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

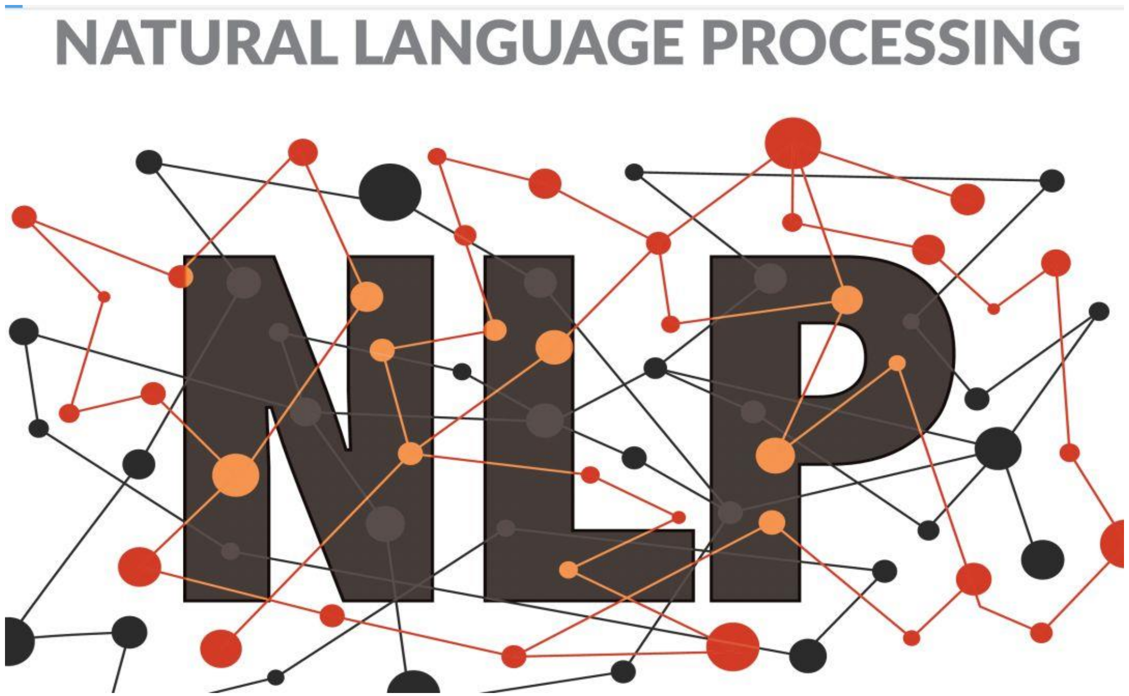
code: https://github.com/wlr737/EMNLP2022-SEEK

