Mutual Conversational Detachment Network for Emotion Recognition in Multi-Party Conversations

Weixiang Zhao, Yanyan Zhao*, Bing Qin
Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, China
{wxzhao, yyzhao, qinb}@ir.hit.edu.cn

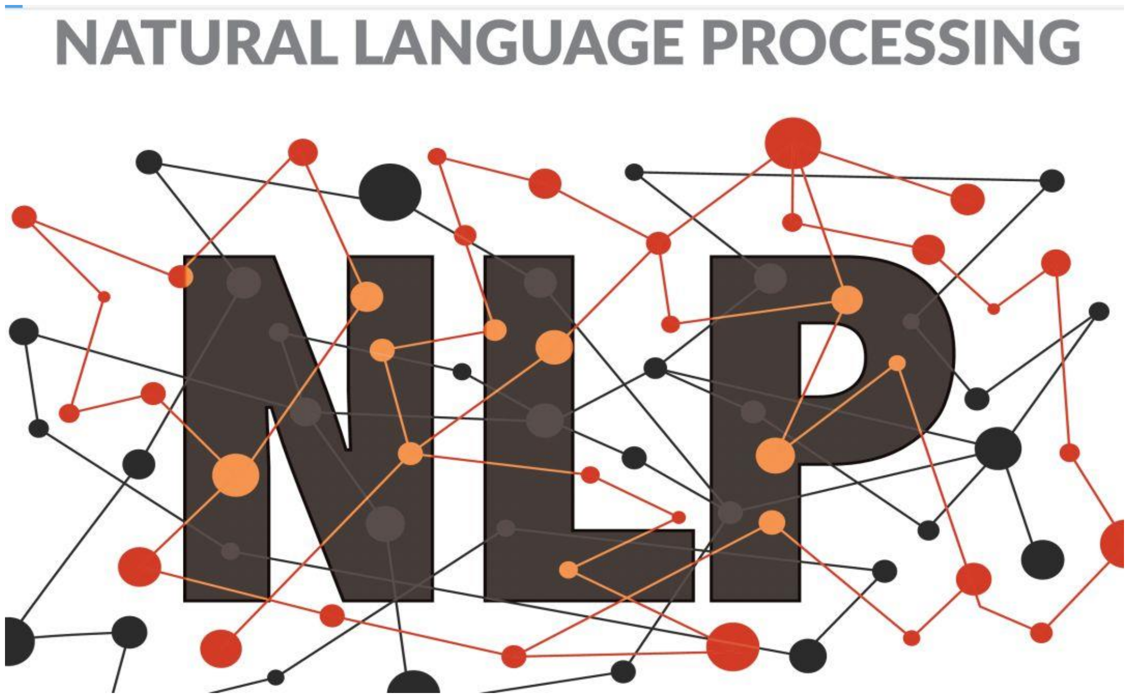
2022.10.15 • ChongQing

2022_COLING

Reported by Yidan Liu

Code: https: //github.com/circle-hit/MuCDN

gesis
Leibniz-Institut für Sozialwissenschaften



1.Introduction

2.Method

3.Experiments

gesis
Leibniz-Institut für Sozialwissenschaften

Introduction

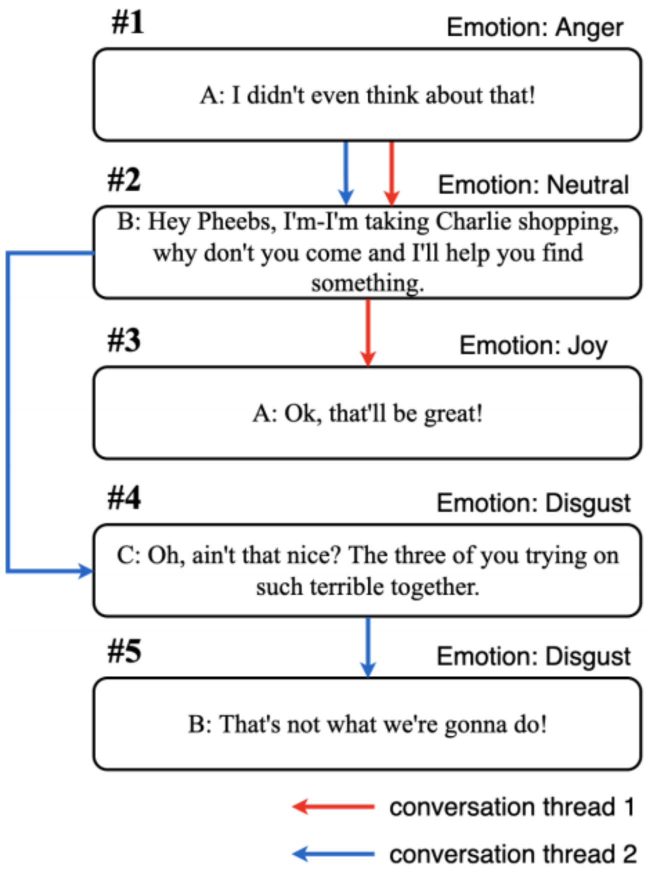


Figure 1: An example of a multi-party conversation from MELD dataset.

However, since emotional interactions among speakers are often more complicated within the entangled multi-party conversations, these works are limited in capturing effective emotional clues in conversational context.

