养天地正气 法古今完人

ChatASU: Evoking LLM's Reflexion to Truly Understand Aspect Sentiment in Dialogues

Yiding Liu, Jingjing Wang , Jiamin Luo, Tao Zeng, Guodong Zhou

School of Computer Science and Technology, Soochow University, Suzhou, China

Motivations

- For task, existing studies on interactive ASU largely ignore the coreference issue for opinion targets.
- For question, how to address the coreference issue for aspects in dialogues is challenging for LLMs, which could assist in precisely predicting the aspect sentiments.
- For model, LLMs usually exhibit factual hallucination problem in their generative and predictive capabilities.

养天地正气 法古今完人

ChatASU Task

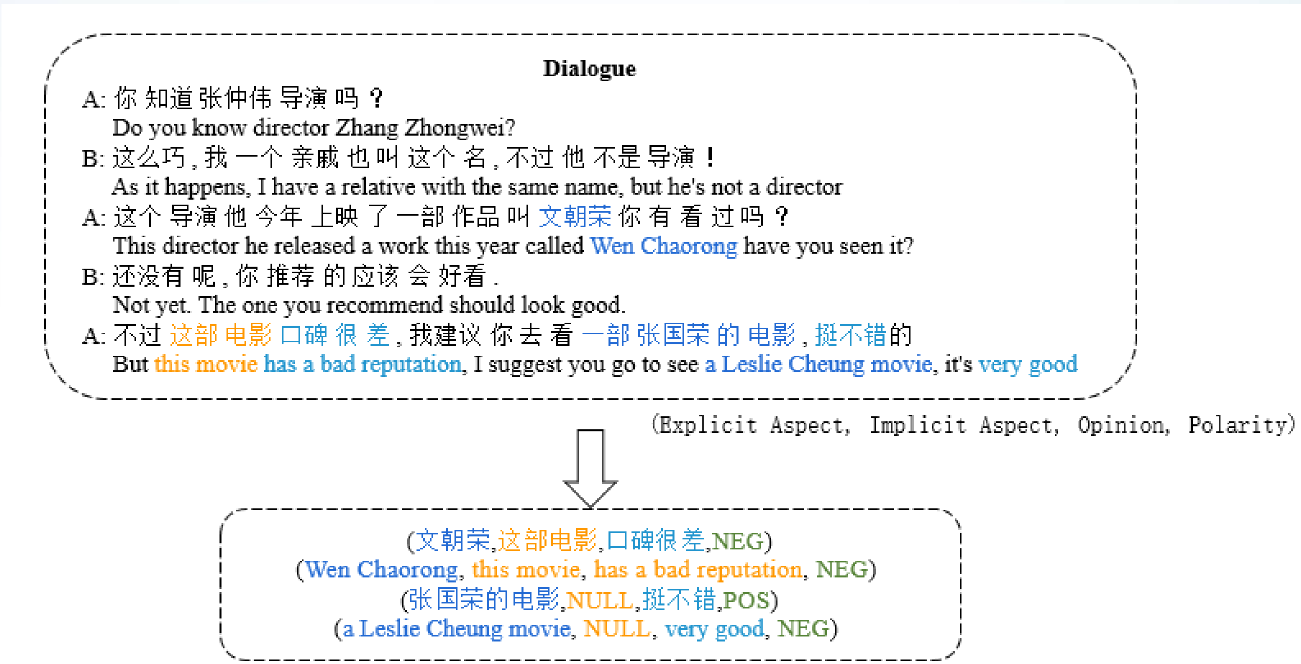
1 Traditional ASU Task: focus on extracting the relationship between aspect entities and opinions in the text.

2 Observation of Corpus: there are a lot of coreference relations between people's conversations in real-life communications and people are accustomed to using pronouns to represent an entity.

3 ChatASU Task: extracting opinion, sentiment polarity of opinion and the corresponding anaphora relationship in the dialogue.

ChatASU Task

Example of ChatASU Task:



ChatASU Dataset

Annotation for ChatASU:

1 Collecting 3000 Chinese dialogues.

2 Annotate Opinion, explicit aspect, implicit and sentiment polarity of opinion in dialogue.

3 The statistics of our constructed dataset as follows, and we randomly split the dataset into train, dev, and test sets with the ratio of 8:1:1 for ChatASU.

