

## 1 Research interests

The author's objective centers around developing a spoken dialogue system (SDS) that can emulate the cognitive and conversational qualities of a human friend. Key attributes such as empathy, knowledge/causality reasoning, and personality are integral components of human interaction. The proposed approach involves the creation of an **Empathy-enriched SDS**, capable of comprehending human emotions and circumstances, thus providing companionship and assistance akin to a trusted friend. Additionally, the **Causality-reasoning for SDS** aims to ground the system in commonsense knowledge and equip it with the ability to reason about causalities, such as predicting user desires/reactions and system intentions/reactions, thereby enhancing the system's intelligence and human-like behavior. Finally, the concept of a **Personality-conditioned SDS** involves enabling systems to exhibit distinct personalities, further enhancing the naturalness of human-robot interaction.

### 1.1 Empathy-enriched SDS

Incorporating empathy into the dialogue system is essential for improving human-robot interaction experiences, as empathy is the emotional bonding among humans; robots expressing empathy would give humans a feeling of being understood and satisfied with the conversation. To produce an empathetic response, the generative models encounter the problem of generating safe responses (generic and meaningless, such as 'I see'), or unnatural responses (have grammatical or logical errors, such as 'that is so sweet. I am sorry to hear that'). Instead, the retrieval-based models are guaranteed to produce natural and empathetic responses, as they are retrieved from external documents, but encounter the problem of producing responses that are not closely relevant to the dialogue context. In order to address the aforementioned challenges, the author proposed to combine a VAE-based response generation model with a retrieval system based on emotion recognition. Additionally, the proposed approach incorporates the use of multi-modal facial expressions by the virtual agent to enhance the vividness of empathy. This combined methodology is subsequently applied in human-robot interaction experiments to evaluate its effectiveness.

### 1.2 Causality reasoning for SDS

Integrating commonsense knowledge into the SDS can significantly enhance the system's expertise and enable it to deliver informative responses, thereby serving as a valuable human life assistant. However, in order to achieve a higher level of human likeness, the causality reasoning capabilities of SDS are also essential. In particular, the ability to generate responses that cater to human satisfaction relies on accurate prediction of user desires/reactions from the user's standpoint, as well as the ability to reason about the system's intentions/reactions from a perspective that closely mimics human behavior.

With the advent of large language models (LLMs) such as GPT-3 and ChatGPT, Bang et al. (2023b) introduced ChatGPT's potential in causal reasoning based on whether the model can make a judgment on correct causes or effects. However, existing evaluations primarily focus on assessing the LLMs' capacity to recognize causes or effects from the user's perspective, rather than generating causality explanations from the view of both the user and the system. In this study, the author initially evaluated the ability of LLMs for causality explanation generation and subsequently proposed an approach to enhance this capability through the integration of in-context learning and commonsense reasoning, which considers the system's intention and reaction, along with the user's desire and reaction.

### 1.3 Personality-conditioned SDS

Personality refers to the unique set of enduring traits, patterns of thoughts, feelings, and behaviors that characterize an individual. A personality-conditioned dialogue system that exhibits distinct personalities, can create a more human-like conversational experience, fostering a sense of rapport and understanding with the users. The expression of personality is contingent upon the situation. For example, people may be more inclined to openly express their thoughts, feelings, and experiences among close friends, while in a formal or initial conversation, such expression tends to be politer and more subtly implied. Furthermore, individuals with diverse personality traits tend to exhibit distinct empathetic styles in their responses (Richendoller et al. et al (1994)). Extroverts, for example, may frequently employ positive emotional language and show perspective-taking compared to intro-

verts. By incorporating personality-based empathetic responses, more gratifying conversations can be achieved. Therefore, the author actively conducts ongoing research in the field of personality-conditioned SDS, which aims to develop spoken dialogue systems that adapt their responses based on individual personality traits.

## 2 Spoken dialogue system (SDS) research

In the forthcoming years, two potential directions for the SDS community could involve the automated evaluation of subjective aspects within SDS and developing personalized SDS to cater to users with diverse personalities.

### 2.1 Trustable evaluations for SDS

In the field of open-domain dialogue systems, evaluation is commonly conducted using automatic metrics and human judgments. However, automatic metrics, such as BLEU, METEOR, and ROUGE, are based on word overlap and struggle to capture the diverse nature of dialogue systems. On the other hand, human judgments are more reliable, but expensive and lack standardized protocols. Hence, there exists a necessity to combine the merits of automated and human evaluations while mitigating their respective drawbacks. Inspired by Giorgi et al. (2023a) who proposed human-centered metrics (such as emotion, and personality) for dialog system evaluation, hierarchical evaluation of spoken dialogue systems (SDS) represents a possible approach to effectively quantify system performance. For instance, at the utterance level within a conversation, to evaluate the "relatedness," "fluency," and "informativeness" of the responses. Furthermore, at the conversation level, it is crucial to evaluate whether the responder demonstrates a distinct personality and exhibits empathy appropriately. Lastly, at the system level, the evaluation should consider the system's ability to maintain robustness across interactions with users possessing diverse personalities. However, since all the above evaluation aspects are subjective, the research of suitable automated metrics requires further exploration.

### 2.2 User-adaptable SDS

Humans with different personalities have varied preferences for systems personalities, therefore, a personalized SDS that is adaptable to the user's personality is essential to improve human-robot interactions. This involves a three-step process: firstly, accurately detecting the user's personality, the accuracy of personality (such as big-five traits) recognition is still not good even with the assistance of LLMs; secondly, exploring the mapping between user personality traits and corresponding system personalities in both chit-chatting and task-oriented domains; and finally, personalized response generation which is tailored to the user's unique personality. Moreover, for the purpose of achieving a higher level of human likeness, it

is crucial to incorporate a fusion of verbal and non-verbal response generation techniques within the system. This entails the inclusion of elements such as backchannels, fillers, pitch variations, facial expressions, and other relevant non-verbal cues. By incorporating these steps, the SDSs are expected to effectively cater to the individual preferences of users, thereby improving the overall interaction experience.

## 3 Suggested topics for discussion

- What are the subsequent advancements in common-sense/knowledge reasoning for SDS, and how can they be effectively applied to various SDS tasks?
- The interplay between personality and emotion in personality recognition and response generation.
- In what ways can an empathy-enriched SDS contribute to the treatment and management of mental disorders?

## Acknowledgements

This work was supported by JSPS KAKENHI (JP19H05691), and JST Moonshot R&D (JPMJPS2011). This work was also supported by JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2123.

## References

- Richendoller Nadine R et al. 1994. Exploring the links between personality and empathic response style. *Personality and Individual Differences* 17(3):303–311.
- Salvatore Giorgi et al. 2023a. Human-centered metrics for dialog system evaluation. *arXiv preprint arXiv:2305.14757*.
- Yejin Bang et al. 2023b. A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint arXiv:2302.04023*.

## Biographical sketch



Yahui Fu is currently pursuing a Ph.D. degree at the Department of Intelligence Science and Technology, Kyoto University, Kyoto, Japan. She received M.S. degrees from both Tianjin University, Tianjin, China, and the Japan Advanced Institute of Science and Technology, Ishikawa, Japan, in 2021. Additionally, she is currently a research intern at rinna Co., Ltd. in Japan. Her research interests include spoken dialogue systems and multimodal emotion recognition.