LaERC-S: Improving LLM-based Emotion Recognition in Conversation with Speaker Characteristics

Yumeng Fu[1], Junjie Wu[2], Zhongjie Wang[1],
Meishan Zhang[3], Lili Shan[1], Yulin Wu[3], Bingquan Liu[1]*,
[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China,
[2]School of Computer Science and Technology, Soochow University, Suzhou, China,
[3]School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China,
24b303004@stu.hit.edu.cn, 20224027010@stu.suda.edu.cn, zjwang@insun.hit.edu.cn,
mason.zms@gmail.com, {shanlili, liubq}@hit.edu.cn, yulinwu@cs.hitsz.edu.cn

Abstract

Emotion recognition in conversation (ERC), the task of discerning human emotions for each utterance within a conversation, has garnered significant attention in human-computer interaction systems. Previous ERC studies focus on speaker-specific information that predominantly stems from relationships among utterances, which lacks sufficient information around conversations. Recent research in ERC has sought to exploit pre-trained large language models (LLMs) with speaker modelling to comprehend emotional states. Although these methods have achieved encouraging results, the extracted speaker-specific information struggles to indicate emotional dynamics. In this paper, motivated by the fact that speaker characteristics play a crucial role and LLMs have rich world knowledge, we present LaERC-S, a novel framework that stimulates LLMs to explore speaker characteristics involving the mental state and behavior of interlocutors, for accurate emotion predictions. To endow LLMs with this knowledge information, we adopt the two-stage learning to make the models reason speaker characteristics and track the emotion of the speaker in complex conversation scenarios. Extensive experiments on three benchmark datasets demonstrate the superiority of LaERC-S, reaching the new state-of-the-art.[1]

1 Introduction

Emotion recognition in conversation (ERC) is a fundamental task in the community of natural language processing (NLP), which targets to automatically identify the emotion of each utterance within a conversation. With the proliferation of conversation data on social media platforms, likewise Twitter and Facebook, detecting human emotions around conversations (Tu et al., 2022; Gao et al., 2024) holds promising potential for a series of real-world applications, such as recommendation (Song et al., 2024) and opinion mining (Kumar et al., 2023). However, unlike sentence-level emotion recognition (Deng et al., 2023; Zhang et al., 2024), conversation involves a process of dynamic interactions, which poses a unique challenge for ERC.

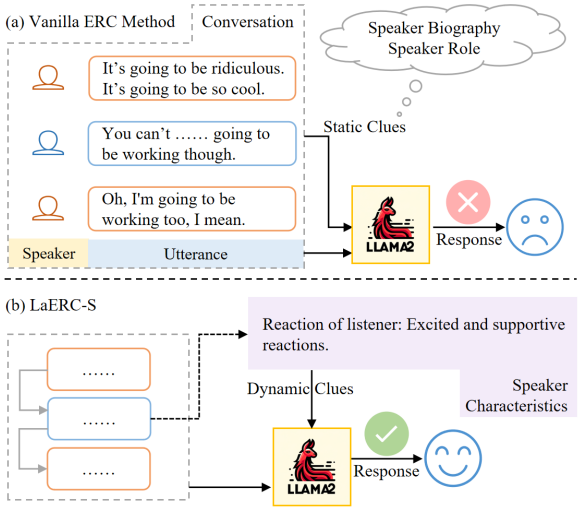

Figure 1: Comparison between existing ERC models and the proposed LaERC-S. (a) The existing ERC methods exploit static clues, such as speaker biography and speaker role, for emotional states. (b) The proposed LaERC-S captures rich and deep clues of emotional dynamics, including the mental state and behavior of interlocutors, to trigger the target emotion.

Faced with such a challenge, initial attempts to analyze the content of conversation relied on converstaional context modelling (Sun et al., 2021; Shen et al., 2021b), while current sophisticated methods (Song et al., 2022b; Lee and Lee, 2022; Zhang et al., 2023; Wang et al., 2024a) start the investigation of speaker-specific information to mitigate emotion ambiguity. However, these methods rely on highly structured paradigms, which make the models overfit to specific data distributions, thereby hampering progress in the realm of ERC.

Apart from above studies, another strand of re-

*Corresponding author
[1]https://github.com/bigcat-1/LaERC-S

Proceedings of the 31st International Conference on Computational Linguistics, pages 6748–6761
January 19–24, 2025.

search resorts to the reasoning and generation capabilities of large language models (LLMs), such as PaLM (Chowdhery et al., 2023) and LLaMA2 (Touvron et al., 2023), for different conversational datasets. A pioneering work by InstructERC (Lei et al., 2023) fine-tunes LLaMA2 by introducing speaker identification. Such paradigm gets significant performance compared to conventional pre-trained language models (PLMs) in ERC. Subsequently, BiosERC (Xue et al., 2024) integrates the biographical information of speakers to intensify LLMs-based ERC systems. As a result, the exploration of speaker characteristics can bring superior performance to their respective models.

Despite the striking results acquired by above works, they are limited by the following dilemmas: (1) Speaker identification can not provide sufficient information. (2) Speaker biography lacks clues of emotional dynamics in complex conversations. These static information makes the models tend to generate biased responses for all the utterances uttered by a certain speaker. However, as reported in (Hwang et al., 2021; Zhao et al., 2022), speaker characteristics including mental state and behavior of interlocutors can provide deep and rich clues of emotional dynamics(Ghosal et al., 2020), thereby triggering the target emotion. Thus, it would be beneficial to exploit such speaker characteristics into LLMs for ERC.

In this paper, we propose LaERC-S, a novel framework devised to exploit large language models and speaker characteristics for the ERC task. Specifically, we design an efficient instruction template to promote LLMs to generate the mental state, behavior and persona of interlocutors around conversations. Afterwards, to supplement LLMs with this knowledge information, we perform two-stage learning, including speaker characteristic injection and emotion recognition, for the final result. A schematic of LaERC-S is depicted in Figure 1.

