Ran Zhao

Language Technologies Institute,
School of Computer Science,
Carnegie Mellon University,
Pittsburgh, PA 15213 USA
rzhao1@cs.cmu.edu
http://www.cs.cmu.edu/~rzhao1/

1 Research Interests

My research interest originated in the study of human-human conversation and interaction. Through leveraging empirical socio-psychological theories and corpuses of human-human communication data, I progressively better understand and discover how humans in dyadic interactions build, maintain and destroy interpersonal relationship.Ultimately, I aim at developing a socially-aware embodied conversational agents that has capability of constructing social bond with human user during the time of assisting him or her to achieve task goals. Specially, my most interested social phenomena in dialogue is rapport. Rapport has been identified as an important function of human interaction.I developed the first computational framework that is able to explain how humans in dyadic interactions build, maintain, and destroy rapport through the use of specific conversational strategies that function to fulfill specific social goals, and that are instantiated in particular verbal and nonverbal behaviors.(Zhao et al., 2014)

1.1 Automatic recognition of conversational strategies

In that work, I focus on automatically recognizing social conversational strategies that in human conversation contribute to building, maintaining or sometimes destroying a budding relationship. These conversational strategies include self-disclosure, reference to shared experience, praise and violation of social norms. By including rich contextual features drawn from verbal, visual and vocal modalities of the speaker and interlocutor in the current and previous turn, I can successfully recognize these dialog phenomena with an accuracy of over 80% and kappa ranging from 60-80%. My findings have been successfully integrated into an end-to-end socially aware dialog system, with implications for virtual agents that can use rapport between user and system to improve task-oriented assistance.(Zhao et al., 2016a)

1.2 Automatically Assessing Dyadic Rapport from Temporal Patterns of Behavior

This work focuses on data-driven discovery of temporally co-occurring and contingent behavioral patterns signaling high and low interpersonal rapport. By utilizing a reciprocal peer tutoring corpus reliably annotated with thin-slice rapport, non-verbals like eye gaze and smiles, conversational strategies like self-disclosure, shared experience, social norm violation, praise and backchannels, I performed fine-grained investigation of how sequences of interlocutor behaviors manifest differently in friends and strangers, uncovering their social intention of facilitating and destroying rapport. I validated the discovered behavioral patterns by predicting rapport against our ground truth via a forecasting model involving two-step fusion of learned temporal associated rules. Our framework performs significantly better than a baseline linear regression method that does not encode temporal information among behavioral features. Implications for the understanding of human behavior and social agent design are discussed.(Zhao et al., 2016b)

1.3 Cognitive-inspired Socially-Aware Hybrid Discourse Planner

The dialogue manager is composed of a task reasoner that focuses on obtaining information to fulfill the user's goals, and a social reasoner that chooses ways of talking that are intended to build rapport in the service of better achieving the user's goals. A task and social history, and a user model, also play a role in dialogue management, but will not be further discussed here. (Matsuyama et al., 2016)

1.3.1 Task Reasoner

The Task Reasoner is predicated on the system maintaining initiative to the extent possible. It is implemented as a finite state machine whose transitions are determined by different kinds of triggering events or conditions such as: user's intents (extracted by the NLU), past and current state of the dialogue (stored by the task history) and other contextual information (e.g., how many sessions the agent has recommended so far). Task Reasoner's output can be either a query to the domain database or a system intent that will serve as input to the Social Reasoner and hence the NLG modules.

1.3.2 Social Reasoner

The Social Reasoner is designed as a network of interacting nodes where decision-making emerges from the dynamics of competence and collaboration relationships among those nodes. That is, it is implemented as a Behavior Network as originally proposed by (Maes,

