Towards a simple Robotic Theory of Mind

Kyung-Joong Kim
Mechanical & Aerospace Engineering
Cornell University, Ithaca, NY 14853, USA
Department of Computer Engineering
Sejong University, Seoul 143-747, Korea
kimkj@sejong.ac.kr

Hod Lipson
Mechanical & Aerospace Engineering
Computing & Information Science
Cornell University
Ithaca, NY 14853, USA
hod.lipson@cornell.edu

ABSTRACT

Theory of mind (ToM) is a cognitive function in which an agent can infer another agent's internal state and intention based on their behaviors. Can robots realize ToM like humans? There are many issues to be tackled to address this challenging problem, such as the representation, discovery and exploitation of an actor's self models. In this paper we study how robots can represent other's self with artificial neural networks and an evolutionary learning mechanism. This framework was tested with simulated and physical robots and a novel prey-predator scenario was introduced to measure the performance of ToM learning. Experimental results showed that the proposed ToM approach can recover other's self models successfully.

Categories and Subject Descriptors

I.2.0 [Artificial Intelligence – General]: Cognitive Simulation

General Terms

Algorithms, Performance, Design, Reliability, Experimentation, and Verification

Keywords

Robotics, Evolutionary Computation, Estimation-Exploration Algorithm, Theory of Mind, Neural Network, Robot Test

1. INTRODUCTION

Theory of Mind (ToM) is a cognitive capability that allows us to understand another's internal states (intention, goal, and belief) and predict future behaviors of others [1]. From the observation of other's behavior, facial expression, and speech, we can infer the person's internal state (emotions, thought, decision making, and plans). It was known that this function is supported by widely distributed areas of human brain [2][3]. For Chimpanzees, they have ToM but it is a bit different with human's one [4].

ToM has gained great interest from an engineering society. Scassellati built "finding faces and eyes and distinguishing animate from inanimate stimuli" functions for humanoid robots [5]. Buchsbaum *et al*. developed an anthropomorphic animated mouse character that uses his own behavior repositories to interpret other's behavior [6]. Hegel *et al*. studied human's theory of minds for different shapes of robots [7]. Ono *et al*. used theory of mind mechanism to improve human's understanding on robot's intention [8].

Implementing ToM has great difficulty because it is a kind of reverse inference based on observation. Other's self model is hidden and it exists inside of objects. It is not possible to see the internal model directly and it is only indirectly observable. The only thing that we can observe is that the reaction of the object to the inputs from environments. The model with continuous input-output signals is more difficult to be discovered than discrete one.

In this paper, each robot has its own self and the problem of ToM is to discover other robot's self as close as possible. In case of human, the self is located inside of human brain and represented with biological neural networks. The problem of ToM for human is to build models inside of my brain that approximate the behaviors originated from other's internal self. Like other's original self, the inferred other's self is also represented as biological neural networks. The problem can be reformulated as finding another biological neural network that shows close behavior with the original one. Robot uses the similar mechanism to do the ToM.

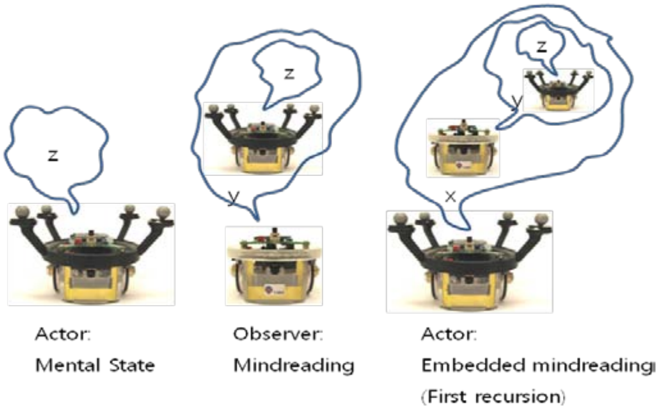
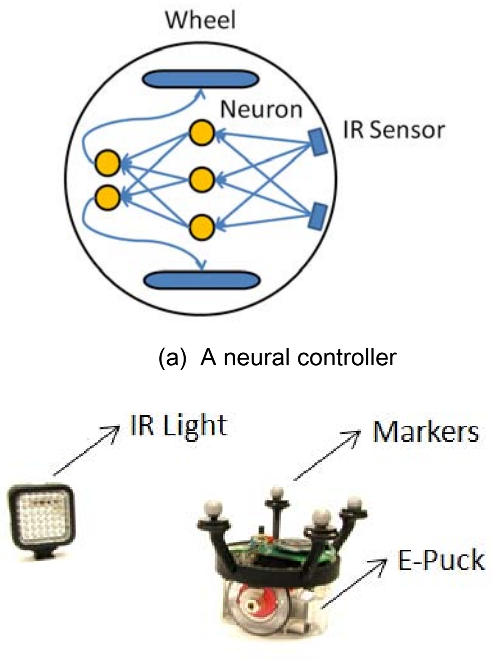


Figure 1. Theory of Mind in robots

In this paper, robots do theory of minds by inferring the neural networks inside other robots based on their movement (Figure 1). An artificial neural network controls the movement of robot's wheels based on sensory inputs. The inference is based on the exploration-estimation algorithm (EEA) used in reverse engineering of nonlinear-dynamical systems [9] and robot's self modeling [10]. After building other's self models, the robot exploits them to predict other robot's behavior. Figure 2 shows a neural network and virtual/physical robots used.

