Amogh Mannekote

University of Florida
Gainesville, FL
United States

amogh.mannekote@ufl.edu
msamogh.github.io

1 Research interests

Over the years, the dialogue systems community has reached broad consensus on **task formulations, modeling approaches, and evaluation methods** for the development of virtual assistants. The community has also benefited from the development of **extensible, open-source frameworks** for task-oriented dialogue (e.g., ParlAI [13], PyDial [18], ConvLab [21], and Rasa [2]). My research interests are focused on the challenges involved in developing similar analogues for **collaborative dialogue systems**, which are severely understudied compared to virtual assistants and hence, remain largely open-ended in terms of these questions. While my motivation to study this area comes from thinking about approaches to build dialogue models for **co-creative and pedagogical agents for teaching computer science** to young learners, I firmly believe that modeling dialogic collaboration, whether in the context of a virtual learning companion or in that of a physical robot assistant, is characterized by a common set of challenges that transcend domain-specific idiosyncrasies. Examples of these challenges are modeling asynchronous turn-taking, reasoning over discrete task streams, and learning from extremely small datasets.

1.1 Developing a Co-Creative Learning Partner Agent

I am currently part of a joint effort in collaboration with Georgia Tech to develop a dialogue agent that plays the role of a *co-creative learning companion*. The companion agent (nicknamed CAI) is situated within EarSketch[1], an online learning platform designed to help high-school students who are learning computer programming by engaging them with the activity of music creation [6]. CAI's goal is to offer students help both in the realm of problem solving (e.g., debugging the code) as well as in that of creative artifact generation (e.g., giving soundtrack suggestions).

With CAI as a representative example of a collaborative dialogue system in mind, I will enumerate some key open challenges that I have identified in the development of collaborative dialogue systems. However, most of these problems are applicable to any collaborative dialogue scenario.

1. **Asynchronous and Mixed-Initiative Turn-Taking.** Unlike simple task-oriented dialogue where the agent is constrained to only respond immediately after a user utterance, collaborative dialogue settings demand mixed-initiative capabilities from the agent [8, 9]. A model of turn-taking behavior needs to take into account all relevant modalities, be robust to noisy examples present in human-human or Wizard-of-Oz corpora, and be evaluated with an appropriate evaluation metric. In addition, intervention models in a learning context also needs to take pedagogical and affective considerations into account [3, 7].

2. **Modeling Discrete Task Streams.** To generate appropriate responses, one needs to fully capture the dynamic task context that accompanies the natural language dialogue (in the case of CAI, the task context includes UI events and the snapshots of the code and generated at specific time intervals). While it is natural to treat the accompanying task stream as "just another modality", it is not, however, obvious as to how one would go about encoding and aligning it with the natural language dialogue. Existing approaches to "multimodal dialogue" dialogue are limited to vision and speech modalities, both of which are continuous and are compositional modalities. Modeling the task stream, therefore, remains an open problem.

3. **Variability in Responses.** The problem of having a large space of valid responses to a given dialogue context precludes the use of common metrics such as BLEU [15] to evaluate a response. One line of approach to address this problem involves using intrinsic or *reference-free* approaches to evaluate model performance [19, 12]. An orthogonal approach is to explicitly model dialogue strategies in an attempt to disentangle it from the "natural flow" of the dialogue [20].

4. **Lack of Large Training Datasets.** Conversations in large datasets of task-oriented dialogue such as

[1] https://earsketch.gatech.edu/

MultiWOZ [4] are set in everyday, familiar settings such as flight booking and restaurant reservations, giving them the advantage of being universally understood. However, collaborative dialogue typically takes place within an external environment, requiring Wizards and users to have extensive familiarity with the environment (e.g., Blocks world, the EarSketch interface) and/or domain expertise (e.g., programming or music theory) often rendering large-scale crowdsourcing through general-purpose platforms such as M-Turk infeasible.