2022.11.27 • ChongQing

2022_EMNLP

Leibniz-Institut für Sozialwissenschaften

Reported by Junhao Cao



1.Introduction

2.Method

3.Experiments

Introduction

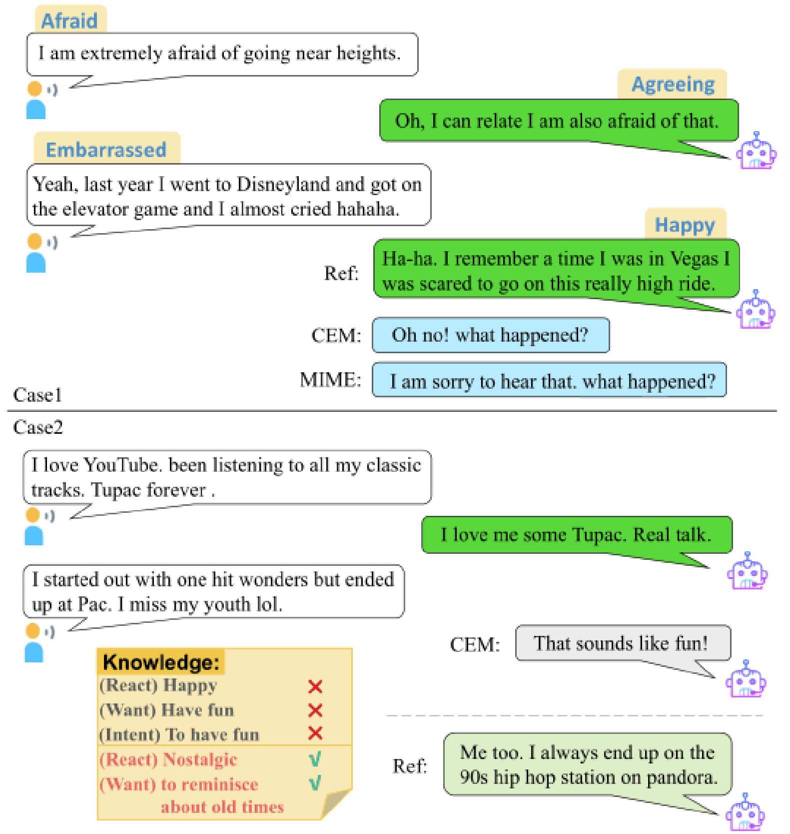


Figure 1: Two cases of multi-turn Empathetic Dialogues. The first case shows the speaker's emotion went from fear at the beginning of the conversation to an embarrassed self-deprecation, ending with a happy mood. And the second case shows that CEM chooses the wrong knowledge leading to inappropriate response.

current approaches to empathetic response generation focus on capturing implicit emotion within dialogue context,where the emotions are treated as a static variable throughout the conversations.

simply importing commonsense knowledge without harmonization may trigger the conflicts between knowledge and emotion.