Without bells and whistles, the proposed LaERC-S surpasses all ERC methods on three benchmark datasets, including IEMOCAP (Busso et al., 2008), MELD (Poria et al., 2019), and EmoryNLP (Zahiri and Choi, 2018). Moreover, LaERC-S provides a unique perspective to capture speaker characteristics in the realm of LLMs-based ERC, which can be reproduced by a sinlge GPU.

In summary, our contributions are three-fold:

- We propose a simple and effective framework, namely LaERC-S, which explores large language models and speaker characteristics for emotion recognition in conversation.
- We design an efficient instruction template to promote LLMs to generate speaker characteristics, and adopt a two-stage learning for capturing emotional dynamics and judging emotional states in conversations.
- Experiments are conducted on three public datasets, including IEMOCAP, MELD, and EmoryNLP, which validates the superiority of LaERC-S over the state-of-the-art methods.

2 Related Work

2.1 Emotion Recognition in Conversation

As an indispensable part of human-interaction systems, the nature of emotion recognition in conversation (ERC) refers to make the models comprehend emotion states of interlocutors within conversations, thereby generating empathy and empathic responses (Majumder et al., 2020). In the literature, existing ERC studies (Poria et al., 2017; Majumder et al., 2019; Ghosal et al., 2019; Li et al., 2021, 2023; Zhao et al., 2022; Zhang et al., 2023; Tu et al., 2023; Jian et al., 2024) can be roughly divided into two ideas. One relys on pre-trained language models (PLMs) to model conversational context and speaker for emotion prediction. Typically, DialogXL (Shen et al., 2021a) introduces an enhanced memory to store conversational contexts, and further captures intra- and inter-speaker dependencies for multi-party structures. CEPT (Gao et al., 2024) devises a mixed prompt template and a label mapping strategy for conversational contexts and comprehensive emotions, respectively. With the advancements of pre-trained large language models (LLMs), another line of research attempts to employ LLMs to the task of ERC. Recently, InstructERC (Lei et al., 2023) transforms ERC into a retrieved-based Seq2Seq form for LLMs adaptation. BiosERC (Xue et al., 2024) leverages speakers' personalities to enhance LLMs.

These methods reveal the statement that speaker characteristics are beneficial for emotion recognition in conversation. However, they lack convincing interpretations for acquiring speaker-specific information, thereby limiting emotional expressions. Therefore, in this paper, we attempt to adopt an explainable way to explore large language models and speaker characteristics for the ERC task.

2.2 Speaker Characteristics

Speaker characteristics involve the mental state, behavior and persona of interlocutors in social interaction (Bosselut et al., 2019; Sap et al., 2019; Hwang et al., 2021). It is beneficial for a human-computer interaction system to comprehend the speaker's intention and purpose, as well as analyze situationally-relevant speaker's reaction and behavior. Motivated by such superiority, a series of works employ speaker characteristics to numerous downstream tasks, such as question answering (Zhang et al., 2022), empathic response generation (Sabour et al., 2022), and emotional gold mining (Wang et al., 2024b). In recent years, scholars have paid attention to making progress in ERC by exploring speaker characteristics. These studies leverage conversational relations expressed by a triplet form, to learn the interaction between speakers. Typically, COSMIC (Ghosal et al., 2020) exploits a commonsense knowledge base to learn commonsense features for emotion prediction. SKAIG (Li et al., 2021) constructs a graph to capture speaker's psychological states. CauAIN (Zhao et al., 2022) regards commonsense knowledge as causal clues to trigger the target emotion.

Our method is different from these methods that achieve speaker characteristics from relationships among utterances. In this paper, we extract rich world knowledge from LLMs by devising an efficient template while making the models reason speaker characteristics and track emotional states. This stimulates the proposed LaERC-S to provide more accurate emotion predictions.

3 The LaERC-S Framework

In this section, we present a framework, namely LaERC-S, which introduces speaker characteristics for adapting LLMs to emotion recognition in conversation, as shown in Figure 2. First, we provide the vanilla model in the task of ERC, followed by the specifics of LaERC-S, including speaker characteristic extraction and injection, emotion recognition. Moreover, LaERC-S can also be extended to any of mainstream large language models.

3.1 Vanilla ERC Model

A conversation data source as $\mathcal{D} = \{(C_i, Y_i)\}_{i=1}^{N}$, where the symbol C_i denotes the i-th conversation, and N is the size of $\mathcal{D}$. Each conversation includes a sequence of utterances $\mathcal{U} = \{u_j\}_{j=1}^{S}$, where the sign S is the number of all utterances. Each utterance in a conversation is assigned with a ground truth label $y_j \in \{e_1, e_2, ..., e_K\}$, where K is the number of emotion categories.

Generally, the ERC model $\mathcal{M}$ based on LLMs is learned from $\mathcal{D}$ to provide a response r over a set of the predefined emotion labels $\mathcal{E} = \{e_k\}_{k=1}^{K}$. The whole process can be expressed as follows:

$$r_{j,i} = \mathcal{M}(u_{<j}, u_j, \mathcal{E}), \tag{1}$$

where, $u_{<j}$ denotes the historical utterances before the target utterance u_j in the i-th conversation.

3.2 Speaker Characteristic Extraction

To extract high quality speaker characteristics in conversation, we adopt prompt engineering for extraction due to the beneficial of this technology (Liu et al., 2023; White et al., 2023; Giray, 2023). Besides, considering the fact that pre-trained LLMs serve as a rich world knowledge base, we design a template to query LLMs to capture speaker characteristics. Besides, we manually verified speaker characteristics extracted from the large model. We provide the generation procedure of available information regarding speaker characteristics in conversational scenarios, as depicted in Figure 2 (a).

Typically, we investigate previous studies (Sap et al., 2019; Hwang et al., 2021), and observe that speaker characteristics cover mental state, behavior and persona. Appendix A.4.1 presents the definitions of the information. Mental state reflects emotional states of interlocutors, containing three relations, i.e., 'oReact', 'xReact' and 'xIntent'. Behavior means a response to an event, including 'xNeed', 'xWant', 'oWant', 'xEffect' and 'oEffect'. Persona indicates the interlocutor's attribute by 'xAttr'.

