotions. In this model, each emotion is positioned within a Valence–Arousal space, where the spatial distance between emotions reflects their degree of similarity. Building on this idea, the proposed grading reward function provides fine-grained, continuous, and emotionally sensitive feedback signals by assigning higher rewards to predictions that are closer to the ground-truth emotions in this affective space. Such a design enables the model to better capture subtle emotional distinctions and handle fuzzy boundaries in multi-level affective expression tasks.

$$R_{\text{grading}} = \begin{cases} 1, & \text{if } A_{\text{pred}} \in \mathcal{C}_0(A_{\text{gt}}) \\ 0.5, & \text{if } A_{\text{pred}} \in \mathcal{C}_1(A_{\text{gt}}) \\ 0.25, & \text{if } A_{\text{pred}} \in \mathcal{C}_2(A_{\text{gt}}) \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

where A_{pred} denotes the predicted emotion label, A_{gt} denotes the ground-truth emotion label, and $\mathcal{C}_i(A_{\text{gt}})$ denotes the set of emotions that are at the i -th distance level from A_{gt} (based on the provided emotion distance hierarchy). For example, when the ground-truth label is $A_{\text{gt}} = \text{happy}$, the corresponding emotion distance tiers are defined as: $\mathcal{C}_0(\text{happy}) = \{\text{happy}\}$, $\mathcal{C}_1(\text{happy}) = \{\text{excited}\}$, $\mathcal{C}_2(\text{happy}) = \{\text{neutral}\}$, and others are $\{\text{angry, frustrated, sad}\}$. The detailed settings are provided in Table 6–8.

4 Experimental Setup

4.1 Datasets and Evaluation Metrics

Datasets. We conduct experiments on the following three benchmarks: IEMOCAP (Busso et al., 2008), MELD (Poria et al., 2018), EmoryNLP (Zahari and Choi, 2018). **IEMOCAP** is a prominent resource for emotion recognition research. Each utterance is annotated with one of six emotion labels: excited, frustrated, sad, neutral, angry, and happy. **MELD** is a multi-party conversation corpus from the TV show Friends. Each utterance has been annotated with one of seven emotion classes: neutral, surprise, fear, sad, joyful, disgust, and anger. **EmoryNLP** dataset is also from the TV series Friends. EmoryNLP utilizes a set of seven emotion categories: neutral, joyful, peaceful, powerful, scared, mad, and sad.

Evaluation Metrics. Following previous works (Lei et al., 2023; Yang et al., 2022; Hu et al., 2023), we choose the weighted average F1 score (w-F1) as the evaluation metric.

4.2 Compared Methods

We compare our method with the following representative non-LLM-based methods and LLM-based methods, as Table 2 shows.

4.3 Implementation Details

In our experiments, we use Llama3-8B (Dubey et al., 2024) as our foundational model. For all experiments in this paper, we identify the best-performing model checkpoint based on the development set results, then use this selected checkpoint

for final evaluation and reporting on the test set. In hyperparameter setting, we set learning rate to $2e-4$. We set the group size for GRPO to 8. We implement all experiments with PyTorch (Paszke et al., 2019) on 80G Nvidia A800 GPUs. More details can be found in Appendix A.3.

5 Experimental Results

Methods	IEMOCAP	MELD	EmoryNLP
<i>non-LLM-based Methods</i>			
DialogueGCN (Ghosal et al., 2019)	64.18	58.10	-
COSMIC (Ghosal et al., 2020)	63.43	65.03	38.49
SKAIG (Li et al., 2021)	66.96	65.18	38.88
DialogXL (Shen et al., 2021a)	66.20	62.41	34.73
SPCL [†] (Song et al., 2022)	69.56	67.24	40.13
HCL [†] (Yang et al., 2022)	65.31	66.18	42.01
CauAIN (Zhao et al., 2022)	65.01	64.89	37.87
DualGATs (Zhang et al., 2023a)	67.68	66.90	40.29
MKFM (Tu et al., 2023)	68.08	65.50	39.76
MFAM (Hou et al., 2023)	70.16	66.65	41.06
CEPT (Gao et al., 2024)	70.53	67.51	-
EACL (Yu et al., 2024)	70.41	67.12	40.24
PEMV (Lin et al., 2025)	70.65	67.95	40.97
<i>LLM-based Methods</i>			
BiosERC (Xue et al., 2024)	69.02	68.72	41.44
InstructERC (Lei et al., 2023)	71.39	69.15	41.37
LaERC-S (Fu et al., 2025)	72.40	69.27	42.08
Baseline	69.95	68.02	39.23
Our method	73.56	70.72	42.35