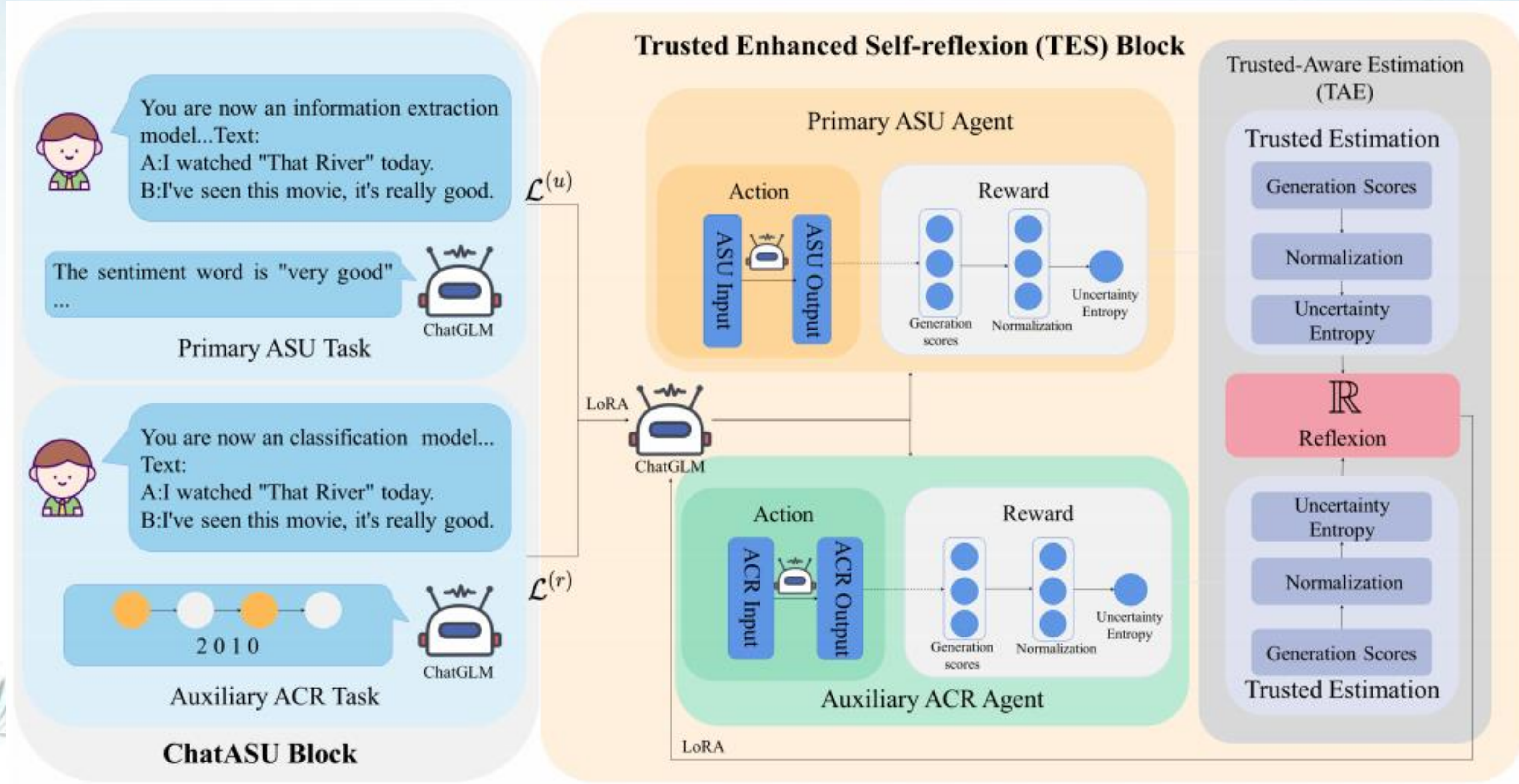
Split	#Utterances(Dialogues)	#Explicit	#Implicit	Aspect Chain		Quadruple			
				#Max	#Avg	#Pos	#Neu	#Neg	#Total
Train	21612(2400)	8959	6172	11	2.40	7234	472	1261	8967
Valid	2727(300)	1161	770	8	2.45	894	57	180	1131
Test	2723(300)	1146	733	9	2.46	987	71	144	1202

Table 1: The statistics for our annotated ChatASU Dataset. #Explicit denotes the number of explicit aspect entities. #Implicit denotes the number of references towards explicit aspects (e.g., the reference "*this movie*" for the explicit aspect "*WenChaorong*" in Figure 1). #Max and #Avg denote the max length and average length of the aspect chain.