These key elements from different perspectives reveal the interaction between utterances, which is intuitively projected into the query template for retrieving available information regarding speaker characteristics. The templates relevant to all the key elements are presented in Appendix A.4.2.

3.3 Speaker Characteristic Injection

Speaker characteristic injection is to learn clues of emotional dynamics in conversation scenarios, which endows the model with speaker characteristics for subsequent emotion analysis. Although pre-trained large language models cover speaker-specific information, they have not yet been activated the perception capability about this under

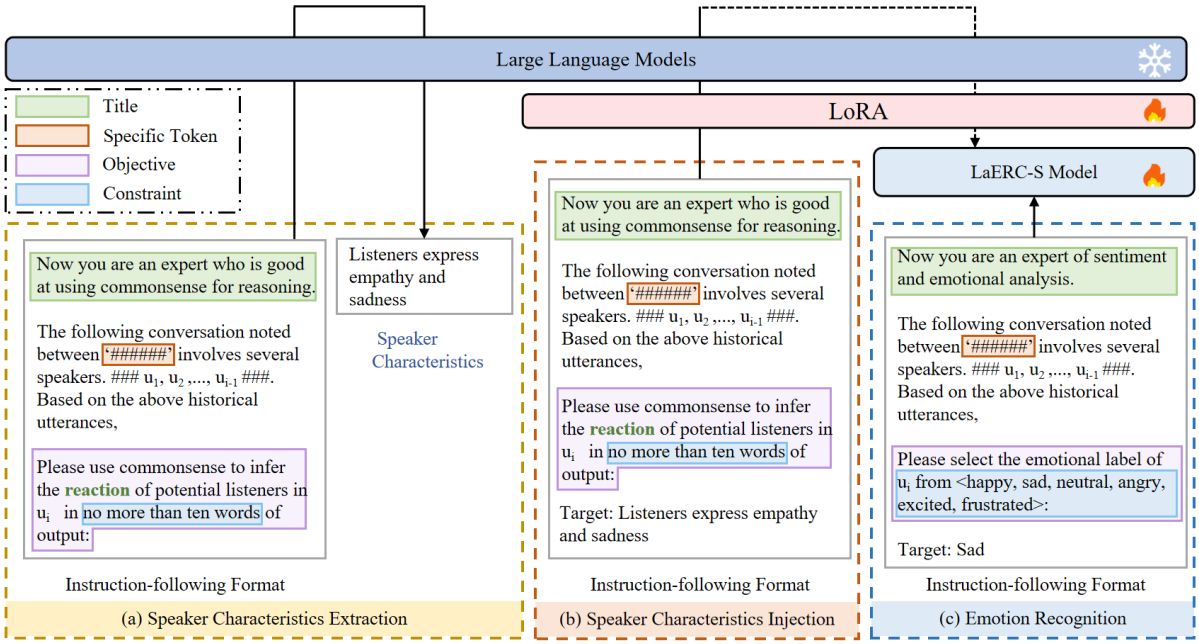

Figure 2: The overview of LaERC-S. LaERC-S includes speaker characteristics extraction and injection, emotion recognition. In the speaker characteristics extraction, speaker characteristics are extracted from LLMs. In the speaker characteristics injection, the generated speaker-characteristics are employed to make the models perceive emotional dynamics. In the emotion analysis, the conversational contents and predefined emotional labels are converted into a formatted input for the final response. As depicted in the instance, LaERC-S bridges the gap between speaker characteristics and the response of "sad".

conversational contexts. To this end, we adopt a instruction-tuning strategy tailored to endow LLMs with speaker characteristics at the initial stage, as shown in Figure 2 (b).

Typically, we design an instruction template with a certain key element and basic elements for knowledge analysis. A key element is one of any relationships provided by above preliminary. The basic elements comprises four aspects, i.e., 'title', 'specific token' and 'objective', 'constraint'. The 'title' indicates that the role of LLMs expert apt in learning emotional clues in conversations. The 'specific token' is to separate conversation contents. The 'objective' refers to a concise elucidation of the task of knowledge analysis, which provides a response based on conversation contexts. The 'constraint' is used to limit the length of the response for avoiding hallucinations. For reference, we construct the input template to align with the instruction-following template of information retrieval at preliminary.

3.4 Emotion Recognition

After the above stage, we achieve an initial model that is available to perceive clues of emotional dynamics in conversations. However, there is a gap between these clues and emotion states. To reach this, we further conduct an instruction-tuning strategy to learn the interplay between emotional tendencies and clues, as depicted in Figure 2 (c).

To aligned with the initial stage, we make adjustments in the initial instruction-following template, i.e., title, objective and constraint. Typically, the "title" presents the role of LLMs as assistant skilled in sentiment and emotion analysis. The "objective" proposes to give a emotional label for the target utterance in a conversation. The "constraint" refers to a set of the predefined emotional labels. Such format can maximize the mutual synergy between multiple tasks, while the generated knowledge information does not need to be added into this template without additional computing resources.

Overall, the objective function for various tasks can be defined as follows:

$$L_k = \sum_{i'}^{j} -\log P(\mu_{(k,i')}|x_k, \theta_k), \qquad (2)$$

where k indicates a certain stage, and x_k is the instruction-following template to the certain stage. $\mu_{(k,i')}$ denotes the generated token. In addition, θ_k denotes the trainable parameters in LLMs.

Finally, after the second stage, the well-trained model is leveraged for inference purposes. We

choose ‘oReact‘ item as the final LaERC-S model for emotion analysis in conversation.

4 Experiments

In this section, we successively present three commonly used conversation datasets, compared baselines and basic experimental settings, and then analyze the experimental results in detail.

4.1 Datasets

We evaluate LaERC-S on three representative datasets which involve IEMOCAP (Busso et al., 2008), MELD (Poria et al., 2019), and EmoryNLP (Zahiri and Choi, 2018). More details about these datasets can be found in Appendix A.2.

