The role that the internal model of the others plays in cooperative behavior

Conference Paper · January 2003

DOI: 10.1109/ROMAN.2003.1251856 · Source: IEEE Xplore

Ikuko Nishikawa
Ritsumeikan University

Proceedings of the 2003 IEEE International Workshop on
Robot and Human Interactive Communication
Millbrae, California, USA, Oct. 31 - Nov. 2, 2003

The role that the Internal Model of the Others Plays in Cooperative Behavior

Keisuke Kondo
Dept. of Economics
Kyoto Gakuen University
Kameoka, Kyoto 621-8555 Japan
kondo@kyotogakuen.ac.jp

Ikuko Nishikawa
Dept. of Computer Science
Ritsumeikan University
Kusatsu, Shiga 525-8577 Japan
nishi@cs.ritsumei.ac.jp

Abstract

Internal model of the others is claimed to be essential for the prediction of the others' behavior to realize a cooperative behavior. " Theory of mind" is known as a leading framework of the internal modeling. The present research consists of two parts. In the former part, experimental results are shown on the cooperation between a human subject and a software model of a mobile robot. The cooperative task given to the human subject is to carry a stick from a start point to a goal point by holding the one edge of the stick, in a simulation world displayed on a computer terminal. A virtual robot holds the other edge, whose behavior is controlled by a simple deterministic rule. The human subject is asked to perform the task under the following two different conditions A and B. In condition A, the subject is not told about the existence of the robot which holds the other edge, and only the stick is displayed. In condition B, the existence of the robot is told, but still not shown on the display. The performance of the carrying task is much better in the condition A, under which the subject can separate the movements of the stick and of the robot for their prediction. This result suggests that a construction of a behavior model of the other robot is helpful to carry out the cooperative task. In the latter part, a neural network replaces the human subject to perform the same cooperative task with the robot. The network outputs a motion to move the stick, and has a simple layered structure with an additional part to learn to predict the next motion of the robot. The experiments show the improvement in the task performance through learning the prediction. This also suggests the explicit modeling of the others is effective for the cooperation.

1 Introduction

Internal model of the others is thought to be essential for the prediction of the others' behavior to realize a cooperative behavior. "Theory of mind" is known as a leading framework of the internal modeling, which is proposed by Premack in 1978. [1]. He reported an experimental finding on a certain behavior of a monkey under a certain circumstance, which suggested that a monkey has an internal model of the other monkey, and he explained it by " theory of mind ". Simon Baron-Cohen proposed the hypothesis that autistic children lack this " theory of mind " in 1995, based on the experimental results that about 20% of the autistic children, who were almost mentally intact, failed in so-called " false belief " task. He explained this failure as an inability to represent the mental states, and thus an inability to predict the behavior of the other. In this research, we experimentally investigate the effectiveness of an internal model of the other for the cooperative behavior with the other. The present paper consists of two parts. In the first part, experimental results are shown on the cooperation between a human subject and a simulation model of a mobile robot. The cooperation task given to the human subject is explained in Section 2, and the details of the experiments and the results are shown in Section 3. The second part is described in Section 4. A neural network replaces the previous human subject to perform the same cooperative task with a robot. The neural network has a part to predict the next motion of the robot. The learning of this prediction part, or an internal model of the robot in a sense, is compared with the improvement of the task performance. Section 5 concludes the paper.

2 The task of the cooperative behavior

2.1 An environment of the experiment

A task is to move a stick from a start point to a goal point by the cooperation of two robots which hold both ends of the stick. The aim is to reach the goal as fast as possible without dropping a stick object. The experiments are based on a computer simulation, and a goal, two robots and a stick are displayed on the computer terminal. There is no obstacle in this 2 dimensional environment, as is shown in Figure 1.