Table 2: Performance comparison between our method and existing ERC methods on three datasets. The p-values are all below 0.05 by using pairwise t-test towards our method and the baseline. The best results are **bolded**. Our evaluation metric is weighted-F1. **Baseline** refers to supervised fine-tuning a large model to output chain of thought and emotion label. The methods marked with [†] are our reproduced findings.

5.1 Main Results

In Table 2, we can note that our method achieves the best results and outperforms all the compared methods across three widely used datasets. This illustrates the superiority of our proposed method for the task of ERC. Specifically, Compared with the LaERC-S, our method achieves 1.16%, 1.45%, 0.27% gains of weighted-F1 on IEMOCAP, MELD, and EmoryNLP, respectively. Additionally, our method outperforms our constructed baseline (**Baseline** refers to supervised fine-tuning a large model to output cot and the emotion label.) by 3.61%, 2.70%, and 3.12% on the three datasets. This strong performance across three benchmark datasets highlights the efficacy of our method and its potential to advance emotion recognition in conversation.

Models	IEMOCAP	MELD	EmoryNLP
Baseline _{Llama3-8b}	69.95	68.02	39.23
Ours _{Llama3-8b}	73.56	70.72	42.35
Baseline _{Llama2-7b}	70.23	67.51	38.23
Ours _{Llama2-7b}	72.51	69.03	41.98
Baseline _{Qwen2.5-7b}	66.71	67.23	37.12
Ours _{Qwen2.5-7b}	70.36	68.51	40.52

Table 3: Experimental results on Llama3-8B, Llama2-7B, Qwen2.5-7B.

Setting	IEMOCAP	MELD	EmoryNLP
Our method	73.56	70.72	42.35
<i>(a) Long CoT Instruction Tuning (sft)</i>			
w/o sft	69.11	66.62	37.13
<i>(b) Human-preference Alignment via DPO (DPO)</i>			
w/o DPO	71.33	69.69	40.10
<i>(c) Perception Curriculum Learning (PCL)</i>			
w/o Context Analyzer	72.89	70.01	41.89
w/o Speaker Analyzer	72.03	69.89	42.21
w/o Commonsense Analyzer	71.82	69.56	41.35
w/o Linguistic Analyzer	71.58	69.33	41.01
w/o Interpretable Analyzer	71.31	69.15	40.23
w/o PCL	71.11	69.50	41.11
<i>(d) GRPO with emotion-level reward (GRPO)</i>			
w/o Format Reward	72.01	69.82	41.56
w/o Output Length Reward	72.75	70.01	41.82
w/o Emotion Grading Reward	71.89	69.65	41.41
w/o GRPO	70.24	68.52	39.89

Table 4: Ablation results on reinforcement learning and perception curriculum learning.

5.2 Results on Different LLMs

To demonstrate the generalizability of our method across different models, we further conduct experiments on open-source models of comparable scale, including the representative LLMs: LLaMA3-8B, LLaMA2-7B, and Qwen2.5-7B. As shown in Table 3, our method consistently outperforms the baseline across three datasets, highlighting its model-agnostic nature and effectiveness.

5.3 Ablation Study

We conduct ablation experiments: removing the following module:

Effect of Long CoT Instruction Tuning. Removing SFT leads to a substantial performance drop across all datasets, indicating that SFT is essential for the model to internalize reasoning patterns encoded in chain-of-thought instructions.

Effect of Human-preference Alignment. Without DPO, performance also degrades notably, demonstrating that aligning model outputs with human-preferred reasoning improves both reason-

ing quality and downstream emotion classification.

Effect of Perception Curriculum Learning. We conduct two ablations on the multi-dimension measurer and dynamic curriculum strategy. In rows 1 to 4 of Table 4 (c), we sequentially remove the difficulty measurer for each dimension. The performance decreases, demonstrating that a multi-dimension difficulty measurer improves curriculum learning by capturing diverse data aspects, reducing bias, and boosting the model’s generalization and ability to distinguish similar emotions. To validate the effectiveness of the dynamic curriculum strategy, we remove this module and observe impacts on training. This suggests that the sampling method for the samples affects subsequent reinforcement learning training.

