TW-NLP at SemEval-2024 Task10: Emotion Recognition and Emotion Reversal Inference in Multi-Party Dialogues.

Wei Tian[1], **Peiyu Ji**[2], **Yuan Zheng**[1], **Lei Zhang**[1], **Yue Jian**[1]
[1]Beijing Smartdot Technology Co., Ltd, China
[2]Zhongyuan University of Technology, China
tianwei@smartdot.com,
2020107223@zut.edu.cn

Abstract

In multidimensional dialogues, emotions serve not only as crucial mediators of emotional exchanges but also carry rich information. Therefore, accurately identifying the emotions of interlocutors and understanding the triggering factors of emotional changes are paramount. This study focuses on the tasks of multilingual dialogue emotion recognition and emotion reversal reasoning based on provocateurs, aiming to enhance the accuracy and depth of emotional understanding in dialogues. To achieve this goal, we propose a novel model, MBERT-TextRCNN-PL, designed to effectively capture emotional information of interlocutors. Additionally, we introduce XGBoost-EC (Emotion Capturer) to identify emotion provocateurs, thereby delving deeper into the causal relationships behind emotional changes. By comparing with state-of-the-art models, our approach demonstrates significant improvements in recognizing dialogue emotions and provocateurs, offering new insights and methodologies for multilingual dialogue emotion understanding and emotion reversal research.

1 Introduction

The EDiReF shared task at SemEval 2024 encompasses three subtasks(Kumar et al., 2024): Task 1 involves Emotion Recognition (ERC) in mixed Hindi-English dialogues, Task 2 focuses on Emotion Flipping Reasoning (EFR) in mixed Hindi-English dialogues, and Task 3 involves EFR in English dialogues. In Task 1 ERC, the goal is to assign emotions to each utterance in the dialogue, while in Task 2 and Task 3 EFR, the aim is to identify trigger utterances leading to emotion flipping in multi-party dialogues. The definitions of these tasks provide a crucial framework for understanding the dynamics of emotions in natural language conversations.

Firstly, we are committed to addressing two subtasks: Emotion Recognition in Conversations (ERC) in Task 1, and Emotion Flipping Reasoning (EFR) in Tasks 2 and 3. For Task 1, we constructed the MBERT-TextRCNN-PL model based on MBERT(Pires et al., 2019) and Prompt Learning to identify emotions in mixed-language dialogues. By leveraging the multilingual capability of the MBERT model and incorporating Prompt Learning, we successfully guided the model to focus on key aspects of emotion recognition. This approach can effectively handle mixed Hindi and English conversations while achieving the sharing of model parameters between different languages. Finally, to improve the robustness of the model, this paper integrates FGM to enhance the model's generalization ability.

Secondly, for Tasks 2 and 3, we proposed the XGBoost-EC (Emotion Capture) method aimed at identifying triggers of emotion flipping. We segmented dialogues into fixed-size windows and extracted emotion encodings from each window. To better encode the emotions of the final speaker, we used -1 to fill in blank emotion states within the window. Then, we employed the XGBoost algorithm(Chen and Guestrin, 2016) for classification to identify windows that could potentially be triggers of emotion flipping. By guiding the model to learn patterns and features of emotion flipping triggers through annotated data during training, we were able to effectively identify triggers of emotion flipping in dialogues.

In the EDiReF shared task at SemEval 2024, our[1] proposed MBERT-TextRCNN-PL model achieved 6th place in Task 1. Additionally, our XGBoost-EC model secured 1st place in Task 2 and 5th place in Task 3.

[1]Our codes are available at https://github.com/TW-NLP/SemEval2024-Task10

Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), pages 311–315

2 Background

2.1 Dataset Description

Task 1, the Emotion Recognition (ERC) task, corresponds to the MASAC-ERC dataset compiled by extracting dialogues from Indian television dramas. The dataset comprises 446 dialogues, with 8 emotion categories: disgust, contempt, anger, neutral, joy, sadness, fear, and surprise. The training set consists of 343 dialogues containing 8,506 sentences, the validation set consists of 45 dialogues containing 1,354 sentences, and the test set consists of 56 dialogues containing 1,580 sentences, The data analysis for Task1 is shown in Table 1.

Set	Dlgs	Utts
Train	343	8506
Val	45	1354
Test	56	1580

Table 1: Data statistics for Task1.

For Task 2, corresponding to MASAC-EFR, the dataset includes 5,667 dialogues. The training set comprises 4,893 dialogues containing 98,777 sentences, the validation set comprises 389 dialogues containing 7,462 sentences, and the test set comprises 385 dialogues containing 7,690 sentences, The data analysis for Task2 is shown in Table 2.

Set	Dlgs	Utts
Train	4893	98777
Val	389	7462
Test	385	7690

Table 2: Data statistics for Task2.