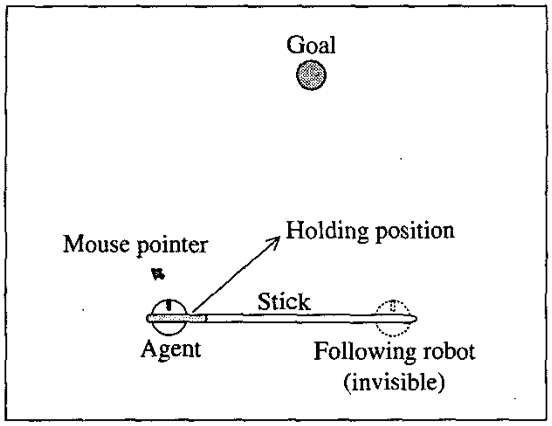

Figure 1: Environment for the cooperative task

The reaching time to the goal is measured for the evaluation of the task performance. The evaluation E is defined as t_0 over t, as the following,

$$E = \frac{t_0}{t} \tag{1}$$

where t_0 is a constant given as the possible shortest reaching time to the goal without carrying any stick, and t is the reaching time at each task trial. When a stick is dropped, the trial ends and E is set to 0.

2.2 A mobile robot

The robot in the simulation moves by two wheels on both sides. Possible movements of the robot are classified into eight actions according to the difference between right and left motor speeds. The eight actions are: 1) move straight ahead, 2) move forward to turn left, 3) move forward to turn right, 4) round anticlockwise, 5) round clockwise, 6) move backward to turn left, 7) move backward to turn right, and 8) move straight backward, which are shown in Figure 2.

The robot can hold only the edge part of the stick, which is indicated by blue color on the display. If the robot detaches the blue part, the stick is judged as dropped. No physical interaction is considered between the robot and the floor or the stick.

Between two robots at both edges, one is controlled by a human subject or a neural network, in the first experiment (Section 3) or in the second experiment (Section 4), respectively. This robot is referred as the agent in the following sections. The behavior of another robot is given by a simple deterministic rule to follow the stick. This robot is referred as the following robot. Action of the following robot is restricted to only three kinds of action, and given by the rules below (Figure 3),

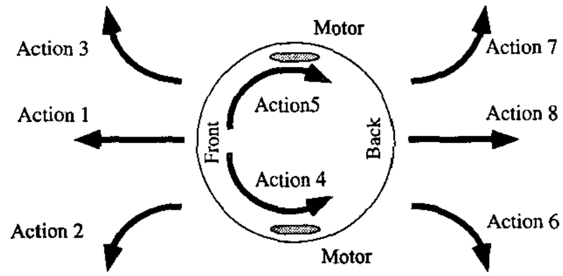

Figure 2: Eight actions of the robot

1. Turn right, when pushed from the left or pulled to the right by the stick.
2. Turn left, when pulled from the left or pushed from the right by the stick.
3. Move straight ahead, otherwise.

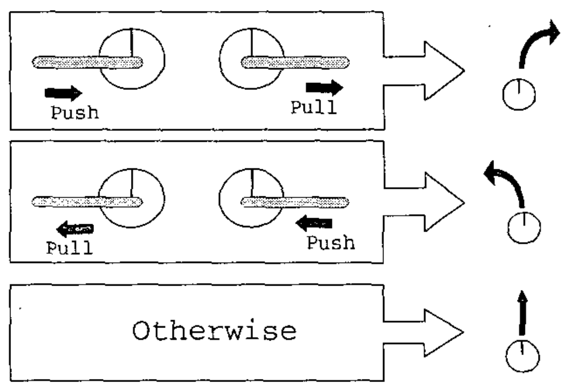

Figure 3: Action rules of the following robot

3 Cooperation task for a human subject

In this first experiment, a human subject uses the mouse to control the agent. For example, when the mouse is located on the front side of robot, then the agent goes straight ahead. A reaching time t is measured for the evaluation. The experiment composed of 3 stages of Practice, Condition A and Condition B.

Practice: A practice environment has three kinds of course without the following robot nor a stick.

The courses are shown Figure 4. The first course has only the goal point. The second course has double circular course. The third course has a right-angled course.

