Empathetic Dialogue Generation via Sensitive Emotion Recognition and Sensible Knowledge Selection

Lanrui Wang[1,2], **Jiangnan Li**[1,2], **Zheng Lin**[1,2]*, **Fandong Meng**[3], **Chenxu Yang**[1,2], **Weiping Wang**[1], **Jie Zhou**[3]
[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3]Pattern Recognition Center, WeChat AI, Tencent Inc, China
{wanglanrui,lijiangnan,linzheng,yangchenxu,wangweiping}@iie.ac.cn
{fandongmeng,withtomzhou}@tencent.com

code: https://github.com/wlr737/EMNLP2022-SEEK

2022.11.26 • ChongQing

2022_EMNLP

Reported by Junhao Cao

Leibniz-Institut für Sozialwissenschaften

Method

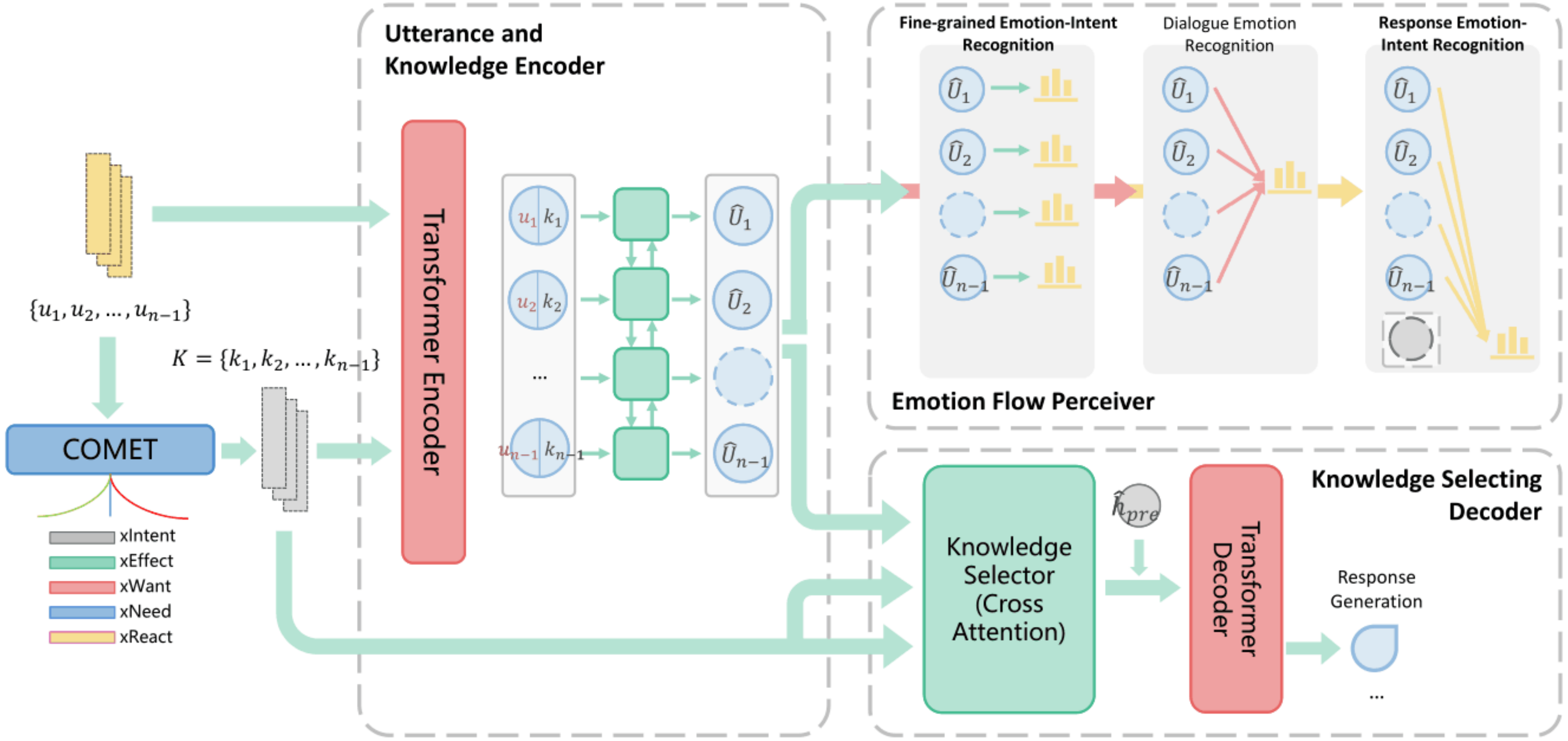


Figure 2: An overall architecture of our proposed model.

Task Formulation

$$C = [C_1, ..., C_{N-1}]$$

$$\boldsymbol{EI} = [\boldsymbol{ei}_1, ..., \boldsymbol{ei}_{N-1}, \boldsymbol{ei}_Y]$$

Utterance and Knowledge Encoder

Utterance Encoding

$$C_i = [w_{CLS}, w_1, w_2, ..., w_{L_i}]$$

$$\boldsymbol{H}_{U_i} = \mathbf{TRS}_{Enc}(EMB_{C_i}), \quad (1)$$

$$\boldsymbol{U}_i = \boldsymbol{H}_{U_i}[0]. \quad (2)$$

Knowledge Encoding

$$\boldsymbol{H}_{K_i} = \mathbf{TRS}_{Enc}(\mathcal{K}_i)$$
$$\boldsymbol{K}_i = \mathbf{Mean}(\boldsymbol{H}_{K_i}) \quad (3)$$