PerMIS'09, September 21-23, 2009, Gaithersburg, MD, USA
Copyright 2009 ACM 978-1-60558-747-9/09/09…$10.00



(a) A neural controller

(b) A real robot

Figure 2. An artificial neural network and robots

2. RELATED WORK

Premack and Woordruff asked “does the chimpanzee have a theory of mind?” in 1978 [1]. Heyes surveyed experimental evidences of non-human theory of mind in 1998 [11]. After 30 years research from the initial question, it was revealed that chimpanzee do have a theory of mind but do not understand others like humans do [12]. Series of experiments were conducted with chimpanzees to know what they know about others [13][14][15]. Two chimpanzees compete for food when only one of them has complete information about the location of food. They concluded that chimpanzees know what other can see and exploit it for food competition.

Childhood autism is related to the lack of theory of mind [16][17]. Baron-Cohen *et al*. compared normal, autistic, and Down’s syndrome subjects using a belief question to test theory of mind. The results for Down’s syndrome and normal subjects were similar (85% and 86% success ratio). On the other hand, 80% of autistic children failed the belief question. Ozonoff *et al*. tested the relationships between autism and first-order, second-order theory of mind [18].

Based on [19], theories for “theory of mind” are classified into four categories: Modular theory, simulation theory, theory-theory, and executive function theory. In modular approach, the theory of mind is functionally dissociable from other cognitive functions [17][20]. They assume that there are one or more neural structures specifically dedicated to theory of mind. In simulation theory [21], there is no general theory guiding the theory of mind. Instead, human’s brain mentally simulates other person’s situation by placing himself to the other person’s place. This perspective-taking view of theory of mind does not support specialized, distinct neural structure for this cognitive skill. In theory theory view, child has a theory about how other minds operate and it evolves over time [22]. Some theorists argue that a distinct theory of mind does not exist and executive functions are sufficient for the cognitive skills [18].

Recently, there are new finding about theory of mind of humans. Herrmann *et al*. compared theory of mind ability among human, chimpanzee, and orangutan with gaze following and intention understanding tasks [23]. They concluded that human outperforms other species in theory of mind. Falck-Ytter *et al*. investigated proactive goal-directed eye movements in 12-month old and 6-month old infants using a specialized system for action perception [24]. They concluded that 12-month old infants do the proactive goal-directed eye movements and this is evidence on the action understanding of infants. False-belief test is a representative method to know whether infants have theory of mind. Onishi *et al*. proposed a novel nonverbal task to examine 15-month old infant’s ToM ability [25]. Rosenbaum *et al*. conducted theory of minds tests for someone with severe impairment of episodic memory and autonoetic consciousness [26]. They reported that there is no difference of the ToM ability between normal and impaired persons. Bloom’s research suggested that theory of mind is important to learn meanings of words [27].

There are works on verifying theories of “theory of mind” with neuroscience knowledge. Gallese *et al*. [28] related to the theory of “theory of mind” and the discovery of mirror neurons in human and monkey’s brain. They argued that the finding supports “simulation theory” but not “theory-theory.” Blakemore *et al*. supports the simulation theory based on the psychophysical and neurophysiological studies [29]. Ramnani *et al*. tested “simulation theory” by comparing human brain’s activation for preparing one’s own actions with one for predicting the future actions of others [30]. The conclusion was that both of them use action control system of the human brain but activate different action sub-systems. This result suggests that a simple form of simulation cannot be the only mechanism involved in ToM [31]. Siegal *et al*. reviewed recent findings on the relationships between brain regions and theory of mind [32]. Some functional components found were not solely dedicated to the theory of mind. However, domain-specific component (centered on the amygdale circuitry) was included in the region. This result supports modularity view. Saxe *et al*. related developmental psychology and functional neuroimaging research and supported the modular approach by arguing the existence of a specialized neural system for ToM [33].

Brain-imaging technology has been widely used to pinpoint region of brain for theory of mind [2]. Frith *et al*. used “story comprehension task” to invoke theory of mind and revealed several active regions (medial prefrontal cortex and posterior superior temporal sulcus) of human brain by ToM [34]. McCabe *et al*. reported that prefrontal cortex is highly activated to the cooperator in “trust and reciprocity” games for cash rewards against human [35]. Gallagher *et al*. reviewed several functional imaging works for theory of mind [36]. Krach *et al*. tested human’s ToM with human-robot game and the activation of brain regions related to ToM is related to the human-likeness (computer<functional robot<anthropomorphic robot<human) [37]. Hampton *et al*. investigated the activation of human brain using fMRI when they play simple two-player strategy game [38]. In their game, players use three different strategies (reinforcement

learning, fictitious play based on history of other players, and sophisticated ToM). They investigated brain activation regarding to the choice of the strategy.

The works that implemented ToM are categorized into two groups based on the level of implementations. Some of them focused on the demonstration only with simulation. A few demonstrated their works in real physical robots. The complexity increases when the work is realized in physical robots.

Christopher developed synthetic vision, memory, and theory of mind module for embodied conversational agents [39]. In his work, agent has three theories to do ToM: "Have they seen me", "Have they seen me looking", and "interest level." Robinson *et al*. invented a mind-reading machine recognizes human's mental states (discrete six states) from video input of human's facial expression [40]. Breazeal *et al*. developed synthetic mouse characters that recognize other mouse's behavior based on their own repositories [6]. Treur *et al*. proposed a two-level BDI (Belief, Desire and Intention) model for ToM [41]. The first level was used to model self's BDI and the other was for reasoning about other agent. Marsella *et al*. developed a social simulation tool, PsychSim whose agents have beliefs about other agents [42]. Arita *et al*. [43] and Zanlungo [44] applied ToM to complex agent-based simulations and discussed about the effect of the level of ToM. Kondo *et al*. used the ToM in "carrying a stick task" for the cooperation of two computer programs [45]. Bringsjord *et al*. created a virtual character with a reasoning engine and they demonstrated that the character can pass the false-belief task by inserting "If someone sees something, they know it and if they don't see it, they don't" statement [46].

