Chongqing University of Technology

ATAI
Advanced Technique of Artificial Intelligence

# Neutral Utterances are Also Causes_Enhancing Conversational Causal Emotion Entailment with Social Commonsense Knowledge


Jiangnan Li[1,2,*] , Fandong Meng[3] , Zheng Lin[1,2,†] , Rui Liu[1,2] , Peng Fu[1,†] , Yanan Cao[1,2] , Weiping Wang[1] and Jie Zhou[3]
1 Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
2School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
3Pattern Recognition Center, WeChat AI, Tencent Inc, China
{lijiangnan,linzheng,liurui1995,fupeng,wangweiping}@iie.ac.cn,
{fandongmeng,withtomzhou}@tencent.com


2022.5.18 • ChongQing

IJCAI 2022

Leibniz-Institut für Sozialwissenschaften

**Reported by Sijin Liu**

Artificial

1.Introduction

2.Method

3.Experiments

gesis
Leibniz-Institut
für Sozialwissenschaften

# Introduction

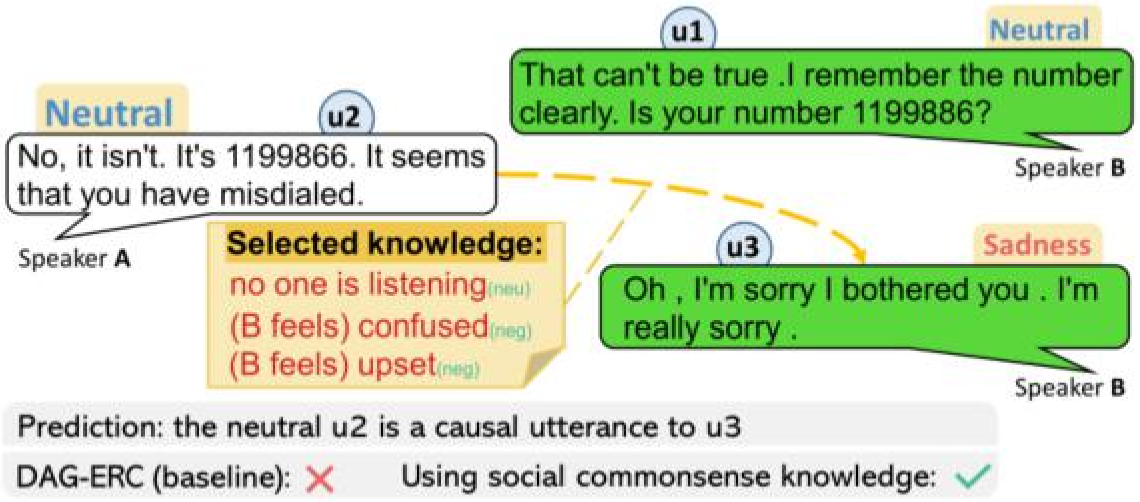



Figure 1: A case that the baseline DAG-ERC fails while the selected commonsense knowledge can help to make a right prediction.

Conversational Causal Emotion Entailment (C2E2) aims to detect causal utterances for a non-neutral targeted utterance from a conversation.

Causal utterances with different emotions, especially neutral ones (neutral causal utterances occupy 87% of this kind of causes), is still hard to detect even with emotion information. Models are limited in reasoning causal clues and passing them between utterances.