CEM: Commonsense-aware Empathetic Response Generation

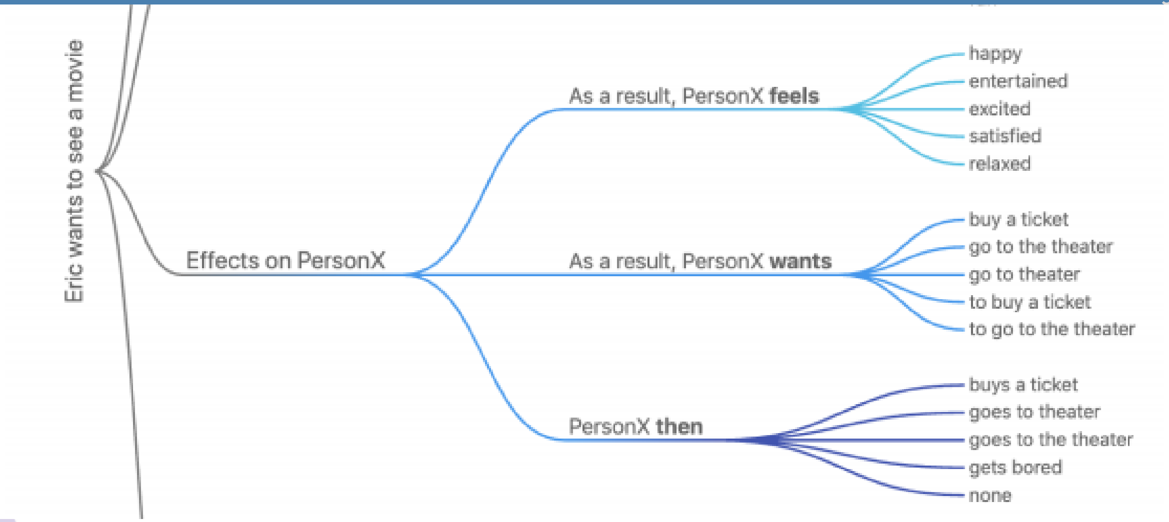
Sahand Sabour, Chujie Zheng, Minlie Huang
The CoAI Group, DCST, Institute for Artificial Intelligence,
State Key Lab of Intelligent Technology and Systems,
Beijing National Research Center for Information Science and Technology,
Tsinghua University, Beijing 100084, China
{sahandfer, chujiezhengchn}@gmail.com, aihuang@tsinghua.edu.cn

——AAAI2022

Code and dataset are available at https://github.com/Sahandfer/CEM

Reported by Yabo Yin

Method



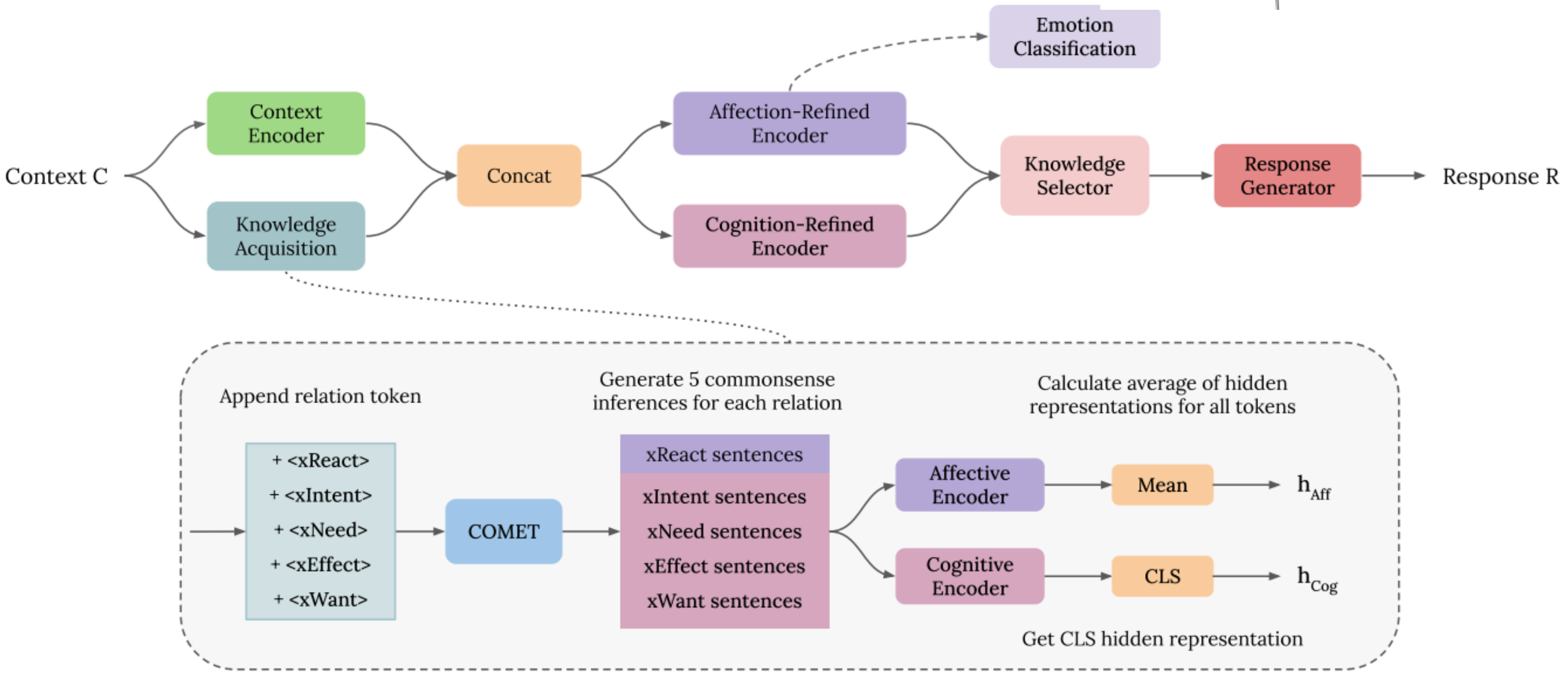


Figure 2: Overview of our model (CEM).

