Other Roles Matter! Enhancing Role-Oriented Dialogue Summarization via Role Interactions

Haitao Lin[1,2], Junnan Zhu[1,2], Lu Xiang[1,2], Yu Zhou[1,3]*, Jiajun Zhang[1,2], Chengqing Zong[1,2]

[1] National Laboratory of Pattern Recognition, Institute of Automation, CAS, Beijing, China
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[3] Fanyu AI Laboratory, Zhongke Fanyu Technology Co., Ltd, Beijing, China
{haitao.lin, junnan.zhu, lu.xiang, yzhou, jjzhang, cqzong}@nlpr.ia.ac.cn,

Code:https://github.com/xiaolinAndy/RODS

2022.05.28 • ChongQing ACL2022

Introduction

(1) Other roles' dialogue utterances could help enhance the informativeness of summaries.

(2) Other roles' summaries could help judge the key content in the dialogue

Dialogue	
0 Q: 我的这个原来绑定的手机号没用了，然后我密码又忘了。(My original phone number was useless, and I forgot my password.) 1 A: 这边为您转接专员您看可以吗？(Can I transfer you to the commissioner here?) 2 Q: 好的。(Yes.) 3 A: 很抱歉转接失败的呢亲，这边为您升级专员[数字]小时回电可以吗亲？(I'm sorry, the transfer failed. Can I help to upgrade the commissioner and call you back in [NUM] hours?) 4 Q: 也可以的啦。(It's OK.) 5 A: 请问还有其他还可以帮到您的吗?(Is there anything else I can help you with?) 6 Q: 对了，**京东是可以微信支付**的吧。(By the way, **Can JD pay via wechat**?) 7 A: 正常是可以的呢亲。(Yes, it is OK normally.)	
User Summary	用户表示绑定手机号不用了，密码忘记。**用户询问京东是否可以微信支付**。(The user said that the mobile phone number was useless, and forgot the password. **The user asked whether Jingdong can pay via wechat**.)
Agent Summary	客服帮助用户转接专员失败后表示[数字]小时内回电。客服表示**京东可以微信支付**。(The customer service helped the user to call back within [number] hours after the transfer specialist fails. The customer service said **JD could pay via wechat**.)

: Information from the user's utterances, could enhance agent summary in

: Information from the agent's summaries, could enhance user summary in

: Information from the user's summaries, could enhance agent summary in

Task Definition

Given a dialogue D containing m utterances $\{u_1, ..., u_m\}$ and p speakers $S = \{s_1, ..., s_p\}$

