Using Computational Cognitive Models to Build Better Human-Robot Interaction

Alan C. Schultz
Naval Research Laboratory
Washington, DC

Introduction

We propose an approach for creating more cognitively capable robots that can interact more naturally with humans. Through analysis of human team behavior, we build computational cognitive models of particular high-level human skills that we have determined to be critical for good peer-to-peer collaboration and interaction. We then use these cognitive models as reasoning mechanisms on the robot, allowing the robot to make decisions that are conducive to good interaction with the human.

Cognitively enhanced intelligent systems

We hypothesize that adding computational cognitive reasoning components to intelligent systems such as robots will result in three benefits:

> Most, if not all, intelligent systems must interact with humans, who are the ultimate users of these systems. Giving the system cognitive models can enhance the human-system interface by allowing more common ground in the form of cognitively plausible representations and qualitative reasoning. For example, mobile robots generally use representations such as rotational and translational matrixes to represent motion and spatial references. However, this is not a natural mechanism for humans, and results in additional computations to translate between these and the qualitative spatial reasoning used by humans. By using cognitive models, reasoning mechanisms and representations, we believe that we can yield a more effective and efficient interface.
>
> Since the resulting system is interacting with the human, giving it behaviors that are more natural and compatible with the human can also result in more natural interactions between the human and the intelligent system. For example, mobile robots that must work collaboratively with humans can actually result in less effective interactions if its behaviors are alien or non-intuitive to the human. By incorporating cognitive models, we can develop systems whose behavior is more expected, natural and therefore compatible with the human team members.
>
> One key interest is in measuring the performance of intelligent systems. We propose that an intelligent system that is cognitively enhanced can be more directly compared to human-level performance. Further, if cognitive models of human performance have been developed in creating the intelligent system, we can directly compare the intelligent systems behavior and performance in the task to the human subject behavior and performance.

Hide and Seek

Our foray into this area started when we were developing computation cognitive models of how young children learn the game of hide and seek (Trafton et al. 2005, Trafton et al. 2006). The purpose was to enable our robots to use human-level cognitive skills to make the decisions about where to look for people or things hidden by people. The research resulted in a hybrid architecture with a reactive/probabilistic system for robot mobility (Schultz, Adams & Yamauchi, 1999), and a high-level cognitive system based on ACT-R (Anderson & Lebiere, 1998) that made the high-level decisions for where to hide or seek (depending on which role the robot was playing). While this work was interesting in its own right, the system led us to the realization that the ability to do perspective taking was a critical cognitive ability for humans, particularly when they want to collaborate.

Spatial perspective taking

To determine just how important perspective and frames of reference were in collaborative tasks in shared space (and also because we were working on a DARPA-funded project to move these capabilities to the NASA Robonaut), we analyzed a series of tapes of two astronauts and a ground controller training in the NASA Neutral Buoyancy Tank facility for an assembly task for Space Station mission 9A. We performed a protocol analysis of these tapes (approximately 800 utterances) focusing on the use of spatial language and commands from one person to another. We found that the astronauts changed their frame of reference (as seen during their dialog) approximately every other utterance. As an example of how prevalent these changes in frame of reference are, consider this following utterance from ground control:

"… if you come *straight down* from where you are, uh, and uh, kind of peek *down under the rail* on the *nadir side*, by *your right hand*, almost *straight nadir*, you should see the…"

Here we see five changes in frame of reference (highlighted in italics) in a single sentence! These rates in the change of reference are consistent with work by Franklin, Tversky & Coon, 1992. In addition, we found that the astronauts had to take other perspectives, or forced others to take their perspective, about 25% of the time (Trafton, Cassimatis, Brock, Bugajska, Mintz & Schultz, 2005). Obviously, the ability to handle changing frames of reference and being able to understand spatial perspective will be a critical skill for robots such as the NASA Robonaut and, we would argue, any other robotic system that needs to communicate with people in spatial contexts (i.e., any construction task, direction giving, etc.).

