1 USER INTENT AND AGENT ACTION ANNOTATION

Inspired by [1, 2] , we categorize user intent and agent action into several labels. Table 1 and Table 2 show the specific categorization rules for user intent and agent action, respectively. Since *reveal* in user intent and *answer* in agent action occupy a large proportion in dataset, we subdivide several subcategories under these two classes.

Due to the complexity of real conversation content, the assessors are allowed to annotate one conversation content with two labels[1]: a primary label and a secondary label. Primary label means that the content of this label is more significant contributing to the conversation than the content of secondary label. For example, the user says "*Jane Eyre sounds interesting! How can I buy it?*" (the last response of agent is "*I recommend Jane Eyre and A Tale of Two Cities*"). The first sentence "*Jane Eyre sounds interesting!*" is a feedback towards *Jane Eyre*, so it belongs to the *others* label. "*How can I buy it?*" express an information need directly inspired by agent, thus it belongs to the *reveal-inspire* label. Furthermore, the expression of this information need motivates the further dialogue, and the agent will continue to answer it. Thus, the appropriate annotation of this example is *reveal-inspire* (primary) + *others* (secondary).

In addition, the *answer* category in agent action involves more complex two-class scenarios:

- If a part of the agent response can already satisfy user's information need completely, then the category of this part is the primary label, while the category of the rest content belongs to is the secondary label.
- If the user's information need contains multiple questions, then the category of the longer response question is the primary label, and the category of the shorter is the secondary label.

In total, we recruit 15 assessors for user intent and agent action annotation. For each assessor, we conduct a training and trial annotation program that lasts about 1 hour. In the first stage, each assessor is asked to conduct annotations for 50 to 150 dialogues after training. To ensure the quality of the final annotation, 6 high-quality assessors are picked for further annotation based on their previous annotation

[1]Single label cases are also allowed to occur.

Table 1: Categorization rules and examples for user intent

Intent		Description	Example	Remark
reveal	initiate	Express information need. Start a new dialogue topic.	(dialogue starts) User: Can you recommend a book? (reveal-initiate) Agent: I recommend ... User: When will the Beijing Olympic Game opens? (reveal-initiate)	First question, or question with the information need that is not related to the previous conversation topic.
	continue	Express information need. Continue the topic user has mentioned above.	User: When will the Beijing Olympic Game opens? Agent: It opens on February 4th, 2022. User: How can I watch the alpine ski race? (reveal-continue)	The information need is relevant to the previous dialogue topic and not directly inspired by prior agent responses.
	inspire	Express information need. The intent is directly inspired by prior agent responses.	User: Can you recommend a book? Agent: I recommend *Jane Eyre*. User: How can I buy it? (reveal-inspire)	The information need must be inspired directly from prior agent responses, with no interruptions from other user questions that utilizes the inspired information.
revise		Revise the intent proactively when the expression of previous information need is wrong or unclear, e.g., typos, grammatical errors or unclear expression.	User: When will the Olympic Game opens? User: I mean Beijing Winter Olympic Game. (revise)	User proactively revises the information need, rather than interpreting the need passively after agent asks clarifying question.
interpret		Interpret or refine the intent by answering the clarification question from the agent.	User: When will the Olympic Game opens? Agent: You mean Beijing Winter Olympic Game? User: Yes. (interpret)	The interpretation is made after agent asks clarifying question.
chitchat		Greetings or other contents that are unrelated to the information need.	User: Thanks for your help. (chitchat) User: Are you there? (chitchat)	
others		Other user intent that are not included in the above mentioned categories.	User: This movie sounds interesting! (others) User: I got it. (others)	

Table 2: Categorization rules and examples for agent action

Intent		Description	Example	Remark
clarify		Ask questions to clarify user intent when the intent is unclear or exploratory.	User: I want to know something about sun. Agent: Do you mean stellar sun or company Sun? (clarify)	The clarifying question could be yes-or-no question, choice question or even open question.
answer	single-fact	Give a unique and unambiguous fact. The answer is objective and certain.	Agent: The price of this ceil phone is 400 dollars. (answer-single-fact) Agent: The three primary colors of light are red, green and blue. (answer-single-fact)	Single-fact does not mean that there is only one entity in the answer, but instead that there is only one answer in the correct answer space.
	multi-fact	Give some elements/aspects or entities of facts. The total answer fact space is certain and large. Some of the facts can satisfy the user's information need.	Agent: Einstein's life experiences are as follows: ..... (answer-multi-fact) Agent: One of the reservation numbers is xxx. (answer-multi-fact)	In contrast to single-fact, multi-fact means that there exist multiple answers in the correct answer space. Either some or all of the answers can satisfy the user's information need.
	opinion	Give instructions, advice, or ideas. The answer can come from personal opinion of the agent or some netizens, or it has been recognized as a social consensus.	Agent: To diminish wrinkles, I recommend that you can ..... (answer-opinion) Agent: The historical meaning of the Renaissance Movement is as follows: ..... (answer-opinion)	If the correctness of answer varies with different people's perceptions, the answer belongs to *opinion*, otherwise it is a *fact*.
no-answer		Notice to the user that the relevant information has not been found	Agent: Sorry, no product has been found to satisfy your requirements. (no-answer)	Notice the difference between *I have not found any products that meet the requirements* (no-answer) and *Apple have not marketed a product that meets the requirements* (answer-single-fact). (User question is *please recommend an iPhone with 2,048GB disk space.*)
chitchat		Greetings or other contents that are unrelated to the information need.	Agent: My pleasure to help you. (chitchat) Agent: Hello! (chitchat)	
others		Other agent actions that are not included in the above mentioned categories.	Agent: Any more questions do you have? (others)	

results. We conduct further training for these high-quality assessors and provide an additional clarification guideline with examples for them. These 6 assessors are divided into 2 groups. Each group of assessors complete annotation task for about 550 conversations, half of the dataset, such that each dialogue is annotated by 3 different annotators.

REFERENCES

[1] Chen Qu, Liu Yang, W Bruce Croft, Johanne R Trippas, Yongfeng Zhang, and Minghui Qiu. 2018. Analyzing and characterizing user intent in information-seeking conversations. In *The 41st international acm sigir conference on research & development in information retrieval*. 989–992.

[2] Pengjie Ren, Zhongkun Liu, Xiaomeng Song, Hongtao Tian, Zhumin Chen, Zhaochun Ren, and Maarten de Rijke. 2021. Wizard of Search Engine: Access to Information Through Conversations with Search Engines. *arXiv preprint arXiv:2105.08301* (2021).