Method

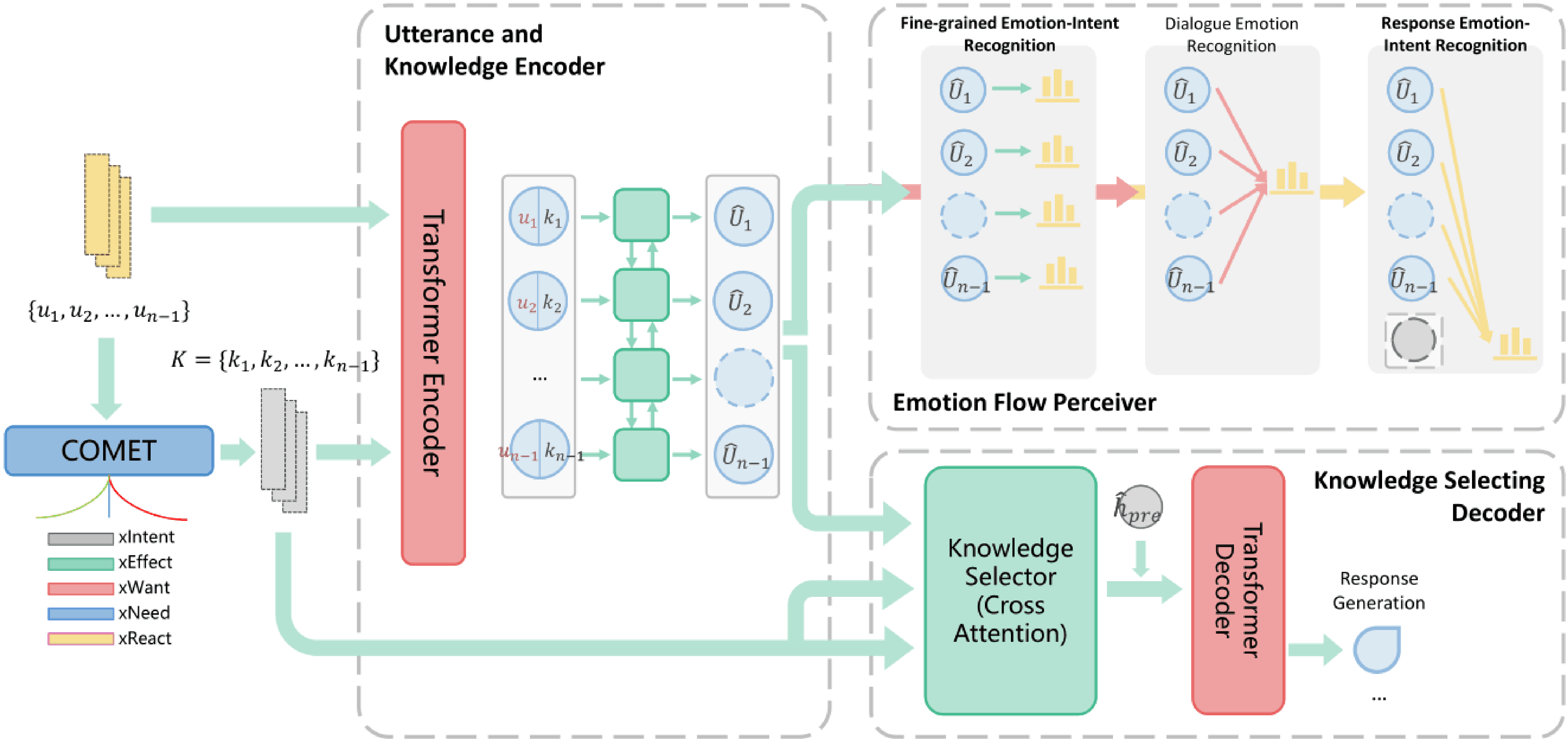


Figure 2: An overall architecture of our proposed model.

