Research Statement

Lizi Liao
School of Computing and Information Systems, Singapore Management University
Tel: (65) 6828-4882; Email: lzliao@smu.edu.sg
22 (Day) 12 (Month) 2023 (Year)

Background

My research explores two questions: What are the underlying principles of humans understanding conversation context as well as making proper responses, and how we can implement them on machine learning models? Research on this topic has to necessarily be at the intersection of Machine Learning, Natural Language Processing and Multimedia. In my lab, we are specifically interested in task-oriented dialogues, proactive conversational agents, and multimodal conversational search and recommendation as the application target.

My research emphasizes on broader types of ‘understand’ the user and ‘respond’ to the user under certain context:

- User Modeling and Interactive Understanding: how to better understand user preference from their past activity history, from their social context; how to accurately solicit user requirements from direct interaction with the user, by performing tasks like dialogue state tracking, multimodal understanding, etc.
- Respond to Context: how to intelligently respond to the user based on world knowledge reasoning; how to effectively complete certain tasks via strategy modeling; how to meet users’ needs, from constrained task completion to more general information consumption and decision support.

Research Areas

1. User Modeling and Interactive Understanding
   Modeling from Static History. The huge amount of static user history left on the web provides a good foundation for performing user modeling. Collaborative filtering has been the most widely adopted principle to infer user preference from many others. As compared to the common paradigm like matrix factorization, we developed neural collaborative filtering to learn embeddings and perform predictions on them [1]. It is among the first to employ deep learning for recommendation systems and gain wide recognition from the community. Beyond collaborative filtering, we also looked into user’s social context to learn more accurate user representations. We proposed a generic social network embedding framework to preserve both the structural proximity and attribute proximity of users [2, 4]. It captures one of the most striking and robust empirical regularities of social life – homophily principle. Furthermore, we also tried to study the lexical variation of people of different ages to better understand them [5].

   Understand Interactive Inputs. Although leveraging static history such as social connection, purchase records, or implicit feedback can work as powerful tool to understand user, it might still return us a biased or incomplete view of user preference. Within a specific situation or context, user preference would be dynamic, detailed and sometimes even contradictory to his or her history. Therefore, tracking user preference through current interaction is essential. Common practice for dialogue state tracking (DST) has been to treat it as classifying the whole dialogue history into a set of pre-defined slot-value pairs, or generating values for different slots. Both have limitations in considering dependencies that occur in dialogues, and are lack of reasoning capabilities over turns. We proposed to track dialogue states turn-by-turn and reason over turns with the help of the back-end database [7]. Moreover, due to the interactive and progressing nature of dialogue, to combat the error accumulation conundrum, we developed a recursive inference mechanism to resolve DST in multi-domain scenarios that call for more robust and accurate tracking ability [11].

Multimodal Understanding. Although natural language might be the primary API for communication with people, understanding signals from other modalities, such as visual context is also critical. Compared to traditional text-based systems, multimodal dialogue enables users to easily provide an image sample instead of racking their minds for an appropriate text description, such as in search of fashion products. At the same time, it is more straight-forward for users to perceive information from system provided images rather than text based on supposition. Correspondingly, we presented the first neural multimodal belief tracker to demonstrate how multimodal evidence can facilitate semantic understanding and dialogue state tracking [13]. It investigates sub-regions of image to learn visual concepts and models user's behavior patterns for more accurate tracking performance, which fills the semantic gap between different modalities and avoids being misled by strong language priors. In the specific fashion domain, we further bridge the gap between textual and visual modalities to achieve interpretable cross-modal retrieval with attribute feedback [3]. We incorporated the structural knowledge from domain taxonomy into the deep learning framework, which facilitated the interactive reasoning of search results and user intent. Moreover, in recognition of the central importance of knowledge to detailed understanding tasks, we further proposed a weak-label modeling module to automatically harvest fashion knowledge from social media [12]. We unified the three tasks of occasion, person, and clothing discovery from online sources in multiple information modalities.

2. Respond to Context

Knowledge-aware Generation. Generating appropriate responses for satisfactory task completion is the ultimate goal of task-oriented dialogue agents. Although multimodal conversational agents show various advantages in helping users, it is non-trivial to make it "smart" in generating substantive answers. Therefore, we presented a knowledge-aware multimodal dialogue model to address the limitation of text-based dialogue systems [9]. It understands fine grained semantics in product images and is aware of fashion style tips. The key idea is that the agent conditions answers based not only on conversation history, but also on the extracted knowledge that are relevant to the current context. Beyond the fashion domain, we also investigated a new solution towards building a crowd-sourced knowledge enhanced multimodal conversational system for travel. It aims to assist users in completing various travel-related tasks, such as searching for restaurants or things to do [6]. We ground this research on the combination of multimodal understanding and recommendation techniques which explores the possibility of a more convenient information seeking paradigm.