Method

$$D = [u_1, u_2, u_3, ..., u_{k-1}]$$

$$u_i = [w_1^i, w_2^i, w_3^i, ..., w_{M_i}^i]$$

Context Encoding

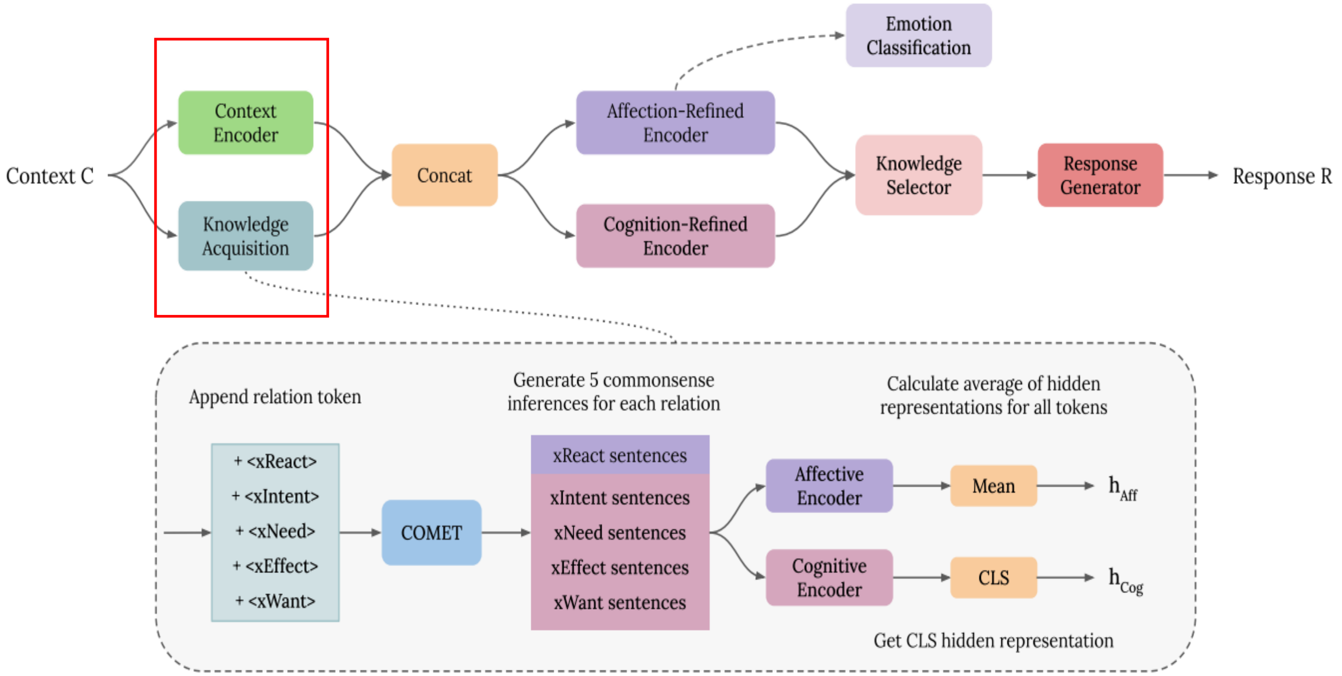
$$C = [\mathrm{CLS}] \oplus u_1 \oplus u_2 \oplus u_3 \oplus ... \oplus u_{k-1}$$

$$\boldsymbol{H}_{CTX} = \mathbf{Enc}_{CTX}(\boldsymbol{E}_C) \quad \boldsymbol{H}_{CTX} \in \mathbb{R}^{L \times d}$$

Knowledge Acquisition

use COMET to generate five commonsense inferences $[cs_1^r, cs_2^r, ..., cs_5^r]$ per relation r.

$$CS_r = cs_1^r \oplus cs_2^r \oplus ... \oplus cs_5^r.$$



$$\boldsymbol{H}_{xReact} = \mathbf{Enc}_{Aff}(\boldsymbol{E}_{CS_{xReact}}) \tag{2}$$

$$\boldsymbol{H}_r = \mathbf{Enc}_{Cog}(\boldsymbol{E}_{CS_r}) \tag{3}$$

$$\boldsymbol{H}_{xReact} \in \mathbb{R}^{l_{xReact} \times d}, \boldsymbol{H}_r \in \mathbb{R}^{l_r \times d},$$

