Ye Liu

Mercedes-Benz AG & Ulm University
Benz-Straße
71063 Sindelfingen
Germany

ye.y.liu@mercedes-benz.com

1 Research interests

My research work centers on how to enable a human-like interaction through generating contextual, emotional or proactive responses, both in task-oriented and in chit-chat spoken dialogue systems (SDSs), because natural language generation (NLG) is an indispensable component in SDSs and can directly affect the user interactive experience of the entire dialogue system. In addition to NLG, I am also interested in natural language understanding (NLU), as it plays a crucial role in SDSs and is a prerequisite for dialogue systems to generate replies.

1.1 Commonsense enabled conversational model

Many pre-trained transformer-based (Vaswani et al., 2017) language models (LMs) have been widely applied in SDSs and shown promising performance. However, the probing experiments in Zhou et al. (2021) demonstrated that pre-trained LMs (Zhang et al., 2020; Roller et al., 2021; Lewis et al., 2020) fail to capture commonsense (CS) knowledge hidden in dialogue utterances, even though they were already pre-trained with numerous datasets.

To improve the CS understanding and reasoning ability of a pre-trained model and to build a dialogue agent like shown in Figure 1, we firstly inject external knowledge into a pre-trained conversational model to establish basic commonsense. Secondly, we leverage this integrated commonsense capability to improve open-domain dialogue response generation so that the dialogue agent is capable of understanding the CS knowledge hidden in dialogue history on top of inferring related other knowledge to further guide response generation (Liu et al., 2022a).

1.2 System-initiated transitions in unified SDSs

SDSs have been separately developed under two different categories, task-oriented and chit-chat. The former focuses on achieving functional goals and the latter aims at creating engaging social conversations without special goals. Creating a unified conversational model that can engage in both chit-chat and task-oriented dialogues is a promising research topic in recent years. We investigate the "initiative" that occurs when there is a transition from chit-chat to task-oriented in one dialogue and develop proactive capabilities for unified models to be able to initiate this transition through generating a transition sentence (Liu et al., 2023b).

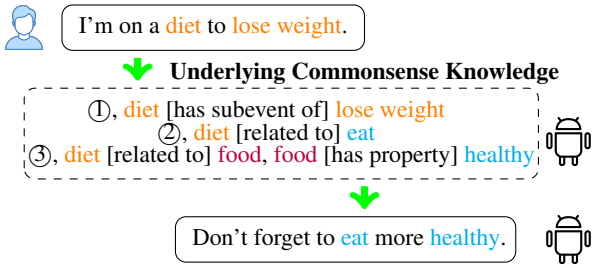

Figure 1: The ideal dialogue agent can understand the CS knowledge hidden in the dialogue history (①), meanwhile, infer the reasonable CS knowledge (② and ③) for further guiding an informative response generation.

We firstly build a *transition info extractor* (TIE) that keeps track of the preceding chit-chat interaction and detects the potential user intention to switch to a task-oriented service. Meanwhile, in the unified model, a *transition sentence generator* (TSG) is extended through efficient Adapter tuning and transition prompt learning. When the TIE successfully finds task-related information from the preceding chit-chat, such as a transition domain ("train") or transition value ("London Kings Cross"), then the TSG is activated automatically in the unified model to initiate this transition by generating a transition sentence under the guidance of transition information extracted by TIE (like "If you want, I can look for a train to London Kings Cross for you."). This proactivity is beneficial for commercial dialogue systems to actively sell their task-related services (Chiu et al., 2022; Liu et al., 2022b).

2 Spoken dialogue system (SDS) research

- **How to tackle hallucinations in large generative models, such as ChatGPT?** Compared with its predecessors, like GPT-2, the ChatGPT improved ability to generate more reasonable replies in various

Proceedings of the 19th Annual Meeting of the Young Reseachers' Roundtable on Spoken Dialogue Systems, pages 59–61
September 11–12, 2023.

contexts. However, it is difficult to completely eliminate the hallucinating generations even with ChatGPT or GPT-4, especially when dealing with complex topics. In the future, reducing hallucinations effectively might be a persistent challenge, as it is related to the inherent properties of neural network architectures. In addition to model development, we can apply some post-processing technologies to identify and remove hallucinations from the generated output.

- **How to enhance LLMs with knowledge graphs (KGs)?** Along with the introduction of LLMs, people are more interested in integrating external knowledge, such as knowledge graphs (KGs), into LLMs to enhance its performance, especially for fact-aware or question answering (QA) tasks. Yang et al. (2023) provides a comprehensive review for KGs enhanced pre-trained LMs and proposes some possible research directions. From my perspective, it is crucial to consider how KGs can be incorporated into dialogue-based generative models. Given the impressive performance of ChatGPT, it is worth to explore to what extent external knowledge can be effectively exploited.

3 Suggested topics for discussion

I suggest discussing the following topics:

- **Chances and challenges to SDS research community along with the launch of ChatGPT:** Since its release at the end of 2022, ChatGPT has received significant attention from both industry and academia. This surge of interest has led to a growing number of researcher to devote themselves into the study of large language models (LLMs). Meanwhile, we have also witnessed many surprising and amazing applications for these models, such as Microsoft 365 Copilot. Despite the promising opportunities, young researchers also encounter various challenges. Because a series of ChatGPT and GPT-4 models are no longer publicly available, they are not easily accessible to young researcher. Even if we have access, do we have sufficient computing resources to run these LLMs? On the other hand, there is a need to reconsider the development of SDS, such as for emotional chatbot, we previously explicitly predict emotions in user utterances and leverage this information to enable empathetic responses. However, the question arises now if it is necessary to put in the effort to explicitly detect user emotions and improve the accuracy of emotion detection. Because with advanced capabilities, ChatGPT have demonstrated the ability to perceive user emotions and generate appropriate responses accordingly (Elyoseph et al., 2023) even without predicting user emotions.