Kelley *et al*. developed a physical robot that uses own learned experience to detect the intentions of the humans [47]. The experience of robot was encoded into Hidden Markov Models. Breazeal *et al*. created animated robot LEONARDO that infers other person's goals based on the simulation theory [48]. It passes a basic false-belief task. Scassellati developed ToM for a humanoid robot COG based on two representative ToM theories [5]. Yokoya *et al*. used a recurrent neural network to model the relationships between robot's movement and actual object's reaction [49]. After building its own model, it observes human's behavior of rotating objects (blocks) and expanded the original self model to model human's one. Demiris *et al*. followed "simulation theory of mind" and used robot's own motor system to understand other robot's behavior [50]. Takanashi *et al*. inferred other robot's behavior based on its own behavior repositories in the game of robot soccer [51].

There are several works targeted to theory of mind. Kuniyoshi *et al*. developed several skills of simulated and embodied robots for theory of minds: "learning by watching," and "imitation" [52]. Kozima *et al*. proposed a framework to implement and exploit theory of mind from indirect experience of infant humanoid robot [53]. Ono *et al*. assumed that human's theory of mind model is organized as Baron-Cohen's modular view and implemented an interface system to help humans understand robot's intention [54]. Agents migrate from physical robots to user's computer for shared attention. Ito *et al*. also focused on factors related to human's ToM in the interaction of artifacts [55][56]. Scassellati *et al*. built a self model from the relationships between visual input and actual motor movement of robot and used it to discriminate others from self [57]. This is an important skill to do theory of mind. Kramer provided with an overview of the theory of mind in communication with virtual humans [58]. McCabe *et al*. introduced the concept of theory of mind to interpret the results of theoretical games played by humans [59]. They mentioned that the form of games is related to the human's theory of mind execution and produce different outcomes. Boella *et al*. stressed the importance of theory of mind in the construction of social reality with multi-agent systems [60]. Akiwa *et al*. recognized that just imitating human's behavior is not interesting to human demonstrator and proposed a system to predict subject's next action based on past experience [61]. The prediction was done based on the difference between current behavior and past one. Flax modeled Leslie's modular view on the theory of mind using first-order modal logic with an example of a scenario [62]. Hall *et al*. used theory of mind assessment of children to evaluate a virtual character system [63].

3. METHODS

In [64], authors tested ToM learning in simulated robots. In this real robot testing, simulation and a real robot was used together to do the ToM. In actor learning, simulation is used. In observer learning, the trajectories were collected from real robots and simulation was used in EEA. In actor exploitation, the position of new light source to seduce actor's robot was determined with simulation and tested in real robots.

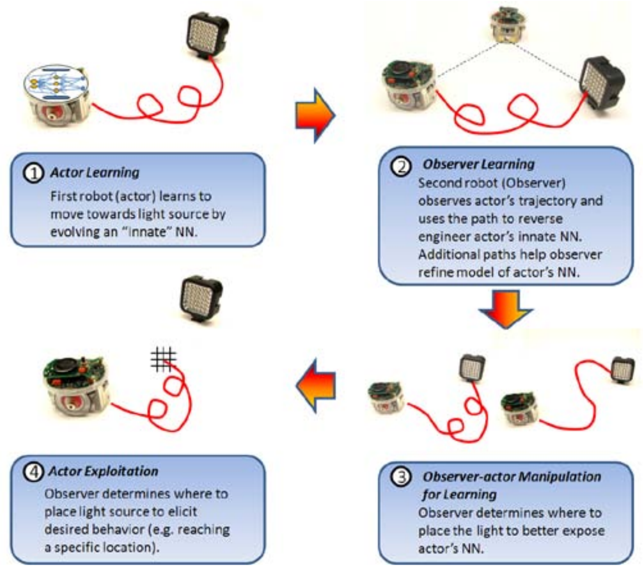

① Actor Learning
First robot (actor) learns to move towards light source by evolving an "innate" NN.
② Observer Learning
Second robot (Observer) observes actor's trajectory and uses the path to reverse engineer actor's innate NN. Additional paths help observer refine model of actor's NN.
④ Actor Exploitation
Observer determines where to place light source to elicit desired behavior (e.g. reaching a specific location).
③ Observer-actor Manipulation for Learning
Observer determines where to place the light to better expose actor's NN.

Figure 3. Overview of ToM learning

3.1 Actor Learning

In the first stage, the neural network controller is evolved for the actor robot. The architecture of neural network is fixed and only the weights are evolved. The sensory inputs (light level) are inputted to the neural network and the output is the movement of wheels. Figure 4 explains the details of the evolutionary algorithm used. Each controller is represented with a vector of weights and each entry has an associated self-adaptive parameter. The mutation operator updates the weights based on the self-adaptive parameter's value. A task is to follow light source and a fitness function is defined based on the distance to the light source.

