IEEE Signal Processing Society

IEEE ICASSP 2022

Paper ID 2539

MODELING INTENTION, EMOTION AND EXTERNAL WORLD IN DIALOGUE SYSTEMS

Wei Peng[1,2], Yue Hu[1,2], Luxi Xing[1,2], Yuqiang Xie [1,2], Xingsheng Zhang[1,2] and Yajing Sun[1,2]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

{pengwei, huyue, xingluxi}@iie.ac.cn

NELIST 信息内容安全技术国家工程实验室 National Engineering Laboratory for Information Security Technologies

Introduction

◆ **Task & Existing problems**

In dialog system, dialog intention recognition and sentiment classification aim to predict the intention and emotion of the user. During the human interaction, intention, emotion and action are important elements. Modeling the interaction process by analyzing the relationships between these elements is a challenging task. However, previous work mainly focused on modeling intention and emotion independently, and neglected of exploring the mutual relationships between intention and emotion.

◆ **RelAtion Interaction Network (RAIN)**

In this paper, we propose a RelAtion Interaction Network (RAIN), consisting of Intention Relation Module and Emotion Relation Module, to jointly model mutual relationships and explicitly integrate historical intention information. Specifically, Intention Relation Module introduces an intention dictionary to explicitly account for the intention recognition task. Then, Emotion Relation Module designs an intention fusion mechanism to explicitly integrate historical intention information for subsequent emotion prediction. The experiments on the dataset show that our model can take full advantage of the intention, emotion and action between individuals and achieve a remarkable improvement over BERTstyle baselines. Qualitative analysis verifies the importance of the mutual interaction between the intention and emotion.

Experimental Results

◆ **Dataset & Evaluation Metrics**

(1) The DailyDialog datasets contain 13,118 multi-turn dialogues. In order to obtain the intention dictionary, we extract 2,046 conversations for the annotation. (2) Macro-average Precision (P), Recall (R) and F1 .

	Intention Recognition			Emotion Prediction		
	P ↑	R ↑	F1 ↑	P ↑	R ↑	F1 ↑
GRU [17]	46.18	41.31	43.61	42.25	41.64	41.94
GRU+Attention [13]	48.66	41.43	44.75	43.74	41.20	42.43
DCR-Net † [12]	52.35	48.56	50.38	47.24	43.91	45.51
BERT [18]	65.84	65.18	65.51	55.35	55.42	55.38
RoBERTa$_{base}$ [15]	68.14	68.53	68.33	56.82	57.89	57.35
RoBERTa$_{large}$ [15]	71.36	70.47	70.91	59.51	58.77	59.13
RAIN	**73.22**	**72.64**	**72.93**	**65.35**	**62.84**	**64.07**

Table 1. Experimental results on the testset for tasks of intention recognition and emotion prediction. † indicates that the performance is reimplemented by ourselves.

	Intention Recognition		Emotion Prediction	
	F1 ↑	$\Delta_{(F1)}$	F1 ↑	$\Delta_{(F1)}$
RoBERTa$_{large}$ [15]	70.91	-	59.13	-
+IntentNet	72.29	+1.38	61.15	+2.02
+Fusion Mechanism	-	-	60.46	+1.33
+Historical Intention	71.77	+0.86	62.11	+2.98
+Multi-task	71.96	+1.05	59.97	+0.84
RAIN	**72.93**	**+2.02**	**64.07**	**+4.94**

Table 2. The results of ablation study on model components.

◆ **Main Experiments**

RAIN can outperform baselines by a large margin hich achieves 2.02% gain on F1 for intention recognition. For the emotion prediction, second column of Table 1, RAIN achieves the best results.

◆ **Ablation Study**

The ablation study has demonstrated that all the components proposed are beneficial for the tasks, including intention dictionary, fusion mechanism, historical intention information and multi-task learning.

◆ **Case Study**

we present dialogue cases and explanations of the emotion to demonstrate how our model performs. the speaker requests for changing the room and intention of the listener is "*reject*" , thus generating *emotion of speaker is anger because his intention is unsatisfied*.

Explanation: Emotion of speaker is anger because his intention is unsatisfied.

Speaker: I'm Bill . I'm in Room 908 . Can you change the room for me ? It's too noisy.