4.2 Baselines

To demonstrate the superiority of LaERC-S in the task of emotion recognition in conversation, we compare LaERC-S with two kinds of mainstream ERC methods as follows.
(i) Conventional ERC methods: COSMIC (Ghosal et al., 2020), SKAIG (Li et al., 2021), DialogXL (Shen et al., 2021a), SPCL (Song et al., 2022a), CauAIN (Zhao et al., 2022), DualGATs (Zhang et al., 2023), MKFM (Tu et al., 2023), MFAM(Hou et al., 2023), and CEPT (Gao et al., 2024).
(ii) LLMs-based ERC methods: ChatGPT (Ouyang et al., 2022), InstructERC (Lei et al., 2023), and BiosERC (Xue et al., 2024).

4.3 Implementation Details

Following current LLMs-based ERC methods (Lei et al., 2023; Xue et al., 2024), we adopt the LLaMA2 (Touvron et al., 2023) as the foundational model in this paper. Consider the expensive training costs and the issue of catastrophic forgeting, we use a lightweight training technique, i.e., LoRA (Hu et al., 2022), to stay the model weights frozen and train a small portion of model parameters for specific subtasks. In detail, we set the learning rate to 2e-4, and the converstaional context window to 12 for all evaluation datasets. In the first stage, the batch size is set to 8. In the second stage, the batch size is set to 16. For the hyper-parameter such as epoch, we tune them on the development dataset. The reported results are an average over five random runs. All the experiments are implemented by using PyTorch (Paszke et al., 2019) on a single NVIDIA Tesla V100 GPUs. We restrict the input length to 1024. More details about parameters analysis of context window can be found in Appendix A.3.

Methods	IEMOCAP	MELD	EmoryNLP	Avg.
COSMIC	63.43	65.03	38.49	55.65
SKAIG	66.96	65.18	38.88	57.01
DialogXL	66.20	62.41	34.73	54.45
SPCL	69.21	66.13	40.25	58.53
CauAIN	65.01	64.89	37.87	55.92
DualGATs	67.68	66.90	40.29	58.29
MKFM	68.08	65.50	39.76	57.78
MFAM	70.16	66.65	41.06	59.29
CEPT	70.53	67.51	-	-
ChatGPT	40.07	54.37	37.55	44.00
BiosERC	69.02	68.72	41.44	59.73
InstructERC	71.39	69.15	41.37	60.64
LaERC	69.95	68.86	40.87	59.89
LaERC-S	**72.40**	**69.27**	**42.08**	**61.25**

Table 1: Performance comparison between our proposed LaERC-S and existing ERC methods on three conversation datasets. LaERC is finetuning Llama2-7B to recognize emotion in conversation. The p-values are all below 0.001 by using pairwised t-test towards our method and the corresponding baselines. The best results are **bolded**. Our evaluation metric is weighted-F1.

4.4 Main Results

To illustrate the effectiveness of LaERC-S framework in the task of ERC, we report the performance of our proposed method and other baseline methods in Table 1, where ‘Avg.’ denote the overall average performance on three benchmark datasets. We can observe that our proposed LaERC-S achieves the best results than other all methods on three public datasets. Such performance demonstrates that LaERC-S has stronger generalization and more accurate predictions for emotion recognition.

Typically, compared to previous ERC paradigms, LLMs-based ERC methods have achieved significant results than them. The reason is the thorough understanding capability of pre-trained large language models. Notably, our proposed method LaERC-S achieves an improvement of 1.01% over InstructERC, 3.38% over BiosERC on the IEMOCAP dataset, respectively. For more complex conversation scenarios, such as MELD and EmoryNLP datasets, LaERC-S still provides meaningful gains in performance. This is due to the efficiency of speaker characteristics explored from the key element ‘oReact’ in the proposed LaERC-S.

Besides, we notice that the results of ChatGPT in

Methods	IEMOCAP	MELD	EmoryNLP	Avg.
w/o S	69.95	68.86	40.87	59.89
w M	70.21	68.52	41.49	60.07
w R	71.43	69.04	40.82	60.43
w S	72.40	69.27	42.08	61.25

Table 2: Performance comparison by speaker characteristics in emotion recognition. 'M' refers to directly introduce the generated speaker characteristics into the stage of emotion analysis. 'R' and 'S' regard speaker identification and speaker characteristics injection as the initial stage.

zero-shot scenarios are far from other methods that trained with the full dataset. It is attributed to the purpose of universality rather than specific tasks. Therefore, consistent with LaERC-S, it is essential to fine-tune the models for the task of ERC. In summary, the above comparative results present that LaERC-S outperforms all the ERC methods.

4.5 Ablation Study

In this section, we demonstrate the superiority of the proposed method LaERC-S from the impact of speaker characteristics. It is to measure the importance of introducing speaker characteristics, and how to sufficiently exploit it in the task of ERC. The experimental results are presented in Table 2, we can achieve the following findings:

- To understand the importance of introducing information around speaker characteristics in conversational scenarios, we present the results of relevant experiments in Table 2, where the first two rows are the one-stage learning, and the last two rows are the two-stage learning. For reference, in the first row of this table, we eliminate any of speaker characteristics, and solely implement the stage of emotion analysis, presenting a lowest result. Next, we directly incorporate the generated speaker characteristics into the stage of emotion analysis, resulting in the performance improvements in the most of datasets. This highlights the importance of speaker characteristics in ERC.
- On the other hand, we adopt the two-stage learning strategy, and regard speaker identification as the initial stage before the stage of emotion analysis. Such method outperforms the first two methods (i.e., one-stage learning), suggesting the efficiency of two-stage learning in the ERC task. In the last row of Table 2, we present the final performance of LaERC-S, which achieves the best results on all the datasets. These experiments demonstrate that LaERC-S can achieve accurate emotion predictions through introducing speaker characteristics, and use the two-stage learning to magnify the efficiency of speaker characteristics to enhance the model in performance.