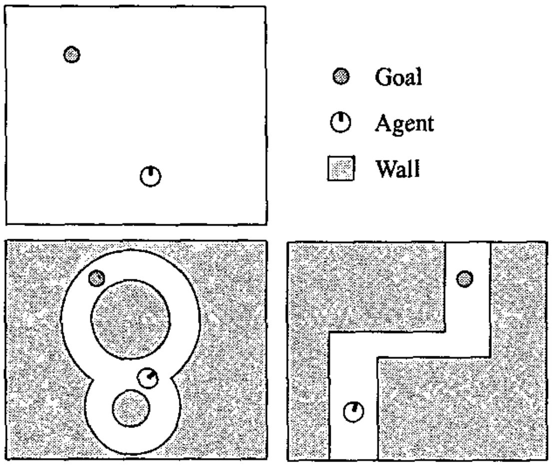

Figure 4: 3 courses in Practice stage

Until the subject feels confident to have fully learned the way to control the motion of the agent, the subject is asked to repeat the trail in this Practice courses. Then, the subject proceeds to the experiments under two different conditions A and B. The difference of the conditions A and B is the information given human subject.

Condition A: The subject is not told about the existence of the following robot, which holds the other edge. It is not displayed on the terminal, neither.

Condition B: The subject is told about the existence of the following robot, which holds the other edge. But it is still not displayed.

3.1 Experimental procedures

As a preliminary result, a human subject tries to find any regularity from the movement of the stick. And a subject seems to discover regularity after about 20 trials, to attain almost similar performance under condition A to condition B. Therefore, the number of trails in the experiments is set to 5 for both conditions A and B, as it seems difficult to find any regularity from only 5 trials, in general.

A different environment is given to each trial, which is shown in Figure 5. As is shown in the figure, the difference is only an initial direction of the following robot. The reaching time t is measured as an evaluation. In addition to conditions A and B, experiments are done also in condition C. In condition C, a human subject is told about the existence of the following robot, and it is also displayed on the terminal. The results obtained in conditions B and C are compared.

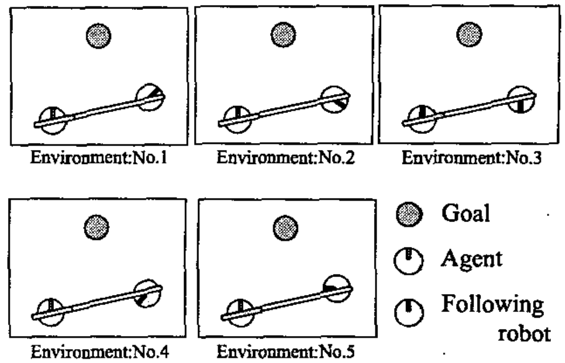

Figure 5: Five environments

The experimental procedures are shown in Figure 6.

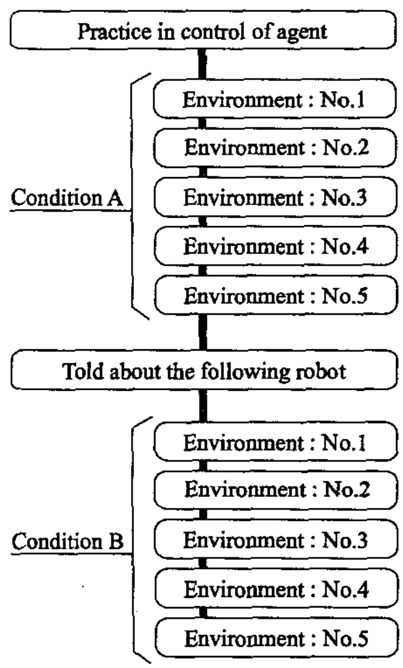

Figure 6: Flowchart

3.2 The result of the Experiments

The number of human subjects is 88. The average of the obtained evaluation value E in each environment are shown in Table 1.