Strategy Modeling. In the interactive process of dialogue for task completion, strategy modeling plays an important role for achieving better completion results. Reinforcement learning has been the most prevalent approach for task-oriented dialogue policy learning. It usually focuses on the target agent policy and simply treat the user's behavior habits or policy as part of the environment. While in real world scenarios, the behavior of the user often exhibits certain patterns or hidden policies, which can be inferred and utilized by the target agent to facilitate its own decision making. This strategy is common in human mental simulation by first imaging a specific action and the probable results before really acting it. We therefore propose a user behavior aware framework for policy learning I task oriented dialogues [14]. We explicitly estimate the user's policy from his past behavior and use this estimation to improve the target agent's policy. By incorporating the estimated user model output as part of the dialogue state, the target agent shows significant improvement on both cooperative and competitive task-oriented dialogues.

To further drive the progress of building multimodal conversational search and recommendation systems using data-driven approaches, we contributed a multimodal multi-domain conversational search (MMConv) dataset [8]. It provides a large-scale multi-turn conversational corpus with fully-annotated dialogues spanning across several domains and modalities. Enlightened by the current progress in dialogue research community, we adopt both state-of-the-art pipeline styled methods and end-to-end styled method from them, and applied to the MCSR scenario systematically. We analyzed the obtained results and discussed what was doable and what was still missing in MCSR. Moreover, under the specific conversational recommendation scenario, we built a topic guided conversational recommender to break the single task constraint in recent works and solicit evidence from back-end database to support candidates reasoning [10].

3. Future Work

MCSR is a new emerging topic which emphasizes on the interactivity coupled with search and recommendation to alleviate the information overload problem of users To make such a paradigm functional and available, there exist many challenges. For example, we need methods about how to handle the ambiguity or indirect inquiry of users; how to model the multimodal context and history; how to design appropriate interaction strategy; how to leverage domain and world knowledge how to generate fair and robust response in diverse situations; and we also see the issues of resources, methodologies and biases in evaluation etc. I am excited to look into these research opportunities and push forward the boundaries of this research area.

Specifically, I would lay out my research in three directions: 1) investigating user state modeling beyond slot filling. Currently, slot filling dominates task-oriented dialogue research to structure user requirements in the practical sense. However, there exists a huge gap between traditional dialogue research setting and MSCR, where users tend to express their intents much more freely. It is unrealistic to fix ontology or even just slots in advance. Also, user intention can be expressed in various ways such as via browsing history, click or purchasing records. Moreover, we also see different interaction behaviors like indirect inquiry which makes the situation even worse. Hence, I plan to explore more accurate and purpose-oriented user state representations that seamlessly bridge user's past (offline) preferences and current (online) requirements in conversation. It would facilitate timely conversation intervention in search sessions and enable accurate response generation in conversation. 2) building efficient and universal user simulators for automatic task-agnostic evaluation. Currently, most of the task-oriented dialogue systems are built and evaluated in a static corpus-based paradigm. It often leads to large performance gap when interacting with real users. Hence, human evaluation is widely applied, yet resulting in huge costs and poor reproducibility. In order to achieve comprehensive, fair and robust evaluation, a viable way is to build efficient and universal user simulators that take the place of humans to accomplish the interaction and judging process. This would largely facilitate dialogue and MCSR research and I am eager to explore along this direction. 3) human-in-the loop optimization for higher intelligence. Current dialogue research heavily relies on training model using human generated dialogue sessions. The intelligence acquired this way generally cannot exceed the intelligence of human. To overcome this, a viable way is to perform human-in-the-loop optimization for MCSR systems. We can vie big data as observation of world, and make use of these to enhance human perception hence tackle complex tasks. The tools such as search or recommendation algorithm itself may not be intelligent enough, but enhanced by collaboration with human to work on these world observation data, it would be able to provide us with better decision support. I plan to delve into this direction and get inspiration from human AI interaction research which has been studied by the HCI community for decades.

Selected Publications and Outputs

[1] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In Proceedings of the 26th international conference on world wide web, pages 173–182, 2017.

[2] Lizi Liao, Xiangnan He, Hanwang Zhang, and Tat-Seng Chua. Attributed social network embedding. IEEE Transactions on Knowledge and Data Engineering, 30(12):2257–2270, 2018.

[3] Lizi Liao, Xiangnan He, Bo Zhao, Chong-Wah Ngo, and Tat-Seng Chua. Interpretable multimodal retrieval for fashion products. In Proceedings of the 26th ACM international conference on Multimedia, pages 1571–1579, 2018.

[4] Lizi Liao, Qirong Ho, Jing Jiang, and Ee-Peng Lim. Slr: A scalable latent role model for attribute completion and tie prediction in social networks. In 2016 IEEE 32nd International Conference on Data Engineering, pages 1062–1073, 2016.

[5] Lizi Liao, Jing Jiang, Ying Ding, Heyan Huang, and Ee-Peng Lim. Lifetime lexical variation in social media. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, pages 1643–1649, 2014.

[6] Lizi Liao, Lyndon Kennedy, Lynn Wilcox, and Tat-Seng Chua. Crowd knowledge enhanced multimodal conversational assistant in travel domain. In International Conference on Multimedia Modeling, pages 405–418, 2020.