Key Ele.	IEMOCAP	MELD	EmoryNLP	Avg.
oReact	72.40	69.27	42.08	61.25
xIntent	71.60	69.56	41.39	60.85
xReact	71.14	69.17	39.91	60.07
xEffect	70.70	68.54	41.94	60.39
oEffect	71.27	68.27	41.64	60.39
oWant	70.81	68.87	43.24	60.97
xWant	71.24	68.65	42.37	60.75
xNeed	71.94	68.50	40.27	60.24
xAttr	70.08	67.82	40.54	59.48

Table 3: Analysis of different elements in the initial stage of LaERC-S. 'oReact' is regarded as the final LaERC-S model for emotion analysis in conversation.

5 In-depth Analysis

5.1 Elements Selection

To investigate the influence of different key elements (Key Ele. for short) within the speaker characteristics extraction and injection stage, we design a more detailed experiment by leveraging just one key elements.

Table 3 shows the results, from which we can observe that apart from 'xAttr', others can efficiently bring performance improvements compared to LaERC-S without the initial stage (the first row of Table 2). These phenomena can be attributed to the fact that 'xAttr' only reflects the personal attribute, which struggles to capture dynamic emotional clues in conversation scenarios. And conversely, the extracted information from the mental state and behavior can provide richer and deeper dynamic emotional clues for emotion prediction (Li et al., 2021; Ghosal et al., 2020).

Notably, in mental state, 'oReact' describes the reaction of listener that refers to the interlocutor of the target utterance in a conversation. It is manifested as dynamic emotional clues provided by the conversational context, capable of revealing emotional states, leading to a significant improvement in performance. Therefore, we choose 'oReact' as the key element in the initial stage.

Models	IEMOCAP	MELD	EmoryNLP	Avg.
Baseline	69.95	68.86	40.87	59.89
Mistral-7B	70.44	69.15	41.25	60.28
Mixtral-7B	70.86	69.32	40.88	60.35
Claude	70.88	69.22	41.77	60.62
Llama2-13B	70.31	69.58	43.19	61.03
Llama2-7B	72.40	69.27	42.08	61.25

Table 4: Performance of LaERC-S with different large language models on three public conversation datasets. 'Claude' represents Claude-3-Haiku.

Templates	IEMOCAP	MELD	EmoryNLP	Avg.
Template 1	71.86	68.32	40.62	60.27
Template 2	71.54	69.05	40.25	60.28
Template 3	71.85	68.04	41.51	60.47
Template 4	72.40	69.27	42.08	61.25

Table 5: Performance of LaERC-S with different templates on three benchmark datasets.

5.2 Different LLMs Impact on speaker Characteristic Extraction

To demonstrate the expansibility of LaERC-S, we make a comprehensive comparison of the generated speaker characteristics from different large language models with parameters ranging from 7B to 13B, as shown in Table 4. Specifically, we employ a series of representative LLMs including Mistral-7B (Jiang et al., 2023), Mixtral-8×7B (Jiang et al., 2024), Claude-3-Haiku, Llama2 (13B, 7B) (Touvron et al., 2023), for evaluation. In the first row of the table, we present the performance of the baseline to intuitively understanding the impact of speaker characteristics in LaERC-S. We can see that various LLMs generate the speaker characteristics that is beneficial to provide the performance improvements of the proposed method. This emphasize the expansibility of the LaERC-S. Moreover, we intuitively think the reason why Llama3-13B performs worse than Llama2-7b is the inconsistency of the adopted models between extraction and injection. The larger-scale language models have not yet provided significant improvements in performance. However, LaERC-S employs Llama2-7B to generate speaker characteristics and further train it for more accurate emotion predictions.

5.3 Different Template Impact on Speaker Characteristics Generation

To explore the impact of various templates in performance, we conduct experiments with four different templates (more details about the templates can be found in Appendix A.4.3), as presented in Table 5. We randomly sampled 100 samples from the training set and generated speaker characteristics for each instance using four different templates. We manually validated the quality of the speaker characteristics produced for sample to determine which template to select. Among the 100 samples, we discovered that 80% selected Template 4, 8% selected Template 3, 7% selected Template 2, and 5% selected Template 1.

Specifically, Although each template solely exists subtle discrepancies, they present different results. For instance, the word "potential" in template 4 is removed in template 2, leading to a 0.97% drop in performance, suggesting the importance of the template in LaERC-S. These experiments proves that LLMs are sensitive to templates, which validates that a good template is important in LaERC-S for emotion recognition task. Therefore, we choose the template 4 in LaERC-S to perform more accurate emotion predictions.

5.4 Robustness Analysis

To validate the robustness capability of LaERC-S, we conduct a cross-dataset validation experiment.

Specifically, we first extract data with the same proportion from the training sets of three datasets, and then merge them into a mixed dataset. Subsequently, we train LaERC-S on the mixed dataset and inference on the test sets of the three original datasets. Finally, we demonstrate the generalization of LaERC-S by comparing its weighted-F1 score to that obtained from training and inference both on the original dataset. Notably, in this experiment, we choose InstructERC as our strong baseline due to its outstanding performance compared to other previous ERC models.

The results are shown in Figure 3, from which we can observe that LaERC-S is less affected by the cross-dataset validation compared to InstructERC. More specifically, in the dataset IEMOCAP and EmoryNLP, the 'Avg' of the proposed LaERC-S surpasses the baseline method InstructERC by significant improvements of 0.29% and 0.39%, respectively. Even in the more complex conversation dataset MELD, LaERC-S presents a better