Method

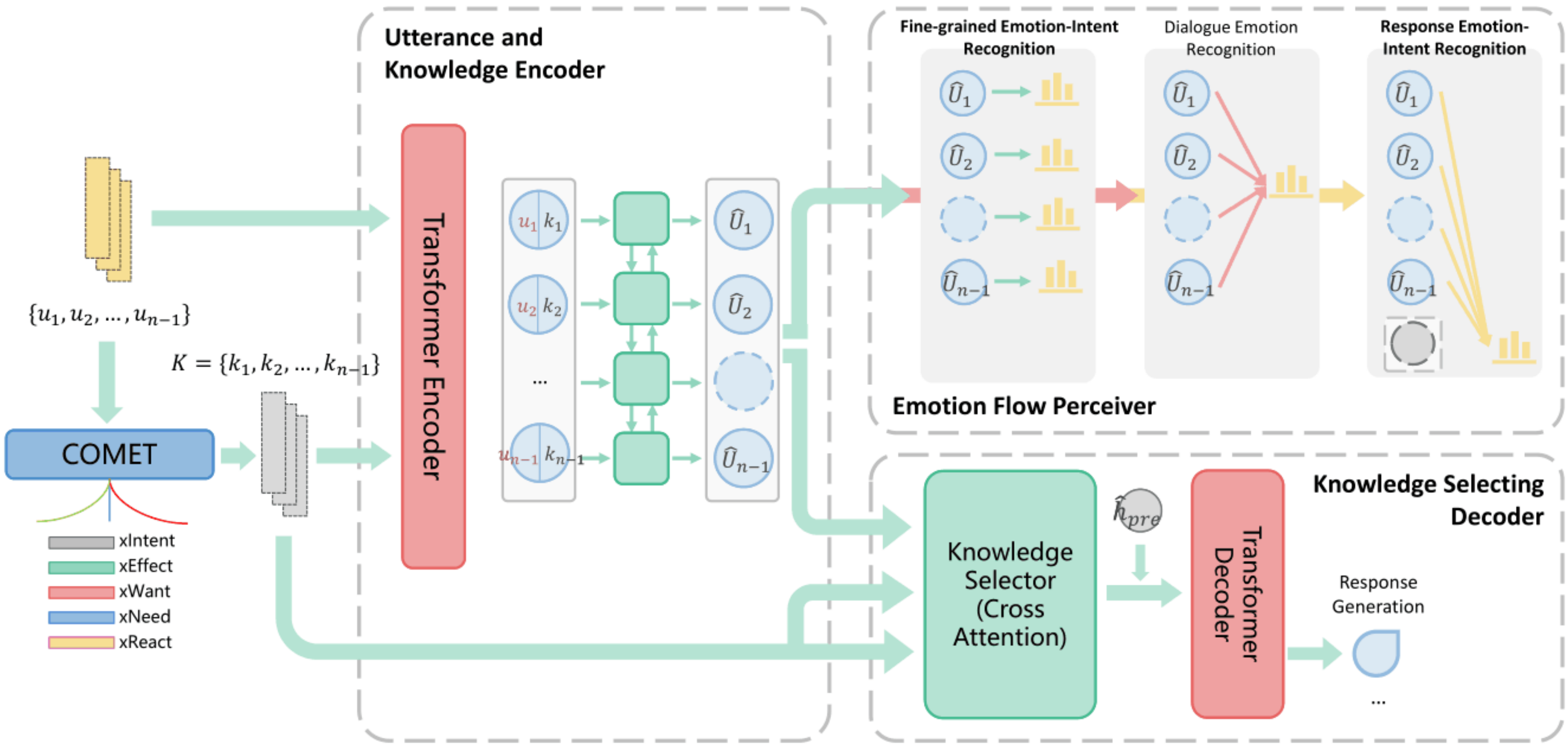


Figure 2: An overall architecture of our proposed model.

Task Formulation

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

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

Utterance and Knowledge Encoder

Utterance Encoding

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