Method

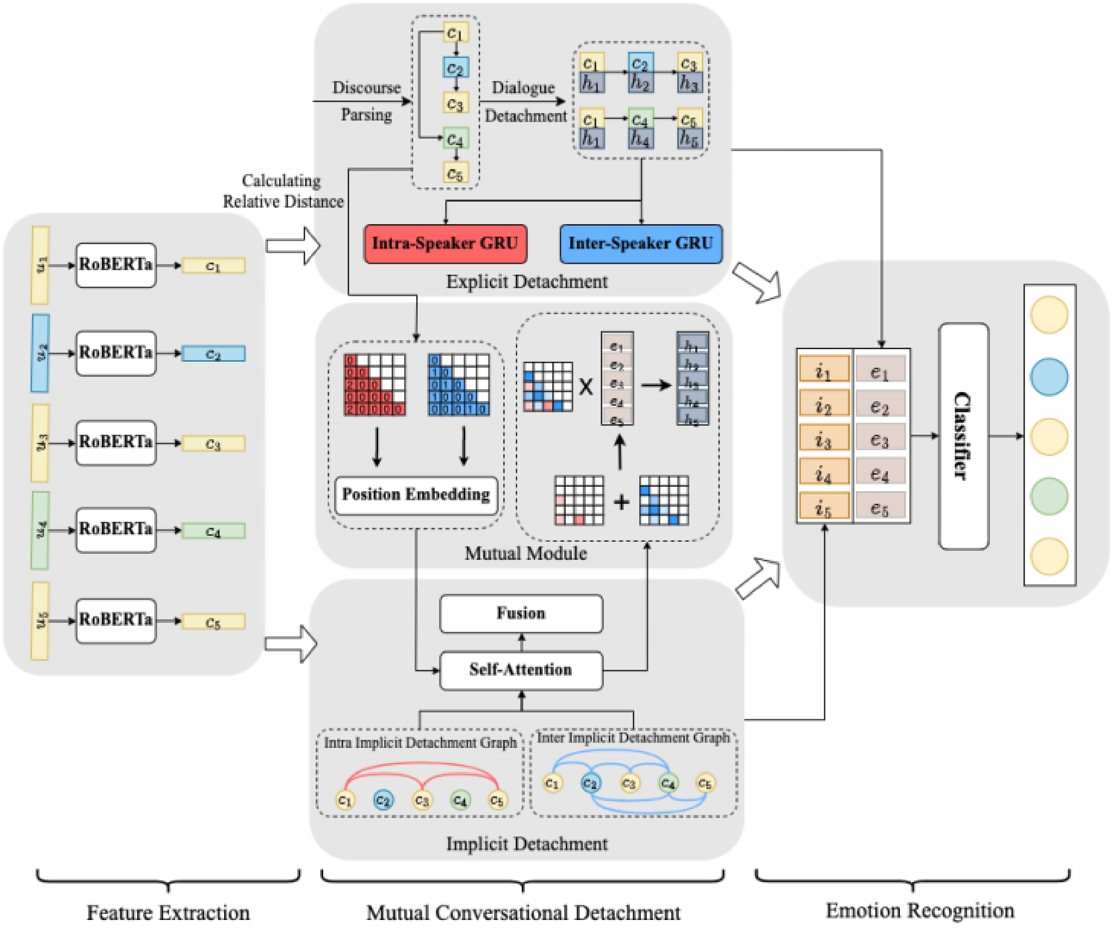
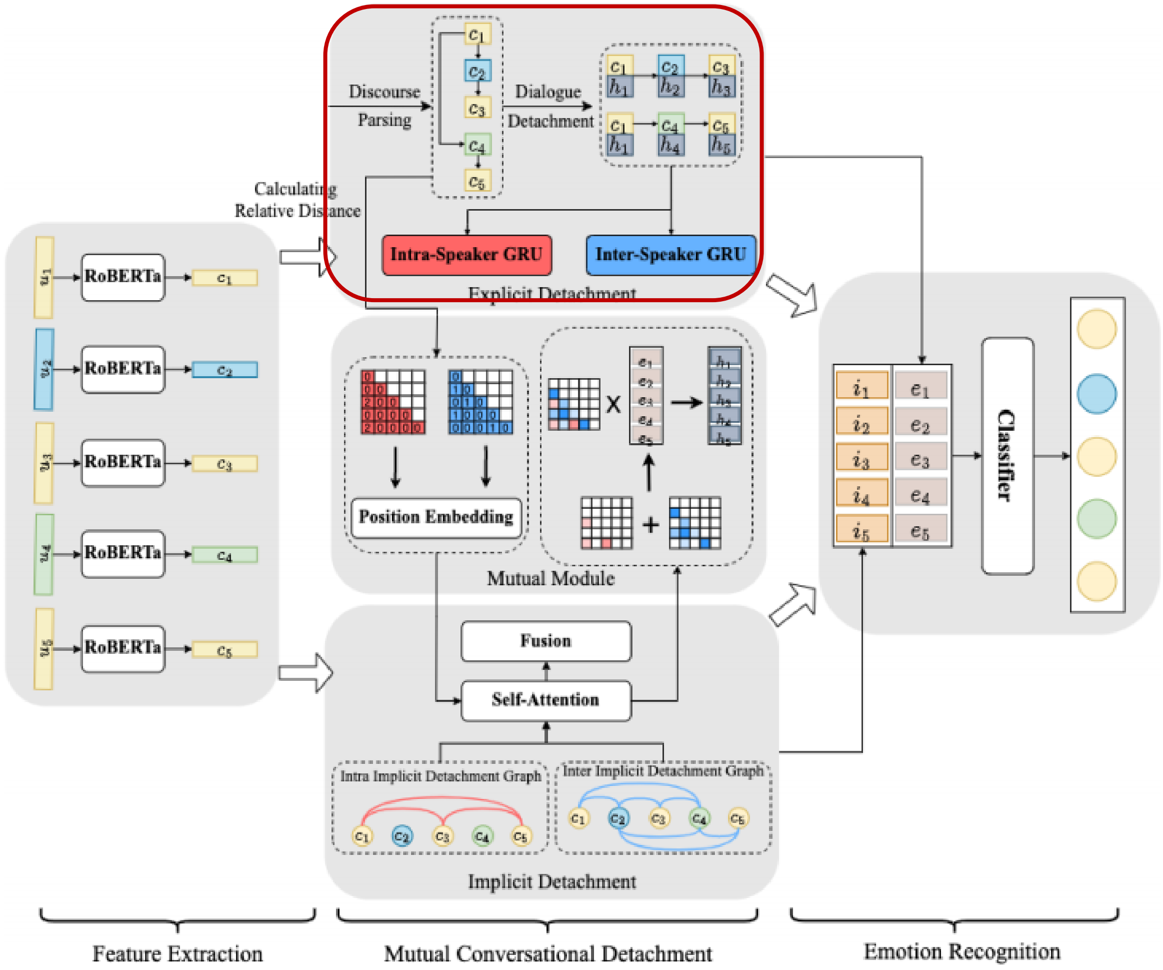


Figure 2: The overall architecture of our proposed model.