# Method

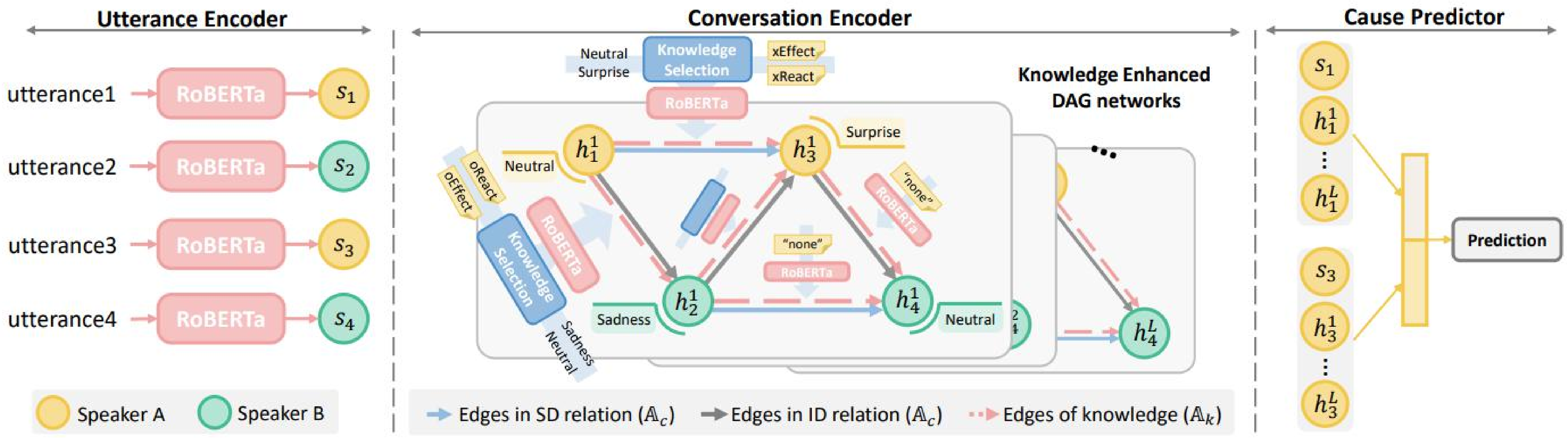
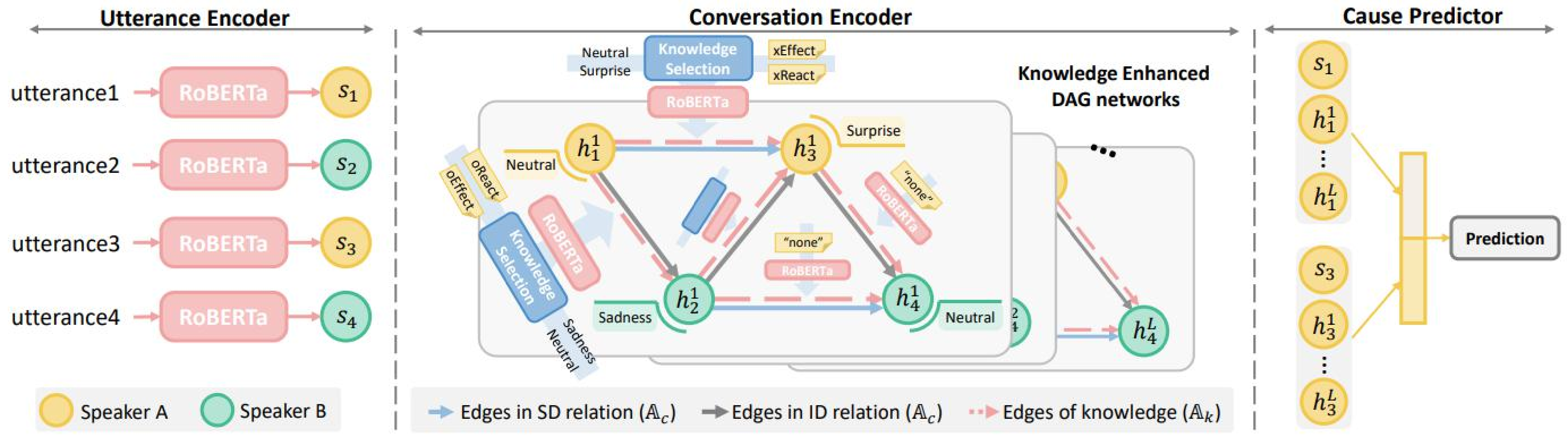



Figure 2: The structure of our model. It contains 3 modules: (1) Utterance Encoder encodes every utterance; (2) KEC graph is constructed from a conversation and knowledge attached in KEC graph is picked up by the knowledge selecting strategy. Conversation Encoder then uses Knowledge Enhanced DAG networks to process KEC graph; (3) Cause Predictor pairs every two utterances to make predictions.

