Towards Comprehensive Situated Language Models for Cognitive Agents Embodied in Complex Worlds

Shiwali Mohan

April 3, 2014

The recent advances in Artificial Intelligence, Robotics, and Cognitive Science are paving way for autonomous computational agents that will collaborate with humans in different capacities. Within the next decade, we can realistically expect smart homes and cars, virtual assistants such *Siri* and *Google Now*, and domestic, assistive robots to become common in human societies. A key design requirement of these agents is that they should be able to interact with non-expert users naturally. The principal modality of communication for humans is language that facilitates social co-ordination in joint tasks and collaborative learning. Therefore, a crucial challenge to designing intelligent collaborators is to develop robust models of language comprehension for agents.

Although there has been substantial research on formal frameworks for representing meaning, it is a non-trivial challenge to extend them to an agent architecture. Most prevalent theories of semantics focus on manipulation of abstract symbols. For intelligent agents living in complex virtual and real worlds, it is important to explain how abstract symbols of language connect to internal representations of concepts. Recent approaches [4, 5, 6] have taken some encouraging but simplistic steps in developing language models for agents. However, these approaches do not yet support the range of complexity in human language and linguistic human-human interactions. Through the following proposed research work, we expect to make progress towards supporting flexible and complex linguistic human-agent interaction.

1. *A theory of task-oriented semantics.* Linguistic communication during collaborative task execution conveys information that is useful for task performance. It identifies and brings to the attention of the collaborators, objects of interests, state features that are useful for successful task execution, useful relationships between objects, goals of the task, and feedback from the environment. To generate meaning from language, the semantics should be aligned with the elements of the agent architecture including goals, policies, perceptual classification, spatial reasoning etc such that accessing these elements during processing is useful for performing the task. We have demonstrated the utility of a preliminary semantic theory in learning new tasks and games from human-robot interaction [1, 2, 3]. The proposed work will extend this theory to cover subject/object/verb plurality and agreement, verb-task polysemy and resolution, and task specification via additional spatial or policy information. The current theory is also limited in making assumptions of complete observability of the environmental state and immediately grounds language to the current perceptual state of the agent. However, complex environments are partially observable and linguistic communication cannot be immediately grounded, requiring the comprehender to access the knowledge of the world.

2. *Contextual language processing models.* Human language is highly contextual and relies on several non-linguistic sources to convey meaning. For successful and effective linguistic communication with human partners, the agent's comprehension and generation models must exploit the *common ground* generated from multiple information sources including perceptions, domain knowledge, common-sensical knowledge, and short and long-term experiences to augment linguistic input. They should also readily incorporate information from non-verbal communication such as gestures and eye-gazes. The use of these *extra-linguistic* sources of information makes comprehension and generation robust to inherent ambiguity in language. Our current work on the Indexical model [4] makes promising advances towards addressing comprehension ambiguities arising from the use of under-specific referring expressions. Significant work

remains to be done on problems arising due to word polysemy, phrase-attachment, and discrepancy in the perceived common ground. For efficient generation, the agent also needs to model the human partner's understanding of the current situation. To date, the problem of efficient generation in situated scenarios has been minimally studied, which we will address in our proposed work.

3. *Evaluation.* Evaluation of situated language models is a challenging problem. Studying the problem of language processing in a task-oriented domain allows the use of information-theoretic objective measures such as utility of linguistic information for task performance and the reduction in entropy of interpretation in presence of extra-linguistic context. We are also interested in conducting HRI/HAI studies for subjective measures such as comfort of the human partner and their perception of the robot.

References

[1] James Kirk and John Laird. Learning Task Formulations through Situated Interactive Instruction. In *Proceedings of the 2nd Conference on Advances in Cognitive Systems*, 2013.

[2] Shiwali Mohan, James Kirk, and John Laird. A Computational Model of Situated Task Learning with Interactive Instruction. In *Proceedings of the 17th International Conference on Cognitive Modeling*, 2013.

[3] Shiwali Mohan, James Kirk, Aaron Mininger, and John Laird. Acquiring Grounded Representations of Words with Situated Interactive Instruction. *Advances in Cognitive Systems*, 2, 2012.

[4] Shiwali Mohan, Aaron Mininger, and John Laird. Towards an Indexical Model of Situated Language Comprehension for Real-World Cognitive Agents. In *Proceedings of the 2nd Conference on Advances in Cognitive Systems*, 2013.

[5] Matthias Scheutz, Rehj Cantrell, and Paul Schermerhorn. Toward Human-Like Task-based Dialogue Processing for HRI. *AI Magazine*, 2011.

[6] Stefanie Tellex, Thomas Kollar, Steven Dickerson, Matthew Walter, Ashish Banarjee, Seth Teller, and Nicholas Roy. Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, 2011.