Regarding Task 3, corresponding to MELD-EFR, the dataset consists of 5,428 dialogues. The training set comprises 4,000 dialogues containing 35,000 sentences, the validation set comprises 426 dialogues containing 3,522 sentences, and the test set comprises 1,002 dialogues containing 8,642 sentences, The data analysis for Task3 is shown in Table 3.

Set	Dlgs	Utts
Train	4000	35000
Val	426	3522
Test	1002	8642

Table 3: Data statistics for Task3.

2.2 Related Work

Emotion Recognition in Conversation. Emotion recognition in dialogues is categorized into monolingual and multilingual dialogue emotion recognition. Under monolingual conditions, the DialogXL(Shen et al., 2021) model utilizes the XL-Net(Yang et al., 2019) architecture for Emotion Recognition in Conversation (ERC). They encode dialogue discourse and leverage dialogue-aware self-attention to incorporate dialogue semantics. Additionally, (Jiao et al., 2019) employs a hierarchical gated recursive unit framework involving two different levels of GRU. The lower-level GRU models word-level inputs, while the higher-level GRU captures contextual information at the discourse level. Furthermore, (Lian et al., 2021) proposes a correction model named "Dialogue Emotion Correction Network (DECN)." The aim of this work is to enhance emotion recognition performance by automatically identifying errors made by emotion recognition strategies. (Shou et al., 2022) employs graph-based approaches to tackle ERC. They introduce a session-level sentiment analysis model that combines dependency parsing and graph convolutional neural networks. Self-attention mechanisms capture the most effective words in the dialogue and then construct a graph. In multilingual settings, (Kumar et al., 2023a) proposes an advanced fusion technique that first translates into a uniform language, followed by the integration of commonsense knowledge with the dialogue comprehension module.

Emotion Flipping Reasoning. (Kumar et al., 2022)introduces a novel Emotional Flip Reasoning (EFR) aimed at identifying the utterance that triggered an emotional state flip in an individual at a certain point in the past. Additionally, a Transformer-based network is proposed to carry out the Emotional Flip Reasoning. To identify emotional instigators, (Kumar et al., 2023b) proposes the TGIF framework for multilingual dialogue data. It utilizes a combination of Transformer and GRU to identify emotional instigators.

3 System Overview

3.1 ERC

In Task 1's data format, we designed the MBERT-TextRCNN-PL model based on prompt learning, as illustrated in Figure 1, to contextualize the conversation. The input data for this model is organized according to the format in Table 4.

Speaker	Utterance	Emotion
Sp1	Aaj to bhot awful day tha!	Sad
Sp2	Oh no! Kya hua?	Sad
Sp1	Kisi ne mera sandwich kha liya!	Sad
Sp2	Me abhi tumhare liye new bana deti hun!	Joy

Table 4: Data instance for ERC task

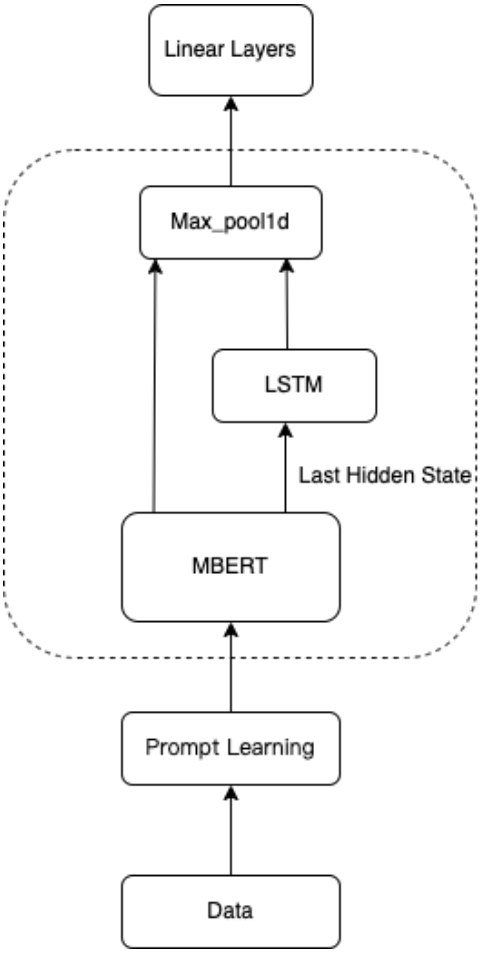

Figure 1: The architecture diagram of MBERT-TextRCNN-PL.

Specifically, we employ prompt learning to construct the model's input, where the inputs for Sp1 and Sp2 are formatted as follows: "The following is 's conversation history. Sp1: Aaj to bhot awful day tha!" and "The following is 's conversation history. Sp1: Aaj to bhot awful day tha!, Sp2: Oh no! Kya hua?". Through this formatted input, the model can better understand the conversation history and context, thereby comprehensively capturing semantic correlations in the text.