Method



Utterance-Level Feature Extraction

$$c_i = \text{RoBERTa}(\,[CLS], w_1, w_2, \cdots, w_L\,) \quad (1)$$

C is $\{c_1, c_2, \cdots, c_N\}$.

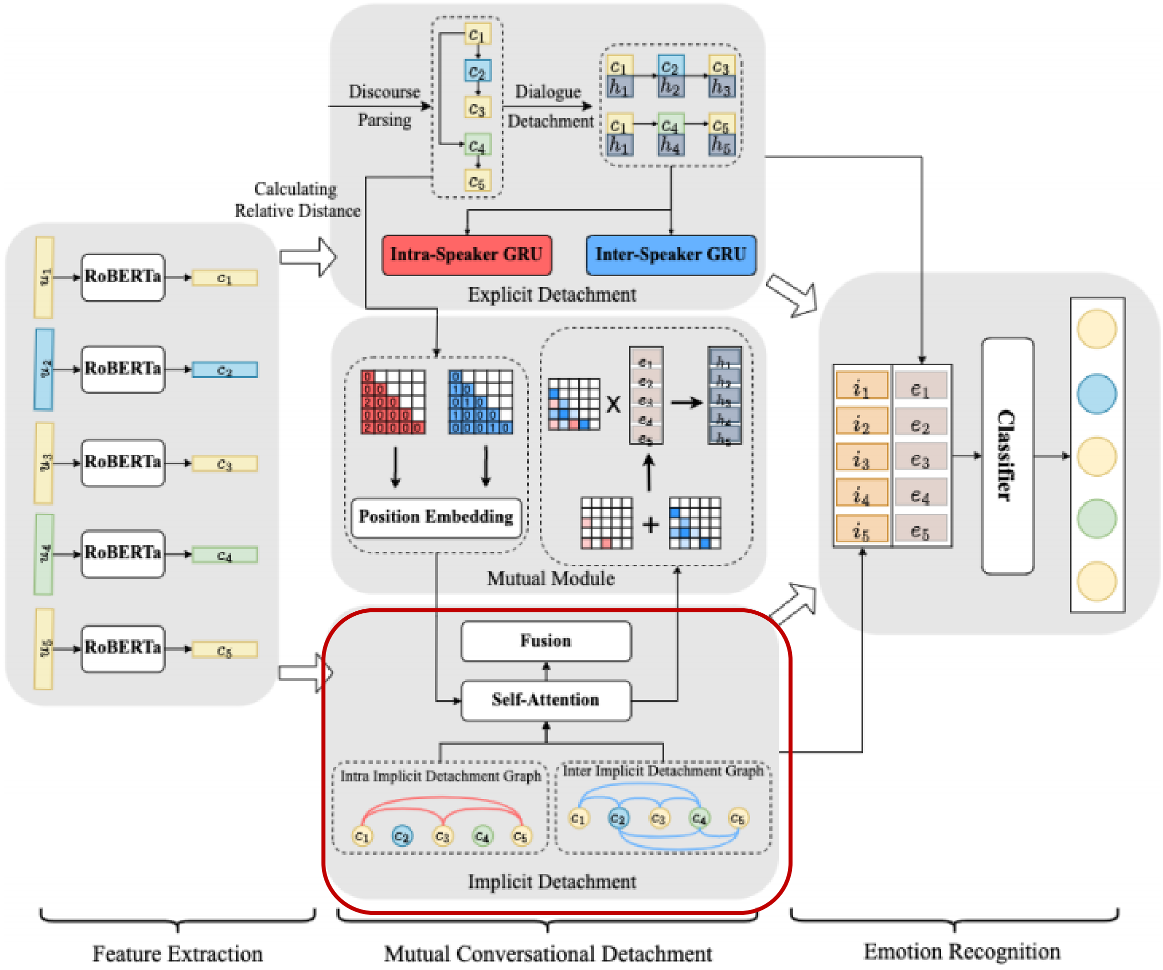
Explicit Detachment

$$\{(i, j, e_{ij}), \cdots\} = \text{Parser}(\{u_1, u_2, \cdots, u_N\}) \quad (2)$$

$$D_{i,j} = \begin{cases} 1, & \text{if } e_{ij} \text{ exists in discourse tree} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

$$e_i = \begin{cases} \text{GRU}^{intra}(c_i,\ e_p), & \text{if } \phi(u_i) = \phi(u_p) \\ \text{GRU}^{inter}(c_i,\ e_p), & \text{otherwise} \end{cases} \quad (4)$$