GRPO with emotion-level reward. In GRPO, we conduct ablation experiments on three different reward functions as shown in Table 4 (d). Removing any reward function degrades performance, confirming that each one guides the model toward more effective learning and decision-making. Among these, the Emotion Grading Reward shows the most significant performance drop, suggesting that it is more effective than the other two in guiding the model towards making correct decisions when handling similar emotions. Completely removing the GRPO module results in performance decreases of 3.32%, 2.20%, and 2.46% on the three datasets, leading to substantial performance drops. This indicates that the GRPO strategy, with its "generation + selection + optimization" fine-tuning approach, can significantly enhance the model’s understanding of emotions.

5.4 Analysis of Perception Curriculum Learning

	IEMOCAP	MELD	EmoryNLP
InstructERC	71.39	69.15	41.37
+ PCL	72.03	69.89	41.39
LaERC-S	72.40	69.27	42.08
+ PCL	72.72	69.53	42.18
Our Method	71.11	69.50	41.11
+ SPCL	71.53	69.99	41.21
+ HCL	72.21	69.75	41.67
+ PCL	73.56	70.72	42.35

Table 5: Performance comparison with other curriculum learning methods across datasets.

(a) Generalizability To evaluate the generalizability of our curriculum learning, we integrate

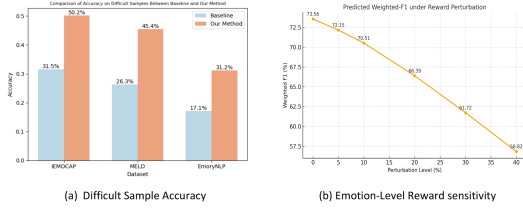


Figure 4: (a) is improvement accuracy on difficult samples. (b) An experiment of Emotion-Level Reward sensitivity on IEMOCAP.

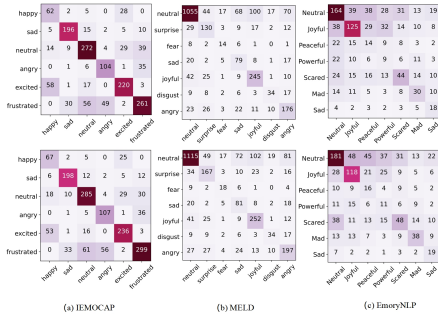


Figure 5: Matrix of datasets. The first line is baseline. The second line is JoPR.

PCL into two representative LLM-based ERC models, InstructERC and LaERC-S. As shown in Table 5, consistent performance improvements are observed across three datasets. Notably, the improvement on the IEMOCAP dataset is more pronounced. We attribute this to the presence of numerous emotion pairs with high similarity in both the IEMOCAP and MELD datasets, such as happy vs. excited and angry vs. frustrated. By leveraging the stable progression provided by curriculum learning, PCL effectively mitigates confusion between similar emotions, leading to enhanced performance.

(b) Stability We reproduce two representative curriculum learning strategies, SPCL (Song et al., 2022) and HCL (Yang et al., 2022), and integrate them into our model to compare with our proposed PCL approach. SPCL measures sample difficulty via distance to class centroids and introduces data progressively from easy to hard, but its single-perspective, rule-based evaluation limits its representational capacity. HCL extends this idea to dialogue and utterance levels but often suffers from instability as difficult samples surge late in training. As shown in Table 5, our PCL consistently outperforms both methods by combining their strengths to enhance difficulty evaluation through multi-dimensional analysis and improve stability with a dynamic sampling strategy. More details can be found in Appendix A.4.

5.5 Analysis of Emotion-Level Reward

(a) Distinction Similar Emotion. We further evaluate performance on the top 10% of hard samples, identified by the multi-dimension difficulty measurer, which mainly come from easily confusable similar emotions, with results shown in Figure 4(a). JoPR achieves a significant improvement over the baseline, with gains of 18.7%, 19.1%, and 14.1% on the three datasets, respectively. This substantial improvement demonstrates the superior performance of our approach on challenging instances, highlighting its effectiveness in capturing subtle emotional distinctions and enhancing robustness in fine-grained emotion classification.