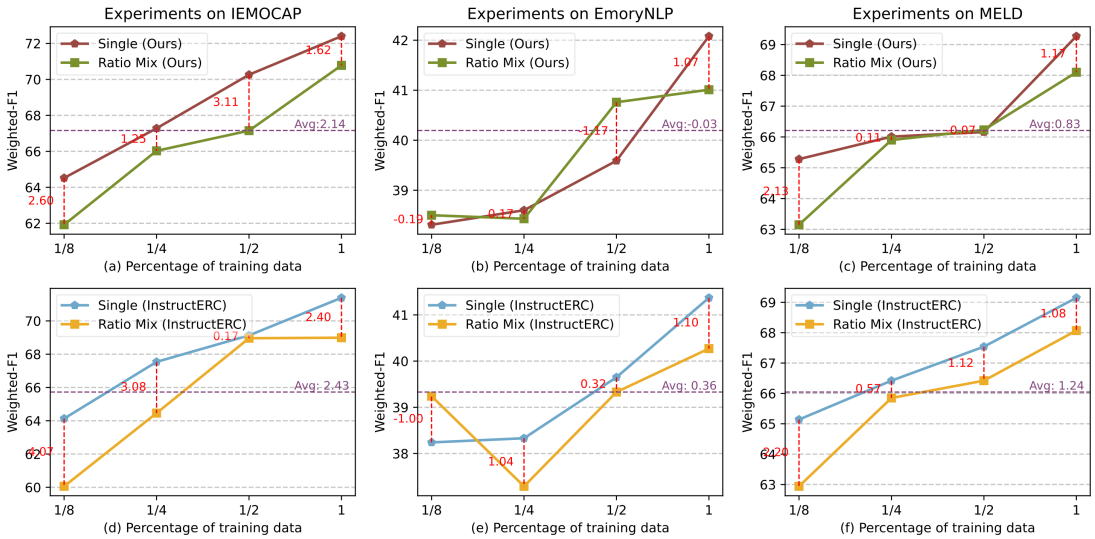

Figure 3: The cross-datasets analysis. 'Single' and 'Mixed Ratio' refer to training on a single and mixed dataset, respectively. We sequentially select data from each dataset in the ratios of 1/8, 1/4, 1/2, and 1. 'Avg' represents the average of the differences between 'Single' W-F1 and 'Ratio mix' W-F1.

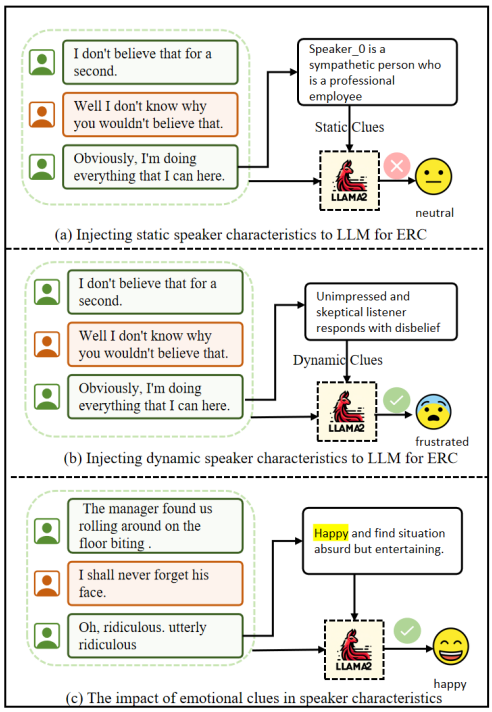

Figure 4: The case study of three samples from IEMOCAP dataset.

robustness (a performance improvement of 0.41%). These phenomena exhibits the exceptional robustness of our model. More details about robustness analysis can be found in Appendix A.1.

6 Case Study

In this section, we present two influence to ERC, including, speaker characteristic categories and emotional clues in speaker characteristics.

The difference between dynamic speaker characteristics and static speaker characteristics. Figure 4 (a) and (b) gives two demonstrations from IEMOCAP dataset about the same sentence affected by static speaker characteristics and dynamic speaker characteristics and then generate different emotional responses. Conversation (a) predict a neutral label due to the fact that speaker character is expressed as sympathy. In the contrast, conversation (b) generates an interactive characteristic of the current listener including some dynamic emotional clues about frustration.

The impact of emotional clues in speaker characteristics. Figure 4 (c) shows the impact of emotional clues in speaker characteristics. We can find that the responses a word 'Happy' of listeners generated will align with the emotional expression of the speaker. It can assist the model in producing accurate results.

7 Conclusion

In this paper, we propose LaERC-S, a novel framework that explores speaker characteristics, such as mental state, behavior and persona, to promote the progress of emotion recognition in conversation (ERC). LaERC-S is well-designed with three imperative parts: speaker characteristics extraction, speaker characteristics injection and emotion analysis, all of which work in harmony to make the model reason emotional dynamics and identify emotional tendencies for each utterance in con-

versations. Extensive experiments on three public conversation datasets demonstrate the effectiveness and superiority of our proposed LaERC-S.

In the future work, we would like to delve into the correlation and discrepancy between speaker characteristics in form of diverse expressions. This reason is that the speaker-specific information under different perspectives presents consistent clues of an identical emotion for the utterance. These properties can make the model possess convincing explanations for emotion analysis.

Limitations

While LaERC-S has made a significant progress in adapting the LLMs for the task of emotion recognition in conversation, the current work can still be improved in the following ways. Firstly, it is important to find effective ways to maintain an efficient running cost for such large-scale embedding models. Secondly, speaker characteristics around the mental state and behavior of interlocutors have potential to be extended to other tasks in the realm of natural language processing.

Acknowledgments

We thank the anonymous reviewers for their insightful comments. This work was supported by the Fundamental Research Funds for the Central Universities (project number: 2022FRFK060002).

References

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. Comet: Commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

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

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. 2023. Palm: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240):1–113.

Xiang Deng, Vasilisa Bashlovkina, Feng Han, Simon Baumgartner, and Michael Bendersky. 2023. Llms to the moon? reddit market sentiment analysis with large language models. In *Companion Proceedings of the ACM Web Conference 2023*, pages 1014–1019.

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. 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)*. Association for Computational Linguistics.

Louie Giray. 2023. Prompt engineering with chatgpt: a guide for academic writers. *Annals of biomedical engineering*, 51(12):2629–2633.

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.

Edward J Hu, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2022. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Jena D Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi. 2021. (comet-) atomic 2020: On symbolic and neural commonsense knowledge graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 6384–6392.