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

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

Knowledge Encoding

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

Method

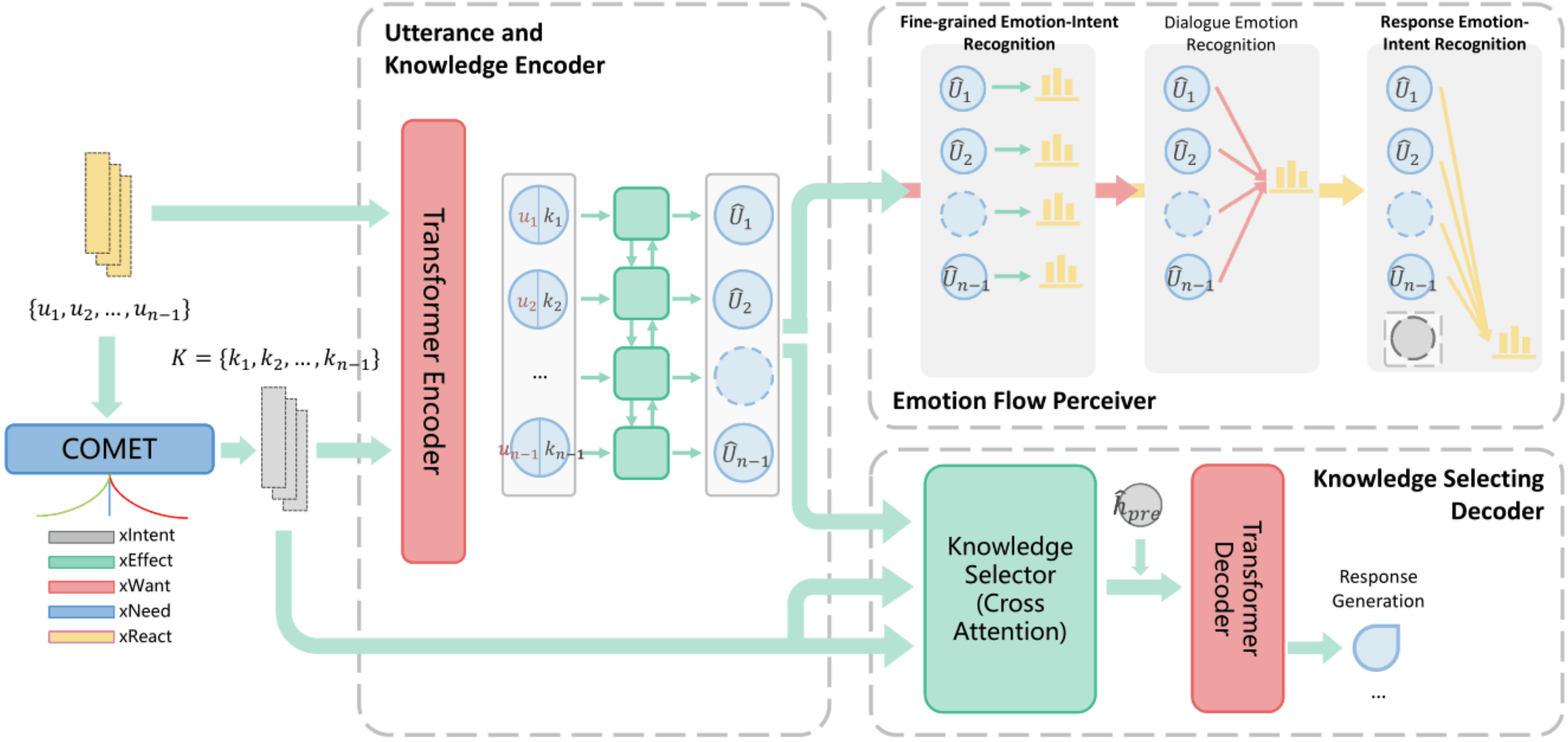


Figure 2: An overall architecture of our proposed model.