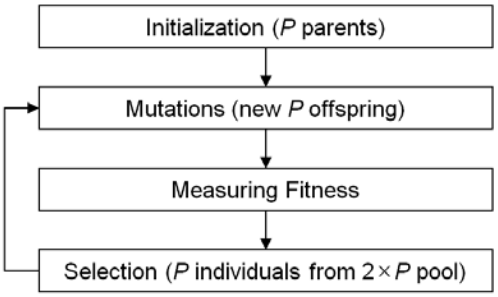

Initialization (*P* parents)
Mutations (new *P* offspring)
Measuring Fitness
Selection (*P* individuals from 2×*P* pool)

Figure 4. The evolutionary procedure to evolve actor robot's controller

3.2 Observer Learning

The goal of this stage is to discover actor robot's self (the neural network evolved) based on their real trajectories. It uses EEA (Estimation-Exploration Algorithm) to learn other's self models [9]. Initially, one trajectory is observed from the actor robot. In Estimation step, it runs learning other's self models multiple times with different random seed and produces multiple candidates (neural networks). In Exploration step, using the candidates, a number of starting points are tested and the EEA chooses the one with the maximum disagreement of the candidates as a next observing point. The next trajectory is observed from the new starting point chosen and the two trajectories are used for the next estimation step. A new population of the estimation step is initialized with the best candidates of the previous estimation step. Evolutionary algorithm is used to learn the other's self model in the estimation step. It is a kind of active incremental learning algorithm.

Figure 5 explains the fitness function in the evolutionary algorithm. The trajectory of the robot is a time-series sequence of the X-Y coordinates. At time t, the robot is placed in (X(t), Y(t)) in the environment and the next position is estimated by a candidate neural network. The fitness was calculated based on the difference between the original position and the estimated one.

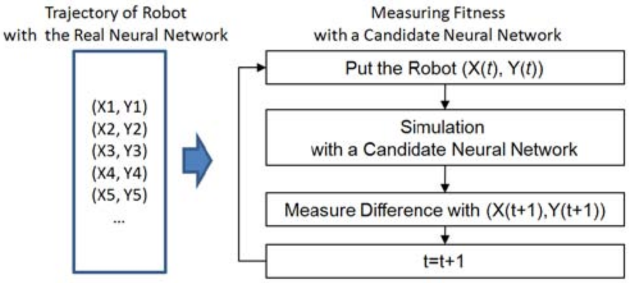

Trajectory of Robot
with the Real Neural Network
(X1, Y1)
(X2, Y2)
(X3, Y3)
(X4, Y4)
(X5, Y5)
...
Measuring Fitness
with a Candidate Neural Network
Put the Robot (X(*t*), Y(*t*))
Simulation
with a Candidate Neural Network
Measure Difference with (X(t+1),Y(t+1))
t=t+1

Figure 5. Fitness measuring of a candidate neural network

3.3 Actor Exploitation

Once actor's self models are discovered, they can be used to predict the robot's trajectory and observer robot can catch it with a trap based on the estimation. A trap is placed in the middle of the light source and a starting position of the other robot. With the actor's self models discovered, a new light position can be estimated to seduce the other robot to the trap. This is called "ToM estimation." The easiest way to predict the other robot's movement is "straight line estimation" assuming that the robot will go straightly to the light source. However, the movement of robot evolved is not straight line and shows several interesting patterns. The two approaches are compared to measure the goodness of our method.

4. EXPERIMENTAL RESULTS

The proposed method was tested in various settings from simulations to real physical robots. In a simulation side, PhysX (A simulator with physics engine) and EnKi (for E-Puck robot) are used. In a physical side, E-Puck robots are used to get results. The robot has two light sensors (left and right) and controls the robot by adjusting the wheels. In PhysX simulation, the neural network outputs are "the rotational angle" and "speed" of wheel. For E-Puck robot, the speed of left and right wheels is outputs of the network. In case of visible trap, the robot can detect the trap located, and left and right sensors digitize the strength of signals from the trap. Each neuron in a neural network has a bias parameter and the arc tangent function is used as a transfer function.

Based on the success of the virtual experiments [64], our experiments were expanded to the real physical robots. In our experiment, E-Puck mobile robots were used. It has two wheels and eight infra-red sensors. Like the virtual cases, only two sensors were used. As a light source, infra-red LED light was used. The trajectory of robot was recorded using Vicon motion capture system. Reflexive balls were attached to robot's custom-built mounting base and the Vicon system recognized the position and angles of the robot based on the balls detected. Our simulator was implemented based on EnKi simulator. In our simulator, a sensor model was built based on sampling data (129 positions ×15 different angles × 8 sensors). Additionally, wheel speed level was readjusted based on real sampled data.

The actor's neural network was evolved at each setting. Figure 6 shows trajectories of the evolved controllers at various starting positions. Their trajectories are not straight line and have a lot of curves. Also, they are very complex and have a lot of rotations to reach the goal position (light source). Although the controllers are evaluated at one starting position in the evolution, they can generalize well for different starting positions.

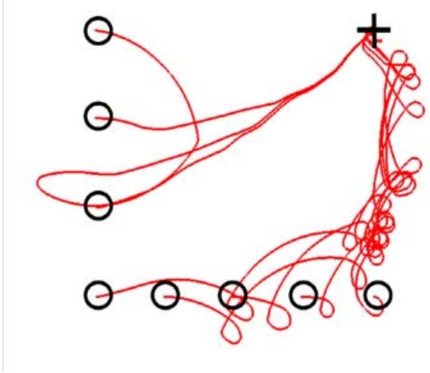


Figure 6. Trajectories of evolved neural controller (Black circle = Initial position, Black cross = Light)

Figure 7, Figure 8 and Figure 10 shows the progress of EEA learning. Figure 9 shows successful exploitation results for real physical robots. In EEA learning, real trajectories were collected from actor's robot. In the exploitation scenario, the new light position was estimated with simulation and tested with real robots. It shows that the reconstructed controllers can be used successfully to seduce the actor robot to the trap.