(b) Emotion-Level Reward Sensitivity Analysis. This experiment aims to assess the sensitivity of the grading reward (Equation 11) to noise. As Figure 4(b) shows, we randomly perturb a certain proportion (0–40%) of emotion reward mappings within the predefined reward matrix, such as swapping the reward scores for “happy” and “excited” and measure the resulting changes in weighted-F1 scores. Model performance gradually declines as the perturbation level increases, from 73.56% at 0% noise to 56.8% at 40% noise, indicating sensitivity to reward distortions and demonstrating the rationality of this reward design.

(c) Visualization Analysis We analyze the model’s prediction performance using the confusion matrices in Figure 5. The matrices show the distribution of predicted versus ground-truth emotion categories. Notably, after incorporating the hierarchical reward via reinforcement learning, the model demonstrates a clearer separation between similar emotions, such as excited and happy. This improvement is reflected in the reduced off-diagonal confusion between these categories, indicating that the reward mechanism encourages more deliberate and fine-grained reasoning

6 Conclusion

In conclusion, we propose JoPR, a framework that combines perception-curriculum learning and emotional reasoning for conversational emotion recognition. By integrating chain-of-thought instruction tuning, human-preference alignment, multi-view curriculum learning, and tiered rewards, JoPR effectively distinguishes nuanced emotions. Experiments on three datasets demonstrate its superior performance and strong human emotion reasoning capabilities.

Limitations

Despite its effectiveness, JoPR has several limitations. On the one hand, its reliance on long chain-of-thought reasoning and multi-view curriculum increases computational cost, which may hinder real-time applications. On the other hand, JoPR's performance may still degrade in highly ambiguous or contextually sparse conversations, indicating room for further improvement in modeling subtle emotional cues.

References

- Helene Ackermann, Anna L Lange, Verena V Hafner, and Rebecca Lazarides. 2025. How adaptive social robots influence cognitive, emotional, and self-regulated learning. *Scientific Reports*, 15(1):6581.
- 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(4):335–359.
- Huayu Chen, Kaiwen Zheng, Qinsheng Zhang, Ganqu Cui, Yin Cui, Haotian Ye, Tsung-Yi Lin, Ming-Yu Liu, Jun Zhu, and Haoxiang Wang. 2025. Bridging supervised learning and reinforcement learning in math reasoning. *arXiv preprint arXiv:2505.18116*.
- Wim De Neys. 2023. Advancing theorizing about fast-and-slow thinking. *Behavioral and Brain Sciences*, 46:e111.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, and 1 others. 2024. The llama 3 herd of models. *arXiv e-prints*, pages arXiv–2407.
- Pavlos Efraimidis and Paul Spirakis. 2008. Weighted random sampling. In *Encyclopedia of algorithms*, pages 1024–1027. Springer.
- Junwei Feng and Xueyan Fan. 2025. Cross-modal context fusion and adaptive graph convolutional network for multimodal conversational emotion recognition. *arXiv preprint arXiv:2501.15063*.
- Yumeng Fu, Junjie Wu, Zhongjie Wang, Meishan Zhang, Lili Shan, Yulin Wu, and Bingquan Liu. 2025. Laerc-s: Improving llm-based emotion recognition in conversation with speaker characteristics. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 6748–6761.
- Qingqing Gao, Jiuxin Cao, Biwei Cao, Xin Guan, and Bo Liu. 2024. Cept: A contrast-enhanced prompt-tuning framework for emotion recognition in conversation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 2947–2957.
- 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. Dialoguecn: 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.
- Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shitong Ma, Peiyi Wang, Xiao Bi, and 1 others. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.
- Guiyang Hou, Yongliang Shen, Wenqi Zhang, Wei Xue, and Weiming Lu. 2023. Enhancing emotion recognition in conversation via multi-view feature alignment and memorization. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 12651–12663.
- Dou Hu, Yinan Bao, Lingwei Wei, Wei Zhou, and Songlin Hu. 2023. 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.
- Jingwen Hu, Yuchen Liu, Jinming Zhao, and Qin Jin. 2021. Mmgcn: Multimodal fusion via deep graph convolution network for emotion recognition in conversation. 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 5666–5675.
- Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, and 1 others. 2024. Openai o1 system card. *arXiv preprint arXiv:2412.16720*.
- Richard S Lazarus. 1991. *Emotion and adaptation*. Oxford University Press.
- Shanglin Lei, Guanting Dong, Xiaoping Wang, Keheng Wang, and Sirui Wang. 2023. Instructerc: Reforming emotion recognition in conversation with a retrieval multi-task llms framework. *CoRR*.
- Jiangnan Li, Zheng Lin, Peng Fu, and Weiping Wang. 2021. Past, present, and future: Conversational emotion recognition through structural modeling of psychological knowledge. In *Findings of the association*