Generate a summary y^k for each speaker s_k

u_k consists of a speaker role $r_k \in S$ and related content

final input $\{x_1, ..., x_n\}$

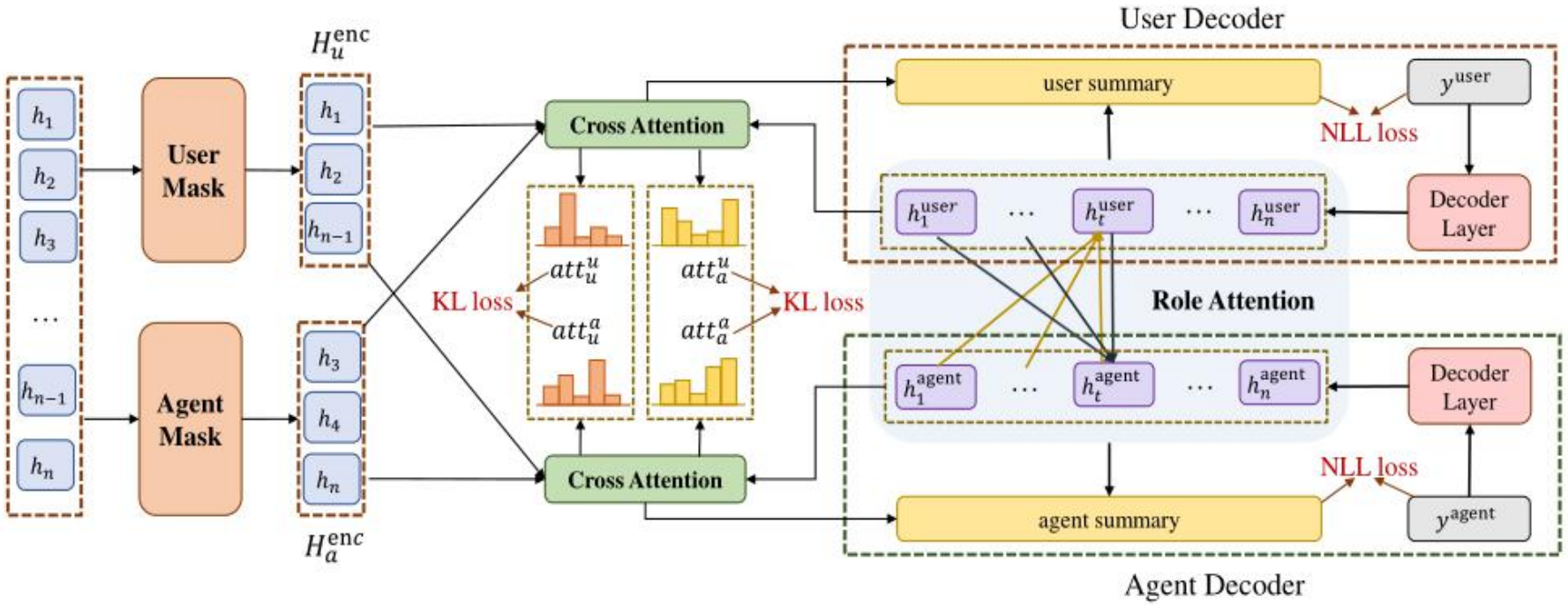

Figure 2: The model structure of our proposed method with role interactions.

Cross Attention Interaction

$$\mathcal{L}_{\text{att-user}} = \text{KL}(\text{Avg}(att_u^a)||\text{Avg}(att_u^u))$$

$$\mathcal{L}_{\text{att-agent}} = \text{KL}(\text{Avg}(att_a^u)||\text{Avg}(att_a^a))$$