Zhongquan Jian, Ante Wang, Jinsong Su, Junfeng Yao, Meihong Wang, and Qingqiang Wu. 2024. Emotrans: Emotional transition-based model for emotion recognition in conversation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5723–5733.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Albert Q Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, et al. 2024. Mixtral of experts. *arXiv preprint arXiv:2401.04088*.

Shivani Kumar, Ishani Mondal, Md Shad Akhtar, and Tanmoy Chakraborty. 2023. Explaining (sarcastic) utterances to enhance affect understanding in multimodal dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 12986–12994.

Joosung Lee and Wooin Lee. 2022. Compm: Context modeling with speaker's pre-trained memory tracking for emotion recognition in conversation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5669–5679.

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

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.

Wei Li, Luyao Zhu, Rui Mao, and Erik Cambria. 2023. Skier: A symbolic knowledge integrated model for conversational emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13121–13129.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9):1–35.

Navonil Majumder, Pengfei Hong, Shanshan Peng, Jiankun Lu, Deepanway Ghosal, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. Mime: Mimicking emotions for empathetic response generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8968–8979.

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.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

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

Soujanya Poria, Erik Cambria, Devamanyu Hazarika, Navonil Majumder, Amir Zadeh, and Louis-Philippe Morency. 2017. Context-dependent sentiment analysis in user-generated videos. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)*, pages 873–883.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. Meld: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

Sahand Sabour, Chujie Zheng, and Minlie Huang. 2022. Cem: Commonsense-aware empathetic response generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 11229–11237.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019. Atomic: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3027–3035.

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 Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1551–1560.

Xiaohui Song, Longtao Huang, Hui Xue, and Songlin Hu. 2022a. Supervised prototypical contrastive learning for emotion recognition in conversation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 5197–5206.

Xiaohui Song, Liangjun Zang, Rong Zhang, Songlin Hu, and Longtao Huang. 2022b. Emotionflow: Capture the dialogue level emotion transitions. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8542–8546. IEEE.

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.

Yang Sun, Nan Yu, and Guohong Fu. 2021. A discourse-aware graph neural network for emotion recognition in multi-party conversation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2949–2958.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

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.

Geng Tu, Jintao Wen, Cheng Liu, Dazhi Jiang, and Erik Cambria. 2022. Context-and sentiment-aware networks for emotion recognition in conversation. *IEEE Transactions on Artificial Intelligence*, 3(5):699–708.

Yan Wang, Bo Wang, Yachao Zhao, Dongming Zhao, Xiaojia Jin, Jijun Zhang, Ruifang He, and Yuexian Hou. 2024a. Emotion recognition in conversation via dynamic personality. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5711–5722.

Zhunheng Wang, Xiaoyi Liu, Mengting Hu, Rui Ying, Ming Jiang, Jianfeng Wu, Yalan Xie, Hang Gao, and Renhong Cheng. 2024b. Ecok: Emotional commonsense knowledge graph for mining emotional gold. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 8055–8074.

Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C Schmidt. 2023. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382*.

Jieying Xue, Minh Phuong Nguyen, Blake Matheny, and Le Minh Nguyen. 2024. Bioserc: Integrating biography speakers supported by llms for erc tasks. *arXiv preprint arXiv:2407.04279*.

Sayyed M Zahiri and Jinho D Choi. 2018. Emotion detection on tv show transcripts with sequence-based convolutional neural networks. In *Workshops at the thirty-second aaai conference on artificial intelligence*.

Duzhen Zhang, Feilong Chen, and Xiuyi Chen. 2023. 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.

Wenxuan Zhang, Yue Deng, Bing Liu, Sinno Pan, and Lidong Bing. 2024. Sentiment analysis in the era of large language models: A reality check. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 3881–3906.

Xikun Zhang, Antoine Bosselut, Michihiro Yasunaga, Hongyu Ren, Percy Liang, Christopher D. Manning, and Jure Leskovec. 2022. Greaselm: Graph reasoning enhanced language models for question answering. In *International Conference on Representation Learning (ICLR)*.

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*.

A Appendix

A.1 The Details of Robustness Experiment

In this section, we will introduce how to construct the custom dataset used in § 5.4. Specifically, since the emotional labels in each original dataset are different, we need to map them to a unified label before the extracting and merging, as shown in Table 6.

A.2 Details of the Datasets

IEMOCAP is a dataset collected from improvisations or scripted scenarios, which contains 12 hours of conversation videos from 10 unique speakers. It has five sessions consisting of 151 conversations and 7,433 utterances. Each utterance is annotated with one of six emotion classes: neutral, happy, sad, excited, angry, and frustrated.

MELD is another dataset including more than 13,000 video snippets from the Friends TV series. It comprises 1,433 conversations and 13,708 utterances in total. Each utterance is labeled as one of seven emotion categories: anger, disgust, fear, joy, neutral, sadness, and surprise.

EmoryNLP is also based on the Friends TV series, which contains 97 episodes, 897 scenes and 12,606 utterances. Each utterances is annotated as one of seven emotion types: neutral, joyful, peaceful, powerful, scared, mad, and sad.

A.3 The Context Window Investigation

To examine the impact of the context window in the performance, we conduct a parametric sensitivity

Number	IEMOCAP	MELD	EmoryNLP	Final Emotion
1	happy	joyful	joyful	joyful
2	sad	sad	sad	sad
3	neutral	neutral	neutral	neutral
4	angry	angry	mad	mad
5	excited	-	-	excited
6	-	surprise	powerful	powerful
7	scared	fear	frustrated	fear
8	-	-	peaceful	peaceful
9	-	disgust	-	disgust

Table 6: Unified Label Mapping

analysis with different context window, as depicted in Table 7. We can notice that LaERC-S achieves the superior performance over InstructERC under any context window settings. This highlights the efficiency of LaERC-S on the task of ERC. For reference, in the first row of the table, LaERC-S provides a 5.01%, 2.79%, and 1.52% improvements over InstructERC on IEMOCAP, MELD, and EmoryNLP, respectively. With the size increasing of the context window, the performance of both methods presents a tendency of improvement. Compared with MELD and EmoryNLP, the models on the dataset IEMOCAP present a significant improvement with the same context window. This is attributed to the average length of conversation in various datasets. The average length of IEMOCAP is longer than that of other datasets, thereby exploiting the larger window providing the necessary historical context for an improvement in performance. Although the performance discrepancy between them gradually decreases, the proposed LaERC-S still achieves significant superiority on three benchmark datasets. Therefore, we set a context window of 12 in LaERC-S to sufficiently capture the historical cotext in a conversation.