- for computational linguistics: *EMNLP 2021*, pages 1204–1214.
- Xinhao Li, Ziang Yan, Desen Meng, Lu Dong, Xiangyu Zeng, Yanan He, Yali Wang, Yu Qiao, Yi Wang, and Limin Wang. 2025. Videochat-r1: Enhancing spatio-temporal perception via reinforcement fine-tuning. *arXiv preprint arXiv:2504.06958*.
- Chen Lin, Fei Li, Donghong Ji, and Chong Teng. 2025. Pemv: Improving spatial distribution for emotion recognition in conversations using proximal emotion mean vectors. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 345–357.
- Peixian Ma, Xialie Zhuang, Chengjin Xu, Xuhui Jiang, Ran Chen, and Jian Guo. 2025. Sql-r1: Training natural language to sql reasoning model by reinforcement learning. *arXiv preprint arXiv:2504.08600*.
- Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. 2019. Dialoguernn: An attentive rnn for emotion detection in conversations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6818–6825.
- Saeed Masoudnia and Reza Ebrahimpour. 2014. Mixture of experts: a literature survey. *Artificial Intelligence Review*, 42(2):275–293.
- Sanmit Narvekar, Bei Peng, Matteo Leonetti, Jivko Sinapov, Matthew E Taylor, and Peter Stone. 2020. Curriculum learning for reinforcement learning domains: A framework and survey. *Journal of Machine Learning Research*, 21:1–50.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, and 1 others. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.
- Emmanouil Antonios Platanios, Otilia Stretcu, Graham Neubig, Barnabás Póczos, and Tom Mitchell. 2019. Competence-based curriculum learning for neural machine translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1162–1172.
- Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2018. Meld: A multimodal multi-party dataset for emotion recognition in conversations. *arXiv preprint arXiv:1810.02508*.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741.
- James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.
- Weizhou Shen, Junqing Chen, Xiaojun Quan, and Zhixian Xie. 2021a. Dialogxl: All-in-one xlnet for multi-party conversation emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13789–13797.
- Weizhou Shen, Siyue Wu, Yunyi Yang, and Xiaojun Quan. 2021b. Directed acyclic graph network for conversational emotion recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1551–1560.
- Huatong Song, Jinhao Jiang, Yingqian Min, Jie Chen, Zhipeng Chen, Wayne Xin Zhao, Lei Fang, and Jirong Wen. 2025. R1-searcher: Incentivizing the search capability in llms via reinforcement learning. *arXiv preprint arXiv:2503.05592*.
- Xiaohui Song, Longtao Huang, Hui Xue, and Songlin Hu. 2022. Supervised prototypical contrastive learning for emotion recognition in conversation. *arXiv preprint arXiv:2210.08713*.
- Xiaotong Song, Huiping Lin, Jiatao Zhu, and Xinyi Gong. 2024. Cagk: Collaborative aspect graph enhanced knowledge-based recommendation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 2612–2621.
- Dhruvitkumar Talati. 2023. Artificial intelligence (ai) in mental health diagnosis and treatment. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, 2(3):251–253.
- Lingxiao Tang, He Ye, Zhongxin Liu, Xiaoxue Ren, and Lingfeng Bao. 2025. Codereasoner: Enhancing the code reasoning ability with reinforcement learning. *arXiv preprint arXiv:2507.17548*.
- Qwen Team. 2024. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*.
- Geng Tu, Bin Liang, Bing Qin, Kam-Fai Wong, and Ruifeng Xu. 2023. An empirical study on multiple knowledge from chatgpt for emotion recognition in conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 12160–12173.
- Xin Wang, Yudong Chen, and Wenwu Zhu. 2021. A survey on curriculum learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):4555–4576.
- Kun Xiang, Zhili Liu, Zihao Jiang, Yunshuang Nie, Runhui Huang, Haoxiang Fan, Hanhui Li, Weiran Huang, Yihan Zeng, Jianhua Han, and 1 others. 2024. Atomthink: A slow thinking framework for

multimodal mathematical reasoning. *arXiv preprint arXiv:2411.11930*.