Method



Implicit Detachment

$$IDG_{i,j}^{intra} = \begin{cases} 0, & \text{if } j <= i \text{ and } \phi(u_i) = \phi(u_j) \\ -\infty, & \text{otherwise} \end{cases} \tag{5}$$

$$IDG_{i,j}^{inter} = \begin{cases} 0, & \text{if } j < i \text{ and } \phi(u_i) \neq \phi(u_j) \\ -\infty, & \text{otherwise} \end{cases} \tag{6}$$

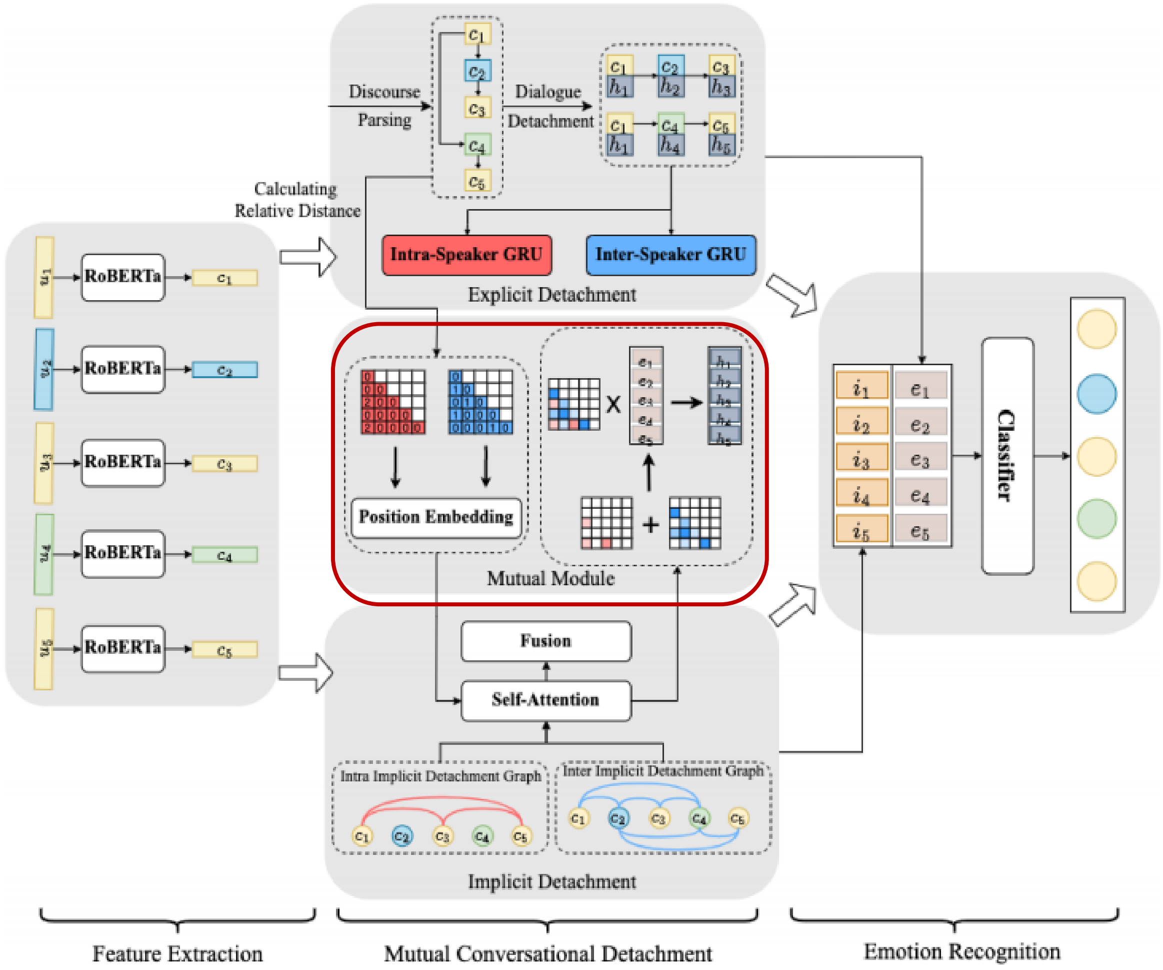
$$G = \mathrm{MHSA}(C,\ IDG^t),$$

$$\mathrm{Att}(Q, K, V, IDG^t) = \mathrm{Softmax}(\frac{QK^T}{\sqrt{d_k}} + IDG^t)V \tag{7}$$