Building upon this, we propose a vector representation approach that combines the features of MBERT and TextRCNN. Firstly, we utilize the MBERT pre-trained model to encode the text sequences into semantic vectors, leveraging its multilingual semantic understanding capability. Subsequently, contextual information of the sequences is further enhanced by capturing it through a bidirectional LSTM layer, strengthening the model's comprehension of the dialogue history. Next, the last hidden state output of MBERT and the output of LSTM are concatenated in the feature dimension to fuse word-level and sequence-level semantic information. Finally, classification of the text data is achieved through a fully connected layer, enabling effective categorization of the text.

This architecture fully leverages the multilingual semantic understanding of MBERT and the sequence modeling capability of TextRCNN, enabling the model to better understand and classify text data. Such design not only enhances the performance and effectiveness of the model in text understanding tasks but also improves its understanding of context, allowing the model to more accurately capture semantic information in the text.

3.2 EFR

For EFR's Task 2 and Task 3, we propose the XGBoost-EC model, as illustrated in Figure 2. The XGBoost-EC model is an emotion recognition model based on XGBoost, designed to utilize emotion feature encoding for training and prediction.

In the XGBoost-EC model, we initially encode the emotion features, converting them into numerical forms for processing by the XGBoost algorithm. Specifically, we map emotion labels to integer values, such as encoding 'joy' as 1, 'sadness' as 2, and so forth. To better capture emotion information, we employ emotion windows for encoding and assign -1 for missing emotions.

By encoding emotions, the XGBoost-EC model can learn the associations between emotions and other features, thereby predicting the emotional states in text or dialogue more accurately. Leveraging the efficient performance of the XGBoost algorithm and its capability to handle large-scale

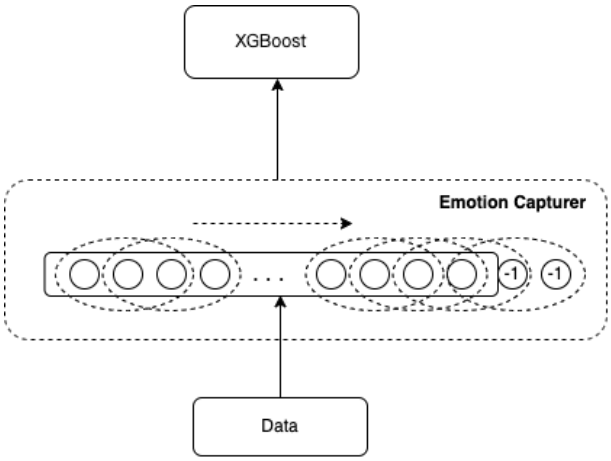

Figure 2: The architecture diagram of XGBoost-EC.

datasets, this model holds significant value for emotion recognition tasks.

4 Experimental Setup

Task 1. For the ERC Task 1, our model was built using the Hugging Face Transformers library, where we directly employed pre-trained tokenizer and language models for further fine-tuning. Specifically, we utilized the MBERT-TextRCNN-PL model with parameter configurations of a learning rate of 3e-6 and a batch size of 16 for training. We opted for the AdamW optimizer to update the model parameters. To ensure the model could handle longer text sequences, we set the maximum sequence length of the tokenizer to 512. During the model's validation evaluation process, we employed the F1 score as the performance metric. Hardware-wise, we utilized the NVIDIA RTX3090 (24G) graphics card to accelerate both model training and inference processes.

Task 2. For the EFR Task 2 and Task 3, we built the XGBoost-EC model based on the XGBoost library. In Task 2, after validation, we chose to set the scale-pos-weight parameter in the XGBClassifier to 1.08 to address the issue of imbalanced samples. Additionally, we fixed the random seed to 42 to ensure reproducibility of the results. To better capture changes in emotion, we set the emotion window size to 4.

Task 3. In Task 3, we adjusted the scale-pos-weight parameter in the XGBClassifier to 1.6 to accommodate different levels of sample imbalance. Similarly, we fixed the random seed to 42 and set the emotion window size to 3, enabling the model to capture trends in emotion changes more effectively.

5 Results

5.1 Task1

Based on Table 5, we observe that the performance of MBERT-TextRCNN-PL is more competitive compared to BERT, RoBERTa, MBert, MURIL, CoMPM, DialogXL, BERT+COFFEE and MBert+COFFEE as contrasted by Shivani Kumar et al(Kumar et al., 2023a). MBERT-TextRCNN-PL, based on prompt learning, provides a better summary of the former's remarks, which aligns more closely with conventional expression norms. Leveraging the multilingual MBERT-TextRCNN model enhances the representation of semantic information, rendering the model state-of-the-art. As a result, it achieved the 6th position on the official leaderboard.