Table 1: *Averages of E in environments No.1 to 5*

	E in each environment					
Condition	No.1	No.2	No.3	No.4	No.5	Ave.
A	0.32	0.33	0.01	0.45	0.49	0.32
B	0.55	0.52	0.02	0.62	0.59	0.46

E in environment No.3 is much lower than the others, because many subjects drop the stick during a trail. This difficulty is caused by the initial direction of the following robot, which is opposite to the goal only in this environment. E is higher under condition B than under condition A, in all the environments. The ratios of 88 subjects with higher E value under condition B than A, and vice versa, are shown in Table 2. 64.8% of 88 subjects show the better performance under condition B, as the average in 5 environments.

Table 2: *Ratios that E is better, equal or worse under condition B compared with A, in 5 environments*

	Ratio in each environment(%)					
	No.1	No.2	No.3	No.4	No.5	Ave.
A<B	87.5	69.3	5.7	86.4	75.0	64.8
A=B	1.1	5.7	90.9	2.3	4.5	20.9
A>B	11.4	25.0	3.4	11.4	20.5	14.3

The distribution of both E values of 88 subjects in environment No.1 is plotted in Figure 7.

The above results indicate that the prediction of the next movement of the stick becomes easier when a subject can model the movements of the stick and of the following robot separately, to attain a shorter reaching time.

4 Cooperation task for a neural network

In this section, the movement of the agent is given by the output from a neural network. The neural network learns for the better performance based on a conventional back-propagation (BP) method. The network outputs a motion to move the stick, and has a simple 2- layered structure with an additional part to learn to predict the next motion of the following robot. The details of the structure will be described in section 4.1. Inputs to the network are a direction of the goal (θ_1), a direction of the stick (θ_2) and a

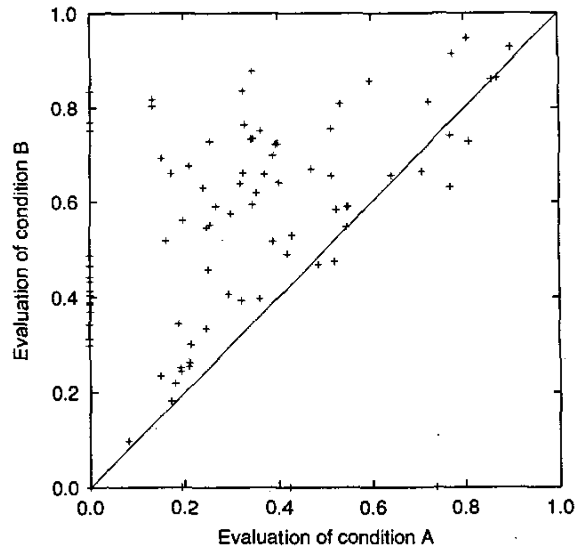

Figure 7: Distribution of E values of 88 subjects in environment No.1

holding position of the stick (dis) (See Figure 8). The following robot is the same as is described in section 3.

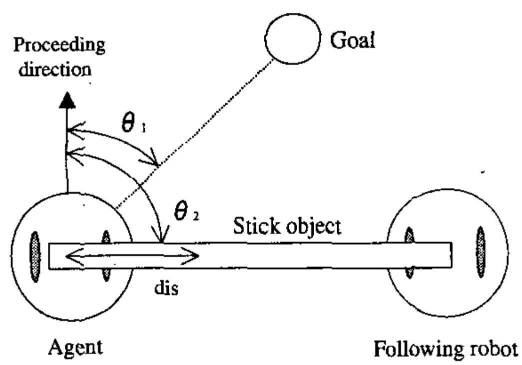

Figure 8: 3 inputs to the neural network

4.1 Learning of a neural network

A structure of the neural network is shown in Figure 9, with input and output values. The network has a simple 2-layered structure with an additional hidden layer for the prediction of the motion of the following robot.