Emotion Flow Perceiver

$$\boldsymbol{a}_i = [\boldsymbol{U}_i; \boldsymbol{K}_i], \quad \hat{\boldsymbol{U}}_i = \textbf{BiLSTM}(\boldsymbol{W}_a \boldsymbol{a}_i), \tag{4}$$

Fine-grained Emotion Recognition

$$P_{tag}(\boldsymbol{ei}_i) = \text{Softmax}(\boldsymbol{W_e} \hat{\boldsymbol{U}}_i) \tag{5}$$

$$\mathcal{L}_{emo} = -\sum_{i=1}^{N-1} log(P_{tag}(\boldsymbol{ei}_i)). \tag{6}$$

Response Emotion-Intent Prediction

$$\hat{\mathbf{h}}_{pre} = \text{attention}([\hat{\boldsymbol{U}}_1, \hat{\boldsymbol{U}}_2, \ldots, \hat{\boldsymbol{U}}_{N-1}]), \quad P_{pre} = \text{Softmax}(\boldsymbol{W_p} \hat{\mathbf{h}}_{pre}), \tag{7}$$

$$\mathcal{L}_{pre} = -log(P_{pre}(\boldsymbol{ei}_N)). \tag{8}$$

Method

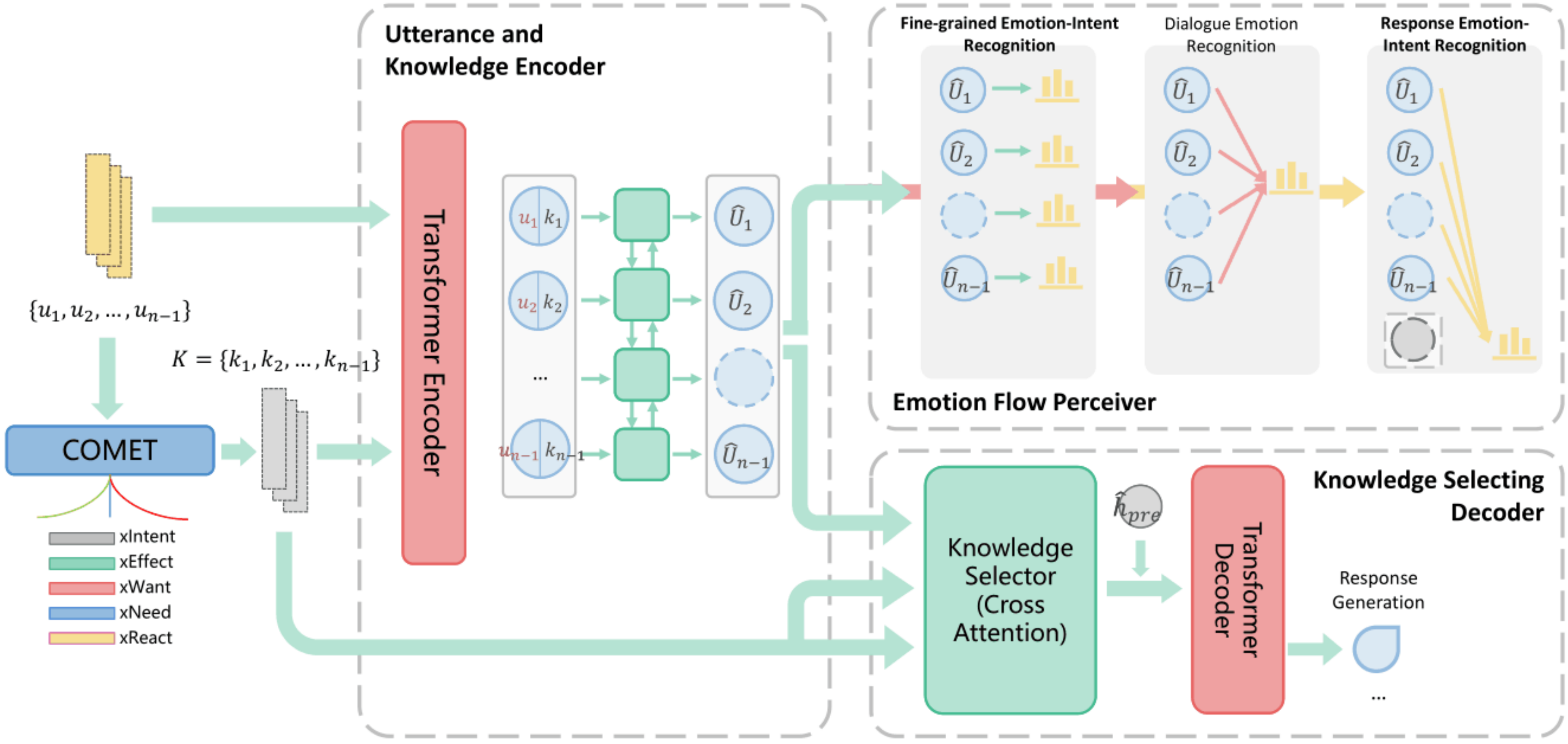


Figure 2: An overall architecture of our proposed model.

Dialogue Emotion Recognition

$$\hat{\mathbf{h}}_{dia} = \text{attention}([\hat{\boldsymbol{U}}_1, \hat{\boldsymbol{U}}_2, ..., \hat{\boldsymbol{U}}_{N-1}]), \quad P_{dia} = \text{Softmax}(\boldsymbol{W_d}\hat{\mathbf{h}}_{dia}), \tag{9}$$

$$\mathcal{L}_{dia} = -log(P_{dia}(\boldsymbol{e}^*)). \tag{10}$$

Knowledge Selecting Decoder

$$\boldsymbol{\mathcal{S}} = \text{Cross-Attention}(\hat{\boldsymbol{U}}, \boldsymbol{\mathcal{K}}, \boldsymbol{\mathcal{K}}), \tag{11}$$

$$\boldsymbol{S} = \text{pooling}(\boldsymbol{\mathcal{S}}). \tag{12}$$

$$[\text{SOS}] = \boldsymbol{W}_k([\boldsymbol{S}; \hat{\mathbf{h}}_{pre}]) \tag{13}$$