Listener: I am sorry , there are no rooms available …

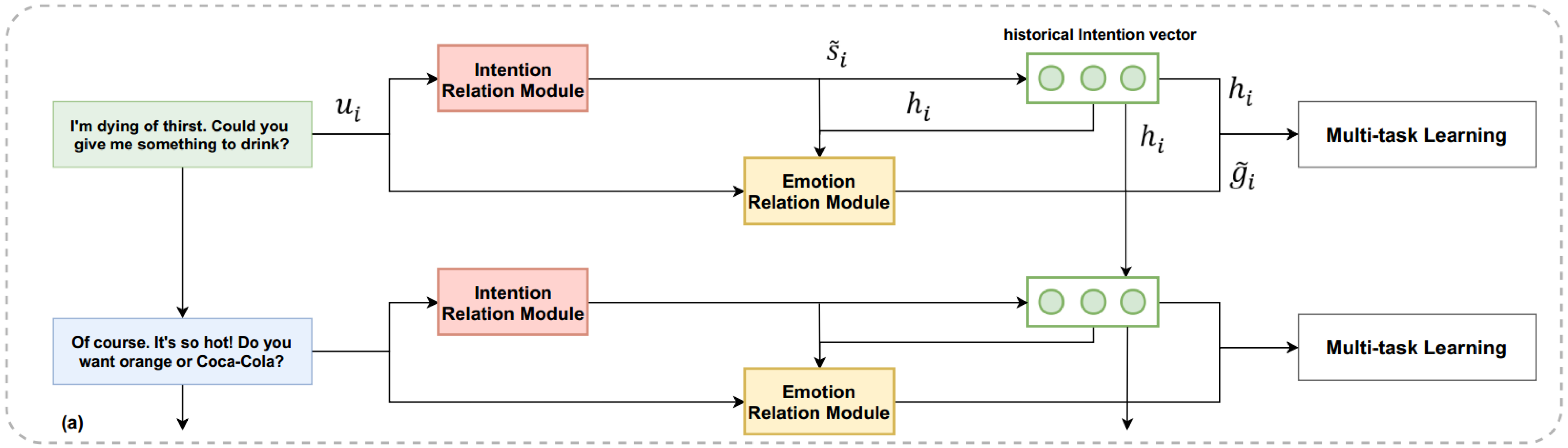
Speaker: Oh! Goodness !

Intention of Speaker: Request.

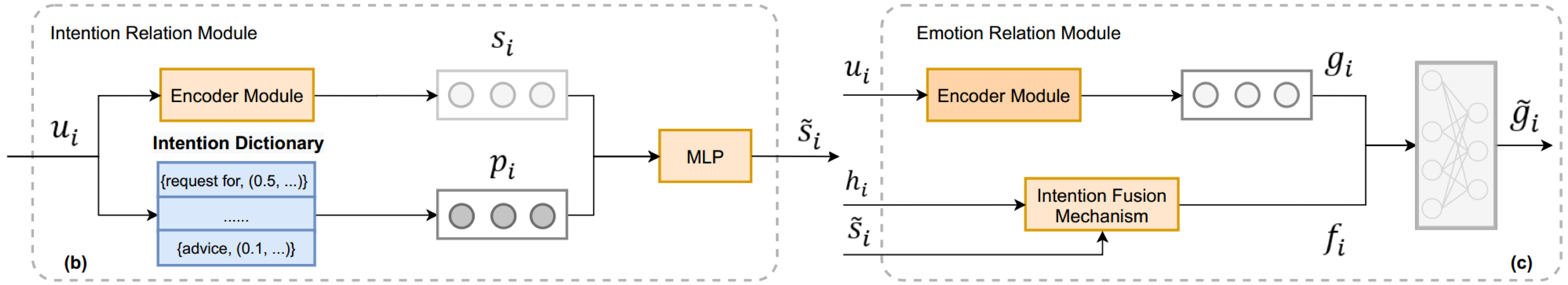
Intention of Listener: Reject.

Emotion of Speaker: Anger.

Architecture



◆ (a) The overview of our RAIN which consists of the Intention Relation Module (b) and Emotion Relation Module (c). Giving an utterance u_i . the two modules jointly model mutual relationships between the intention and emotion, and the updated representations of the intention and emotion are used for multi-task learning.



◆ **Intention Relation Module**

- *Encoder Module*

$$s_i = \text{RoBERTa}\,([CLS], x_{i,1}, \ldots, x_{i,T}, [SEP])[0]$$

- *Intention Dictionary*

Dictionary is to make an interpretable intention recognition with symbolic representations. We mark the labeled feature words, and count the word frequency of each feature word over different intentions, and normalize it to get the probability distribution.

$$\tilde{s}_i = \text{MLP}\,(\text{ReLU}(W_s^T s_i + p_i))$$

◆ **Emotion Relation Module**

- *Encoder Module*

$$g_i = \text{RoBERTa}\,([CLS], x_{i,1}, \ldots, x_{i,T}, [SEP])[0]$$

- *Intention Fusion Mechanism*
  - *Historical Intention Modeling*

$$h_i = \text{LSTM}\,(\tilde{s}_i, h_{i-1}))$$

$$f_i = \text{Fuse}\,(\tilde{s}_i, h_i)$$

$$\text{Fuse}(\tilde{s}_i, h_i) = \tanh(W_f^T[\tilde{s}_i; h_i; \tilde{s}_i \circ h_i; \tilde{s}_i - h_i] + b_f)$$

$$\tilde{g}_i = \text{ReLU}\,(W_g^T[f_i; g_i] + b_g)$$

◆ **Emotion Relation Module**

$$\hat{y}_i^m = \text{Softmax}(W_m^T \tilde{s}_i + b_m), \qquad \hat{y}_i^e = \text{Softmax}(W_e^T \tilde{g}_i + b_e),$$

Conclusion

(1) We present a RelAtion Interaction Network (RAIN) to jointly model mutual relationships between the intention and emotion.

(2) Intention dictionary and historical intention information are explicitly modeled in Intention and Emotion Relation Module.

(3) The proposed RAIN is very effective, which outperforms the previous methods with a single model, as well as making an explanation of the two tasks.

(4) Some other psychological factors will be considered for human activities, such as personal character, educational background and so on.

中国科学院 信息工程研究所 INSTITUTE OF INFORMATION ENGINEERING,CAS