$$\begin{aligned} F^t &= \mathrm{ReLU}(\mathrm{FC}([C, G^t, C - G^t, C \odot G^t])), \\ g &= \mathrm{Sigmoid}(\mathrm{FC}[F^{intra}, F^{inter}]), \\ I &= g \odot F^{intra} + (1 - g) \odot F^{inter} \end{aligned} \tag{8}$$

where $I \in R^{N \times d_h}$ and FC is the fully-connected layer.

Method



Mutual Module

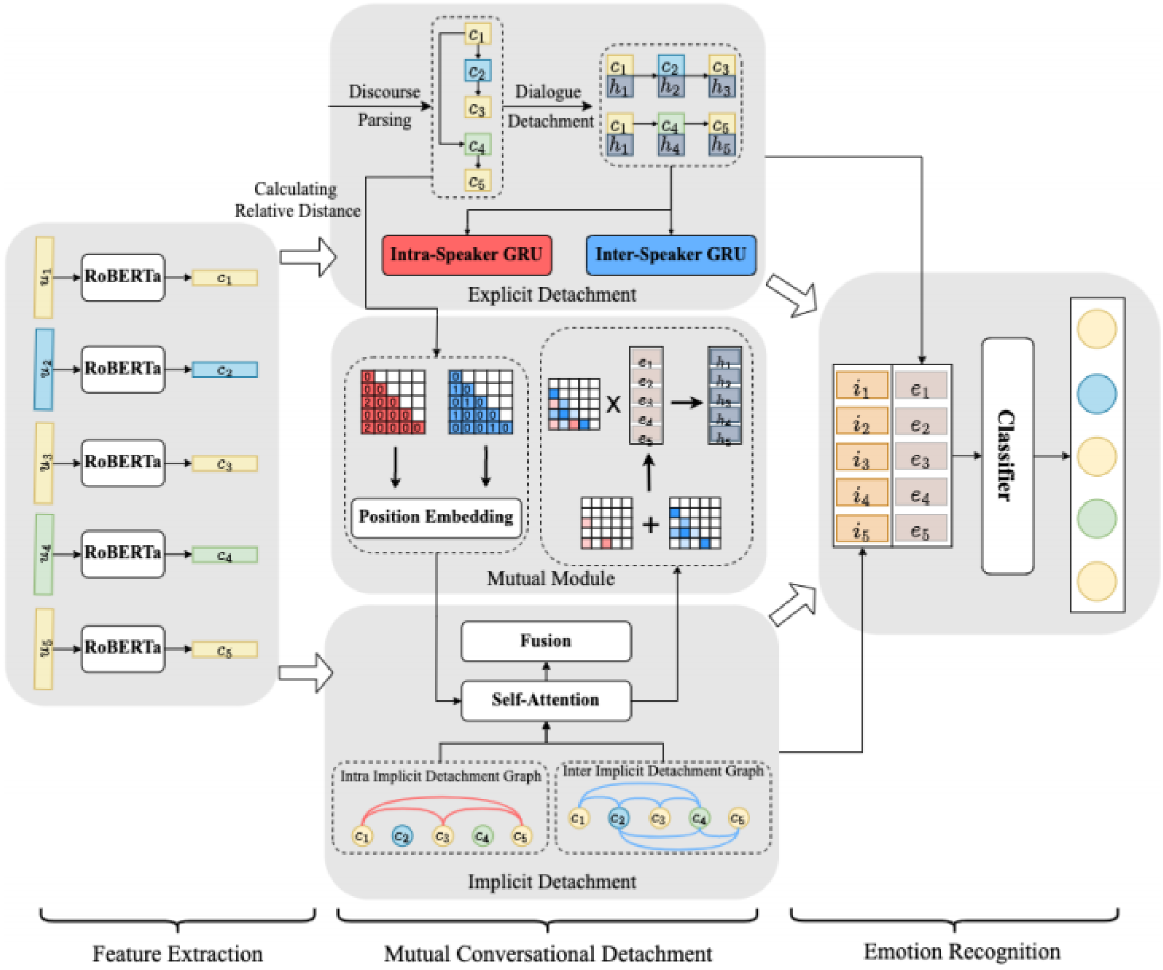
$$h_i = A^{joint}_{i,<i} \times E_{<i} \tag{9}$$

