Abstract

Our research encompasses primarily two interrelated areas: Emotion Recognition in Conversations (ERC) and multilingual multi-label emotion detection. The former aims to identify the emotional state of each utterance in a dialogue, while the latter addresses the detection of multiple emotions across languages within a given sample.

Within the ERC domain, we explore several critical yet underexplored dimensions: In emotion context modeling, traditional sequential models often capture only local emotional dependencies, overlooking long-range emotional influences that may arise between distant parts of a conversation. We argue that emotional states can be affected and transmitted by speakers and utterances throughout the conversation, regardless of their positional distance. In response to this limitation, we propose the *L*ong-range dependenc*Y* emotion*S* *M*odel (LYSM), which employs self-attention mechanisms to capture emotional dynamics throughout entire conversations, allowing the system to integrate emotional dependencies from both nearby and distant utterances. Experimental results on four benchmark datasets confirm its strong generalizability and effectiveness.

In utterance representation, unlike isolated sentence-level emotion tasks, the core challenge in dialogue emotion recognition lies in effectively representing a target utterance within its conversational context. Although pre-trained language models (PLMs) such as RoBERTa offer strong capabilities for context modeling, existing PLM-based approaches often fail to fully exploit their potential for fine-grained contextual encoding. To overcome these limitations, we introduce *Acc*umulating Word *R*epresentations in Multi-level Context Integration for ERC Task (AccWR), which aggregates multi-level contextual word information before inputting it into the PLM. This enriched contextual word aggregation enhances both semantic understanding and the model's focus on the target utterance. Experimental results show that AccWR consistently outperforms baseline models on four benchmark datasets and demonstrates strong potential for broader applications such as response generation and semantic parsing.

For speaker modeling, ERC typically relies on spoken dialogues transcribed by automatic speech recognition systems. Individual traits such as linguistic style and personality significantly influence emotional expression, yet prior work often depends on implicitly learned speaker

features, limiting interpretability and cross-domain generalization. To address this, we propose BiosERC: Integrating *Bio*graphy-Based *S*peaker Representations with Large Language Models for *E*motion *R*ecognition in *C*onversations, which employs LLMs with prompt-based techniques to extract explicit speaker profiles as external knowledge. These biography-based representations enhance the emotional understanding of each speaker, leading to more accurate and nuanced recognition, especially in complex or multi-party conversations. BiosERC achieves competitive or state-of-the-art (SOTA) performance on three benchmark datasets.

In multi-level context modeling, the spontaneous nature of conversations makes it difficult to capture transient and dynamically evolving emotional states using only contextual discourse and static speaker profiles. To address this, we propose TraceERC: *T*racking *R*elational *A*wareness of *C*ontextual, Character, and *E*motional States in *E*motion *R*ecognition in *C*onversations, which leverages LLMs to jointly encode dialogue context, speaker personality, and dynamic emotional cues. These enriched representations enable emotion predictions that are sensitive to both conversational flow and individual speaker characteristics, enhancing emotional understanding and adaptability. As one of the first LLM-based ERC models to incorporate contextual learning and in-context fine-tuning, TraceERC achieves SOTA results on MELD and strong performance across benchmarks.

In the multilingual domain, our team participated in SemEval-2025 Task 11, tackling both multi-label classification (Track A) and emotion intensity detection (Track B). We developed a generation-based framework leveraging multilingual PLMs and LLMs to support both high- and low-resource languages.

Keywords: Emotion Recognition in Conversations, Multilingual Multi-label Emotion Detection, Large Language Models, Speaker Personality Modeling, Emotion Context Modeling