# Method




## Task Definition

$$C = [u_1, ..., u_N] \;\; E = [e_1, ..., e_N]$$

$$P = [p_1, ..., p_N]$$

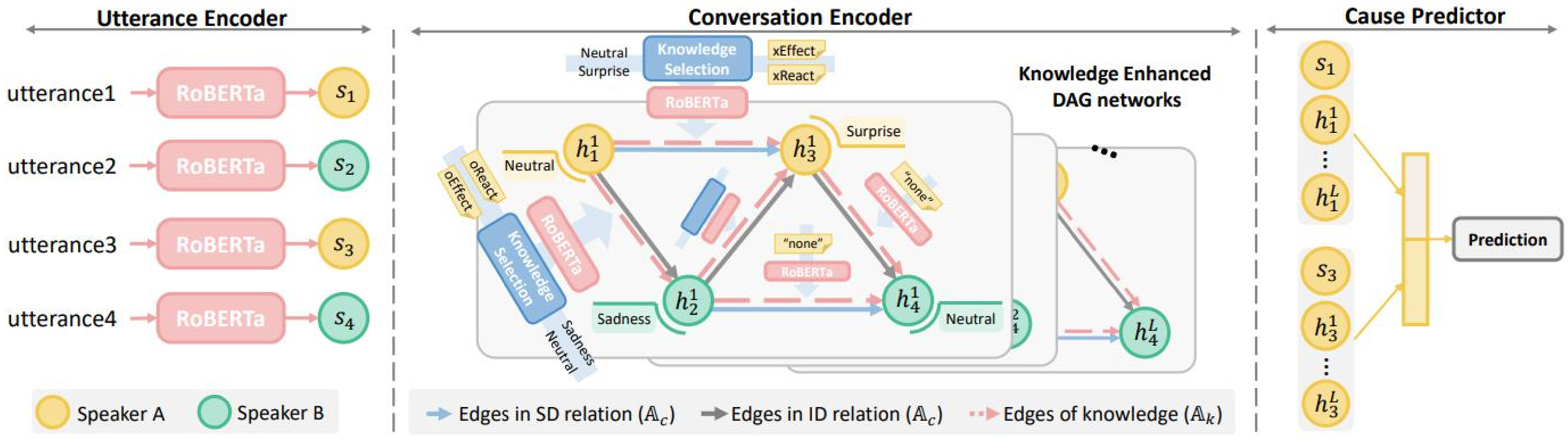
Every utterance $u_i$ is paired with its contextual utterances $u_j$ ($j \leq i$). If $u_i$ is a non-neutral utterance and $u_j$ is its causal utterance, the pair ($u_i$ , $u_j$ ) is labeled with 1. Otherwise, ($u_i$ , $u_j$ ) is labeled with 0.

## Utterance Encoder

$$s_i = \text{Linear}(\text{Maxpooling}(\text{RoBERTa}(u_i))), \qquad (1)$$

where $s_i \in \mathbb{R}^{d_u}$ and $d_u$ is the dimension of utterance representation.

# Method




## Knowledge Enhanced Conversation Graph

$$\mathcal{G} = (\mathbb{V}, \mathbb{A}_c, \mathbb{A}_k)$$

### Utterance nodes

$\mathbb{V}$ models all utterances as nodes $v_i$ in a conversation. $v_i$ contains the attribute rep to store the utterance representation (i.e., $v_i.\mathbf{rep} = s_i$).

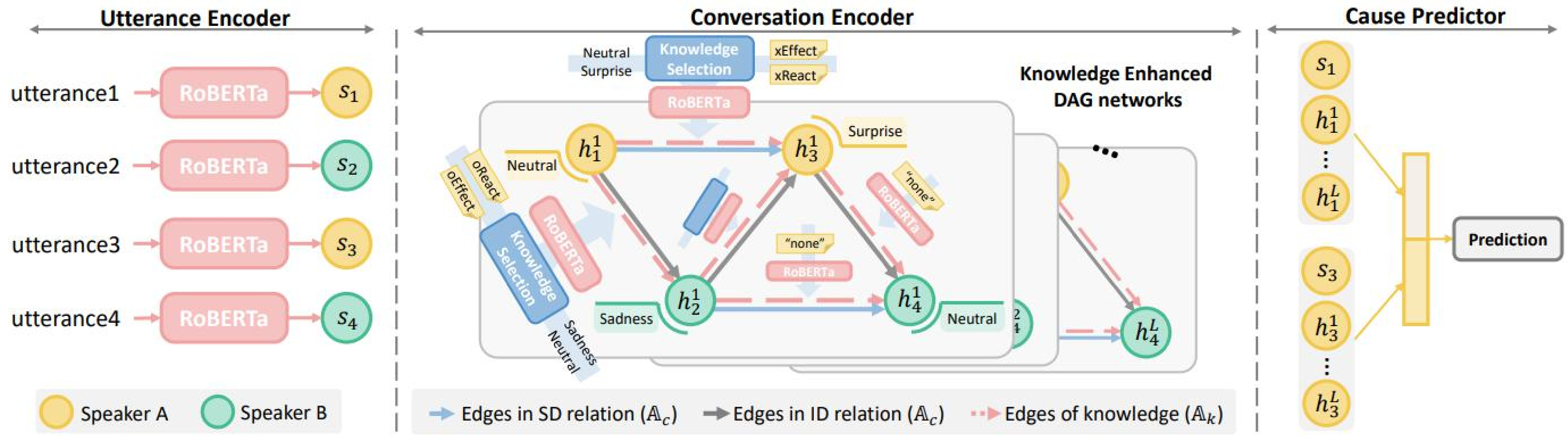
### Utterance interaction adjacency matrix

$\mathbb{A}_c$ contains two attributes: item and rel, where item stores 0 or 1 to denote the existence of an edge and rel stores the relation type of an edge.