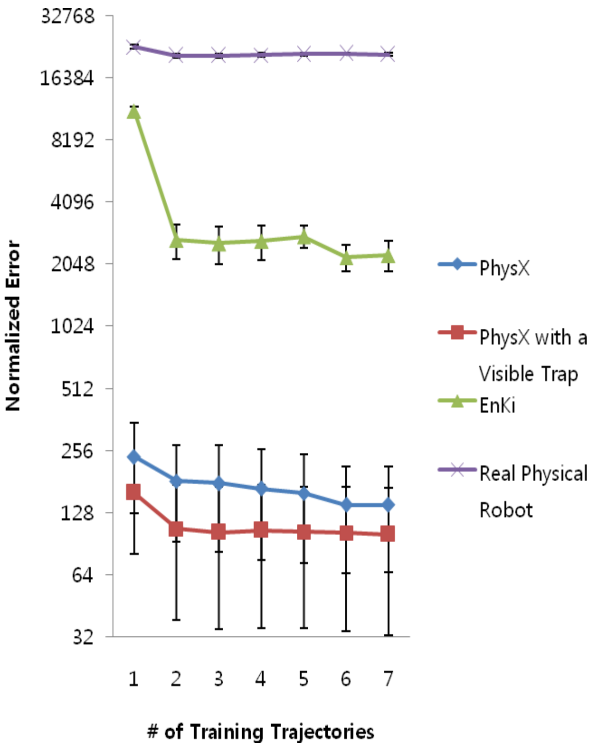


Figure 7. The progress of the observer learning in various environments

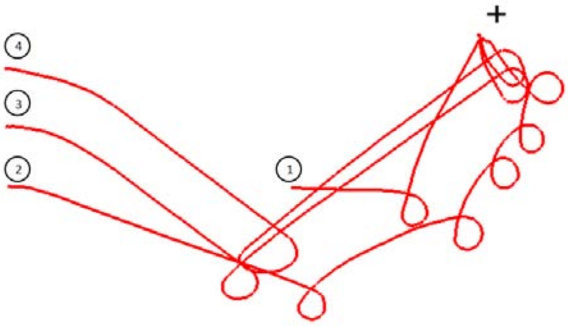


Figure 8. The trajectories actively chosen by the observer learning

Table 1 summarizes errors of all experimental environments. The ToM was compared with straight line estimation (assume that the robot will go straightly to the light source). For all cases, the ToM method can beat the straight line estimation method. In PhysX case, the ensemble of five candidate neural networks was successful and outperforms the single best neural network candidate and the straight line estimation. However, it is not true for the EnKi case and the ensemble method was not used for real robots.

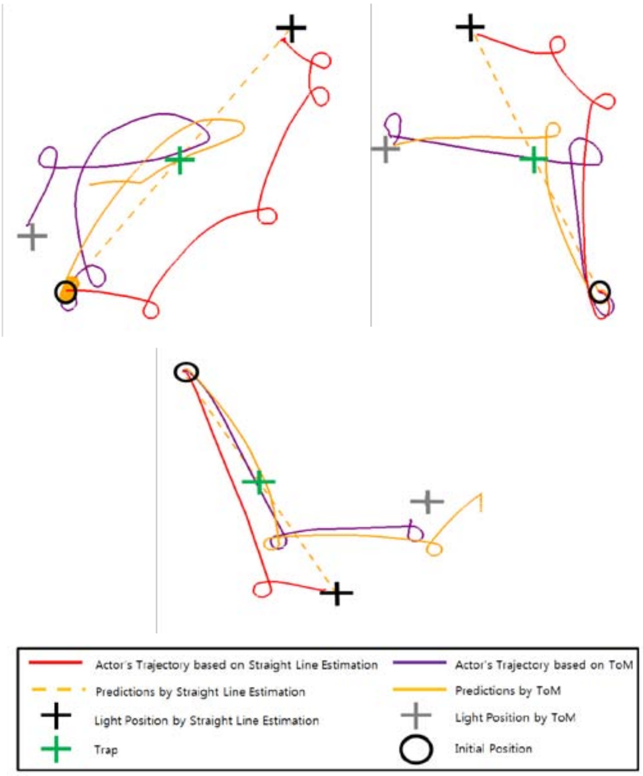

Actor's Trajectory based on Straight Line Estimation
Predictions by Straight Line Estimation
Light Position by Straight Line Estimation
Trap
Actor's Trajectory based on ToM
Predictions by ToM
Light Position by ToM
Initial Position

Figure 9. An example of exploitation for real robots.

Table 1. Statistical summary

	Straight Line Estimation	ToM (Single neural network)	ToM (Ensemble of 5 neural networks)
PhysX[1]	5.93 ± 0.54	3.87 ± 0.69	1.10 ± 0.28
PhysX with a Visible Trap[1]	18.21 ± 2.60	18.29 ± 2.72	11.97 ± 2.24
EnKi[1]	10.75 ± 1.26	0.89 ± 0.30	29.08 ± 5.75
Real Robots[1] **(Simulation)**	54.28 ± 2.84	35.37 ± 3.35	-
Real Robots[2]	34.80 ± 7.66	26.59 ± 9.33	-

1: Average of 100 points
2: Average of 10 points

5. CONCLUSIONS

In this paper, a variety of experiments were conducted to show the possibility of theory of mind implementation for robots. Each robot can model other robot's internal self model (neural network) based on their observation using EEA learning algorithm. Once the model was built, they can be used to predict other robot's future behavior. In these experiments, several virtual experiments and real physical robot testing successfully show the benefit of the other's self modeling.

In this paper, it is assumed that the neural structure of an actor robot is the same with the one of observer and there is no process to identify the fundamental structure. The number of the input-output neurons has to be analyzed to determine the structure of neural networks. After then, there are also many structural considerations: The number of layers, the number of hidden nodes for each layer, and the existence of recurrent links. The solution might be evolving topology and weights of neural networks simultaneously.