$$r \in \{xWant, xNeed, xIntent, xEffect\}.$$

$$\boldsymbol{h}_{xReact} = \mathrm{Average}(\boldsymbol{H}_{xReact}) \tag{4}$$

$$\boldsymbol{h}_r = \boldsymbol{H}_r[0] \tag{5}$$

$$h_{xReact}, h_r \in \mathbb{R}^d.$$

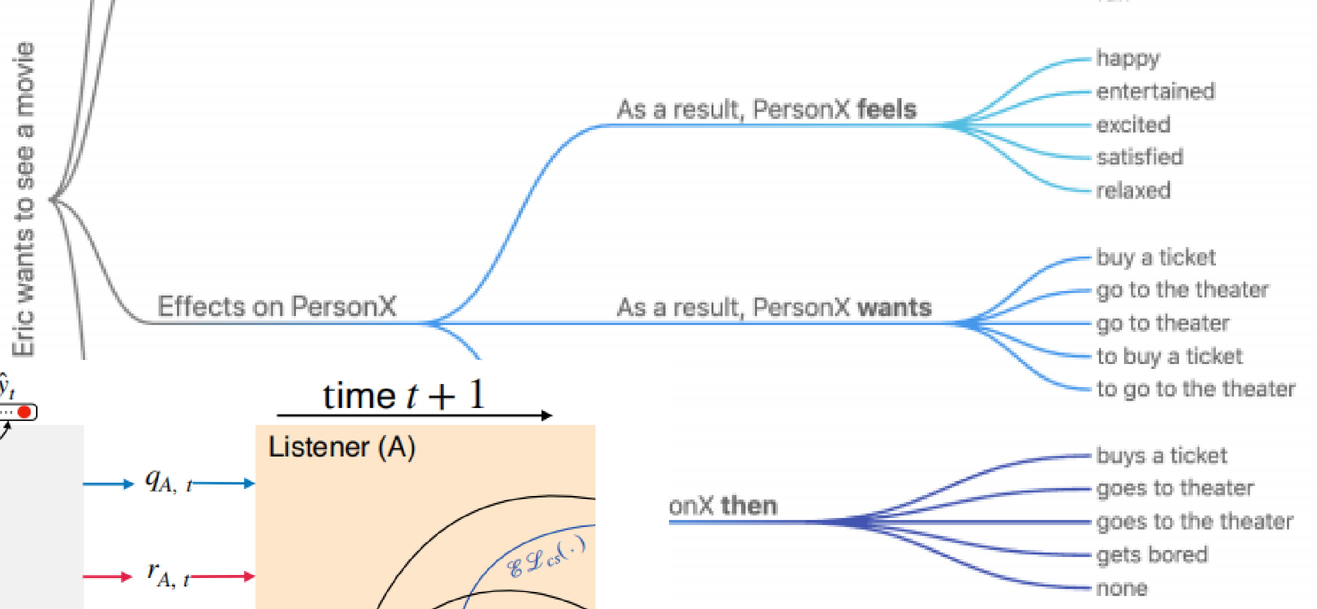
COSMIC: COmmonSense knowledge for eMotion Identification in Conversations

Deepanway Ghosal† , Navonil Majumder† , Alexander Gelbukh,
Rada Mihalcea4, Soujanya Poria†
† Singapore University of Technology and Design, Singapore
CIC, Instituto Polit
´
ecnico Nacional, Mexico
4 University of Michigan, USA
{deepanway ghosal@mymail.,navonil majumder@, sporia@}sutd.edu.sg,
gelbukh@cic.ipn.mx, mihalcea@umich.edu