A.4 Prompts

A.4.1 Definitions of key elements

We give the definition of key elements in Table 8. This key elements include nine categories.

A.4.2 Prompts for key elements

The key elements are used in template for speaker characteristics extraction and injection. As illustrated in Table 8 and Table 9, we design the instruction-following templates for speaker characteristic extraction and injection, respectively. These templates provide precise descriptions for basic elements, such as “title”, “specific token”, “objective” and “constraint”, to promote LLMs in performing the ERC task. Such a design is essential to guarantee clarity and accuracy in each stage.

A.4.3 Details of Various Templates Design on Speaker Characteristics Extraction

In the different template design shown as Table 10, we have designed different textual expressions for each key element of speaker characteristics. For example, the key element "oReact" can be expressed as "the reaction of potential listeners", "the reaction of listeners", "the oReact of listeners ", and "the reaction of listeners to the event". We find that we use template with "the reaction of potential listeners" word can better extract accurate speaker characteristics.

A.5 The analysis of different emotion label’ s performance

Compared with InstructERC, our method achieves improvements in most emotion label, and presents sub-optimal performance in rare cases. (1) As for IEMOCAP, our method is superior to InstructERC across all emotion classes. The highest gain is 6.63% on “Happy”. (2) As for remaining two datasets, our method still maintains consistent improvements, and achieves sub-optimal results on “Disgust” due to its few samples (2.6% of the total dataset).

Context	IEMOCAP		MELD		EmoryNLP	
Window	InstructERC	LaERC-S	InstructERC	LaERC-S	InstructERC	LaERC-S
1	56.12	61.13 (5.01↑)	65.91	68.70 (2.79↑)	38.32	39.84 (1.52↑)
5	68.65	69.97 (1.32↑)	66.97	69.21 (2.24↑)	40.48	41.96 (1.48↑)
12	71.39	72.40 (1.01↑)	69.15	69.27 (0.12↑)	41.37	42.08 (0.71↑)

Table 7: Parameter analysis of the context window in the proposed method LaERC-S on three widely-used benchmark datasets. The symbol ↑ represents an improvement in performance over the compared method InstructERC.

Key element		**Description**
Mental-state	xIntent	The reason why the speaker would cause the event
	xReact	The reaction that the speaker would have to the event
	oReact	The reaction of listeners to the event
Event	xWant	What the speaker may want to do after the event
	oWant	What the listener may want to do after the event
	xEffect	The effect the event would have on the speaker
	oEffect	The effect the event has on the listener
	xNeed	What the speaker might need to do before the event
Persona	xAttr	How the speaker might be described given their part in the event

Table 8: Definitions of different key elements.

Key element	**Prompt**	**Speaker characteristics**
xIntent	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker1:"Okay, so big news." Speaker0:"What?" ### Please use common sense to infer the intention of <Speaker0:"What?" >:	Expecting explanation or clarification
xReact	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker1:"Okay, so big news." Speaker0:"What?" ### Please use common sense to infer the reaction of speaker in <Speaker0:"What?" >:	Surprised and curious about the news
oReact	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker1:"Okay, so big news." Speaker0:"What?" ### Please use common sense to infer the reaction of potential listeners in <Speaker0:"What?" >:	Listener looks surprised and excited.
xEffect	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker1:"Okay, so big news." Speaker0:"What?" ### Please use common sense to infer the effect on speaker in <Speaker0:"What?" >:	Speaker 0 looks excited about the news
oEffect	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker1:"Okay, so big news." Speaker0:"What?" ### Please use common sense to infer effect of potential listeners in <Speaker0:"What?" >:	Expectation arises; curious minds eagerly await details
oWant	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker1:"Okay, so big news." Speaker0:"What?" ### Please use common sense to infer the wanted by listeners in <Speaker0:"What?" >:	Exciting development or surprise event
xAttr	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker1:"Okay, so big news." Speaker0:"What?" ### Please use common sense to infer the attribute of speaker in <Speaker0:"What?" >:	Speaker 0 is a curious person
xwant	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker1:"Okay, so big news." Speaker0:"What?" ### Please use common sense to infer the wanted by speaker in <Speaker0:"What?" >:	Want to know the big news
xneed	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker1:"Okay, so big news." Speaker0:"What?" ### Please use common sense to infer the need of speaker in <Speaker0:"What?" >:	Expecting important information or reaction

Table 9: Prompts of different key elements.

Template	Prompt
Template1	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker0 : "Hey."Speaker1 : "Hey."Speaker0 : "Esmeralda, guesswhat?"### Based on the above historical utterances, please use common sense to infer the reaction of listeners to the event in <Speaker1 : "What?" >in no more than ten words of output :
Template2	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker0 : "Hey."Speaker1 : "Hey."Speaker0 : "Esmeralda, guesswhat?"### Based on the above historical utterances, please use common sense to infer the reaction of listeners in <Speaker1 : "What?" >in no more than ten words of output :
Template3	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker0 : "Hey."Speaker1 : "Hey."Speaker0 : "Esmeralda, guesswhat?"### Based on the above historical utterances, please use common sense to infer the oReact of listeners in <Speaker1 : "What?" >in no more than ten words of output :
Template4	Now You are an expert who is good at using common sense for reasoning. The following conversation noted between '### ###' involves several speakers. ### Speaker0 : "Hey."Speaker1 : "Hey."Speaker0 : "Esmeralda, guesswhat?"### Based on the above historical utterances, please use common sense to infer the reaction of potential listeners in <Speaker1 : "What?" >in no more than ten words of output :

Table 10: The samples of different templates.