$$Y = [[\text{SOS}], y_1, ..., y_T]$$

$$\mathcal{L}_{nll} = -\sum_{t=1}^{T} log(P(y_t|C, y_{<t}). \tag{14}$$

Method

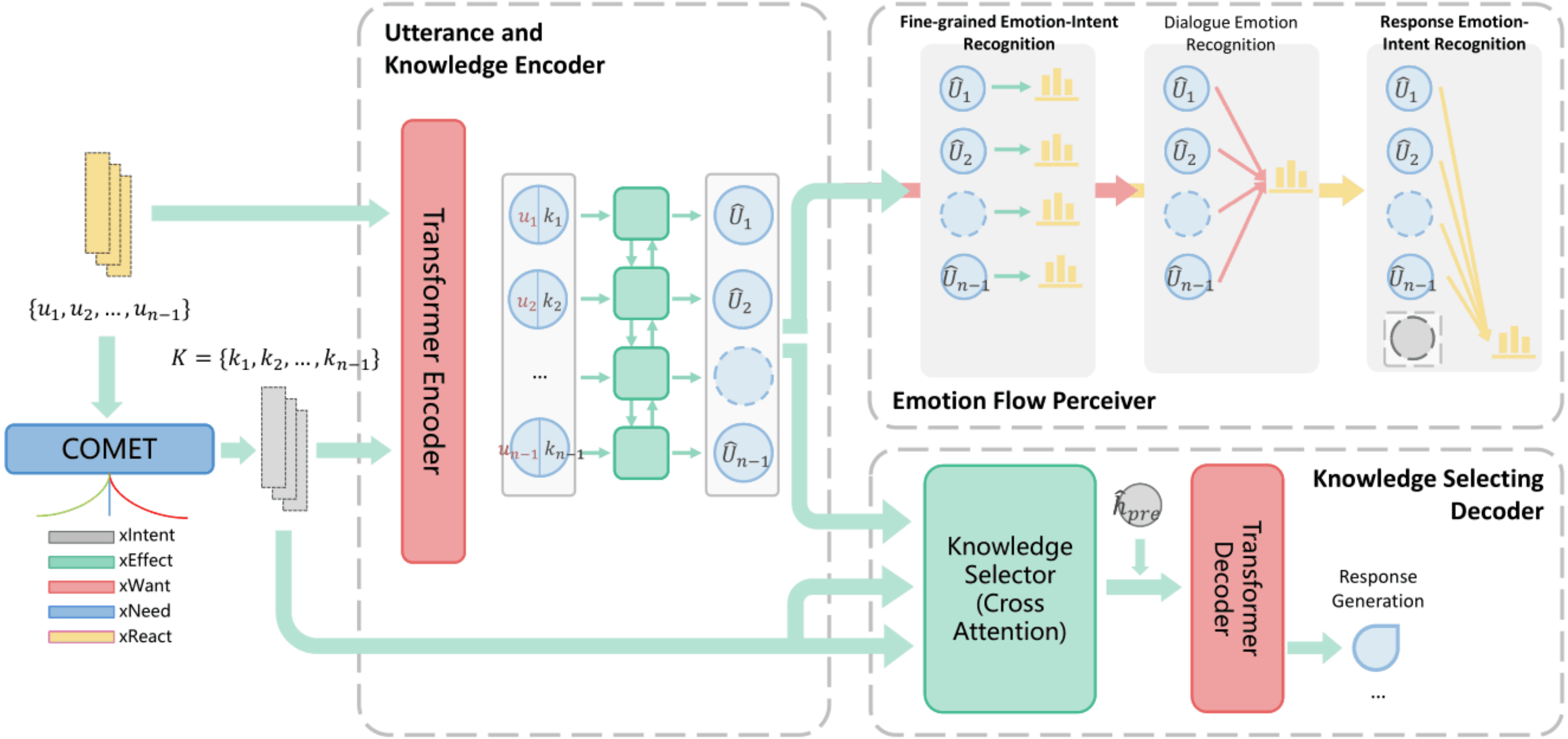


Figure 2: An overall architecture of our proposed model.

Training Objectives

$$\mathcal{L}_{cls} = \mathcal{L}_{\cancel{tag}} + \mathcal{L}_{pre} + \mathcal{L}_{dia}. \qquad (15)$$

emo

$$\mathcal{L}_{div} = -\sum_{t=1}^{T}\sum_{i=1}^{V} w_i \delta_t(c_i) log(P(y_t|C, y_{<t}), \qquad (16)$$

$$\mathcal{L} = \alpha\mathcal{L}_{nll} + \beta\mathcal{L}_{cls} + \gamma\mathcal{L}_{div}, \qquad (17)$$

Experiment

Models	PPL	Dist-1	Dist-2	DE Acc.	UEI Acc.	REI Acc.
MIME	37.08	0.31	1.03	29.38	-	-
EmpDG	37.77	0.59	2.48	30.03	-	-
KEMP	**36.89**	0.61	2.65	37.58	-	-
CEM	37.03	0.66	2.99	36.44	-	-
SEEK	37.09	**0.73**	**3.23**	**41.85**	**34.08**	**25.67**

Table 1: Automatic Evaluation results of baselines and our model. The improvement of SEEK to four strong baselines is statistically significant (paired t-tests with p-values < 0.05).

Experiment

Models	PPL	Dist-1	Dist-2	DE Acc.	UEI Acc.	REI Acc.
SEEK	**37.09**	**0.73**	**3.23**	**41.85**	34.08	25.67
w/o Utter	37.37	0.70	3.13	38.9	-	**30.41**
w/o Res	37.97	0.63	2.74	40.82	**50.48**	-
w/o Utter & Res	38.48	0.60	2.70	39.7	-	-
w/o Emo	37.67	0.61	2.66	41.27	35.88	23.37
w/o Know	37.35	0.31	1.19	41.07	33.53	25.58
+ Others know	37.50	6.90	2.88	38.25	34.43	24.32
+ Context Enc	38.68	0.67	2.60	41.81	32.86	24.45

Table 2: Ablation study of our proposed model SEEK. The best results are marked with bold.