养天地正氣 法古今完人

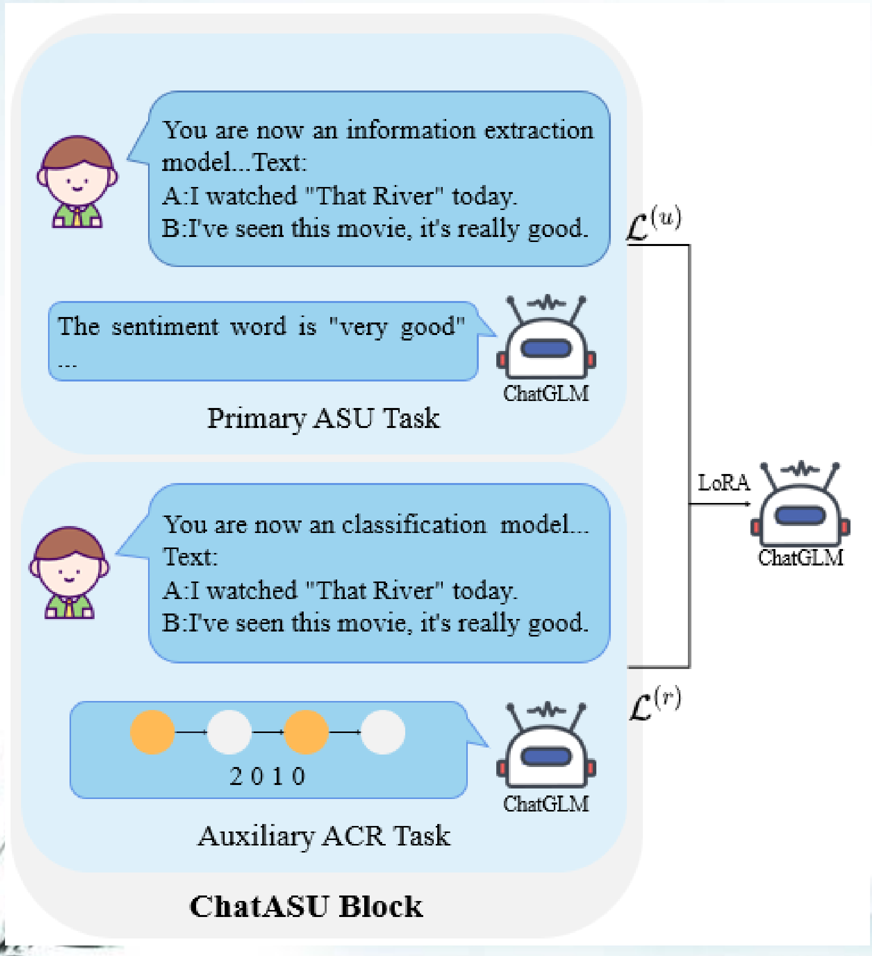
Trusted Self-reflexion Approach

The overall framework of our Trusted Self-reflexion Approach (TSA) to ChatASU:



Trusted Self-reflexion Approach

Step 1: ChatASU Block



Step 1 consists of two tasks:

1. Primary ASU Task.
2. Auxiliary ACR Task.

Trusted Self-reflexion Approach

Step 1: Primary ASU Task

1 Input: Instruction [·] text

Instruction: ``You are now an information extraction model. Please help me to extract opinions from the input and tell me the sentiment polarity of the opinions, what the explicit aspect referred to by the opinion is, and what pronoun is used for the explicit aspect in the utterance where the opinion occurs. "

2 Taking the text ``I watched That River today, the movie is very good to recommend you to see"

Output: The opinion is ``very good". The sentiment tendency is ``POS". The opinion refers to the explicit aspect ``That River". The pronoun of ``That River" is ``the movie".