Collaborative dialogue systems have a long history, starting with the Collaborative Problem Solving (CPS) Model [1], which was proposed in the early 2000s. The CPS framework was designed to overcome the limitations imposed by the simplistic "Intent+Slot" formulation of dialogue moves in task-oriented dialogue [5] by disentangling the dialogue state into individual and collaborative levels. The CPS framework also explicitly modeled the process of negotiation through dialogue acts corresponding to the proposal, acceptance, and rejection of goals.

More recently, collaborative dialogue has seen renewed interest, with multiple datasets and accompanying task formulations being released every year. These recent formulations of the collaborative dialogue problem have leaned towards Leader-Follower setups within game-like environments. In these setups, the Leader is typically privy to some exclusive information that the Follower is not. The Leader's objective is to communicate this information to the Follower through natural language dialogue. The Follower's objective is to execute these instructions to reach the common goal. Examples of recently published datasets in this area include CerealBar [16, 11], Minecraft Dialogue Corpus [14], OneCommon [17], and CoDraw [10]. Common tasks in these setups include retrieving/generating the next utterance of the Leader, retrieving/generating the next utterance of the Follower, and predicting the next action of the Follower based on the dialogue history.

2 Spoken dialogue system (SDS) research

2.1 Challenges in the Near Future

Over the next five years, I see the use of zero-shot natural-language prompts on large language models becoming the de-facto mode though which dialogue systems get developed. Two central research questions will be:

1. What is the best way to design a development interface to ensure that a dialogue agent's *actual* behavior when deployed is aligned with the the developer's *expectations* of its behavior (assuming the latter is specified using ambiguous natural language)?
2. How to facilitate multimodal continual/lifelong learning of dialogue agents (e.g., lifelong learning from verbal and non-verbal cues and sparse reward signals)?

For both the above questions, I suspect that apart from general advances in areas such as active learning, inverse reinforcement learning, it is crucial that communities such as SIGDIAL actively develop a better understanding of human-robotic interaction (HRI) principles in the context of dialogue.

2.2 What do you think this generation of young researchers could accomplish in that time?

In short, I think we should work towards minimizing the fragmentation in the dialogue systems community. To this end, I believe the single most important thing the current generation of dialogue researchers should focus on is the consolidation of the advances in dialogue systems research into well-documented open-source tools and frameworks that are re-usable and extensible (in the spirit of communities such as HuggingFace[2]) to bring about standardization of semantics and interfaces within the community.

3 Suggested topics for discussion

- Dialogue systems in education
- Methods to bootstrap dialogue systems and learn dialogue models from small datasets
- Barriers to the democratization of dialogue systems development
- Novel task formulations for dialogue models beyond "simple slot-filling agents" (e.g., collaborative and situated dialogue settings)

References

[1] Nate Blaylock. "A Collaborative Problem-Solving Model of Dialogue". In: *SIGDIAL*. 2005.

[2] Tom Bocklisch et al. "Rasa: Open source language understanding and dialogue management". In: *arXiv preprint arXiv:1712.05181* (2017).

[3] Anthony F. Botelho et al. "Developing Early Detectors of Student Attrition and Wheel Spinning Using Deep Learning". In: *IEEE Transactions on Learning Technologies* 12 (2019), pp. 158–170.

[2] https://huggingface.co/

[4] Paweł Budzianowski et al. “MultiWOZ - A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling”. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, Oct. 2018, pp. 5016–5026. DOI: 10.18653/v1/D18-1547. URL: https://aclanthology.org/D18-1547.

[5] Philip Cohen. “Foundations of Collaborative Task-Oriented Dialogue: What’s in a Slot?” In: *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*. Stockholm, Sweden: Association for Computational Linguistics, Sept. 2019, pp. 198–209. DOI: 10.18653/v1/W19-5924. URL: https://aclanthology.org/W19-5924 (visited on 12/25/2021).

[6] Jason Freeman et al. “EarSketch: Engaging Broad Populations in Computing through Music”. In: *Commun. ACM* 62.9 (Aug. 2019), pp. 78–85. ISSN: 0001-0782. DOI: 10.1145/3333613. URL: https://doi.org/10.1145/3333613.