——EMNLP2020

gesis
Leibniz-Institut
für Sozialwissenschaften

Method



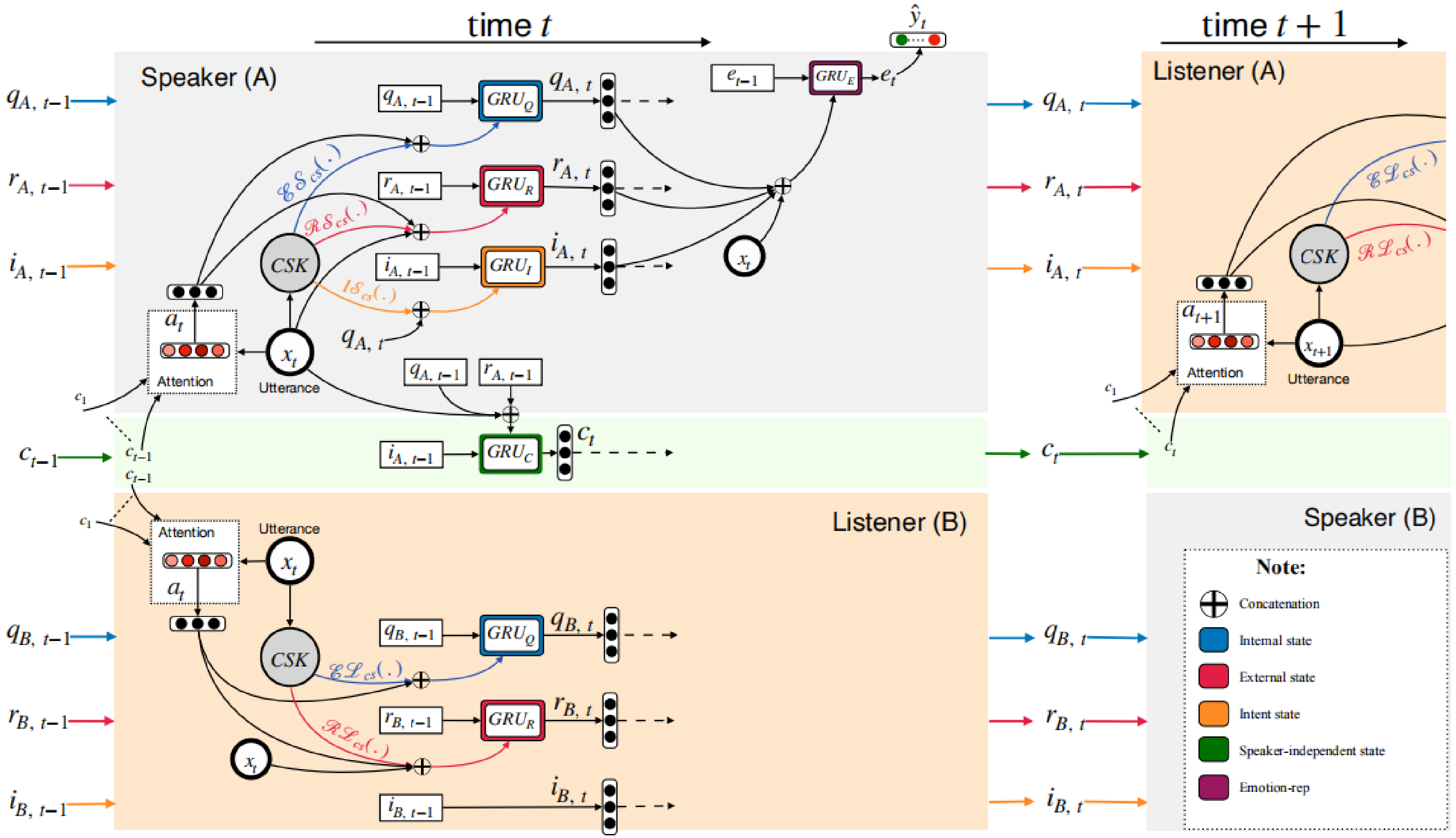


Figure 2: Illustration of COSMIC framework. *CSK*: Commonsense knowledge from COMET. In practice we use Bidirectional GRU cells. However, for clarity unidirectional cells are shown in the sketch.

Method

we take the pretrained COMET model on ATOMIC knowledge graph and discard the phrase generating decoder module. We treat utterance U as the subject

核心:为一个句子和对应的5种关系通过Comet的Encoder来获取5个不同的向量表示.并拼接到句子向量上

$$q_{s(u_t),t} = GRU_Q(q_{s(u_t),t-1}, (a_t \oplus \mathcal{ES}_{cs}(u_t))) \tag{3}$$

$$q_{j,t} = GRU_Q(q_{j,t-1}, (a_t \oplus \mathcal{EL}_{cs}(u_t))); \forall j \neq s(u_t) \tag{4}$$

$$r_{s(u_t),t} = GRU_R(r_{s(u_t),t-1}, (a_t \oplus x_t \oplus \mathcal{RS}_{cs}(u_t))) \tag{5}$$

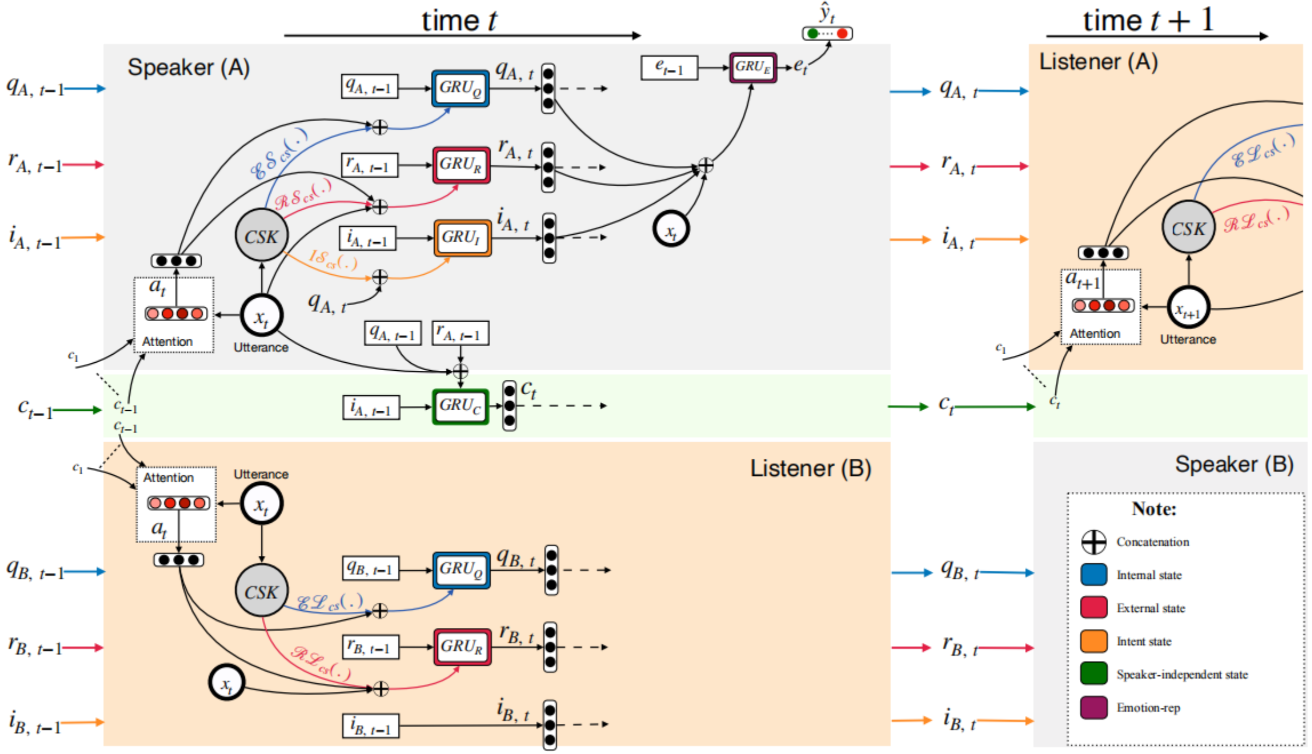


Figure 2: Illustration of COSMIC framework. *CSK*: Commonsense knowledge from COMET. In practice we use Bidirectional GRU cells. However, for clarity unidirectional cells are shown in the sketch.