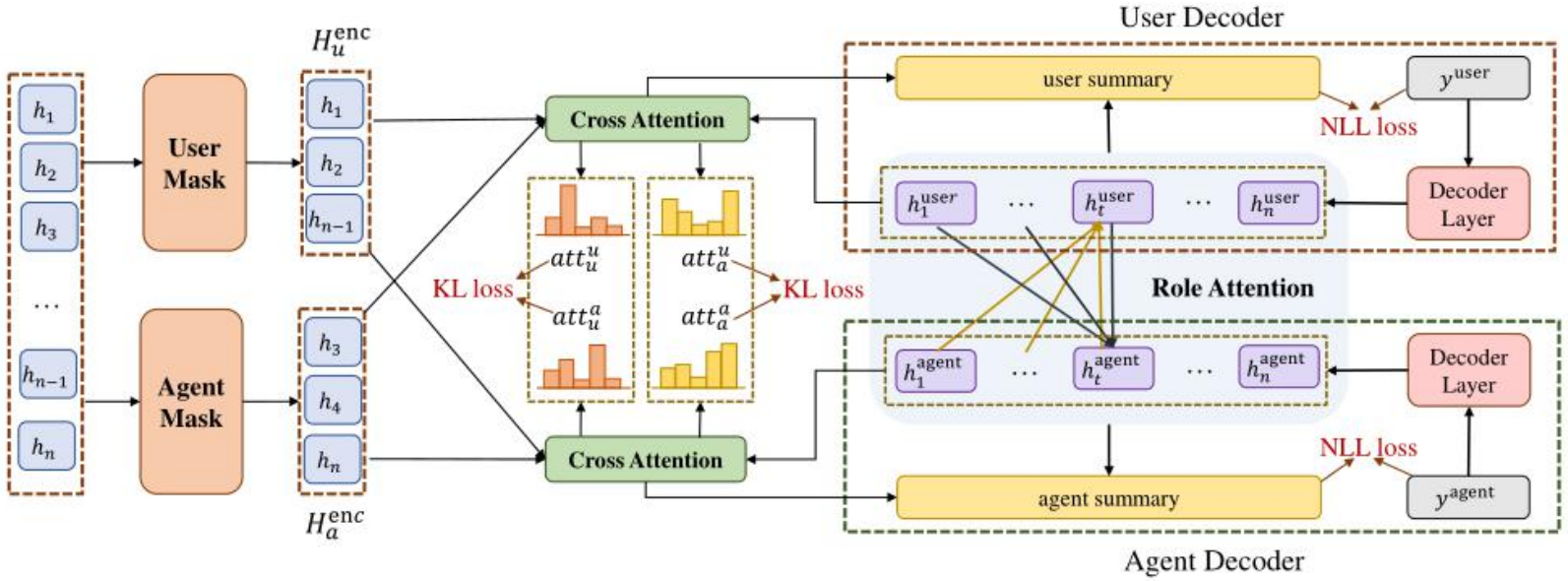

Figure 2: The model structure of our proposed method with role interactions.

Decoder Self-Attention Interaction

$$r_t^{\text{user}} = \text{Attn}(h_t^{\text{user}}, h_{1:t}^{\text{agent}})$$

$$p(\hat{y}_t^{\text{user}}) = \mathcal{F}(h_t^{\text{user}}, r_t^{\text{user}}, c_{u,k}^u, c_{a,k}^u)$$