Model	F1
BERT	0.40
RoBERTa	0.41
MBert	0.30
MURIL	0.35
CoMPM	0.35
DialogXL	0.41
BERT+COFFEE(Kumar et al., 2023a)	0.41
MBERT+COFFEE(Kumar et al., 2023a)	0.31
Ours(MBERT-TextRCNN-PL)	0.46

Table 5: The results of Task 1.

5.2 Task2

As shown in Table 6, the proposed XGB-EC achieved an F1 score of 0.79 in the evaluation of Task 2, securing the first position in this task.

Model	F1
TECHSSN Team	0.1
IASBS Team	0.12
IITK Team	0.56
Knowdee Team	0.66
FeedForward Team	0.77
Ours(XGB-EC)	0.79

Table 6: The results of Task 2.

5.3 Task3

As indicated in Table 7, the proposed XGB-EC outperformed AGHMN, TL-ERC, DGCN, DialogXL, and BERT by a significant margin, exhibiting a substantial improvement compared to the TGIF

framework with a 38-point increase in F1 score. In Task 3, it obtained the 5th position. The XGB-EC framework not only considers the current emotion but also captures emotion variations through emotion windows, enabling a better understanding of emotional provocations.

Model	P	R	F1
AGHMN	0.15	0.17	0.16
TL-ERC	0.07	0.33	0.13
DGCN	0.10	0.67	0.17
DialogXL	0.09	0.34	0.15
BERT	0.14	0.55	0.21
TGIF(Kumar et al., 2023b)	0.26	0.55	0.33
Ours(XGB-EC)	0.71	0.71	0.71

Table 7: The results of Task 3.

6 Conclusion

In this paper, we focus on addressing the challenges of multilingual conversation emotion recognition and emotion flipping reasoning tasks. To this end, we propose two models to tackle these tasks.

Firstly, for the multilingual conversation emotion recognition task, we introduce the MBERT-TextRCNN-PL model. This model combines prompt learning and the MBERT-TextRCNN approach to better recognize emotions from multiple speakers. Through prompt learning, we can provide richer contextual information, thereby accurately capturing the emotional content in the text. Leveraging the features of MBERT-TextRCNN, this model effectively utilizes the capabilities of multilingual semantic understanding and sequence modeling to enhance the accuracy and effectiveness of emotion recognition.

Secondly, for Task 2 and Task 3 of EFR, we propose the XGBoost-EC (emotion capturer) model, aimed at identifying emotion instigators. This model, employing emotion windows and XGBoost classifier, captures emotion instigators more effectively. The setting of emotion windows allows the model to consider the historical trend of emotion changes, leading to a more comprehensive analysis of emotional data. The application of the XGBoost classifier enables efficient classification and reasoning of emotional data, thereby enhancing the model's ability to identify instigators.

The introduction of these two models provides effective solutions for multilingual conversation emotion recognition and emotion flipping reasoning tasks, offering new insights and methods for research and applications in related fields.

References

Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Wenxiang Jiao, Haiqin Yang, Irwin King, and Michael R Lyu. 2019. Higru: Hierarchical gated recurrent units for utterance-level emotion recognition. *arXiv preprint arXiv:1904.04446*.

Shivani Kumar, Md Shad Akhtar, Erik Cambria, and Tanmoy Chakraborty. 2024. Semeval 2024–task 10: Emotion discovery and reasoning its flip in conversation (ediref). *arXiv preprint arXiv:2402.18944*.

Shivani Kumar, Md Shad Akhtar, Tanmoy Chakraborty, et al. 2023a. From multilingual complexity to emotional clarity: Leveraging commonsense to unveil emotions in code-mixed dialogues. *arXiv preprint arXiv:2310.13080*.

Shivani Kumar, Shubham Dudeja, Md Shad Akhtar, and Tanmoy Chakraborty. 2023b. Emotion flip reasoning in multiparty conversations. *arXiv preprint arXiv:2306.13959*.

Shivani Kumar, Anubhav Shrimal, Md Shad Akhtar, and Tanmoy Chakraborty. 2022. Discovering emotion and reasoning its flip in multi-party conversations using masked memory network and transformer. *Knowledge-Based Systems*, 240:108112.

Zheng Lian, Bin Liu, and Jianhua Tao. 2021. Decn: Dialogical emotion correction network for conversational emotion recognition. *Neurocomputing*, 454:483–495.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual bert? *arXiv preprint arXiv:1906.01502*.

Weizhou Shen, Junqing Chen, Xiaojun Quan, and Zhixian Xie. 2021. Dialogxl: All-in-one xlnet for multiparty conversation emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13789–13797.

Yuntao Shou, Tao Meng, Wei Ai, Sihan Yang, and Keqin Li. 2022. Conversational emotion recognition studies based on graph convolutional neural networks and a dependent syntactic analysis. *Neurocomputing*, 501:629–639.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32.