$$r_{j,t} = GRU_R(r_{j,t-1}, (a_t \oplus x_t \oplus \mathcal{RL}_{cs}(u_t))); \forall j \neq s(u_t) \tag{6}$$

$$i_{s(u_t),t} = GRU_I(i_{s(u_t),t-1}, (\mathcal{IS}_{cs}(u_t) \oplus q_{s(u_t),t})) \tag{7}$$

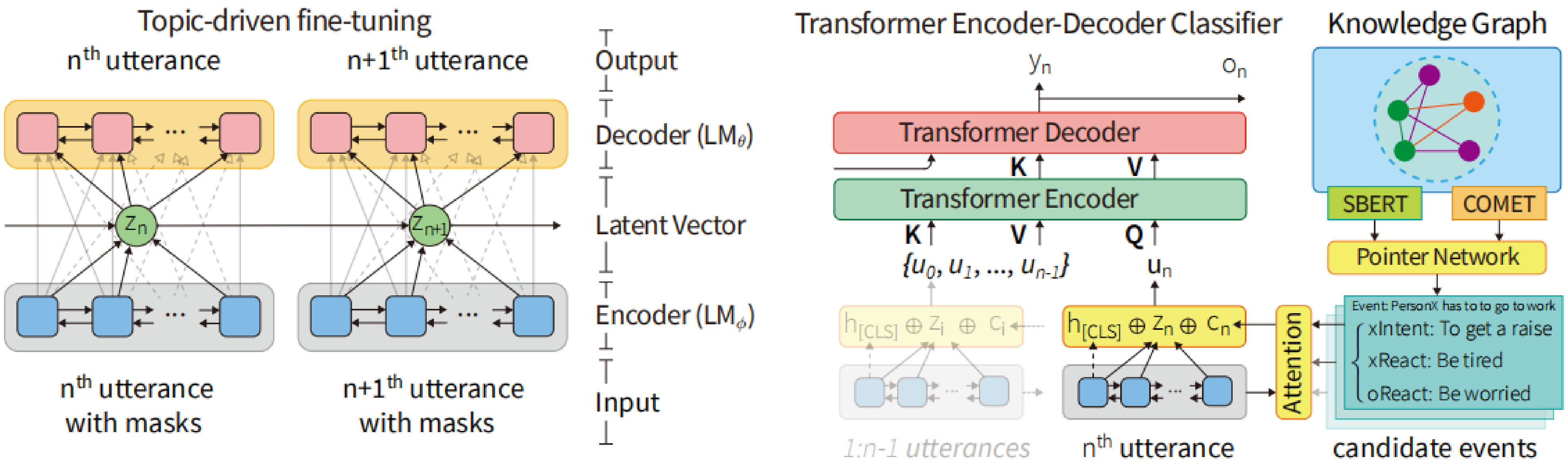
Topic-Driven and Knowledge-Aware Transformer for Dialogue Emotion Detection

Lixing Zhu†, Gabriele Pergola†, Lin Gui†, Deyu Zhou § , Yulan He
† †Department of Computer Science, University of Warwick, UK
§ School of Computer Science and Engineering, Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, China
{lixing.zhu,gabriele.pergola,lin.gui,yulan.he}@warwick.ac.uk
d.zhou@seu.edu.cn

——ACL2021

gesis
Leibniz-Institut für Sozialwissenschaften

Method



(a) Topic-driven fine-tuning of a pre-trained LM.

(b) Knowledge-aware transformer.

Figure 2: TOpic-Driven and Knowledge-Aware Transformer (TODKAT).

Method

tuned language model. We use COMET to generate the K most likely events, each with respect to the three event relation types. The produced events are denoted as $\{g_{n,k}^{sI}, g_{n,k}^{sR}, g_{n,k}^{oR}\}, k = 1, \ldots, K$.

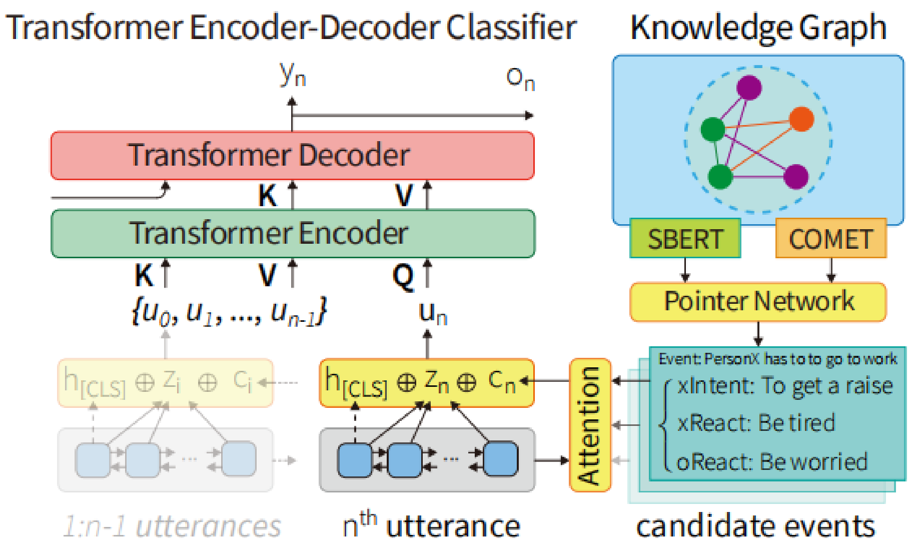
核心:从Comet生成的句子要与原始的Utterance 进行相关性计算充当注意力权重,从而聚合出一个Cn,再将二者拼接起来,进行encode

