Hiroki Onozeki

The University of Electro-Communications
Chofu, Tokyo
Japan

`o2330023@edu.cc.uec.ac.jp`

1 Research interests

My research interests lie in the area of **building a dialogue system to generate interesting and entertaining responses**, with a particular focus on knowledge-grounded dialogue systems. Study of open-domain dialogue systems seeks to maximize user engagement by enhancing specific dialogue skills. To achieve this goal, much research has focused on the generation of empathetic responses, personality-based responses, and knowledge-grounded responses (Algherairy and Ahmed, 2024). In addition, interesting responses from the open-domain dialogue systems can increase user satisfaction and engagement due to their diversity and ability to attract the user's interest. Interesting responses are defined here as those that contain facts not generally well known but that provide surprise and engage the user. It has also been observed in task-oriented dialogue, user engagement can be increased by incorporating interesting responses into the dialogue. For example, Vicente et al. (2023) incorporated interesting responses into the spoken dialogue systems (SDSs) to support the user in performing complex tasks, making the experience pleasant and enjoyable for the user. However, even in the case of interesting responses, if the dialogue is incoherent, user engagement is likely to be significantly reduced. To create a dialogue system that is consistent and interesting in a dialogue context, I am working on using knowledge-grounded response generation methods to select interesting knowledge that is relevant to the dialogue context and to make responses that are based on that knowledge.

1.1 Introducing interesting knowledge into dialogue systems

Several studies have been conducted to investigate the use of interesting knowledge in dialogues to achieve engaging dialogues. Konrád et al. (2021) built a dialogue system that uses interesting knowledge obtained from crawling from Reddit with high similarity to the dialogue context in the dialogue for improving user engagement in open-domain dialogue. To incorporate the acquired knowledge in a conversational format, the method generated a follow-up question and connected it to the knowledge. However, in this approach, the timing of the insertion of interesting knowledge into the dialogue is deter-

mined by rules, and the content response does not take into account the dialogue context, resulting in unnatural dialogue. Vicente et al. (2023) proposed a method of introducing interesting knowledge into a spoken dialogue system to help users perform complex tasks using templates. In this approach, interesting knowledge gathered from web searches is introduced into the dialogue via a template, producing a dialogue that lacks naturalness and coherence for the dialogue context. To address these challenges, I am working on selecting appropriate knowledge, taking into account both the dialogue context and interestingness, and generating responses based on knowledge without using templates. This will enable the generation of natural and interesting responses that are consistent with the dialogue and will improve user satisfaction.

1.2 Knowledge-grounded dialogue systems

Knowledge-grounded dialogue systems are approaches that generate responses that are based on external knowledge relevant to the dialogue, and can generate diverse and informative responses. Knowledge-grounded dialogue systems basically consist of two modules: knowledge selection and response generation (Wang et al., 2023). The knowledge selection module selects knowledge for use in the next response from the candidate knowledge related to the dialogue, and the response generation module generates a response that is based on the content of the retrieved knowledge and the dialogue context. Kim et al. (2020) built a model of knowledge selection with continuous latent variables modeling past knowledge selection. Zhao et al. (2020) proposed an unsupervised approach to jointly optimize knowledge selection and response generation using a prior learning model. However, most existing methods mainly perform knowledge selection by considering only the dialogue context, resulting in responses that contain much general information and are uninteresting (Xu et al., 2023). To address this, Xu et al. (2023) modeled a shift in dialogue topics and built a model for selecting a variety of knowledge while remaining consistent with the dialogue context. Generating responses that are consistent and interesting in the dialogue context is considered necessary to build a dialogue system that is close to human and engaging. To this end, I am working on a model that es-

timates the appropriate interestingness of the knowledge for use in a response to select knowledge that is based on this interestingness and the context of the dialogue. As responses that contain general content are generally preferred to interesting content at the beginning of a dialogue and topic switches, it is important to capture topic switches in a dialogue to estimate the appropriate interestingness of the knowledge used in a response.

1.3 Trade-off between fidelity and consistency

A dialogue system that produces responses that are engaging and interesting to the user must provide responses faithful to knowledge and consistent with the context of the dialogue. However, there is a recognized trade-off between generating responses that are consistent with the dialogue context and faithful to knowledge (Chawla et al., 2024). Rashkin et al. (2021) proposed a method of improving fidelity to knowledge by adding control tokens to the beginning of the model input. The results showed that improving fidelity to knowledge may sacrifice consistency in the dialogue context. Chawla et al. (2024) built an approach to generating responses that balance fidelity and consistency by planning the content of the responses for generation and then creating a response generation model. It is more important to generate responses that balance fidelity to knowledge and consistency with the dialogue context than to focus solely on generating responses that reduce the generation of hallucinations and are faithful to knowledge.