The inputs $x_i(i = 0, \cdots, 7)$ consist of $\sin\theta_1$, $\cos\theta_1$, $\sin\theta_2$, $\cos\theta_2$, $\delta\theta_1$, $\delta\theta_2$, dis and δdis. The outputs $y_j(j = 0, \cdots, 7)$ consist of eight actions of the agent (See Figure 2 in section 2.2). The hidden units $z_k(k =$

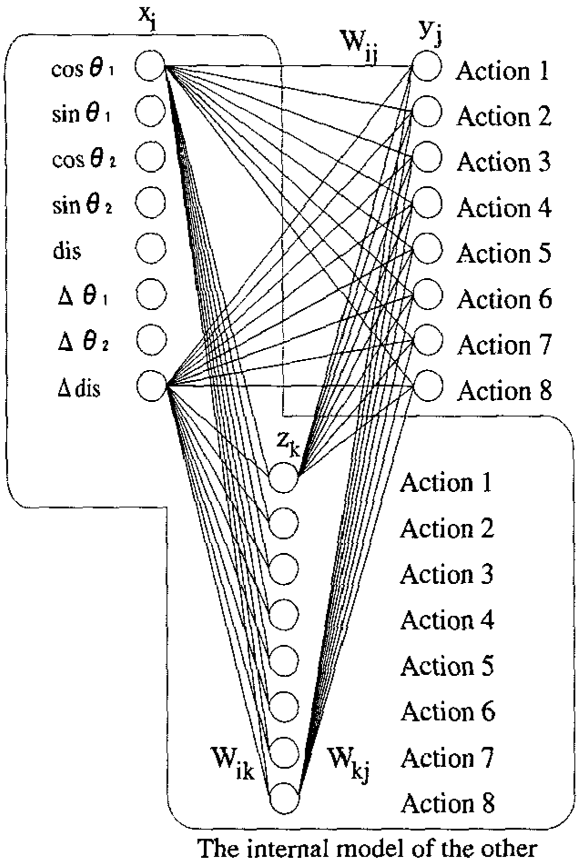

Figure 9: Structure of a neural network

$0, \cdots, 7$) correspond to the same eight actions of the following robot. The outputs from the input values are given by the following formula;

$$z_k(t) = \sum_{i=0}^{7} w_{ik} x_i(t), \ k = 0, \cdots, 7 \quad (2)$$

$$y_j(t) = \sum_{i=0}^{7} w_{ij} x_i(t) + \sum_{k=0}^{7} w_{kj}(t) z_k(t), \ j = 0, \cdots, 7 \quad (3)$$

And the winner of 8 output units determines the next behavior of the agent, according to the followings;

$$K_f(t) = \arg\max_k z_k(t) \quad (4)$$

$$J_a(t) = \arg\max_j y_j(t) \quad (5)$$

where $K_f(t)$ is the network prediction of the next motion of the following robot, and $J_a(t)$ is the motion of the agent which will be taken. In a sense, these are the decisions by a hidden layer and an output layer, respectively.

The weights w_{ij} and w_{kj} are updated by Hebbian type learning according to an obtained reward. A learning rule is given as follows [4, 5, 6, 7];

$$w_{ij}(t+1) = w_{ij}(t) + \alpha r(t) \sum_{n=0}^{100} \gamma^n x_i(t-n) \delta_{j J_a(t-n)} \quad (6)$$

$$w_{kj}(t+1) = w_{kj}(t) + \alpha r(t) \sum_{n=0}^{100} \gamma^n z_k(t-n) \delta_{j J_a(t-n)} \quad (7)$$

where α is a learning rate, γ is a discount ratio, and δ_{ij} is Kronecker's delta. r is a reward, which is given to the agent on approaching or arriving to a goal (a positive reward), and also according to a possibility or a realization of dropping a stick (a negative reward as a penalty).

The weights w_{ik} are updated by Delta rule;

$$w_{ik}(t+1) = w_{ik}(t) + \beta(z_{teach}(t) - z_k(t)) x_i(t) \quad (8)$$

, where $z_{teach}(t)$ is a real motion taken by the following robot, and β is a learning rate.

