Learning to Perceive Human Intention and Assist in A Situated Context

Hee-Tae Jung*, Shichao Ou† and Roderic Grupen*
*Department of Computer Science, University of Massachusetts Amherst, Amherst, Massachusetts 01003, USA
Email: {hjung, grupen}@cs.umass.edu
†Network Equipment Technologies, New Jersey 07724, USA
Email: chao_ou@net.com

***Abstract*— In this paper, we propose that a robot can create a series of monitors describing its own pointing gesture and reaching behavior and that these models can be used to infer human intention delivered by a pointing gesture and assist accordingly. This extends our previous work where we demonstrated that an intrinsically motivated robot can be designed to seek controllable relationships with the world and employ them to solicit assistance from a nearby human via expressive gestures. Preliminary experimental results demonstrate that our approach enables a robot to respond appropriately after learning a receptive quality of gesture.**

I. Introduction

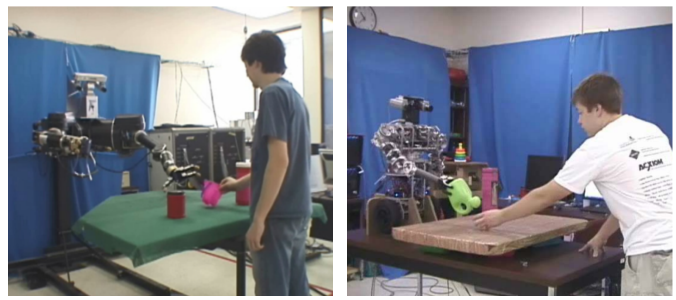
It is envisioned that personal robots will assist human daily activities in various settings [1]. In these scenarios, robots should be able to parse communicative signals and understand the intention delivered by evocative gestures so that even naive users and those with disabilities can receive aid when necessary. Drawing inspiration from the psychology literature [2], Ou and Grupen studied how a pointing gesture can emerge from a reaching behavior in under-actuated contexts and demonstrated that the pointing gesture can successfully solicit human assistance [3] as in Fig. 1a. In this work, we address how this expressive behavior knowledge can be exploited to perceive human intention via a pointing gesture and assist accordingly in the same context associated with the very behavior. This coincides with the recent hypothesis in biology literature such that there may exist a common coding for expressing and perceiving behaviors [4].

In Section II, we first visit the psychology and biology literature that inspired this work. Then, in Section III, we discuss how a robot learns a receptive behavior which perceives human intention and assists in a situated context using the control basis and action schema frameworks. Section IV concludes this paper.

II. Background

A. Manual Behavior and Communicative Gestures

Psychologists acknowledge a tight connection between manual behavior and communicative gestures. For instance, Vygotsky noted that a pointing gesture originated from an unsuccessful attempt to grasp certain objects [2]. As infants attempt to reach for out-of-reach objects, even though they inevitably fail, in the presence of a caregiver, the action is recognized and interpreted as the "intention" to acquire the objects. Thus, the action becomes a means of conveying the intention to the caregiver. In this work, this insight is applied to robotics wherein a robot learns to perceive the pointing gesture of a human as an evocative gesture and assists by handing the object to the human.



(a) A learned pointing gesture to solicit help in the presence of a human. (b) A learned assisting behavior responding to a human pointing gesture.

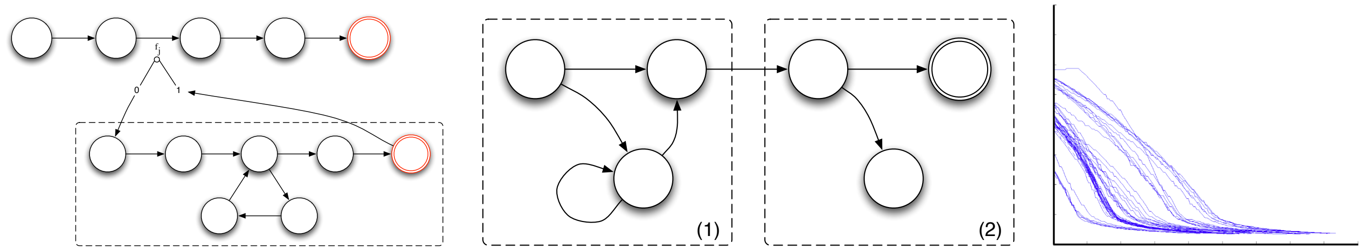
Fig. 1: A robot executes its learned behaviors (a) to solicit human assistance and (b) to assist human.

B. Common Coding for Expressive and Receptive Behaviors

Mirror neurons were first found in the inferior frontal gyrus (region F5) and the inferior parietal lobule of the macaque monkeys [4]. It is observed that mirror neurons discharge both when a monkey performs an action and when it observes others do a similar action. Recently, a similar research result on human subjects have been reported [5]. Specifically, mirror neurons which discharge when a human grasps an object would also discharge when the human perceives another person grasp the object. Inspired in part by this finding, we designed the frameworks and a representation that can support the expressive and receptive symmetry that some models of mirror neurons propose.