1989) and extended by (Romero, 2011). Such a network is ideal here as it can efficiently make both short-term decisions (real-time or reactive reasoning) and long-term decisions (deliberative reasoning and planning). The network's structure relies on observations extracted from data-driven models (in this case the collected data referenced above). Each node (behavior) corresponds to a specific conversational strategy (e.g., SD, PR, QE, etc.) and links between nodes denote either inhibitory or excitatory relationships which are labeled as pre-condition and post-condition premises. As pre-conditions, each node defines a set of possible system intents (generated by the Task Reasoner, e.g., "self_introduction", "start_goal_elicitation", etc.), rapport levels (high, medium or low), user conversational strategies (SD, VSN, PR, etc.), visuals (e.g., smile, head nod, eye gaze, etc.), and system's conversational strategy history (e.g., system has performed VSN three times in a row). Post-conditions are the expected user's state (e.g., rapport score increases, user smiles, etc.) after performing the current conversational strategy, and what conversational strategy should be performed next. For instance, when a conversation starts (i.e., during the greeting phase) the most likely sequence of nodes could be: [ASN, SD, PR, SD ... VSN ...] i.e., initially the system establishes a cordial and respectful communication with user (ASN), then it uses SD as an icebreaking strategy, followed by PR to encourage the user to also perform SD. After some interaction, if the rapport level is high, a VSN is performed.

The Social Reasoner is adaptive enough to respond to unexpected user's actions by tailoring a reactive plan that emerges *implicitly* from the forward and backward spreading activation dynamics and as result of tuning the network's parameters which determine reasoner's functionality.

2 Future of Spoken Dialog Research

People pursue multiple conversational goals in dialog (Tracy and Coupland, 1990). Contributions to a conversation can be divided into those that fulfill *propositional* functions, contributing informational content to the dialog; those that fulfill *interactional* functions, managing the conversational interaction; and those that fulfill *interpersonal* functions, managing the relationship between the interlocutors (Cassell and Bickmore, 2003; Fetzer, 2013).There are extensive studies of developing techniques to support human user's propositional goal and interactions goal in dialogue. I think the future of spoken dialogue system should have human-like capabilities that enable to build social bond with the human through generating appropriate behaviors.

3 Suggestions for Discussion

- Evaluation metrics of social dialogue system.
- Cloud-based dialogue system and construct a shared service platform which will enable us work more collaborative.
- Data resource of social dialogue and application of deep learning techniques in dialogue system.

References

Justine Cassell and Timothy Bickmore. 2003. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and User-Adapted Interaction*, 13(1-2):89–132.

Anita Fetzer. 2013. 'no thanks': a socio-semiotic approach. *Linguistik Online*, 14(2).

Pattie Maes. 1989. How to do the right thing. *Connection Science*, 1(3):291–323.

Yoichi Matsuyama, Arjun Bhardwaj, Ran Zhao, Oscar J. Romero, Sushma Akoju, and Justine Cassell. 2016. Socially-aware animated intelligent personal assistant agent. In *17th Annual SIGdial Meeting on Discourse and Dialogue*.

Oscar J. Romero. 2011. An evolutionary behavioral model for decision making. *Adaptive Behavior*, 19(6):451–475.

Karen Tracy and Nikolas Coupland. 1990. Multiple goals in discourse: An overview of issues. *Journal of Language and Social Psychology*, 9(1-2):1–13.

Ran Zhao, Alexandros Papangelis, and Justine Cassell. 2014. Towards a dyadic computational model of rapport management for human-virtual agent interaction. In *Intelligent Virtual Agents*, pages 514–527.

Ran Zhao, Tanmay Sinha, Alan Black, and Justine Cassell. 2016a. Automatic recognition of conversational strategies in the service of a socially-aware dialog system. In *17th Annual SIGdial Meeting on Discourse and Dialogue*.

Ran Zhao, Tanmay Sinha, Alan Black, and Justine Cassell. 2016b. Socially-aware virtual agents: Automatically assessing dyadic rapport from temporal patterns of behavior. In *International Conference on Intelligent Virtual Agents*. Springer.

Ran Zhao is a doctoral Student in the Language Technologies Institute at Carnegie Mellon University, where he is supervised by Prof. Justine Cassell. He is currently working on Rapport project, which focuses on developing an Autonomous System for Embodied Conversational Agents(ECAs). His scientific research interests lie in both designing general architecture of multimodal behavior generation and exploring human behavior. Before joining Articulab, he received his B.S. in Computer Science from University of Illinois,Urbana-Champaign working with Prof. Dan Roth in Natural Language Processing and M.S. from Yale University.