Jieying Xue, Minh-Phuong Nguyen, Blake Matheny, and Le-Minh Nguyen. 2024. Bioserc: Integrating biography speakers supported by llms for erc tasks. In *International Conference on Artificial Neural Networks*, pages 277–292. Springer.

Lin Yang, Yi Shen, Yue Mao, and Longjun Cai. 2022. Hybrid curriculum learning for emotion recognition in conversation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 11595–11603.

Fangxu Yu, Junjie Guo, Zhen Wu, and Xinyu Dai. 2024. Emotion-anchored contrastive learning framework for emotion recognition in conversation. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 4521–4534.

Sayyed M Zahiri and Jinho D Choi. 2018. Emotion detection on tv show transcripts with sequence-based convolutional neural networks. In *AAAI Workshops*, volume 18, pages 44–52.

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

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

Weixiang Zhao, Yanyan Zhao, and Xin Lu. 2022. Cauain: Causal aware interaction network for emotion recognition in conversations. In *International Joint Conference on Artificial Intelligence*.

Yaoyu Zhu, Di Huang, Hanqi Lyu, Xiaoyun Zhang, Chongxiao Li, Wenxuan Shi, Yutong Wu, Jianan Mu, Jinghua Wang, Yang Zhao, and 1 others. 2025. Codev-r1: Reasoning-enhanced verilog generation. *arXiv preprint arXiv:2505.24183*.

A Appendix

A.1 Prompt for Emotion Long Cot Generation

As shown in Figure 6, we use the prompt to distill the DeepSeek-R1-Distill-Qwen32B (Team, 2024) model and obtain high-quality reasoning processes. The prompt content includes role-playing, context, speaker and objective.

Prompt for Emotion Long Cot Generation.

Now you are expert of sentiment and emotional analysis. The following conversation noted between '### ###' involves several speakers. ### {CONVERSATION} ### Based on the true emotional label **{EMOTION_LABEL}** of this utterance **<{UTTERANCE}>**, output the thought process behind obtaining this emotional label step by step in no more than 200 words.

Figure 6: Prompt for Emotion Long CoT Generation.

A.2 Prompts for Difficulty Measurer

As shown in Figures 7–11, we fine-tune five separate LLaMA3-8B models with prompts designed for context, speaker, commonsense, linguistic features, and interpretable, respectively, to serve as difficulty measurers. In particular, the commonsense extraction follows the approach proposed in LaERC-S (Fu et al., 2025).

Prompt for evaluating from the context dimension.

Now you are expert of sentiment and emotional analysis. The following conversation noted between '### ###' involves several speakers. ### {CONVERSATION} ### Pay attention to **Contextual Factors** and select the emotional label of {UTTERANCE} from <happy, sad, neutral, angry, excited, frustrated >:

Figure 7: The prompt for evaluating from the context dimension.

Prompt for evaluating from the speaker dimension.

Now you are expert of sentiment and emotional analysis. The following conversation noted between '### ###' involves several speakers. ### {CONVERSATION} ### Pay attention to **Speaker Identity and Personality** and select the emotional label of {UTTERANCE} from <happy, sad, neutral, angry, excited, frustrated >:

Figure 8: The prompt for evaluating from the speaker dimension.

A.3 Implementation Settings.

We provide additional experimental settings. For the long cot instruction tuning, we set the learning rate as $2e-4$ and batch size as 16. The model is

Emotion	Reward = 1	Reward = 0.5	Reward = 0.25	Reward = 0
happy	happy	excited	neutral	angry, frustrated, sad
excited	excited	happy	neutral	angry, frustrated, sad
neutral	neutral	–	–	happy, excited, angry, frustrated, sad
angry	angry	frustrated	neutral	happy, excited, sad
frustrated	frustrated	angry	neutral	happy, excited, sad
sad	sad	–	neutral	happy, excited, angry, frustrated

Table 6: Grading reward for IEMOCAP dataset.