6. ACKNOWLEDGMENTS

This work has been supported in part by the U.S. National Science Foundation (NSF) Creative-IT program, grant #0757478. K.-J.K. was supported by the Korea Research Foundation Grant (KRF-2007-357-D00220) funded by Korean Government (MOEHRD) and Korea Health 21 R&D Project, Ministry for Health, Welfare and Family Affairs (A040163)

7. REFERENCES

[1] D. G. Premack, and G. Woodruff, “Does the chimpanzee have a theory of mind?,” Behavioral and Brain Sciences, vol. 1, pp. 515-526, 1978.

[2] C. Zimmer, “How the mind reads other minds,” Science, vol. 300, pp. 1079-1080, 16 May 2003.

[3] M. Siegal, and R. Varley, “Neural systems involved in ‘theory of mind’,” Nature Reviews-Neuroscience, vol. 3, pp. 463-471, June 2002.

[4] J. Call, and M. Tomasello, “Does the chimpanzee have a theory of mind? 30 years later,” Trends in Cognitive Sciences, vol. 12, no. 5, pp. 187-192, 2008.

[5] B. Scassellati, Foundations for a Theory of Mind for a Humanoid Robot, Ph.D. Thesis, Massachusetts Institute of Technology, 2001.

[6] D. Buchsbaum, B. Blumberg, C. Breazeal and A. N. Meltzoff, “A simulation-theory inspired social learning system for interactive characters,” IEEE International Workshop on Robots and Human Interactive Communication, pp. 85-90, 2005.

[7] F. Hegel, S. Krach, T. Kircher, B. Wrede, and G. Sagerer, “Theory of mind (ToM) on robots: A functional neuroimaging study,” Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, pp. 335-342, 2008.

[8] T. Ono, and M. Imai, “Reading a robot’s mind: A model of utterance understanding based on the theory of mind mechanism,” Proceedings of the 17th National Conference on Artificial Intelligence, pp. 142-148, 2000.

[9] J. Bongard, and H. Lipson, “Automated reverse engineering of nonlinear dynamical systems,” Proceedings of the National Academy of Science, vol. 104, no. 24, pp. 9943-9948, 2007.

[10] J. Bongard, V. Zykov, and H. Lipson, “Resilient machines through continuous self-modeling,” Science, vol. 314, no. 5802, pp. 1118-1121, 2006.

[11] C. M. Heyes, “Theory of mind in nonhuman primates,” Behavioral and Brain Sciences, vol. 21, pp. 101-148, 1998.

[12] J. Call and M. Tomasello, "Does the chimpanzee have a theory of mind? 30 years later," Trends in Cognitive Sciences, vol. 12, no. 5, pp. 187-192, 2008.

[13] B. Hare, J. Call, and M. Tomasello, “Do chimpanzees know what conspecifics know?,” Animal Behaviour, vol. 61, pp. 139-151, 2001.

[14] B. Hare, J. Call, B. Agnetta, and M. Tomasello, “Chimpanzees know what conspecifics do and do not see,” Animal Behaviour, vol. 59, pp. 771-785, 2000.

[15] M. Tomasello, J. Call and B. Hare, “Chimpanzees understand psychological states – The question is which ones and to what extent,” Trends in Cognitive Sciences, vol. 7, no. 4, pp. 153-156, 2003.

[16] S. Baron-Cohen, A. M. Leslie, and U. Frith, “Does the autistic child have a “theory of mind”?,” Cognition, vol. 21, pp. 37-46, 1985.

[17] S. Baron-Cohen, Mindblindness, MIT Press, 1997.

[18] S. Ozonoff, B. F. Pennington, and S. J. Rogers, “Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind,” Journal of Child Psychology and Psychiatry, vol. 32, no. 7, pp. 1081-1105, 1991.

[19] G. L. Youmans, Theory of Mind in Individuals with Alzheimer-Type Dementia Profiles, Ph.D. Thesis of College of Communication at the Florida State University, 2004.

[20] A. M. Leslie, O. Friedman, and T. P. German, “Core mechanisms in ‘theory of mind’,” Trends in Cognitive Sciences, vol. 8, no. 12, pp. 528-533, 2004.

[21] R. Langdon, and M. Coltheart, “Visual perspective taking and schizotypy: Evidence for a simulation-based account of mentalizing in normal adults,” Cognition, vol. 82, 1-26, 2001.

[22] A. Gopnik and H. Wellman, “Why the child’s theory of mind really is a theory,” Mind and Language, vol. 7, pp. 145-171, 1995.

[23] E. Herrmann, J. Call, M. V. Hernandez-Lloreda, B Hare, and M. Tomasello, “Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis,” Science, vol. 317, pp. 1360-1366, 2007.

[24] T. Falck-Ytter, G. Gredeback and C. von Hofsten, “Infants predict other people’s action goals,” Nature Neuroscience, vol. 9, no. 7, pp. 878-879, 2006.

[25] K. H. Onishi, and R. Baillargeon, “Do 15-month-old infants understand false beliefs?,” Science, vol. 308, pp. 255-258, 2005.

[26] R. S. Rosenbaum, D. T. Stuss, B. Levine and E. Tulving, “Theory of mind is independent of episodic memory,” Science, vol. 318, p. 1257, 2007.

[27] P. Bloom, “Precis of how children learn the meanings of words,” Behavioral and Brain Sciences, vol. 24, no. 6, pp. 1095-1103, 2001.

[28] V. Gallese and A. Goldman, “Mirror neurons and the simulation theory of mind-reading,” Trends in Cognitive Sciences, vol. 2, no. 12, pp. 493-501, 1998.