Experiment

Models	Coh.	Emp.	Flu.
MIME	2.84	2.97	2.87
EmpDG	2.85	2.78	2.76
KEMP	2.73	2.80	2.80
CEM	2.82	2.99	2.75
SEEK	**2.91**	**3.02**	**3.07**

Table 3: Human evaluation results. We apply Fleiss's Kappa, denoted as κ, to measure inter-annotator agreement, where $0.4 < \kappa < 0.6$ indicates moderate agreement.

Comparisons	Aspects	Win	Lose	Tie
SEEK vs. MIME	Coh.	**24.3**	17.1	58.6
	Emp.	**31.4**	22.2	46.4
	Flu.	**28.6**	25.9	45.5
SEEK vs. EmpDG	Coh.	**32.1**	26.3	41.6
	Emp.	**35.5**	27.4	37.1
	Flu.	**26.9**	22.3	50.8
SEEK vs. KEMP	Coh.	**29.2**	25.2	45.6
	Emp.	**28.8**	19.9	51.3
	Flu.	**38.7**	15.6	45.7
SEEK vs. CEM	Coh.	**27.3**	24.8	47.9
	Emp.	**33.4**	27.5	39.1
	Flu.	**35.7**	21.6	42.7

Table 4: Human A/B test (%) on the three aspects: coherence, empathy, and fluency. The comparison results directly shows SEEK outperforms the four baselines models.

Experiment

Context	*Speaker:* I love YouTube. I've been listening to all my **classic** tracks. Tupac forever. (**Nostalgic**)
	Listener: I love me some Tupac. Real talk. (**Acknowledging**)
	Speaker: I started out with *One Hit Wonders* but ended up at Pac. **I miss my youth lol**. (**Nostalgic**)
Speaker Emotion	Nostalgic
MIME	I am sure you will be able to get a new one.
EmpDG	I am sure you will be fine.
KEMP	I am sure you will be fine.
CEM	That sounds like fun!
SEEK	I do not like that, but i think they will be able to get the old times . (**Nostalgic**)
Golden	Me too. I always end up on the 90s hip hop station on Pandora. (**Nostalgic**)

Context	*Speaker:* Yeah about 10 years ago I had a horrifying experience. It was 100% their fault, but they hit the water barrels and survived. They had no injuries, but they almost ran me off the road. (**Guilty**)
	Listener: Did you suffer any injuries? (**Questioning**)
	Speaker: No, I was not hit. It turned out they were drunk. I felt guilty, but realized it was his fault. (**Guilty**)
Speaker Emotion	Guilty
MIME	I would be too. I would be so scared.
EmpDG	I am sorry to hear that.
KEMP	I am sorry to hear that.
CEM	That is so sad.
SEEK	That is pretty scary! I am glad that you were able to get in out! (**Sympathizing**)
Golden	Why did you feel guilty? People really shouldn't drive drunk. (**Questioning**)

Table 5: Two cases of generated responses by SEEK and the baselines. We annotated each turn with the emotional or intentional labels at the end of the utterances. The words relevant to the predicted labels in SEEK's response are highlighted in red.

Experiment

Type	x_intent	x_need	x_want	x_effect	x_react
Knowledge	to see the baby	to have an ultrasound	to see what the baby is	to see the baby	happy
	to know the gender	to see the ultrasound	to show it to their friends	to see the gender	excited
	to know the sex	to have the ultrasound	to show it to everyone	to see the ultrasound	surprised
	to be informed	to have a baby	to show it to others	to be happy	joyful
	none	to get the ultrasound	to see the baby	we get excited	relieved
Context	We asked the doc to put the ultrasound in an envelope so we could record our reaction to the gender reveal. I was very happy when I finally saw it! (**Excited**)				
SEEK	**Congratulations!**				
Gold	**Congrats!** what gender did your child end up being?				

Table 6: The visualization of the cross-attention weights of selecting knowledge in SEEK.

Thank you!

Leibniz-Institut
für Sozialwissenschaften