Emotion	Reward = 1	Reward = 0.5	Reward = 0.25	Reward = 0
surprise	surprise	neutral	angry, joyful	fear, disgust, sad
angry	angry	neutral	–	joyful, surprise, fear, disgust, sad
neutral	neutral	–	–	joyful, surprise, fear, disgust, angry, sad
joyful	joyful	neutral	–	surprise, fear, disgust, angry, sad
sad	sad	neutral	angry	joyful, surprise, fear, disgust
fear	fear	neutral	angry, sad, surprise	joyful, disgust
disgust	disgust	neutral	angry	sad, surprise, joyful, fear

Table 7: Grading reward for MELD dataset.

Prompt for evaluating from the common-sense dimension.

Now you are expert of sentiment and emotional analysis. The following conversation noted between '###' involves several speakers.

{CONVERSATION}

Pay attention to {COMMONSENSE1}{COMMONSENSE2}...{COMMONSENSE9} and select the emotional label of {UTTERANCE} from <happy, sad, neutral, angry, excited, frustrated >:

Figure 9: The prompt for evaluating from the common-sense dimension.

trained for 15 epochs. For the DPO, we set the learning rate as $4e-4$ and batch size as 8. For the GRPO, we set the learning rate as $4e-4$ and batch size as 8. In curriculum learning, we set m as 1000 and γ as 1.0.

A.4 Effect of Dimensionality Num on Difficulty Mesurer.

In Figure 12, we analyze the effect of varying the number of dimensions on performance using a line chart and observed that five dimensions (context, speaker, commonsense, linguistic features, and interpretability) yield the best results. These five dimensions were subsequently used to assess sample difficulty, indicating their complementary contribution to the evaluation.

Prompt for evaluating from the linguistic features dimension.

Now you are expert of sentiment and emotional analysis. The following conversation noted between '###' involves several speakers. ### {CONVERSATION} ### Pay attention to Linguistic Features and select the emotional label of {UTTERANCE} from <happy, sad, neutral, angry, excited, frustrated >:

Figure 10: The prompt for evaluating from the linguistic features dimension.

A.5 Template for Context-Guided Emotion Label Modification

In the SFT stage, the model first predicts an initial emotion label for the target utterance. In the reinforcement learning stage, we further require the model to determine, based on the dialogue context, whether this initial label should be modified. If so, it outputs the revised label, otherwise it retains the original prediction.

Emotion	Reward = 1	Reward = 0.5	Reward = 0.25	Reward = 0
Joyful	Joyful	–	Neutral	Powerful, Mad, Peaceful, Sad, Scared
Powerful	Powerful	Joyful, Neutral	Mad	Peaceful, Sad, Scared
Mad	Mad	Neutral	Joyful, Powerful	Peaceful, Sad, Scared
Peaceful	Peaceful	Neutral, Joyful	–	Sad, Scared, Powerful, Mad
Neutral	Neutral	–	–	Peaceful, Joyful, Sad, Scared, Powerful, Mad
Sad	Sad	Mad, Scared	Neutral	Joyful, Peaceful, Powerful
Scared	Scared	Mad, Sad	Neutral	Joyful, Peaceful, Powerful

Table 8: Grading reward for EmoryNLP dataset.

Prompt for evaluating from the interpretable dimension.

Now you are expert of sentiment and emotional analysis. The following conversation noted between ‘###’ involves several speakers.
 ### {CONVERSATION} ### Pay attention to Cause, analyze step by step and select the emotional label of {UTTERANCE} from <happy, sad, neutral, angry, excited, frustrated >:

Figure 11: The prompt for evaluating from the interpretable dimension.

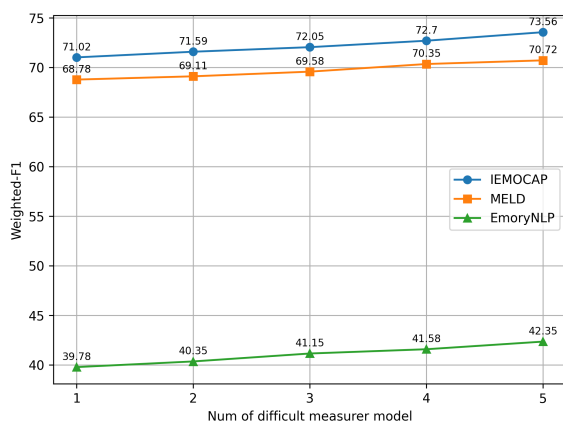


Figure 12: Experiment of dimensionality on difficulty measurement.