Xulin Zhou

Nagoya University
Nagoya, Aichi
Japan

zhou.xulin.j3
@s.mail.nagoya-u.ac.jp

1 Research interests

I am interested in dialogue systems where the user and the system collaboratively work on tasks through conversation. Currently, dialogue systems primarily aim to respond to human requests, functioning mainly as assistants or tools. However, if a dialogue system that engages in creative collaboration on an equal footing with humans can be developed, it would lead to more dynamic cooperation between humans and machines, fostering new types of interactions. To this end, I aim to create a dialogue system that focuses on cooperative collaboration and the expansion and development of ideas through dialogue.

1.1 Collection and Analysis of Dialogue Data for Collaborative Dialogue Systems

I analyze dialogues where two parties collaborate through conversation. Specifically, I focus on collaborative work that produces outcomes without a correct answer.

Several tasks have been used in studying dialogue systems that engage in collaborative work through conversation (Mitsuda et al., 2022). In the map task, one party holds a complete map while the other holds a partial map, and they complete the map through dialogue (Meena et al., 2014). In the OneCommon task, each speaker views one of two images cut from the same original image and identifies the common element by discussing the arrangement of dots of different colors and sizes (Udagawa and Aizawa, 2021). These tasks involve minimal creative elements and focus on finding a predetermined answer through collaboration. However, several real-world tasks do not have a single correct answer. I believe that building dialogue systems capable of collaboratively tackling such open-ended tasks will enhance their applicability. Therefore, I focus on tasks where there is no predetermined answer, aiming to develop dialogue systems for collaborative work.

Our group focuses on the collaborative creation of taglines (Zhou et al., 2024). Specifically, we are working on what we call the tagline co-writing task, where participants discuss and collaboratively edit taglines for given products.

I have created a tagline co-writing dialogue corpus that aims to gain insights useful for building collaborative dialogue systems and provide data for fine-tuning large language models. The corpus includes dialogues of humans performing the tagline co-writing task, the state of collaborative work during the conversations, and self-evaluations of the participants regarding the created taglines, their work, and their feelings through questionnaires.

I have analyzed the corpus by clustering utterances, extracting frequently occurring phrases in tasks with high self-evaluations, examining the workflow of utterances and edits over the entire task duration, and analyzing the interplay between utterances and taglines. The results indicate that tasks where utterances and tagline edits are conducted in parallel throughout the task tend to receive higher self-evaluations from the participants. Additionally, expressions of gratitude, positive evaluations, conveying understanding, and seeking agreement were found to be important in the tagline co-writing task.

1.2 Building a System for Collaborative Operations through Dialogue

As a prototype for a co-writing dialogue system, we developed a dialogue system using large language models (Zhou et al., 2023). This system combines a next-utterance generation model and a tagline generation model, both trained on data from the tagline co-writing dialogue corpus with a tagline evaluation model trained on data where third parties evaluated the taglines. We performed a dialogue experiment using the system and found that the system exhibited behaviors that are not normally observed between humans, such as overwriting changes the other party has made to the taglines without stating that to the other party. Additionally, issues such as the lack of coherence between utterances and edits, insufficient diversity in taglines, and hallucinations were observed. As future work, we intend to integrate highly advanced generative models such as GPT-4 to construct dialogue systems and to conduct evaluations on the components necessary for a dialogue system designed for the tagline co-writing task, considering that current issues may be resolved by utilizing models with higher performance. Additionally, we aim to consider new metrics for evaluating the tagline co-writing dialogue systems, such as the similarity to human-human interactions in terms of

The 20th Annual Meeting of the Young Researchers' Roundtable on Spoken Dialogue Systems, pages 48–49
September 16–17, 2024.

conversational dynamics.

The timing of utterances and edits is also considered. In the tagline co-writing dialogue corpus, we asked each participant to make utterances and edit taglines at arbitrary timings to promote natural interactions. In contrast, the prototype system we developed employs a turn-based system where turns for utterances and tagline edits alternate between the human and the system, which creates a gap between the system's behavior and actual human behavior. By aligning the system's behavior more closely with the natural, non-turn-based behavior of humans, we aim to build a dialogue system that can act in a more timely manner.

2 Spoken dialogue system (SDS) research

I believe that in the near future, the domain that will lead to the broader and more widespread use of dialogue systems is customer service interactions. Research on dialogue systems for customer service covers various types, including product recommendations and customer support (Gao et al., 2021; Jia et al., 2022).

One of the current shortcomings of customer service dialogue systems is their lack of understanding of the surrounding environment. Enabling these systems to utilize the information that humans can naturally observe will lead to dialogues grounded in the physical world. Since vision plays a significant role in human cognition, utilizing multimodal information such as video is crucial for the advancement of dialogue systems. For instance, when considering a customer service dialogue system providing tour guidance, it could potentially offer services based not only on general information such as written guide books, but also on changing surroundings.

3 Suggested topics for discussion

I would like to discuss the following topics:

- What are the efficient methods for evaluating constructive discussions?
- What are the important elements required for a dialogue system to collaborate with humans effectively?
- How should we evaluate aspects of dialogue systems where the inter-annotator agreement among human evaluators is not high?

4 Acknowledgments

This work was supported by JSPS KAKENHI Grant number 19H05692 and JST Moonshot R&D Grant number JPMJMS2011.

References

Chongming Gao, Wenqiang Lei, Xiangnan He, Maarten de Rijke, and Tat-Seng Chua. 2021. Advances and challenges in conversational recommender systems: A survey. *AI Open* 2:100–126.

Meihuizi Jia, Ruixue Liu, Peiying Wang, Yang Song, Zexi Xi, Haobin Li, Xin Shen, Meng Chen, Jinhui Pang, and Xiaodong He. 2022. E-ConvRec: A large-scale conversational recommendation dataset for E-commerce customer service. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. pages 5787–5796.

Raveesh Meena, Gabriel Skantze, and Joakim Gustafson. 2014. Data-driven models for timing feedback responses in a map task dialogue system. *Computer Speech & Language* 28(4):903–922.

Koh Mitsuda, Ryuichiro Higashinaka, Yuhei Oga, and Sen Yoshida. 2022. Dialogue collection for recording the process of building common ground in a collaborative task. In *Proceedings of the 13th Conference on Language Resources and Evaluation*. pages 5749–5758.

Takuma Udagawa and Akiko Aizawa. 2021. Maintaining common ground in dynamic environments. *Transactions of the Association for Computational Linguistics* 9:995–1011.

Xulin Zhou, Takuma Ichikawa, and Ryuichiro Higashinaka. 2023. A prototype dialogue system for co-writing taglines with users. In *Proceedings of the Human-Agent Interaction Symposium 2023*. (in Japanese).

Xulin Zhou, Takuma Ichikawa, and Ryuichiro Higashinaka. 2024. Collecting and analyzing dialogues in a tagline co-writing task. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*. pages 3507–3517.

Biographical sketch

Xulin Zhou is a Ph.D. student at the Graduate School of Informatics, Nagoya University. She is supervised by Prof. Ryuichiro Higashinaka. She is interested in dialogue systems that can collaborate with humans.