养天地正气 法古今完人

Trusted Self-reflexion Approach

Step 1: Auxiliary ACR Task

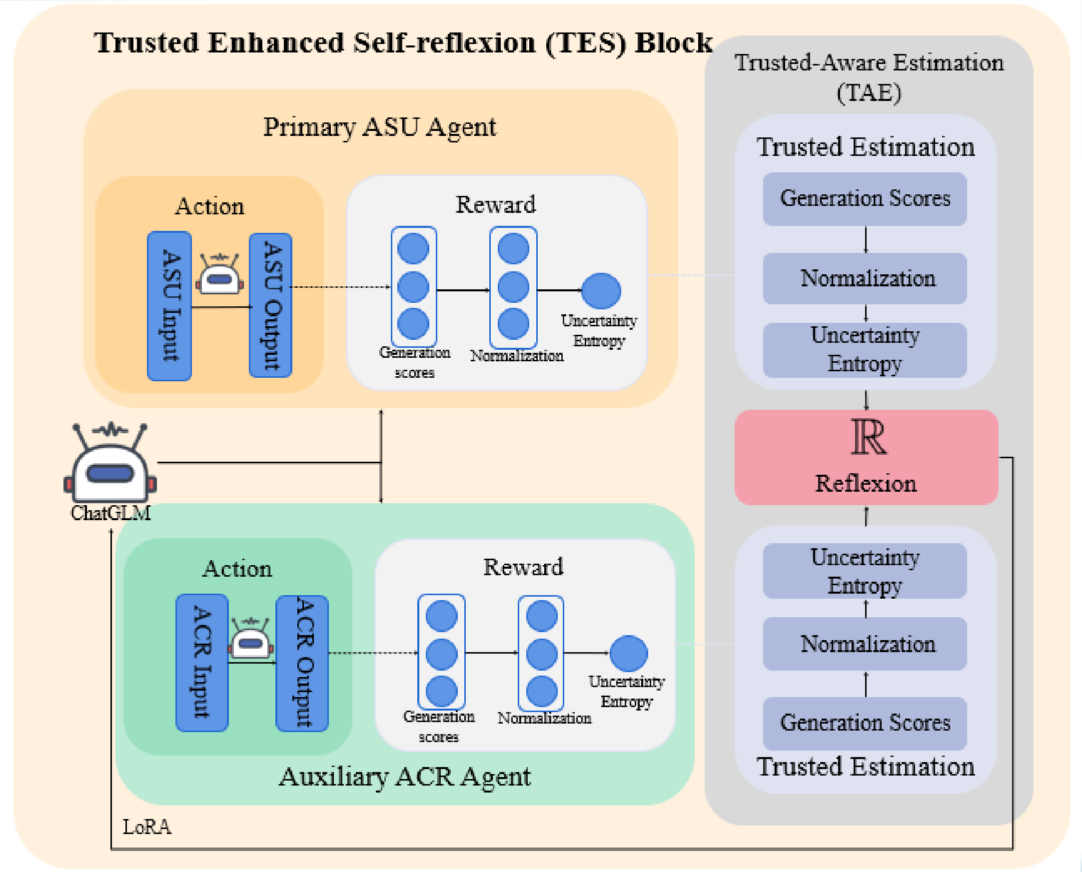
1 Input: Instruction [·] text

Instruction: ``You are now an classification model to judge which utterance in this dialogue appears to be the coreference of ei, outputs 2 if it is an explicit aspect, 1 if it is an implicit aspect, and otherwise 0. Output a sequence of 0, 1, and 2, the length of which is the number of dialogues."

2 Output: The output of the ACR task is a sequence consisting of 0, 1 and 2, with the length of the sequence equal to the number of utterances in the dialogue.

Trusted Self-reflexion Approach

Step 2: Trusted Enhanced Self-reflexion (TES) Block



Step 2 consists of three components:

1.ASU Agent

2.ACR Agent

3.Trusted-Aware Estimation (TAE).

Trusted Self-reflexion Approach

Step 2: ASU Agent and ACR Agent

Action:

$$\text{Output}_{\text{ASU}} = \text{ChatGLM}(\text{Input}_{\text{ASU}})$$

$$\text{Output}_{\text{ACR}} = \text{ChatGLM}(\text{Input}_{\text{ACR}})$$

Reward: We get the generation scores from Output.

$$\mathbb{R}_{ACR} = \mathbb{R}_{TE}(Score_{ACR})$$

$$\mathbb{R}_{ASU} = \mathbb{R}_{TE}(Score_{ASU})$$

養天地正氣 法古今完人

Trusted Self-reflexion Approach

Step 2: Trusted-Aware Estimation (TAE)

The TAE Block is consist of two parts, Trusted Estimation and Trusted Reflexion.

