

# Group Behavior Learning in Multi-Agent Systems Based on Social Interaction among Agents

Kun Zhang\*, Yoichiro Maeda\*\* and Yasutake Takahashi\*\*

\*Dept. of System Design Engineering, Graduate School of Engineering, University of Fukui

\*\*Dept. of Human and Artificial Intelligent Systems, Graduate School of Engineering, University of Fukui

3-9-1, Bunkyo, Fukui, 910-0017 Japan

email: (kzhang, maeda, yasutake)@ir.his.u-fukui.ac.jp

**Abstract**—The research of multi-agent systems which autonomous agents are able to learn the cooperative behavior is expected in recent years. In our laboratory, we aim at the group behavior generation of the multi-agents who have excellent autonomous learning ability like human through the social interaction between agents to acquire the cooperative behavior. The sharing of environment state can improve the cooperation ability and changing the environment state to the information among agents will get the better cooperation ability. On this basis, we use the reward redistribution among agents to reinforce the group behavior and propose a construction method of the multi-agent system with an autonomous group creation ability, which is able to strengthen the cooperative behavior of the group, as a social agent.

## I. INTRODUCTION

As one of the cooperative distributed systems, the multi-agent system (MAS) is actively researched [1]. In MAS, more than one autonomous robots cooperate to achieve the objective. On the other hand, the reinforcement learning (RL) is one of machine learning methods and learns the suitable behavior to solve the problem autonomously by repeating trial and error without teaching signals [2]. It is easy for RL to adjust to a dynamic environmental transformation flexibly. The construction of MAS is expected that is able to learn the cooperative behavior and the group strategy by the application of the RL to the autonomous robots [3].

However, RL is a technique developed for a single agent. If it's used for a MAS like that, the problems like incomplete perception problem, the problem of learning at the same time, and the reward distribution problem will be generated [4]. In a cooperative social environment, it's an effective method for agents to learn by sharing sensing information, episodes, learned policies or exchanging advice [5]. The sharing of environment state can improve the cooperation ability. But one of the main problem concerned learning in a MAS environment is: How can agents benefit from mutual interaction during the learning process [6]? Some researchers undertake the interaction influence by handling the loss of stationarity based on game theory, such as Fierend-or-foe q-learning in general-sum games [7] and Correlated-q learning [8]. But these researchers maybe fail to learn best-response policies even against simple non-learning opponents. Besides, best-response learners to learn best-responses by adjusting

self-strategies or modeling strategies of opponents, but still considering whether the resulting algorithm converges in some form [9].

In this paper, agents can exchange the information about states of itself related to target. It's different from exchanging information having no relationship with target. In addition, it's possible for exchanging rewards among agents to reinforce the coordinative behavior autonomously when the expected actions appeared. Therefore, when an agent got some environment state that other agents can't get for their restricted perception, it will change the environment state to the special signal information among agents to tell other agents. Under the hint of special signal information, the overall performance of agents will be gradually increased than sharing of environment state only especially in the complex joint tasks.

Furthermore, in MAS although the purpose of each agent is to obtain more desired rewards, it doesn't mean that the whole system can obtain more desired rewards. If the reward can't be distributed appropriately, it not only has a bad influence for the behavior of each agent, but also declines the overall performance of system [10]. In this paper, we proposed a method by exchanging the reward among agents to solve the reward distribution problem. All the agents not only get reward from environment but also exchange the reward among each others based on signal information hints and their actions. Agents are able to get excellent cooperative behavior by taking reward as the social interaction strategy like human in society.

## II. MULTI-AGENT REINFORCEMENT LEARNING

Multi-agent systems (MAS) have been widely used to solve problems which are difficult or impossible for an individual agent in a monolithic system by enormous autonomous agents. On the other hand, reinforcement learning (RL) is an attractive method, as it allows the agent to learn behavior on the basis of sparse, delayed reward signals provided only when the agent reaches desired goals. The popular reinforcement learning methods are Q-Learning and Profit Sharing [11]. MAS may be formed by adaptive agents which interact and cooperate for the resolution of certain tasks using RL algorithms [2].

### A. Multi-Agent System

A multi-agent system (MAS) is composed of multiple autonomous agents to solve problems by the cooperation of each other autonomously. Through the communication and cooperation between intelligent agents, it can be used to deal with the incomplete and uncertain problems [12]. The actions of agents are unrestricted by each other in MAS, but the actions of any one agent will have a significant impact to other agents not only to affect the environment. So the interaction of agents becomes very important especially in the joint tasks that can't be completed by a single agent.

### B. Reinforcement Learning in MAS

The reinforcement learning (RL) is a computational paradigm of learning in which an algorithm attempts to maximize a performance measurement based on the reward that it gets upon interacting with an environment [13]. The cooperation learning of multi-agents becomes possible by using RL to MAS. In the reinforcement learning of MAS, the state of environment will change to a new environment by the influence of multi-agents not like as a single agent in the normal RL. The learning of agents becomes difficult than the single environment because the actions of a single agent brings unexpected affect to other agents and environment [14].

As compared with the normal RL, the RL of MAS keeps the previous problems in normal RL and also has its unique problems that are indicated by Arai [4] as follows: 1) When the state space is large, the situation is unable to learn because the agent's perception is limited or the perceptive information is imperfect. 2) When two or more agents learn independently, it's difficult to judge the learning result should be due to which. 3) Although the task was completed, how to distribute the reward etc.