[7] Bradley A. Goodman et al. “Using Dialogue Features to Predict Trouble During Collaborative Learning”. In: *User Modeling and User-Adapted Interaction* 16 (2006), pp. 83–84.

[8] Joseph F. Grafsgaard et al. “Predicting Learning and Affect from Multimodal Data Streams in Task-Oriented Tutorial Dialogue”. In: *EDM*. 2014.

[9] Martin Johansson and Gabriel Skantze. “Opportunities and Obligations to Take Turns in Collaborative Multi-Party Human-Robot Interaction”. In: *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Prague, Czech Republic: Association for Computational Linguistics, Sept. 2015, pp. 305–314. DOI: 10.18653/v1/W15-4642. URL: https://aclanthology.org/W15-4642.

[10] Jin-Hwa Kim et al. “CoDraw: Visual Dialog for Collaborative Drawing”. In: *ArXiv* abs/1712.05558 (2017).

[11] Noriyuki Kojima, Alane Suhr, and Yoav Artzi. “Continual Learning for Grounded Instruction Generation by Observing Human Following Behavior”. In: *Transactions of the Association for Computational Linguistics* 9 (2021), pp. 1303–1319. DOI: 10.1162/tacl_a_00428. URL: https://aclanthology.org/2021.tacl-1.77.

[12] Shikib Mehri and Maxine Eskénazi. “USR: An Unsupervised and Reference Free Evaluation Metric for Dialog Generation”. In: *ArXiv* abs/2005.00456 (2020).

[13] Alexander H Miller et al. “Parlai: A dialog research software platform”. In: *arXiv preprint arXiv:1705.06476* (2017).

[14] Anjali Narayan-Chen, Prashant Jayannavar, and Julia Hockenmaier. “Collaborative Dialogue in Minecraft”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, July 2019, pp. 5405–5415. DOI: 10.18653/v1/P19-1537. URL: https://aclanthology.org/P19-1537.

[15] Kishore Papineni et al. “Bleu: a Method for Automatic Evaluation of Machine Translation”. In: *ACL*. 2002.

[16] Alane Suhr et al. “Executing Instructions in Situated Collaborative Interactions”. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 2119–2130. DOI: 10.18653/v1/D19-1218. URL: https://aclanthology.org/D19-1218 (visited on 12/03/2021).

[17] Takuma Udagawa and Akiko Aizawa. “A Natural Language Corpus of Common Grounding under Continuous and Partially-Observable Context”. In: *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*. AAAI’19/IAAI’19/EAAI’19. Honolulu, Hawaii, USA: AAAI Press, 2019. ISBN: 978-1-57735-809-1. DOI: 10.1609/aaai.v33i01.33017120. URL: https://doi.org/10.1609/aaai.v33i01.33017120.

[18] Stefan Ultes et al. “Pydial: A multi-domain statistical dialogue system toolkit”. In: *Proceedings of ACL 2017, System Demonstrations*. 2017, pp. 73–78.

[19] Chen Zhang et al. “D-Score: Holistic Dialogue Evaluation Without Reference”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021), pp. 2502–2516.

[20] Ran Zhao et al. “Automatic Recognition of Conversational Strategies in the Service of a Socially-Aware Dialog System”. In: *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Los Angeles: Associa-

tion for Computational Linguistics, Sept. 2016, pp. 381–392. DOI: 10.18653/v1/W16-3647. URL: https://aclanthology.org/W16-3647.

[21] Qi Zhu et al. "Convlab-2: An open-source toolkit for building, evaluating, and diagnosing dialogue systems". In: *arXiv preprint arXiv:2002.04793* (2020).

Biographical sketch

Amogh Mannekote is a second-year PhD student at the Learn-Dialogue group at the University of Florida. He is supervised by Dr. Kristy Elizabeth Boyer and co-supervised by Dr. Bonnie Dorr. He has previously worked in both industrial and academic settings in roles spanning Software Development, Machine Learning, NLP, and Robotics. His current research interests include dialogue systems and human-centered computing. Amogh also remains an active contributor to the open-source community.