$$e_i = \begin{cases} \mathrm{GRU}^{intra}([c_i, h_i],\ e_p), & \text{if } \phi(u_i) = \phi(u_p) \\ \mathrm{GRU}^{inter}([c_i, h_i],\ e_p), & \text{otherwise} \end{cases} \tag{10}$$

$$\begin{aligned} Pos^t &= \mathrm{Embedding}(P^t), \\ G &= \mathrm{MHSA}(C,\ IDG^t, Pos^t), \end{aligned}$$

$$\mathrm{Att}(Q, K, V, IDG^t, Pos^t) = \mathrm{Softmax}(\frac{QK^T}{\sqrt{d_k}} + IDG^t + Pos^t)V \tag{11}$$

Method



Emotion Recognition

$$\hat{y} = \mathrm{Softmax}(W_e[C, E, I] + b_e) \qquad (12)$$

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{Emo} \hat{y}_i^{\,j} \cdot log(y_i^j) \qquad (13)$$

Experiment

Dataset	Dialogues			Utterances		
	Train	Val	Test	Train	Val	Test
EmoryNLP	713	99	85	9,934	1,344	1,328
MELD	1,039	114	280	9,989	1,109	2,610

Table 1: Dataset statistics

Experiment

Model	EmoryNLP	MELD
ERMC Methods		
ConGCN	-	57.40
DialogXL	34.73	62.41
ERMC-DisGCN	36.38	64.22
ERC Methods with CSK		
KET	34.39	58.18
KAITML	35.59	58.97
KI-Net	-	63.24
SKAIG	38.88	65.18
COSMIC	38.11	65.21
COSMIC w/o CSK	37.10	64.28
ERC Methods without CSK		
DialogueRNN	31.7	57.03
DialogueGCN	-	58.1
IEIN	-	60.72
RGAT	34.42	60.91
DialogueCRN	-	58.39
DAG-ERC	39.02	63.65
MuCDN (Ours)	**40.09**	**65.37**

Table 2: Comparison of our model against state-of-the-art baselines. CSK represents the commonsense knowledge utilized in COSMIC. Weighted F1 score is adopted as evaluation metrics.

Model	EmoryNLP	MELD
MuCDN	**40.09**	**65.37**
w/o explicit detachment	38.45	64.45
w/o implicit detachment	38.84	64.47
w/o E2I interaction	39.28	64.61
w/o I2E interaction	39.54	64.56

Table 3: Results of ablation study on the two ERMC datasets. E2I interaction is the relative position embedding provided by explicit detachment, while I2E interaction is the complementary global information from implicit detachment.

Experiment

Model	EmoryNLP	MELD
MuCDN	**40.09**	**65.37**
sequence	39.05	64.51
randomness	38.72	64.71

Table 4: Results of our model replaced with different types of dependency structure connecting utterances in Explicit Detachment module.

Model	EmoryNLP	MELD
MuCDN	**40.09**	**65.37**
w/o intra and inter GRU	39.42	64.49
w/o intra and inter graph	38.91	64.46

Table 5: Results of our model without speaker-specific modeling.

Thank you!