$$v_k = \tanh([\boldsymbol{c}_{n,k}, z_{n,k}]\mathbf{W}_\alpha), \quad (9)$$

$$\alpha_k = \frac{\exp(v_k[z_n, u_n]^\top)}{\sum_k \exp(v_k[z_n, u_n]^\top)}, \quad (10)$$

$$\boldsymbol{c}_n = \sum_{k=1}^{K} \alpha_k \boldsymbol{c}_{n,k}. \quad (11)$$

Here, we abuse $\boldsymbol{c}_n$ to represent the aggregated knowledge phrases. We further aggregate $\boldsymbol{c}_n$ by event relation types using a self-attention and the final event representation is denoted as c_n.



(b) Knowledge-aware transformer.

edge-Aware Transformer (TODKAT).

Past, Present, and Future: Conversational Emotion Recognition through Structural Modeling of Psychological Knowledge

Jiangnan Li1,2, Zheng Lin1,2∗, Peng Fu1, Weiping Wang1

1Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
2School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
{lijiangnan,linzheng,fupeng,wangweiping}@iie.ac.cn

——EMNLP2021

gesis
Leibniz-Institut für Sozialwissenschaften

Method

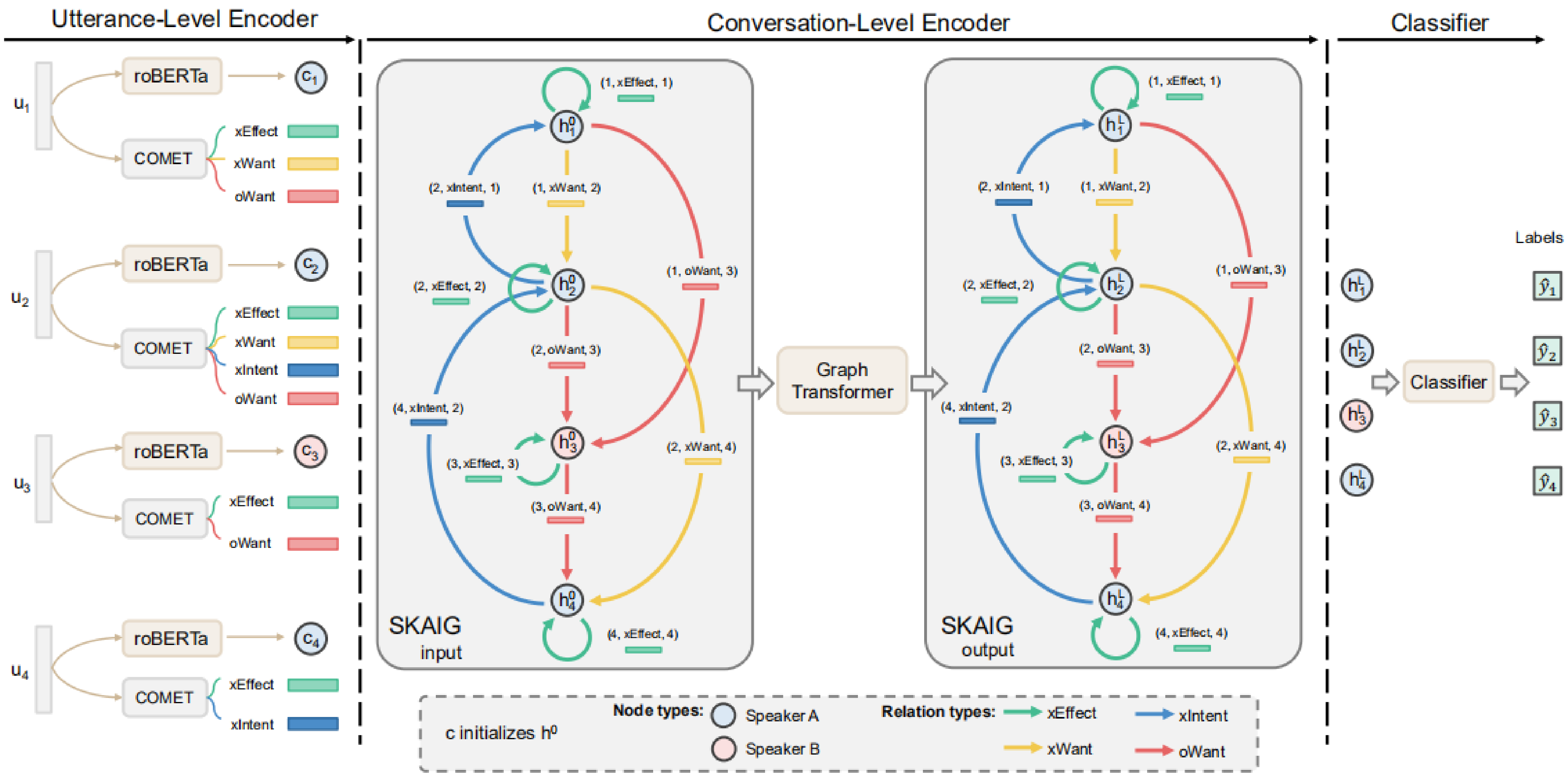
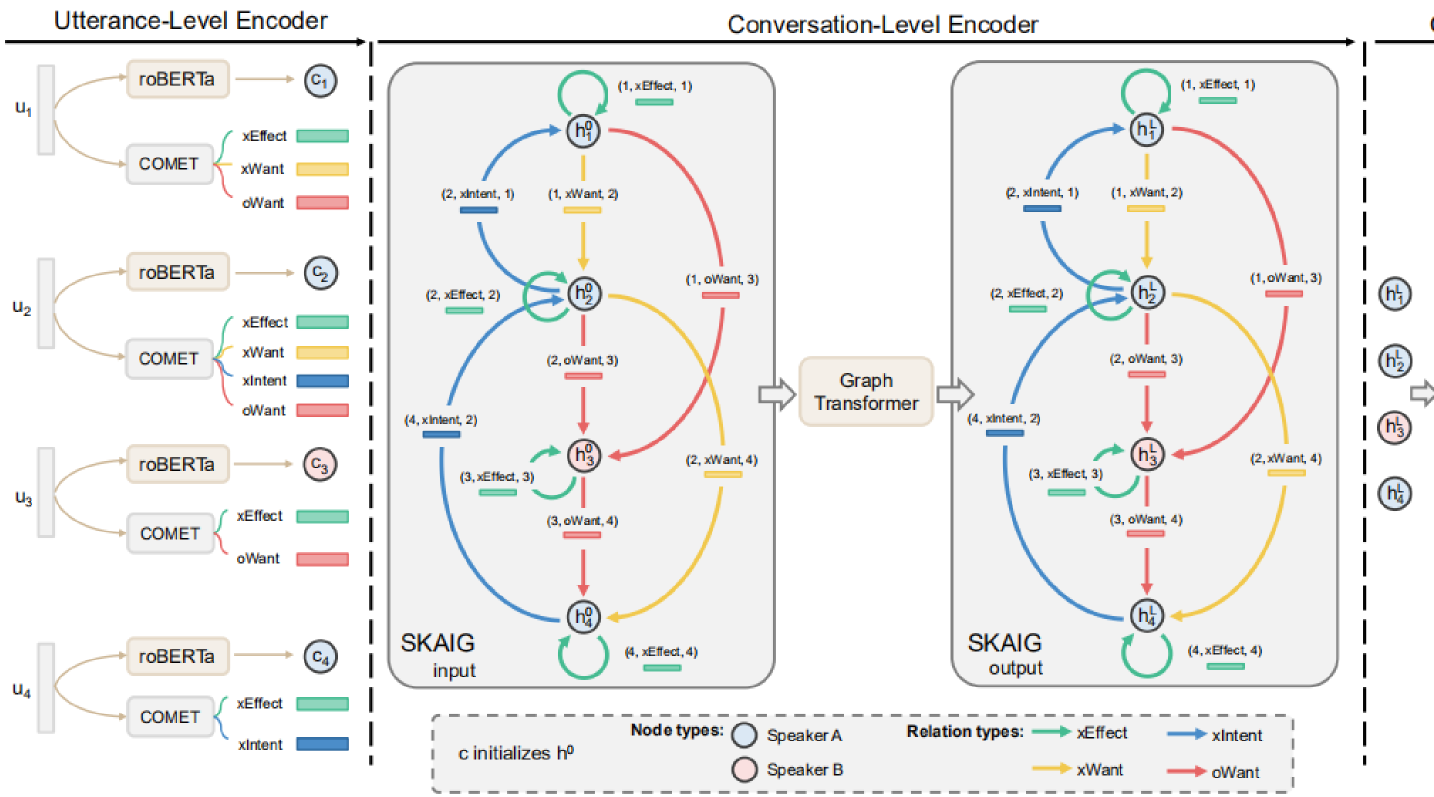


Figure 2: The framework of our model. The utterances are encoded by the utterance-level encoder to produce the utterance representations and the edge representations. The conversation-level encoder processes