Trusted Estimation: Trusted Estimation (TE) utilizes the difference in generation scores obtained from beam search as a measure of confidence, encouraging ChatGLM to produce trustworthy results.

$$\hat{\boldsymbol{g}}_i = \mathbf{Normalization}(\mathbf{G}) = \frac{\boldsymbol{g}_i - min(\mathbf{G})}{max(\mathbf{G}) - min(\mathbf{G})}$$

$$\mathbb{R}_{TE}(\hat{G}) = -\sum_{j=1}^{m} \frac{1}{\sum_{i=1}^{n} \boldsymbol{m}\hat{\boldsymbol{g}}_i \log \hat{\boldsymbol{g}}_i}$$

Trusted Self-reflexion Approach

Step 2: Trusted-Aware Estimation (TAE)

Trusted Reflexion: Trusted Reflexion is the final reward function.

$$\mathbb{R}=\begin{cases}\alpha\mathbb{R}_{ACR}+\beta\mathbb{R}_{ASU}+\gamma(\mathbb{R}_{Rp}+\mathbb{R}_{Ra}), & \text{if } \boldsymbol{p}=0\\ \alpha\mathbb{R}_{ACR}-\beta\boldsymbol{p}+\gamma(\mathbb{R}_{Rp}+\mathbb{R}_{Ra}), & \text{else}\end{cases}$$

$$\mathbb{R}_{Rp}=F1=\frac{2\cdot \mathrm{Pr}_{\mathrm{asu}}\mathrm{Re}_{\mathrm{asu}}}{\mathrm{Pr}_{\mathrm{asu}}+\mathrm{Re}_{\mathrm{asu}}}$$

$$\mathbb{R}_{Ra}=F1=\frac{2\cdot \mathrm{Pr}_{\mathrm{acr}}\mathrm{Re}_{\mathrm{acr}}}{\mathrm{Pr}_{\mathrm{acr}}+\mathrm{Re}_{\mathrm{acr}}}$$

Experiments

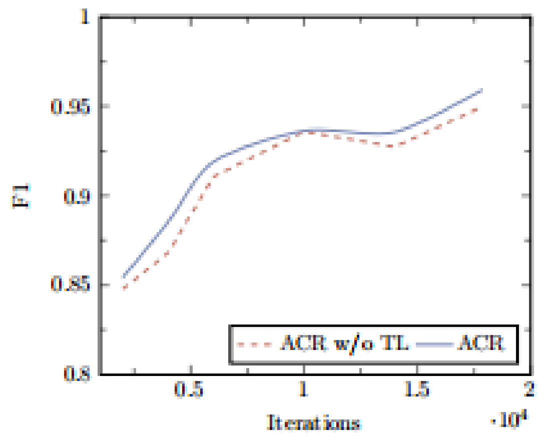
Experimental results as follows:

	Approach	Single				Pair			Quadruple
		Explicit	Implicit	Opinion	Polarity	E-O	E-I	I-O	Extraction
PLM	ASQP(Zhang et al., 2021a)	73.08	60.36	61.35	81.52	49.88	50.45	44.59	36.66
	DiaASQ(Li et al., 2022)	49.42	41.7	43.24	56.11	38.61	36.78	34.23	29.09
LLM	ChatGPT(zero-shot)	47.65	55.6	41.88	64.99	27.43	31.05	26.71	22.38
	ChatGPT(In-context learning)	47.82	56.52	43.47	69.57	37.13	34.78	43.48	30.43
	ChatGLM(Du et al., 2022)	67.49	73.70	61.70	82.21	51.32	57.44	52.00	43.49
	Reflexion(Shinn et al., 2023)	68.84	74.11	65.37	86.59	53.32	57.61	53.98	44.62
	TSA	**70.58**	74.45	**65.98**	**86.85**	**55.05**	**58.92**	**54.81**	**46.34**
	- w/o Trusted Learning	69.84	74.22	64.88	86.59	53.66	57.8	53.50	44.88
	- w/o ACR Task	68.98	**75.59**	64.90	86.45	53.14	58.45	54.53	44.98

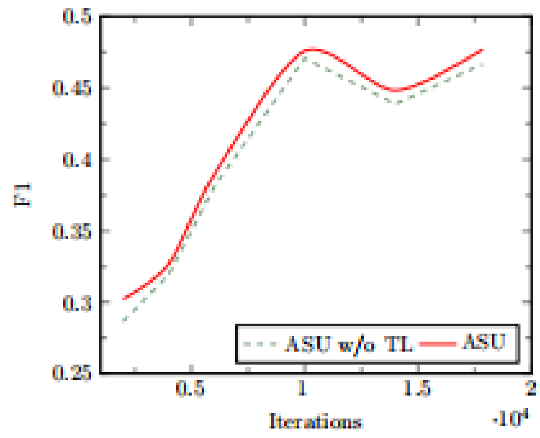
Table 2: Comparison of several state-of-the-art approaches on ASU task, where "Single" denotes the F1 score extracted separately for each element inside quadruple and "Pair" denotes the F1 score for a pair of two elements inside quadruple. E, I, O denote explicit aspect, implicit aspect, opinion, respectively.