### Knowledge passing adjacency matrix

$\mathbb{A}_k$ contains two attributes: item and klg, where item plays the same role as $\mathbb{A}_c$.item and klg stores the knowledge attached on an edge.

# Method




## Knowledge Enhanced Conversation Graph

*xEffect*, *xReact*, *oEffect*, and *oReact*

We denote the generation as CT($\mathbf{u}_i$ , xE), where CT and xE are the abbreviations of COMET and *xEffect* respectively.

The knowledge selection realizes two factors:

**Speaker realization:** If $p_i$ of the target $\mathbf{u}_i$ equals to $p_j$ of the source $\mathbf{u}_j$ , knowledge of *xEffect* and *xReact* is chosen, otherwise knowledge of *oEffect* and *oReact* picked.

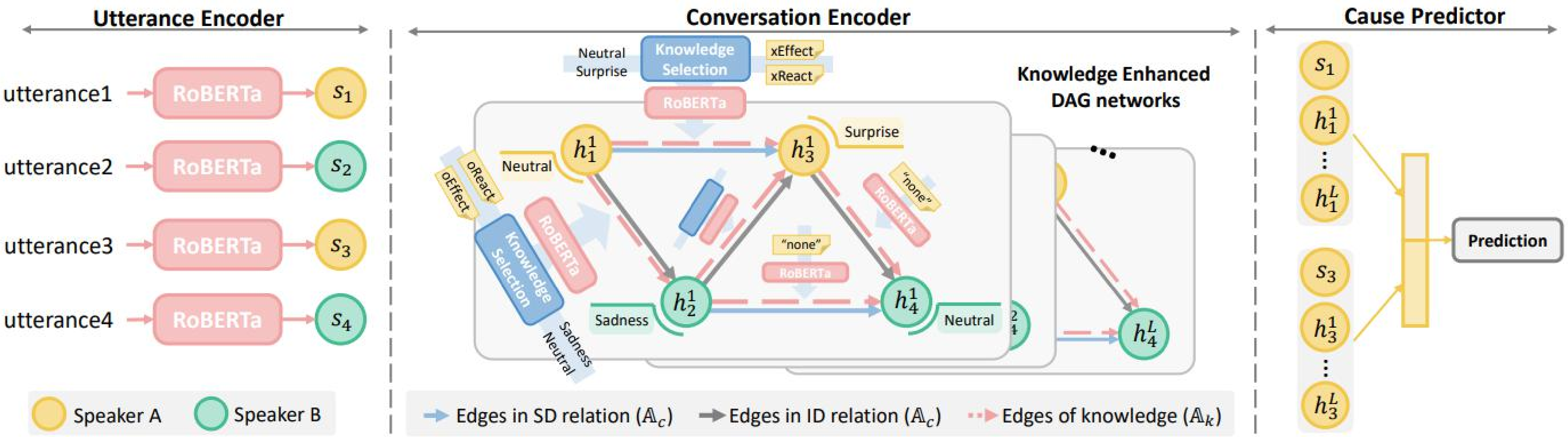
## Sentiment realization:

$$r_s = \begin{cases} r_{pos} - r_{neg}, & |r_{pos} - r_{neg}| > r_{neu} \\ 0, & \text{else} \end{cases} \qquad (2)$$

$$\text{posK}_\text{x}, \text{negK}_\text{x}, \text{neuK}_\text{x} = \textit{split}(\texttt{CT}(u_j, \texttt{xE/xR}))$$
$$\text{posK}_\text{o}, \text{negK}_\text{o}, \text{neuK}_\text{o} = \textit{split}(\texttt{CT}(u_j, \texttt{oE/oR}))$$