the SKAIG whose window size is 1. Finally, the classifier predicts the emotion for every utterance. Especially, edges and their representations with different relations are in different colors.

Method

核心: 自己自定义节点之间的关系,关系的表征使用Comet进行学习关系的representation

We concatenate u_n and a relation with mask tokens (e.g. u_n [MASK] <xWant>) in the inputting format of COMET, and then COMET processes the input. Following Ghosal et al. (2020), we take the hidden state of the relation token from the last layer of COMET transformer encoder as the relation's representation. For an edge



1,不同speaker之间的utterances关系被定义为Owant
2,speaker的自环被定义为Oeffect
3,同一speaker的不同utterances,当前一个utterance指向后一个utterance时 被定义为Xwant
4,同一speaker的不同utterances,后一个utterance指向当前一个utterance时 被定义为Xintent

coder as the relation's representation. For an edge $e_{i,j} = (u_i, \texttt{xWant}, u_j)$, the corresponding representation is $a_{i,j}$, whose dimension is mapped from 768 to d_u with a following linear unit.

We update the node representation $h_i^{(l)} \in \mathbb{R}^{d_u}$ of each node $u_i \in \mathcal{V}$ by:

$$h_i^{(l+1)} = (1-\beta_i)\left(\sum_{j\in\mathcal{N}(i)} \alpha_{i,j} m_j\right) + \beta_i \mathrm{W}_s h_i^{(l)} \tag{3}$$

Eq3利用节点特征和边的特征进行聚合操作,丰富节点的特征表示