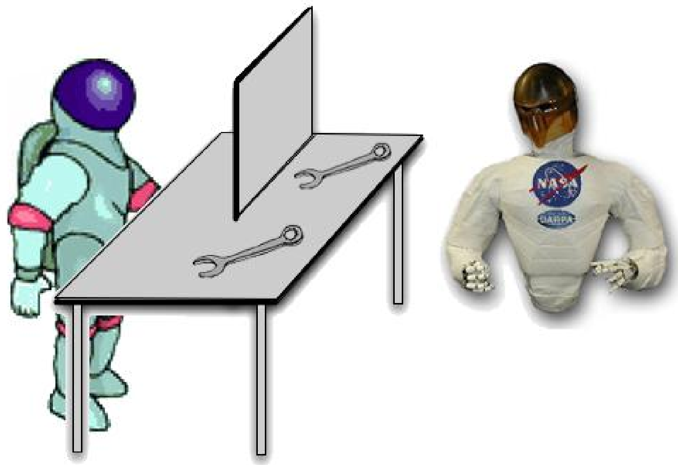

Figure 1: A scenario of an astronaut and a robot; the astronaut asks the robot to "Pass me the wrench."

Models of perspective taking

Imagine the following task, as illustrated in Figure 1. An astronaut and his robotic assistant are working together to assemble a structure in shared space. The human, who because of an occluded view can see only one wrench, says to the robot, "Pass me the wrench." Meanwhile, from the robot's point of view, two wenches are visible. What should the robot do? Evidence suggests that humans, in similar situations, will pass the wench that they know the other human can see (Clark, 1996) since this is a jointly salient feature.

We developed two models of perspective taking that could handle the above scenario in a general sense. The first approach used the ACT-R/S system (Harrison & Schunn, 2002) to model perspective taking using a cognitively plausible spatial representation. The second approach used Polyscheme (Cassimatis, Trafton, Bugajska and Schultz, 2004) and modeled the cognitive process of mental simulation; humans tend to mentally simulate situations in order to resolve problems.

Using these models we have demonstrated a robot being able to solve problems similar to the wench problem.

Future work

We are now exploring other human cognitive skills that seem important for peer-to-peer collaborative tasks and that are appropriate for building computational cognitive models for adding to our robots. One new skill we are considering is non-visual, high-level focus of attention. This skill helps to focus a person's attention to appropriate parts of the environment or situations based on the current environment, task, expectations, models of other agents in the environment and other factors. Another human cognitive skill we are considering involves the role of anticipation in human interaction and decision-making.

Conclusion

It is clear that if humans are to work as peers with robots in shared space, the robot must be able to understand the natural human tendency to use different frames of reference and to take the human's perspective. To create robots with these capabilities, we propose using computational cognitive models as opposed to more traditional programming paradigms for robots. First, a natural and intuitive interaction results in reduced cognitive load. Second, more predictable behavior engenders trust. Finally, more understandable decisions allow the human to recognize and more quickly repair mistakes.

We believe that using computational cognitive models will give our robots the cognitive skills necessary to interact more naturally with humans, particular in peer-to-peer relationships.

References

J. G. Trafton & Alan C. Schultz, N, L. Cassimatis, L. Hiatt, D. Perzanowski, D. P. Brock, M. Bugajska, and W Adams, (2005). "Using Similar Representations to Improve Human-Robot Interaction," Agents and Architectures (Ed. Ron Sun), Erlbaum, 2005.

J. G. Trafton, Alan C. Schultz (2006). "Children and robots learning to play hide and seek," ACM conference on Human-Robot Interaction, March 2006.

N. Cassimatis, J. G. Trafton, M. Bugajska, and A. C. Schultz (2004). "Integrating Cognition, Perception, and Action through Mental Simulation in Robots." *Robotics and Autonomous Systems*, 49(1-2), Elsevier, Nov. 2004, pp. 13-23.

Trafton, J. G., Cassimatis, N. L., Brock, D. P., Bugajska, M. D., Mintz, F. E., & Schultz, A. C. (2005). "Enabling effective human-robot interaction using perspective-taking in robots," IEEE Transactions on Systems, Man, and Cybernetics---Part {A}: Systems and Humans, 35(4), 460-470.

Franklin, N., Tversky, B., & Coon, V. (1992). Switching points of view in spatial mental models. Memory & Cognition, 20(5), 507-518.

Schultz, A., Adams, W., and Yamauchi, B. (1999). "Integrating Exploration, Localization, Navigation and Planning Through a Common Representation," *Autonomous Robots*, 6(3), June, 1999.

Anderson, J. R., & Lebiere, C. (1998). Atomic components of thought. Mahwah, NJ: Erlbaum.

Clark, H. H. (1996). Using language. Cambridge University Press.

Harrison, A. M., & Schunn, C. D. (2002). ACT-R/S: A computational and neurologically inspired model of spatial reasoning. In W. D. Gray & C. D. Schunn (Eds.), Proceedings of the Twenty-Fourth Annual Meeting of the Cognitive Science Society (pp. 1008). Fairfax, VA: Lawrence Erlbaum Associates.