Speech, Gesture, and Space: Investigating Explicit and Implicit Communication in Multi-Human Multi-Robot Collaborations

Aaron St. Clair, Amin Atrash, Ross Mead, and Maja J Matarić

Interaction Lab
Computer Science Department
University of Southern California
Los Angeles, CA 90089-0641
{astclair, atrash, rossmead, mataric}@usc.edu

Abstract

Effective communication is often required for agents to properly handle collaborative multi-agent tasks. This is particularly true when humans are working alongside synthetic agents and traditional wireless communication modes are impractical. A framework for communication must allow for both explicit communication, where actions are directly execute to convey information, and implicit communication, where the agent projects information indirectly as a consequence of actions taken to achieve the tasks. We propose a Theory of Mind-based approach to communication that allows an agent to reason about its own state, the states of the other agents, and the other agents' beliefs about each other's state.

Introduction

Communication is often required between agents as they attempt to solve collaborative multi-agent tasks. This is particularly true in conditions in which an agent is working alongside a human—clearly, conventional electronic communication is not feasible in this scenario; rather, these agents, including humans, must take advantage of physical communication in the shared context to confer necessary information. As an agent observes the actions of the others, it must modify its own behavior accordingly.

We examine the use of communication between multiple, potentially heterogeneous, agents, including people, during performance of a coordinated task as a mechanism for transmitting data. Each agent must monitor both implicit communication (in which an agent infers the intentions of other agents given observable actions) and explicit communication (in which an agent issues or receives spoken commands, gestures, or other directed communication). Since the physical signaling modalities of robotic agents and humans may be significantly different, we investigate signaling models that exploit the shared

Background

Collaborative scenarios are those in which two or more agents are working together to achieve one or more shared goals. The study of human-machine collaboration has largely been motivated by the goal of developing systems capable of estimating or predicting the intentions of a user and modifying or adapting behavior to improve task performance or user satisfaction. Collaboration in general, human-robot collaboration, and other related topics have all been previously investigated to varying degrees.

Human Collaboration

It has been demonstrated that people have a tendency to adapt both their linguistic representations and physical actions in response to those they are interacting with, i.e., they tend to formulate behavior and speech that will be salient and sensible to a collaborating partner (Whittaker 2003). Collaboration in humans occurs via a process in which people align their linguistic representations of the environment allowing for more effective communicative behavior. This alignment is achieved via a process in which local alignment of environmental representations, i.e., specific speech and gesture, are implicitly adopted and propagated to global representations via a priming mechanism (Pickering04). There is however some debate over how deeply people model their interaction partners and how they integrate with language and gesture production (Hanna04).

Human-Machine Collaboration

Prior work on human-machine collaboration includes approaches from human-computer and human-robot interaction as well as cognitive science and linguistics. There is an extensive body of work on top-down deliberative approaches aimed at establishing and maintaining alignment to assure coherent discourse (Grosz 1999). There has also been extensive work on applying perspective-taking and theory of mind-inspired models to allow a robot to recognize intentional behavior through observation (Crick 2008, Kelley 2008, Ullman 2010). Work by Breazeal and Hoffman has demonstrated the ability for a robot to learn simple tasks through human tutelage and collaborate effectively via turn-taking or pre-emptive action.

Approach and Discussion

In order to effectively coordinate a robot's actions with those of its human counterpart, the robot must be able to accurately estimate the human's planned actions from the context or from explicit communication. Analogously, the robot must be able to effectively convey its planned actions clearly to a human.