III. Learning Receptive Behaviors

A receptive behavior, in this work, consists of recognizing human pointing and assisting human grasping. Throughout expriments, a robot is situated in a naturally cluttered lab environment where a human stands across a table and points at an out-of-reach object as in Fig. 1b. Human detection was done by using [6]. Due to the lack of space, detailed explanation



(a) A simplified pointing and reaching behavior schema [3], which corresponds to Fig. 1a. The terminating states are marked with red color.

(b) A simplified schema learned by a robot to (1) recognize human intention and (2) hand the object to a human to get rewarded. This corresponds to Fig. 1b.

(c) An error dynamics between an object and a robot hand observed in the terminating state of Fig.2a.

Fig. 2: The (a) expressive and (b) receptive behavior schemas that a robot learned, and (c) an error dynamics observed by C_M.

on the basic concept of the control basis and action schema frameworks is omitted here. Interested readers are encouraged to read the work done by Platt and Hart [7, 8].

A. Recognizing Human Pointing

Within control basis and action schema frameworks, expressive behaviors are represented as state transition diagrams. A transition from one state to the next occurs by executing a controller and each state contains the execution result of controllers. For learning to recognize a human pointing gesture, a robot first studies the terminating states of its own behavior schema because the robot receives rewards when it reaches those states, i.e. Fig. 2a. The robot searches a pool of randomly generated monitors for the one that best describes the terminating states. After learning such a monitor C_M, the robot receives a reward when the monitor observes a similar error dynamics that was observed in the learning phase (Fig. 2c) as well as it reaches terminating states of the executing schema (Fig.2a). Then the robot observes the error dynamics of a human pointing gesture and learns that C_M describing its own behavior is most relevant to the error dynamics between an object and a part of a human's body. Thereafter the robot uses the learned monitor to infer that the intention of the human is to reduce the error. This is synonymous with achieving a "touch" reward in the intrinsically motivated system.

B. Assisting Human Grasping

When learning to assist human grasping, a robot is given a set of behaviors learned in the previous work [8] along with the learned monitors and the corresponding state space. Due to the learned C_M, the robot is rewarded when the distance error between a target object and a human hand stays less then a certain threshold. During the training phase, 2 objects were used to ensure that the robot was exposed to sufficient positive experience for behavior acquisition. Within about 20 episodes, in the presence of a human, the robot learned a behavioral schema that recognizes a pointing gesture and hands the referenced object to the human using Q-learning algorithm. After the robot acquired a stable schema, i.e. Fig. 2b, either 2 or 4 objects were placed on the table to evaluate the performance of the learned behavior schema. During the testing phase, 10 human subjects participated in the experiment. Subjects were asked to solicit a robot assistance by using a gesture of their own choice. When the robot correctly hands an object to the human, the occasion was counted as success. The success rate was 78% for 4 objects and 96% for 2 objects respectively. It is important to note that the robot learned how to observe intention and how to respond directly by interacting with humans rather than simply executed a preprogrammed response.

IV. Conclusion

In this paper, we address the problem where a robot learns the intention of a human pointing gesture and assists accordingly. The preliminary experimental results demonstrate that a robot can learn assisting behavior through intrinsically motivated learning. In the future, we plan to conduct further experiments where a robot uses a suite of monitors to recognize diverse human gestures more robustly.

Acknowledgment

This research was funded in part by a gift from Microsoft Corporation and award N00014-07-1-0749 from Massachusetts Institute of Technology/ONR Muri. The first author gracefully acknowledges Robin Popplestone Fellowship and Cambridge Culture Foundation Scholarship.

References

[1] P. Deegan, R. A. Grupen, A. Hanson, E. Horrell, S. Ou, E. Riseman, S. Sen, B. Thibodeau, A. Williams, and D. Xie, "Mobile manipulators for assisted living in residential settings," *Autonomous Robots*, vol. 24, no. 2, pp. 179–192, 2007.

[2] L. Vygotsky, *Mind in society*. Cambridge, MA: Harvard University Press, 1930.

[3] S. Ou and R. A. Grupen, "From manipulation to communicative gesture," in *5th ACM/IEEE International Conference on Human-Robot Interaction*, Osaka, Japan, March 2010.

[4] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi, "Premotor cortex and the recognition of motor actions," *Cognitive Brain Research*, vol. 3, no. 2, pp. 131–141, 1996.

[5] R. Mukamel, A. D. Ekstrom, J. Kaplan, M. Iacoboni, and I. Fried, "Single-neuron responses in humans during execution and observation of actions," *Current Biology*, vol. 20, pp. 1–7, 2010.

[6] S. Ou, R. Lee, and R. A. Grupen, "Learn to detect and recognize humans using small data sets," University of Massachusetts Amherst, Amherst, MA, Tech. Rep. UM-CS-2009-042, 2009.

[7] R. Platt, "Learning and generalizing control based grasping and manipulation skills," Ph.D. dissertation, Department of Computer Science, University of Massachusetts, Amherst, 2006.

[8] S. Hart, "The development of hierarchical knowledge in robot systems," Ph.D. dissertation, Department of Computer Science, University of Massachusetts, Amherst, 2009.