[29] S.-J. Blakemore, and J. Decety, “From the perception of action to the understanding of intention,” Nature Reviews – Neuroscience, vol. 2, pp. 561-567, 2001.

[30] N. Ramnani, and R. C. Miall, “A system in the human brain for predicting the actions of others,” Nature Neuroscience, vol. 7, no. 1, pp. 85-90, 2004.

[31] N. Sebanz and C. Frith, “Beyond simulation? Neural mechanisms for predicting the actions of others,” Nature Neuroscience, vol. 7, no. 1, pp. 5-6, 2004.

[32] M. Siegal, and R. Varley, “Neural systems involved in ‘theory of mind’,” Nature Reviews-Neuroscience, vol. 3, pp. 463-471, June 2002.

[33] R. Saxe, S. Carey, and N. Kanwisher, “Understanding other minds: Linking developmental psychology and functional neuroimaging,” Annual Review of Psychology, vol. 55, pp. 87-124, 2004.

[34] C. D. Frith, and U. Frith, “Interacting minds-A biological basis,” Science, vol. 286, pp. 1692-1695, 1999.

[35] K. McCabe, D. Houser, L. Ryan, V. Smith, and T. Trouard, “A functional imaging study of cooperation in two-person reciprocal exchange,” Proceedings of the National Academy of Sciences, vol. 98, no. 20, pp. 11832-11835, 2001.

[36] H. L. Gallagher, and C. D. Frith, “Functional imaging of ‘theory of mind’,” Trends in Cognitive Sciences, vol. 7, no. 2, pp. 77-83, 2003.

[37] S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher, “Can machines think? Interaction and perspective taking with robots investigated via fMRI,” PLOS One, vol. 3, no. 7, e2597, 2008.

[38] A. N. Hampton, P. Bossaerts, and J. P. O’Doherty, “Neural correlates of mentalizing-related computations during strategic interactions in humans,” Proceedings of the National Academy of Sciences, vol. 105, no. 18, pp. 6741-6746, 2008.

[39] C. Peters, “A perceptually-based theory of mind for agent interaction initiation,” International Journal of Humanoid Robotics, vol. 3, no. 3, pp. 321-339, 2006.

[40] R. E. Kaliouby, and P. Robinson, “Mind reading machines: Automated inference of cognitive mental states from video,” IEEE International Conference on Systems, Man and Cybernetics, pp. 682-688, 2004.

[41] T. Bosse, Z. A. Memon, and J. Treur, “A two-level BDI-agent model for theory of mind and its use in social manipulation,” Proceedings of the Artificial and Ambient Intelligence Conference, pp. 335-342, 2007.

[42] D. V. Pynadath, and S. C. Marsella, “PsychSim: Theory of mind with decision-theoretic agents,” Proceedings of the International Joint Conference on Artificial Intelligence, pp. 1181-1186, 2005.

[43] M. Takano, and T. Arita, “Asymmetry between even and odd levels of recursion in a theory of mind,” Proceedings of ALIFE X, pp. 405-411, 2006.

[44] F. Zanlungo, “A collision-avoiding mechanism based on a theory of mind,” Advances in Complex Systems, vol. 10, no. 2, pp. 363-371, 2007.

[45] K. Kondo, and I. Nishikawa, “The role that the internal model of the others plays in cooperative behavior,” Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, pp. 265-270, 2003.

[46] S. Bringsjord, A. Shilliday, D. Werner, M. Clark, E. Charpentier, and A. Bringsjord, “Toward logic-based cognitively robust synthetic characters in digital environments,” Proceedings of the First Artificial General Intelligence, pp. 87-98, 2008.

[47] R. Kelley, C. King, A. Tavakkoli, M. Nicolescu, M. Nicolescu, and G. Bebis, “An architecture for understanding intent using a novel hidden markov formulation,” International Journal of Humanoid Robotics, vol. 5, no. 2, pp. 1-22, 2008.

[48] C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby, and B. Blumberg, “Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots,” Artificial Life, vol. 11, pp. 31-62, 2005.

[49] R. Yokoya, T. Ogata, J. Tani, K. Komatani, and H. G. Okuno, “Discovery of other individuals by projecting a self-model through imitation,” IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1009-1014, 2007.

[50] Y. Demiris, and M. Johnson, “Distributed, predictive perception of actions: A biologically inspired robotics architecture for imitation and learning,” Connection Science, vol. 15, no. 4, pp. 231-243, 2003.

[51] T. Takanashi, T. Kawamata, M. Asada, and M. Negrello, “Emulation and behavior understanding through shared values,” IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3950-3955, 2007.

[52] Y. Kuniyoshi, Y. Yorozu, Y. Ohmura, K. Terada, T. Otani, A. Nagakubo, and T. Yamamoto, “From humanoid embodiment to theory of mind,” Lecture Notes in Artificial Intelligence, vol. 3139, pp. 202-218, 2004.

[53] H. Kozima, and J. Zlatev, “An epigenetic approach to human-robot communication,” Proceedings of the 2000 IEEE International Workshop on Robot and Human Interactive Communication, pp. 346-351, 2000.

[54] T. Ono, and M. Imai, “Reading a robot’s mind: A model of utterance understanding based on the theory of mind mechanism,” Proceedings of the 17th National Conference on Artificial Intelligence, pp. 142-148, 2000.

[55] K. Terada, T. Shamoto, H. Mei, and A. Ito, “Reactive movements of non-humanoid robots cause intention attribution in humans,” IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3715-3720, 2007.