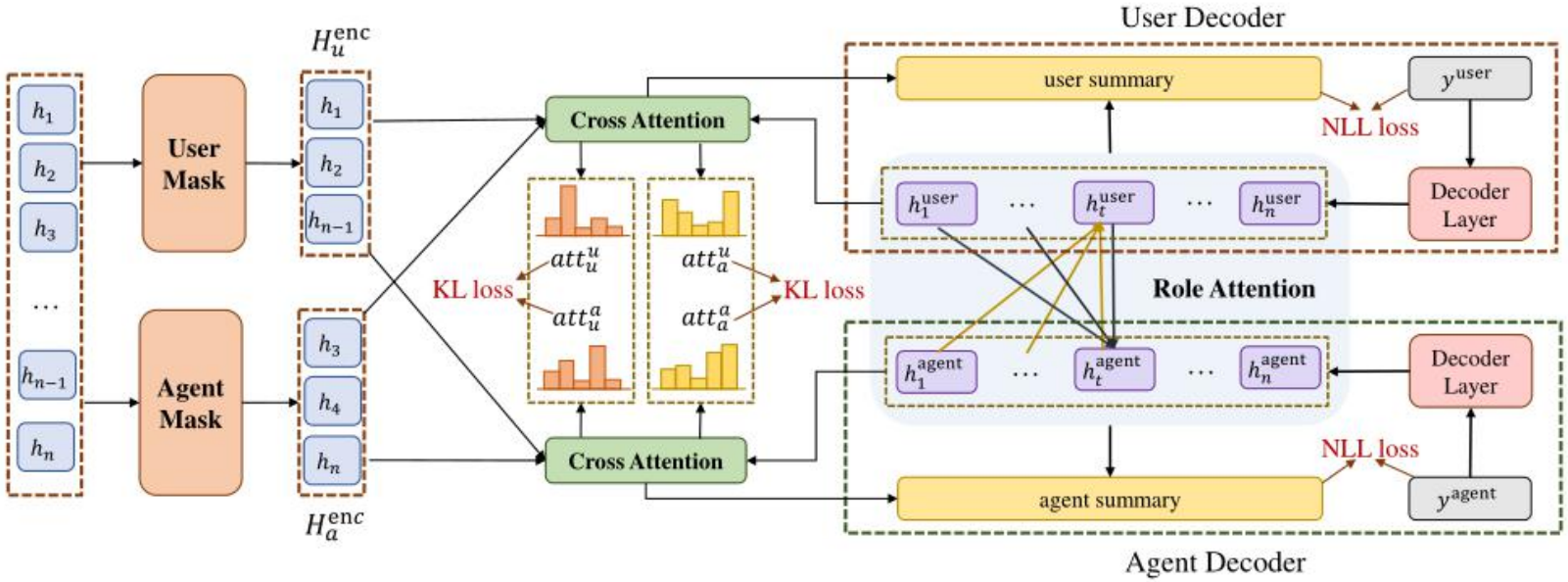

Figure 2: The model structure of our proposed method with role interactions.

Training and Inference

$$\mathcal{L}_{\text{nll}} = -(\alpha \cdot \sum_{i=1}^{|y^{\text{user}}|} \log P(y_i^{\text{user}} | y_{<i}^{\text{user}}, y_{<i}^{\text{agent}}, D) + (1-\alpha) \cdot \sum_{i=1}^{|y^{\text{agent}}|} \log P(y_i^{\text{agent}} | y_{<i}^{\text{agent}}, y_{<i}^{\text{user}}, D))$$

$$\mathcal{L} = \mathcal{L}_{\text{nll}} + \beta(\mathcal{L}_{\text{att-user}} + \mathcal{L}_{\text{att-agent}})$$

Experiment

	CSDS	MC
Train Size	9,101	29,324
Val Size	800	3,258
Test Size	800	8,146
Turns	25.92	18.48
Dial. Length	321.92	292.21
User Sum. Length	37.28	22.37
Agent Sum. Length	48.08	95.32

Table 1: Statistics of CSDS and MC. All the lengths are counted on Chinese characters.

Experiments

CSDS	ROUGE-1		ROUGE-2		ROUGE-L		BLEU		BERTScore		MoverScore	
	user	agent	user	agent	user	agent	user	agent	user	agent	user	agent
PGN-single	53.55	50.20	37.06	35.12	51.05	47.59	29.64	28.25	78.68	76.13	26.68	25.13
PGN-multi	54.01	49.94	37.38	34.78	51.95	48.20	30.04	29.09	78.78	75.95	27.16	24.90
PGN-cross	54.34	50.80	37.75	35.89	51.95	48.20	31.19	30.40	78.97	76.51	27.89	25.60
PGN-self	55.49	51.00	38.75	35.70	53.08	48.52	31.84	**30.47**	79.37	76.48	27.74	25.55
PGN-both	**56.08***	**51.62***	**39.10***	**36.50***	**53.89***	**49.12***	**33.54***	29.78*	**79.52***	**76.74***	**28.28***	**26.25***
BERT-single	52.72	49.57	36.39	33.82	50.44	46.83	30.17	26.99	79.23	76.39	24.96	23.87
BERT-multi	56.09	50.49	39.91	35.17	54.02	48.08	26.91	25.39	80.50	76.65	27.19	23.71
BERT-cross	57.29	50.35	**41.03**	35.27	**55.29**	48.09	30.70	24.19	**80.90**	76.65	28.55	23.70
BERT-self	56.94	50.96	40.37	35.24	54.85	48.40	30.61	27.13	80.53	76.80	28.24	24.83
BERT-both	**57.36***	**51.92***	40.70	**36.37***	55.17*	**49.52***	**32.04***	**29.23***	80.70	**77.23***	**28.66***	**25.48***

Table 2: The automatic metric results for CSDS. * indicates that the improvement of applying two interactions (PGN-both, BERT-both) over *single* and *multi* are both statistically significant ($p < 0.01$).

