

Xiaolong Li

Department of Computer & Information &
Science & Engineering,
University of Florida
432 Newell Dr,
Gainesville, FL 32611

xiaolongli@ufl.edu

1 Research Interests

In the dialogue systems research community, there is growing recognition that dialogue systems need to support users in increasingly complex tasks. To move in this direction, dialogue systems must perform natural language understanding within richer and richer contexts, including semantic interpretation of user utterances. My research interests are generally in the area of dialogue understanding, specifically for situated dialogues, which happen in a situated environment. My previous work has focused on reference resolution in situated dialogues, specifically tutorial dialogues for computer science problem solving. I proposed a semantic parser using a conditional random field to parse referring expressions in situated dialogues. I combined semantic parsing results with dialogue history as well as user action history to perform reference resolution in Java programming tutoring sessions. I am currently working on unsupervised approaches for reference resolution in situated dialogues, which solves the problem by employing the temporal relationship between user utterances and user actions in the environment. The goal is to automatically learn the relationship between referring expressions and their referents in the environment.

1.1 Reference Resolution for Situated Dialogues

The content of a situated dialogue is very closely related to the environment in which it happens (Grosz and Sidner, 1986). As dialogue systems move toward assisting users in increasingly complex tasks, these systems must understand users language within the environment of the tasks. To achieve this goal, dialogue systems must perform reference resolution, which involves identifying the referents in the environment that the user refers to (Iida et al., 2010; Liu et al., 2014; Liu and Chai, 2015). Imagine a dialogue system that assists a novice student in solving a programming problem. To understand a question or statement the student poses, such as, Should I use the 2 dimensional array?, the system must link the referring expression the 2 dimensional array to an object in the environment.

Reference resolution in situated dialogue is challeng-

ing because of the ambiguity inherent within dialogue utterances and the complexity of the environment. Prior work has leveraged dialogue history and task history information to improve the accuracy of reference resolution (Iida et al., 2010; Iida et al., 2011; Funakoshi et al., 2012). However, these prior approaches have employed relatively simple semantic information from the referring expressions, such as a manually created lexicon, or have operated within an environment with a limited set of pre-defined objects. I have created an approach that addresses those challenges by combining the learned semantic structure of referring expressions with dialogue history into a ranking-based model. I evaluated the new technique on a corpus of human-human tutorial dialogues for computer programming and have shown that it performs significantly better than the previous state of the art (Li & Boyer, In press). An excerpt dialogue is shown in Figure 1.

Tutor:	table = new int[10][5];
Tutor:	that is where they initialize the size of the 2 dimensional array
...	...
	[student adds line of code: arra = new int[s.length()];]
Tutor:	great!
...	...
	[student adds line of code: new2=Integer.parseInt(parse1);]
Student:	does my array look like it is set up correctly now
Tutor:	umm..... in the for loop , what should you be storing in the array ?
Student:	:)

Figure 1: Excerpt of tutorial dialogue illustrating referring expressions refer to referents in the Java code.

1.2 Future Work

For a dialogue system in a situated environment, extra-linguistic information from the environment should also be considered to make proper decisions for the dialogue manager, since the content of the dialogue is closely related to the state of the environment. I plan to further improve the performance of my reference resolution models by incorporating that rich information. Afterward, I plan to investigate the extent to which this higher performing reference resolution module improves outcomes in an end-to-end evaluation of a tutorial dialogue system for introductory computer programming. My study, which will feature my reference resolution model in one version of the system compared to a control condition version of the system, will shed light on the open question of the extent to which student learning and user satisfaction are improved by higher fidelity reference resolution models.

2 Future of Spoken Dialog Research

I believe that there needs to be more research on situated dialogues and dialogue systems that support complex tasks. For several decades, researchers have created effective dialogue systems for a variety of domains. Based on the current state of research work, I think there will be more research on building dialogue systems that can support complex real-world tasks such as problem solving. These kinds of dialogue systems need to employ information from the system's situated environment. There have been several commercial products, such as Amazon Echo, released to work as smart assistants to human users using spoken language. Although these systems are examples of a limited subset of dialogue systems and do not currently extensively employ active awareness of their environments, it is natural for these systems and others (such as assistance robots) to move on to this type of new functionality in the future. To achieve such a goal, dialogue systems must be able to understand user utterances in the situated environment, which requires the systems to leverage multimodal information from the environment.

3 Suggestions for Discussion

- What are the current best practices for building an NLU module for a dialogue system? What are possible ways to improve the NLU module's performance by employing information from the dialogue system's situated environment?
- What are the current best practices to encode simple and complex domain knowledge respectively for a dialogue system? In what ways does the design of the dialogue manager depend upon this encoding?
- What are some of the challenges in making a dialogue system that can learn from the interactions with its users and adapt based on what it has learned?

References

- Funakoshi, Kotaro and Nakano, Mikio and Tokunaga, Takenobu and Iida, Ryu. 2012. A Unified Probabilistic Approach to Referring Expressions. *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 237-246.
- Grosz, Barbara J and Sidner, Candace L. 1986. Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 175-204.
- Iida, Ryu and Kobayashi, Shumpei and Tokunaga, Takenobu. 2010. Incorporating Extra-linguistic Information into Reference Resolution in Collaborative Task Dialogue. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 1259-1267.
- Iida, Ryu and Yasuhara, Masaaki and Tokunaga, Takenobu. 2011. Multi-modal Reference Resolution in Situated Dialogue by Integrating Linguistic and Extra-Linguistic Clues. *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP 2011)*, 84-92.
- Liu, Changsong and Chai, Joyce Y. 2015. Learning to Mediate Perceptual Differences in Situated Human-Robot Dialogue. *Proceedings of AAAI 2015*, 2288-2294.
- Liu, Changsong and She, Lanbo and Fang, Rui and Chai, Joyce Y. 2014. Probabilistic Labeling for Efficient Referential Grounding Based On Collaborative Discourse. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 13-18.
- Xiaolong Li, Kristy E. Boyer 2016. Reference Resolution in Situated Dialogue with Learned Semantics. *Proceedings of SIGDial 2016*

Biographical Sketch



Xiaolong Li is a Ph.D. student in the Department of Computer & Information Science & Engineering at the University of Florida. He received his master's degree in Computer Science from Zhejiang University, Hangzhou, China, and his bachelor's degree in Computer Engineering from Northwest Polytechnical University, Xi'an China. His research interests lie in situated dialogue understanding.