[56] K. Terada, T. Shamoto, and A. Ito, “Utilizing theory of mind on human agent interaction,” IEEE International Symposium on Robot and Human Interactive Communication, pp. 757-762, 2006.

[57] P. Michel, K. Gold, and B. Scassellati, “Motion-based robotic self-recognition,” Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2763-2768, 2003.

[58] N. C. Kramer, “Theory of mind as a theoretical prerequisite to model communication with virtual humans,” Lecture Notes in Artificial Intelligence, vol. 4930, pp. 222-240, 2008.

[59] K. A. McCabe, V. L. Smith, and M. Lepore, “Intentionality detection and “mindreading”: Why does game form matter?,” Proceedings of the National Academy of Sciences, vol. 97, no. 8, pp. 4404-4409, 2000.

[60] G. Boella, L. van der Torre, “From the theory of mind to the construction of social reality,” Proceedings of CogSci, pp. 298-303, 2005.

[61] Y. Akiwa, Y. Suga, T. Ogata, and S. Sugano, “Imitation based human-robot interaction-roles of joint attention and motion prediction,” Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication, pp. 283-288, 2004.

[62] L. Flax, “Logical modeling of Leslie’s theory of mind,” Proceedings of 5th IEEE International Conference on Cognitive Informatics, pp. 25-30, 2006.

[63] L. Hall, S. Woods, R. Aylett, and A. Paiva, “Using theory of mind methods to investigate empathic engagement with synthetic characters,” International Journal of Humanoid Robotics, vol. 3, no. 3, pp. 351-370, 2006.

[64] K.-J. Kim and H. Lipson, “Towards a “theory of mind” in simulated robots,” *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation Conference*, pp. 2071-2076, 2009.

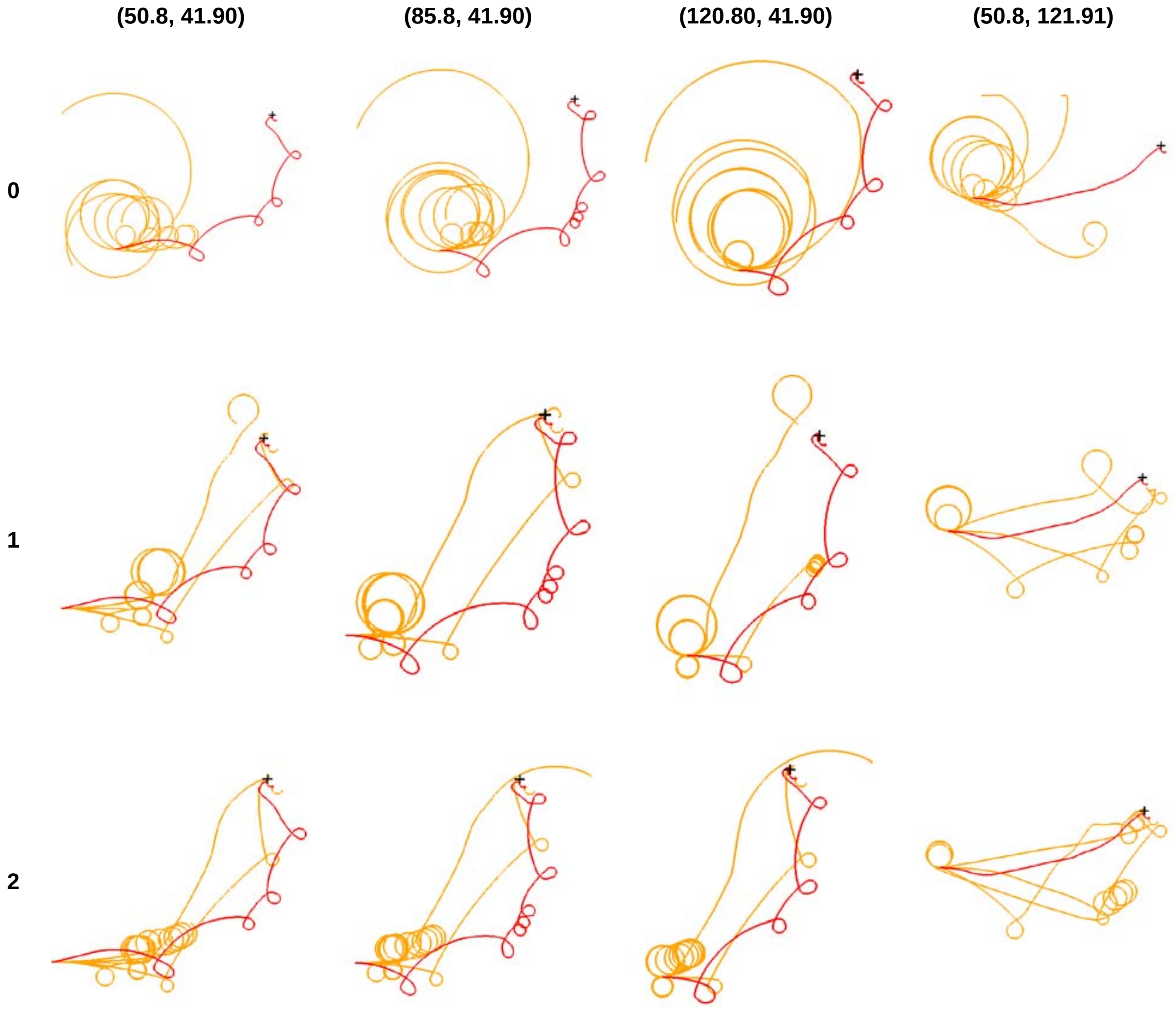


Figure 10. Progress of observer learning for real robots (Red line = Real Trajectory, Yellow line = Predicted trajectory)