Each weight is also normalized by the followings to avoid a divergence during the learning;

$$w_{ij}^{new} := C_{w_j} \cdot \frac{w_{ij}^{old}}{|w_j|} \quad (\ C_{w_j}^2 = 1600) \quad (9)$$

$$w_{kj}^{new} := C_{w_j} \cdot \frac{w_{kj}^{old}}{|w_j|} \quad (\ C_{w_j}^2 = 1600) \quad (10)$$

$$w_{ik}^{new} := C_{w_k} \cdot \frac{w_{ik}^{old}}{|w_k|} \quad (\ C_{w_k}^2 = 800) \quad (11)$$

4.2 The result of the Experiments

The network is trained by the carrying task in 5 environments shown in Figure 5. One learning step consists of one trail at each environment. Figure 10 shows a learning process during 100 learning steps. An evaluation value E is obtained as the average over 5 environments at each learning step, and a hitting ratio means a rate of the correct prediction on the motion of the following robot. At the beginning of the learning, the hitting ratio is about 1/8, which means no correlation with a real motion, while it is improved by the learning to attain a greatly increased E value.

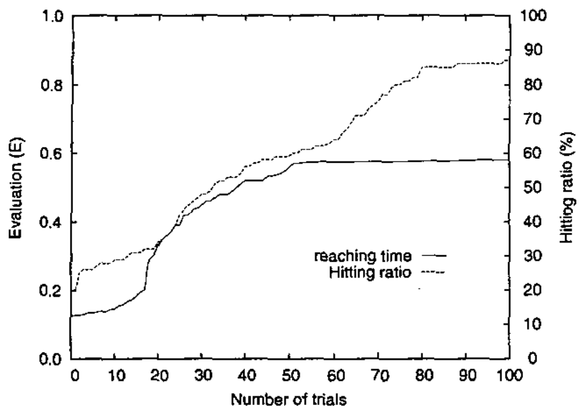

Figure 10: The result obtained by a learning network

5 Conclusion

Modeling of the others and the prediction of their behavior are thought essential to attain an effective cooperation with the others. In the present research, we showed some experimental results which suggest that an explicit modeling of the others is effective for the cooperation. In the first experiment, a human subject showed a better cooperation performance when the information is given on the existence of the other robot to cooperate with. In the second experiment, a neural network attained a higher performance through the improvement of the prediction of the other robot to cooperate with.

In the future, we will tackle more difficult cooperation such as a task in the environments with obstacles, an adaptive cooperation in the varying environments, or an adaptive cooperation with a varying rule of the others' motion, as a basic research to design an effective interface for a robot-human system.

Acknowledgement

A part of this research was by Grant-in-Aid for Young Scientists (B) of the Ministry of Education, Culture, Sports, Science and Technology of Japan. (KAKENHI #15700224)

References

[1] David Premack, Guy Woodruff,"Does the chimpanzee have a theory of mind?", The Behavior and Brain Science. No.4, pp. 512-526 (1978)

[2] Baron-Cohen,"Mindblindness: An Essay on Autisum and Theory of mind", The MIT press (1995)

[3] Masuo Koyasu, "Iwanami Libraryof science: Theory of mind", Iwanami press, 2000.

[4] R. S. Sutton and A. G. Barto : *Reinforcement Learning, An Introduction*, MIT Press (1998)

[5] Kondo, Nishikawa and Tokumaru, "A study on an autonomous mobile robot control with behavior network", Int. Symp. on AROB, pp. 80–83 (1997)

[6] K. Kondo, I. Nishikawa, " Hebbian learning of behavior network and its application to a collision avoidance by an autonomous agent", *Transactions of the institute of system, control and information engineers*, Vol. 15, No. 7, pp. 350-358, 2002.

[7] Kondo, Nishikawa, "The internal model of the other for learning the cooperative behavior", Int. Symp. on AROB, pp. 28–31 (2003)