- **Understanding ability of ChatGPT:** ChatGPT and its predecessor GPT-2 are both auto-regressive generation models. However, the ChatGPT has shown impressive capability in understanding a wide range of topics, which underlies its remarkable performance on generating human-like responses. Some academic studies have started to investigate and evaluate the logical reasoning ability of ChatGPT and GPT-4 (Liu et al., 2023a; Zhong et al., 2023; Zhao et al., 2023). Hence, there are some follow-up questions, like how can we accurately evaluate the understanding ability of these large generative models? Furthermore, do we underestimate the performance of these large generative models in terms of its understanding ability?

- **Evaluation in LLMs:** To assess the performance of LLMs, many researchers subject ChatGPT to various Benchmarks (Zhong et al., 2023; Bang et al., 2023). Zhao et al. (2023) explores the emotional dialogue capabilities in ChatGPT and finds that metric results may not necessarily reflect its poor understanding. One potential reason is significant discrepancy between its prediction standard and annotation standard. When it comes to generation tasks, human evaluation is commonly viewed as the best reliable way to evaluate NLG systems, but come with many issues, such as costly and time consuming and human judgement bias (Celikyilmaz et al., 2020). However, some papers (Chiang and Lee, 2023) investigate the possibility of using LLMs to be an alternative to human evaluation.

References

Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, et al. 2023. A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint arXiv:2302.04023* .

Asli Celikyilmaz, Elizabeth Clark, and Jianfeng Gao. 2020. Evaluation of text generation: A survey. *arXiv preprint arXiv:2006.14799* .

Cheng-Han Chiang and Hung-yi Lee. 2023. Can large language models be an alternative to human evaluations? *arXiv preprint arXiv:2305.01937* .

Ssu Chiu, Maolin Li, Yen-Ting Lin, and Yun-Nung Chen. 2022. Salesbot: Transitioning from chit-chat to task-oriented dialogues. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. pages 6143–6158.

Zohar Elyoseph, Dorit Hadar-Shoval, Kfir Asraf, and Maya Lvovsky. 2023. Chatgpt outperforms humans in emotional awareness evaluations. *Frontiers in Psychology* 14:1199058.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. pages 7871–7880.

Hanmeng Liu, Ruoxi Ning, Zhiyang Teng, Jian Liu, Qiji Zhou, and Yue Zhang. 2023a. Evaluating the logical reasoning ability of chatgpt and gpt-4. *arXiv preprint arXiv:2304.03439* .

Ye Liu, Wolfgang Maier, Wolfgang Minker, and Stefan Ultes. 2022a. Conceptnet infused dialogpt for underlying commonsense understanding and reasoning in dialogue response generation. In *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue - Poster Abstracts*. SEMDIAL, Dublin, Ireland.

Ye Liu, Stefan Ultes, Wolfgang Minker, and Wolfgang Maier. 2023b. Unified conversational models with system-initiated transitions between chit-chat and task-oriented dialogues. *arXiv preprint arXiv:2307.01664* .

Ye Liu, Yung-Ching Yang, Wolfgang Maier, Wolfgang Minker, and Stefan Ultes. 2022b. On system-initiated transitions in a unified natural language generation model for dialogue systems. In *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue - Poster Abstracts*. SEMDIAL, Dublin, Ireland.

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, et al. 2021. Recipes for building an open-domain chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. pages 300–325.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30.

Linyao Yang, Hongyang Chen, Zhao Li, Xiao Ding, and Xindong Wu. 2023. Chatgpt is not enough: Enhancing large language models with knowledge graphs for fact-aware language modeling. *arXiv preprint arXiv:2306.11489* .

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and William B Dolan. 2020. Dialogpt: Large-scale generative pre-training for conversational response generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. pages 270–278.

Weixiang Zhao, Yanyan Zhao, Xin Lu, Shilong Wang, Yanpeng Tong, and Bing Qin. 2023. Is chatgpt equipped with emotional dialogue capabilities? *arXiv preprint arXiv:2304.09582* .

Qihuang Zhong, Liang Ding, Juhua Liu, Bo Du, and Dacheng Tao. 2023. Can chatgpt understand too? a comparative study on chatgpt and fine-tuned bert. *arXiv preprint arXiv:2302.10198* .

Pei Zhou, Pegah Jandaghi, Hyundong Cho, Bill Yuchen Lin, Jay Pujara, and Xiang Ren. 2021. Probing commonsense explanation in dialogue response generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*. pages 4132–4146.

Biographical sketch

Ye Liu is a PhD candidate at Mercedes-Benz AG, in cooperation with Ulm University. She is supervised by Dr. Wolfgang Maier, Prof.Dr.-Ing. Stefan Ultes (University of Bamberg) and Prof.Dr.Dr.-Ing. Wolfgang Minker (Ulm University). She holds double master of science degrees from Tongji University, Shanghai and Technical University of Munich, Germany. She has project experience in automatic speech recognition(ASR) and neural machine translation (NMT) before PhD study. Now her expertise spans over natural language understanding (NLU) and natural language generation (NLG) in spoken dialogue systems (SDSs).