**if** $e2s(e_i) \neq \text{neu}$ **then**
$\quad \mathbb{A}_\text{k}[i, j].\text{item} = 1$
$\quad$ **if** $e2s(e_j) == \text{neu}$ **then**
$\quad\quad \mathbb{A}_\text{k}[i, j].\text{klg} = \text{K}[\text{neu}] + [\text{sep}] + \text{K}[e2s(e_i)]$
$\quad$ **else**
$\quad\quad \mathbb{A}_\text{k}[i, j].\text{klg} = \text{K}[e2s(e_i)]$

# Method




$$h_i^0 = \text{Linear}([s_i || eemb_{e_i}])$$

## Knowledge Enhanced DAG Networks

$$\alpha_{i,j} = \text{Softmax}_{j \in \mathcal{N}_i}(\mathrm{W}_{\mathrm{w}}^l [h_i^{l-1} || (h_j^l + \mathrm{W}_{\mathrm{k}}^l k_{i,j})]), \quad (3)$$

$$msg_i = \sum_{j \in \mathcal{N}_i} \alpha_{i,j} \mathrm{W}^l_{\mathbb{A}_{c[i,j].\mathtt{rel}}} h_j^l, \quad (4)$$

$$nlg_i = \sum_{j \in \mathcal{N}_i} \alpha_{i,j} \mathrm{W}_{\mathrm{k}}^l k_{i,j}, \quad (5)$$

$$nod_i = \text{GRU}_n(h_i^{l-1}, msg_i), \quad (6)$$

$$cxt_i = \text{GRU}_c(msg_i, h_i^{l-1}). \quad (7)$$

$$ckg_i = \text{GRU}_k(nlg_i, h_i^{l-1}), \quad (8)$$

$$skg_i = \text{GRU}_s(k_{i,i}, h_i^{l-1}). \quad (9)$$

Finally, the node representation of utterance $u_i$ in the $l^{th}$ layer is updated by summing the four types of information, i.e. $h_i^l = nod_i + cxt_i + ckg_i + skg_i$.

After the encoding of two-level encoders, the final utterance representation is computed by $h_i = ||_{l=0}^{L} h_i^l$ [Shen *et al.*, 2021b]. Whether $u_j$ is the cause of $u_i$ is then computed by:

$$p_{i,j} = sigmoid(\text{MLP}([h_i || h_j])), \quad (10)$$

# Experiments

| | | Train | Dev | Test |
|---|---|---|---|---|
| Num. of Causal Pair | Positive | 7027 | 328 | 1767 |
| | Negative | 45392 | 2842 | 14052 |
| Num. of Dialgoue | | 834 | 47 | 225 |
| Num. of Utterance | | 8206 | 493 | 2405 |
| Avg. Len. of Utterance | | 14 | 16 | 15 |

Table 1: Statistics of RECCON-DD. "Positive" means the true causal pair.

| | Model | Neg. F1 | Pos. F1 | macro F1 |
|---|---|---|---|---|
| 1 | ECPE-2D$^{\triangle}$ | 94.96 | 55.50 | 75.23 |
| | ECPE-MLL$^{\triangle}$ | 94.68 | 48.48 | 71.59 |
| | RankCP$^{\triangle}$ | 97.30 | 33.00 | 65.15 |
| 2 | KAG | $94.49_{(0.22)}$ | $55.52_{(2.39)}$ | $75.02_{(1.13)}$ |
| | Adapted | $95.67_{(0.24)}$ | $62.47_{(4.72)}$ | $79.07_{(2.44)}$ |
| | SKAIG | $95.26_{(0.12)}$ | $63.15_{(1.00)}$ | $79.21_{(0.49)}$ |
| 3 | Entail | $94.83_{(0.47)}$ | $58.59_{(3.78)}$ | $76.66_{(1.66)}$ |
| | DAG-ERC | $95.33_{(0.25)}$ | $63.56_{(2.10)}$ | $79.44_{(1.16)}$ |
| 4 | KEC (ours) | $95.74_{(0.05)}$ | $\mathbf{66.76}^{*}_{(0.33)}$ | $\mathbf{81.25}^{*}_{(0.17)}$ |

Table 2: Results of all models on RECCON-DD. $^{*}$ denotes that our method is significant against the best baseline DAG-ERC (p-value<0.05) with the paired T-test. $^{\triangle}$ denotes the results referred from Poria et al. [2021].