To accomplish this we propose a Theory of Mind-inspired model in which the robot contains estimates of its own state, the state of third parties, and those third parties' estimates of the robot's state. These states contain information relevant to the task including a world model and a partial task allocation i.e. assignments of various agents to sub-tasks. Previous work has demonstrated the viability of similar frameworks to learn social skills such as deictic reference and joint attention from the bottom up (Scasselati 2002). Our work is aimed instead at integrating social communication with task control, learned or otherwise, to support coordination in collaborative environments. This specific context allows for the simplifying assumption of a shared goal among all participants and enables the robot to evaluate other's actions using its own task controller to determine whether planned policies are aligned and detect assistive opportunities i.e. when things go wrong. One existing approach to perspective taking is employed in (Breazeal 2009) where transformed sensor input is used to learn tasks by demonstration and participate in a turn-taking game.

Given this framework, the agent can consider the consequences of its actions, both in terms of manipulating the environment as well as conveying information to the collaborative partners. Modeling sub-goal dynamics allows the robot to convey information implicitly to the collaborative agent by using its spatial positioning to project the information. In addition, it can compare the estimates of itself with that of others to: detect and handle discrepancies, identify assistive situations, and recognize agents using a different plan than itself potentially allowing it to adapt to or instruct others. Finally, this framework allows the robot to explicitly communicate information by selecting actions such as gesturing and vocalization.

We consider a probabilistic approach to reasoning over this framework (Kaelbling 1998). The agent maintains a probabilistic belief over its own state, over the states of the third parties, and over the third party estimates of the robot's state. This allows it to perform robust decision-making given environment noise, uncertainty in the estimates of its own state, the state of others, and the potential outcomes of its actions.

This approach will be validated on a challenging cooperative task involving a humans and robots, possibly multiple agents of each, who must communicate effectively to achieve a collaborative goal in a very dynamic environment. We will demonstrate the ability of the framework to capture both the implicit and explicit communication required to complete the task. Since the system will be targeted mainly at human-robot interaction and may rely on detailed sensory information about the human such as head direction estimates, it is unlikely to scale for scenarios with many people or where the number of humans is much greater than the number of robots.

Acknowledgements

This work was supported in part by National Science Foundation (NSF) grants CNS- 0709296, IIS-0803565, and IIS-0713697 and ONR MURI (N00014-09-1-1031).

References

C. Breazeal, J. Gray, and M. Berlin, "An embodied cognition approach to mindreading skills for socially intelligent robots," The International Journal of Robotics Research, vol. 28, p. 656, May 2009.

C. Crick and B. Scassellati, "Inferring narrative and intention from playground games," Proceedings of the 7th IEEE Int'l Conf. on Development and Learning, pp. 13–18, August 2008.

B. Grosz, "Collaborative systems," AI magazine, vol. 17, no. 2, pp. 67–86, 1996.

J. Hanna and M. Tanenhaus, "Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements," Cognitive Sci., vol. 28, no. 1, pp. 105–115, 2004.

G. Hoffman and C. Breazeal, "Effects of anticipatory perceptual simulation on practiced human-robot tasks," Autonomous Robots, vol. 28, pp. 403–423, May 2010.

L. Kaelbling, M. Littman, and A. Cassandra. "Planning and acting in partially observable stochastic domains." Artificial Intelligence pp. 99–134. 1998

R. Kelley, C. King, A. Tavakkoli, M. Nicolescu, M. Nicolescu, and G. Bebis, "An architecture for under- standing intent using a novel hidden markov formulation," International Journal of Humanoid Robotics, Special Issue on Cognitive Humanoid Robots, vol. 5, no. 2, pp. 1–22, 2008.

M. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," Behavioral and Brain Sciences, vol. 27, no. 02, pp. 169–190, 2004.

B. Scassellati, "Theory of mind for a humanoid robot," Autonomous Robots, vol. 12, no. 1, pp. 13–24, 2002.

S. Whittaker, "Things to talk about when talking about things," Human–Computer Interaction, vol. 18, no. 1, pp. 149–170, 2003.

T. Ullman, C. Baker, O. Macindoe, O. Evans, N. Goodman, and J. Tenenbaum, "Help or hinder: Bayesian models of social goal inference," Advances in Neural Information Processing Systems (NIPS), vol. 22, 2010.