2 Spoken dialogue system (SDS) research

Due to the advent of large language models, text dialogue systems can now generate natural and fluent responses that are close to those of humans. Thus, it is expected that SDSs can be studied more actively and used in a wide range of aspects of society, such as restaurant reservations, product recommendations, and counseling. In particular, multimodal dialogue systems have attracted particular attention recently because they enable close communication with humans by using user information such as facial expressions and voice information. However, several challenges have to be resolved before SDSs can be used in many fields of society.

First, there are insufficient datasets to train SDSs. Audio data collection is much more costly and time-consuming than text data, resulting in a smaller number of datasets and smaller-size datasets. In particular, there is a lack of non-English audio datasets. This lack of datasets directly prevents SDSs from being used in a wide variety of situations. It is necessary to find a way to train each module using independent text, audio, and video data and combine them to build a model.

Further, SDSs need to reduce response time relative to text dialogue. Long response times are unnatural and pro-

vide the user a sense of distrust. The larger the model that generates the response, the more fluent it is, but the larger the computational resources required, the longer it takes to generate the response. Also, multimodal dialogue systems need to process voice and image information rather than text information, which takes time to respond. To increase user engagement, we must find ways to reduce response time.

Furthermore, in the use of SDSs in society, it is essential to reduce hallucinations. Responses containing incorrect information when making restaurant reservations or recommending products are a serious problem. The occurrence of hallucinations is thought to be the main reason that SDSs are not still widely used in society today. Methods that do not produce hallucinations with high reliability and methods that detect responses including hallucinations will be particularly important in the future.

3 Suggested topics for discussion

I suggest discussing the following topics:

- What methods can reduce the time to generate knowledge-grounded responses in SDSs?

- How should multimodal information of users such as facial expressions be used in knowledge-grounded dialogues?

- Are existing methods for assessing response diversity adequate? How can response diversity be appropriately and automatically assessed?

References

Atheer Algherairy and Moataz Ahmed. 2024. A review of dialogue systems: current trends and future directions. *Neural Computing and Applications* 36(12):6325–6351.

Kushal Chawla, Hannah Rashkin, Gaurav Singh Tomar, and David Reitter. 2024. Investigating content planning for navigating trade-offs in knowledge-grounded dialogue. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, pages 2316–2335.

Byeongchang Kim, Jaewoo Ahn, and Gunhee Kim. 2020. Sequential latent knowledge selection for knowledge-grounded dialogue. In *International Conference on Learning Representations*.

Jakub Konrád, Jan Pichl, Petr Marek, Petr Lorenc, Van Duy Ta, Ondrej Kobza, Lenka Hýlová, and Jan Sedivý. 2021. Alquist 4.0: Towards social intelligence using generative models and dialogue personalization. *CoRR* abs/2109.07968.

Hannah Rashkin, David Reitter, Gaurav Singh Tomar, and Dipanjan Das. 2021. Increasing faithfulness in knowledge-grounded dialogue with controllable features. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, pages 704–718.

Frederico Vicente, Rafael Ferreira, David Semedo, and Joao Magalhaes. 2023. The wizard of curiosities: Enriching dialogues with fun facts. In *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, pages 149–155.

Ming Wang, Bo Ning, and Bin Zhao. 2023. A review of knowledge-grounded dialogue systems. In *2023 8th International Conference on Image, Vision and Computing (ICIVC)*. pages 819–824.

Lin Xu, Qixian Zhou, Jinlan Fu, and See-Kiong Ng. 2023. Cet2: Modelling topic transitions for coherent and engaging knowledge-grounded conversations. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 31:3527–3536.

Xueliang Zhao, Wei Wu, Can Xu, Chongyang Tao, Dongyan Zhao, and Rui Yan. 2020. Knowledge-grounded dialogue generation with pre-trained language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, pages 3377–3390.

Biographical sketch

Hiroki Onozeki is a master's student at the Graduate School of Informatics and Engineering, The University of Electro-Communications. He is interested in knowledge-grounded dialogue systems. He has participated in several competitions building dialogue systems, including Dialogue Robot Competition 2023, Dialogue System Live Competition 6, and AIWolfDial2024jp.