# Experiments

| Method | Pos. F1 | macro F1 |
|---|---|---|
| KEC | **66.76**$_{(0.33)}$ | **81.25**$_{(0.17)}$ |
| – *contextual knowledge unit* | 64.93$_{(0.69)}$ −1.83$_{(\downarrow)}$ | 80.25$_{(0.37)}$ −1.00$_{(\downarrow)}$ |
| – *self-loop knowledge unit* | 65.00$_{(0.92)}$ −1.76$_{(\downarrow)}$ | 80.21$_{(0.54)}$ −1.04$_{(\downarrow)}$ |
| – neutral knowledege | 65.80$_{(0.96)}$ −0.96$_{(\downarrow)}$ | 80.67$_{(0.52)}$ −0.58$_{(\downarrow)}$ |
| – emotion embs | 63.19$_{(0.25)}$ −3.56$_{(\downarrow)}$ | 79.35$_{(0.17)}$ −1.90$_{(\downarrow)}$ |
| – emoition embs & CSK | 46.66$_{(0.76)}$ −20.1$_{(\downarrow)}$ | 69.90$_{(0.48)}$ −11.4$_{(\downarrow)}$ |

Table 3: Ablation Study

# Experiments

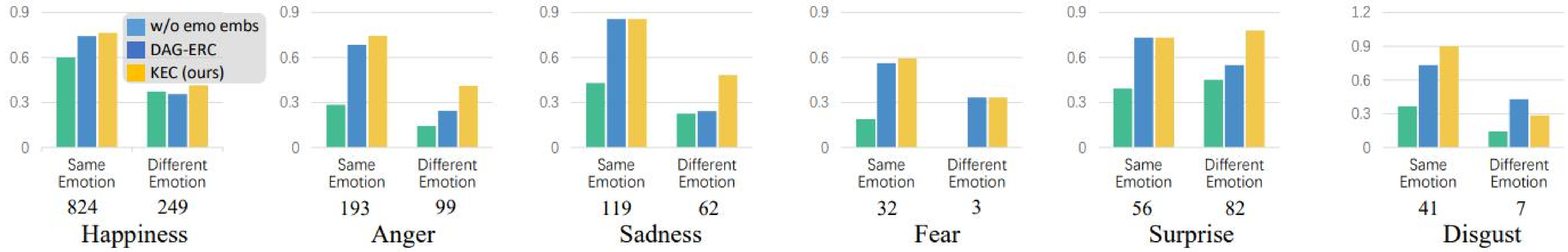



Figure 3: "Same Emotion" reports the recall of causal pairs whose causal utterances are with the same emotion as the targeted utterance. "Different Emotion" refers to different emotions from the targeted utterance. The number of a type of pairs is presented below the x-axis. Green bars denote DAG-ERC without emotion embeddings, blue bars for DAG-ERC with emotion embeddings, and orange bars for KEC.

# Experiments

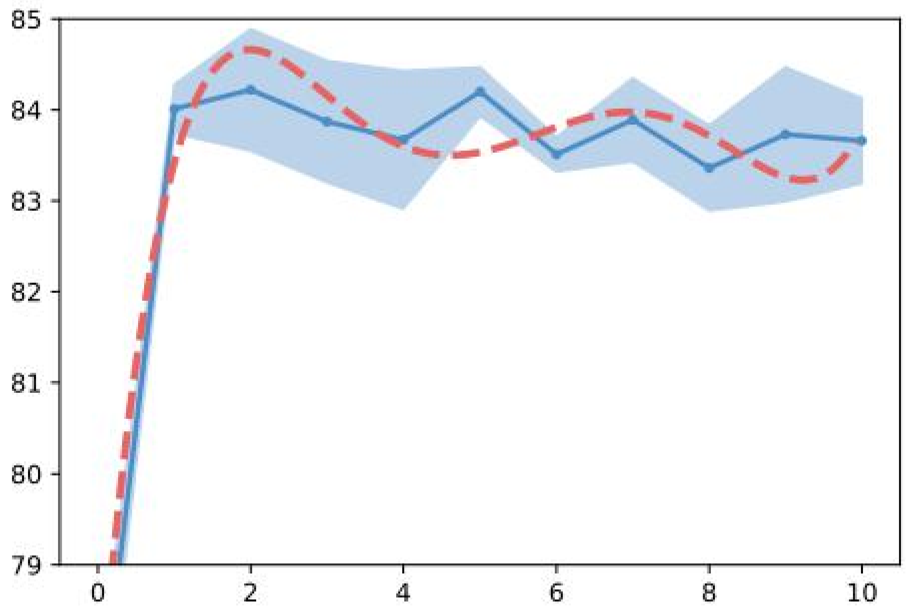



Figure 4: The effect of the window size. The y-axis denotes the macro F1 on the development set. To better fit the trend, we draw the trend line of fifth-order polynomial as the red dotted line.