Experiments

MC	ROUGE-1		ROUGE-2		ROUGE-L		BLEU		BERTScore		MoverScore	
	user	agent	user	agent	user	agent	user	agent	user	agent	user	agent
(Song et al., 2020)	92.80	83.31	88.97	75.48	92.80	83.29	-	-	-	-	-	-
PGN-single	94.83	82.63	94.32	77.83	94.78	81.51	87.66	68.10	97.60	91.74	90.28	67.95
PGN-multi	94.58	83.16	93.98	78.33	94.53	81.96	87.23	**69.96**	97.49	91.92	89.87	68.42
PGN-cross	**95.12**	83.40	**94.63**	78.60	**95.07**	82.18	**87.99**	69.61	**97.75**	92.07	**90.73**	69.06
PGN-self	95.08	83.17	94.59	78.48	95.04	82.00	87.90	69.29	97.70	91.99	90.64	68.54
PGN-both	95.11*	**83.48***	94.59*	**78.73***	95.06*	**82.28***	87.82*	69.63	97.71*	**92.15***	90.66*	**69.24***
BERT-single	95.13	81.66	94.50	76.73	95.08	80.42	87.20	64.09	97.86	91.71	90.31	68.29
BERT-multi	95.18	81.20	94.61	76.37	95.13	79.97	87.38	64.83	97.90	91.51	90.71	67.55
BERT-cross	95.18	81.75	94.61	77.04	95.13	80.55	87.40	**65.63**	97.89	91.70	90.67	68.28
BERT-self	95.18	81.61	94.61	77.01	95.13	80.49	87.37	65.01	97.89	91.72	90.69	68.37
BERT-both	**95.19**	**82.11***	**94.63**	**77.49***	**95.14**	**80.92***	**87.40**	65.40*	**97.90**	**91.91***	**90.72**	**68.95***

Table 3: The automatic metric results for MC. * represents the same with the one in Table 2.

Experiments

CSDS	Info	Non-Red	Flu	Overall
PGN-multi	**0.69**/0.65	0.54/0.55	0.70/0.79	0.64/0.66
PGN-both	0.66/**0.69**	**0.58/0.59***	**0.73/0.81**	**0.66/0.70***
BERT-multi	0.58/0.56	**0.66/0.61**	0.84/**0.87**	0.69/0.68
BERT-both	**0.62*/0.60***	0.62/0.60	**0.85/0.87**	**0.70/0.69**

Table 4: The human evaluation results for CSDS. Two values in each block represents user summary and agent summary. All the values are in range 0 to 1. * indicates that the improvement of applying two interactions over the *multi* baseline is statistically significant ($p < 0.05$).

Experiments

CSDS	ROUGE-1 Type A/B	ROUGE-2 Type A/B	ROUGE-L Type A/B	BLEU Type A/B	BERTScore Type A/B	MoverScore Type A/B
PGN-multi	55.13/59.45	37.76/41.22	52.73/56.20	30.66/28.29	76.40/77.64	23.74/25.47
PGN-both	**56.00/62.28**	**38.58/43.88**	**53.66/58.99**	**31.06/29.14**	**76.84/78.59**	**24.41/27.15**
BERT-multi	46.59/50.07	32.33/34.59	44.49/47.65	23.45/26.37	75.03/75.64	22.47/24.34
BERT-both	**50.96/54.62**	**35.72/37.93**	**48.82/51.93**	**27.47/30.10**	**76.27/76.94**	**24.19/26.17**

Table 5: The performance on different types of samples. Type A represents agent summaries that need to be integrated, and Type B represents for those that do not. Here all the metrics here are recall scores except for BLEU and MoverScore since they do not have a recall version. We use their available results instead.

Experiments

Methods	Precison	Recall	F1
PGN-multi	28.61/18.86	28.87/19.67	28.74/19.27
PGN-both	**31.79/21.06**	**30.85/21.58**	**31.31/21.32**
BERT-multi	**40.16/23.99**	30.26/18.81	34.51/21.09
BERT-both	37.37/22.09	**32.17/20.66**	**34.57/21.35**

Table 6: Sub-summary matching ratio for baselines and our methods. Two values in each block represents user summary and agent summary.

Thanks