[7] Lizi Liao, Le Hong Long, Yunshan Ma, Wenqiang Lei, and Tat-Seng Chua. Dialogue state tracking with incremental reasoning. Transactions of the Association for Computational Linguistics, 9:557–569, 2021. 3

[8] Lizi Liao, Le Hong Long, Zheng Zhang, Minlie Huang, and Tat-Seng Chua. Mmconv: An environment for multimodal conversational search across multiple domains. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021.
[9] Lizi Liao, Yunshan Ma, Xiangnan He, Richang Hong, and Tat-Seng Chua. Knowledge-aware multimodal dialogue systems. In Proceedings of the 26th ACM international conference on Multimedia, pages 801–809, 2018.
[10] Lizi Liao, Ryuichi Takanobu, Yunshan Ma, Xun Yang, Minlie Huang, and Tat-Seng Chua. Topic-guided relational conversational recommender in multiple domains. IEEE Transactions on Knowledge and Data Engineering, 2020.
[11] Lizi Liao, Tongyao Zhu, Le Hong Long, and Tat-Seng Chua. Multi-domain dialogue state tracking with recursive inference. In Proceedings of the Web Conference 2021, pages 2568–2577, 2021.
[12] Yunshan Ma, Xun Yang, Lizi Liao, Yixin Cao, and Tat-Seng Chua. Who, where, and what to wear? extracting fashion knowledge from social media. In Proceedings of the 27th ACM International Conference on Multimedia, pages 257–265, 2019.
[13] Zheng Zhang, Lizi Liao, Minlie Huang, Xiaoyan Zhu, and Tat-Seng Chua. Neural multimodal belief tracker with adaptive attention for dialogue systems. In The World Wide Web Conference, pages 2401–2412, 2019.
[14] Zheng Zhang, Lizi Liao, Xiaoyan Zhu, Tat-Seng Chua, Zitao Liu, Yan Huang, and Minlie Huang. Learning goal-oriented dialogue policy with opposite agent awareness. In Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, pages 122–132, 2020.
[15] Y. Wu, L. Liao, X. Qian, and T.S. Chua, Semi-supervised New Slot Discovery with Incremental Clustering, in Findings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.
[16] C. Huang, Z. Zhang, H. Fei and L. Liao, Conversation Disentanglement with Bi-Level Contrastive Learning, in Findings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.
[17] D. Wan, Z, Zhang, Q. Zhu, L. Liao, M. Huang, A Unified Dialogue User Simulator for Few-shot Data Augmentation, in Findings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.
[18] C. Ye, L. Liao, F. Feng, W. Ji, and T.S. Chua, Structured and Natural Responses Co-generation for Conversational Search, in Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2022.
[19] Huy Quang Dao, Lizi Liao, Dung D. Le, Yuxiang Nie, Reinforced Target-driven Conversational Promotion, in Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.
[20] Jinggui Liang, Lizi Liao, ClusterPrompt: Cluster Semantic Enhanced Prompt Learning for New Intent Discovery, in Findings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.
[21] Yang Deng, Lizi Liao, Liang Chen, Hongru Wang, Wenqiang Lei, Tat-Seng Chua, Proactive Dialogue Systems in the Era of Large Language Models: Evaluating from a Prompting Perspective, in Findings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.
[22] Libo Qin, Wenbo Pan, Qiguang Chen, Lizi Liao, Zhou Yu, Yue Zhang, Wanxiang Che, Min Li, End-to-end Task-oriented Dialogue: A Survey of Tasks, Methods, and Future Directions, in Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.
[23] Bobo Li, Hao Fei, Lizi Liao, Yu Zhao, Chong Teng, Tat-Seng Chua, Donghong Ji, Fei Li, Revisiting Disentanglement and Fusion on Modality and Context in Conversational Multimodal Emotion Recognition, in Proceedings of the 31th ACM International Conference on Multimedia, 2023.

[24] Wei Ji, Renjie Liang, Lizi Liao, Hao Fei, Fuli Feng, Partial Annotation-based Video Moment Retrieval via Iterative Learning, in Proceedings of the 31th ACM International Conference on Multimedia, 2023.
[25] Cheng Chen, Yong Wang, Lizi Liao, Yueguo Chen, Xiaoyong Du, REAL: A Representative Error-Driven Approach for Active Learning, in Proceedings of the ECML/PKDD, 2023.
[26] Lizi Liao, Grace Hui Yang, Chirag Shah, Proactive Conversational Agents in the Post-ChatGPT World, in Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2023.
[27] Bobo Li, Hao Fei, Fei Li, Yuhan Wu, Jinsong Zhang, Shengqiong Wu, Jingye Li, Yijiang Liu, Lizi Liao, Tat-Seng Chua and Donghong Ji, DiaASQ: A Benchmark of Conversational Aspect-based Sentiment Quadruple Analysis, in Findings of the Association for Computational Linguistics (ACL), 2023.