Experiments

Importance and Robustness Study for ACR Task and Trusted Learning



(a) ACR



(b) ASU

Experiments

Qualitative Study

Input

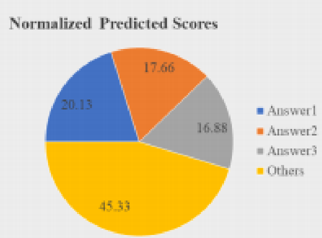
Instruction:
You are now an information extraction model. Please help me to extract opinions from the input and tell me the sentiment polarity of the opinions, what the explicit aspect referred to by the opinion is, and what pronoun is used for the explicit aspect in the utterance where the opinion occurs.
Dialogue:
U1 A: Have you seen the movie "Wind Chimes"? U2 B: No, who starred in it?
U3 A: I feel that Hong Jiantao's acting is superb, so let's go and see him in a movie sometime. **U4** B: Well, he's a great actor.
U5 A: And Hong Jiantao also has a masterpiece called Wulin Outreach, which was released five years ago! **U6** B: I've seen this one, it is really good.
U7 A: The director is Shang Jing, and online reviews say the movie is just nostalgia. **U8** B: I used to love this when I was a kid.

Ground-Truth

(Hong Jiantao,him,acting is superb,POS), (Hong Jiantao,he,great actor,POS), (Wulin Outreach,it,really good,POS), (Wulin Outreach,this,love,POS).

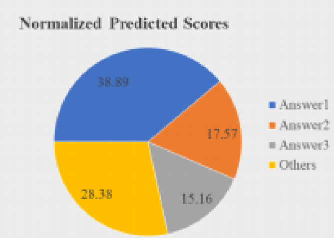
Prediction(ChatGLM)

Answer1:
(Hong Jiantao,null,acting is superb,POS),
(Hong Jiantao,null,great actor,POS),
(Wulin Outreach,null,really good,POS),
(Wulin Outreach,this,love,POS).
Answer2:
(Hong Jiantao,null,acting is superb,POS),
(Hong Jiantao,he,great actor,POS),
(Hong Jiantao,he,great actor,POS),
(Wulin Outreach,null,really good,POS),
(Wulin Outreach,this,love,POS).
Answer3 (GT):
(Hong Jiantao,him,acting is superb,POS),
(Hong Jiantao,he,great actor,POS),
(Wulin Outreach,it,really good,POS),
(Wulin Outreach,this,love,POS).



Prediction(Our TSA)

Answer1 (GT):
(Hong Jiantao,him,acting is superb,POS),
(Hong Jiantao,he,great actor,POS),
(Wulin Outreach,it,really good,POS),
(Wulin Outreach,this,love,POS).
Answer2:
(Hong Jiantao,null,acting is superb,POS),
(Hong Jiantao,he,great actor,POS),
(Wulin Outreach,null,really good,POS),
(Wulin Outreach,this,love,POS).
Answer3:
(Hong Jiantao,null,acting is superb,POS),
(Hong Jiantao,null,great actor,POS),
(Wulin Outreach,null,really good,POS),
(Wulin Outreach,this,love,POS)



Contributions

- We propose a new ChatASU task with a specially-designed ACR sub-task to address the coreference issue of aspects in dialogue ASU scenarios, which may open up a promising avenue for research in this direction.
- We incorporate both reflexion mechanisms and trusted learning for better understanding aspect chain and alleviating hallucinations problems, thereby enhancing the ability and credibility of LLMs in understanding aspect sentiments.
- We meticulously annotate a high-quality Chinese dataset ChatASU to evaluate the aspect sentiments comprehension ability of LLMs within dialogue ASU scenarios. Our work marks the first of its kind, shedding light on coreference issue in dialogue ASU scenarios and contributing to the evaluation and enhancement of LLMs' performance.

养天地正气 法古今完人

Thank you.