## III. REINFORCEMENT LEARNING BASED ON SOCIAL INTERACTION IN MAS

In this paper, the agents learn the cooperative behavior through the communication of the environment state and reward among each agent as shown in Fig.1. In the normal RL, an individual agent can repeat only the trial and error through the interaction with the environment. In our proposed method, all the agents can learn cooperative behavior by the trial and error among agents autonomously through the interaction information not only between the environment.

We aim at the formation of group behavior of multi-agents who not only act based on own profit but also considering a joint profit with other agents through the communication of the reward as social agents.

### A. Reinforcement Learning Sharing Environment State (RL-SES)

In the normal RL, the agent selects its action based on the environment state. But in the MAS environment, the normal RL has some problems such as incomplete perception problem indicated by Arai. In a cooperative social environment, the agents learn not only by trial and error but also through

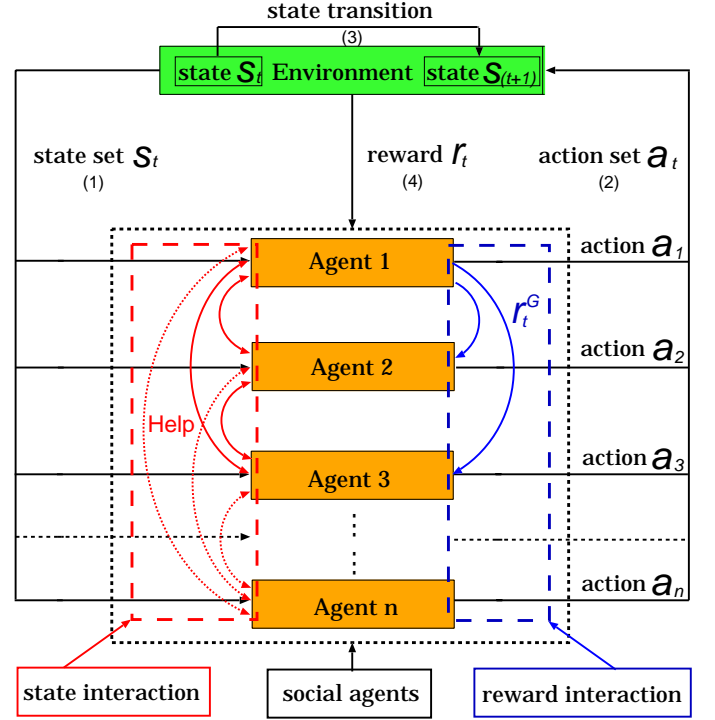


Fig. 1. The RL model of social agents with the interaction model

cooperation by sharing instantaneous information [5]. So we verify this idea by sharing environment state (we call RL-SES). When an agent gets some new environment state, the new environment state will be shared by other agents who can't perceive it through  $S_t^s$ , which means the state told by the agent who can perceive some new state. In RL method, the environment state is described by  $S_t$  not including  $S_t^s$ . In RL-SES method, Q-value is described as follows:

$$Q(S_t \vee S_t^s, a_t) \leftarrow (1 - \alpha) \cdot Q(S_t \vee S_t^s, a_t) + \alpha(r_t + \gamma \max_{a \in A} Q(S_{t+1} \vee S_{t+1}^s, b)), \quad (1)$$

where,

$S_t$  : environment state in current moment  $t$ ,

$S_t^s$  : shared state informed by other agent in current moment  $t$ ,

$a_t$  : current action in current moment  $t$ ,

$\alpha$  : learning rate,

$\gamma$  : discount rate,

$r_t$  : immediate reward received after performing of action  $a_t$  at current state  $s_t$ .

### B. Interactive Reinforcement Learning (I-RL)

In generally, it's necessary for two or more agents to achieve a task through the autonomous cooperation when the task can't be achieved only by one agent in MAS. Therefore, in our paper when the target was discovered, each agent send a special signal information (called Help) to other agents if the task couldn't be achieved by itself in a dynamic environment.

Concretely, once the target was discovered, the agent send Help information to other agents even they are outside the range of itself perception. At the same time, other agents receive the coordinates of generated Help and compute the relative direction from itself position to Help. The emergence of Help and the relative direction will be recorded as the environment change information. All the agents will perceive the emergence of task through the interaction among agents. So this I-RL method based on the emergence of Help and the relative direction can solve the problem of incomplete perception in MAS. Q-value of I-RL is changed as follows:

$$Q(S_t \vee S_t^h, a_t) \leftarrow (1 - \alpha) \cdot Q(S_t \vee S_t^h, a_t) + \alpha(r_t + \gamma \max_{a \in A} Q(S_{t+1} \vee S_{t+1}^h, b)), \quad (2)$$

where,  $S_t^h$  is the special information state of environment changing called help state informed by other agent in current moment  $t$ . It includes the relative direction from own position to Help position in this paper. Efficient group behaviors will be expected under the influence of interaction of agents.

### C. Social Interactive Reinforcement Learning (SI-RL)

In the approach of I-RL, the communication among agents is possible through the interaction about the emergence of Help and the relative direction from itself position to Help position. But it just tell the partial information to each other not giving any evaluation for the action under this information. In our human society, if person A likes an action performed by person B, then A will praise B. The possibility of perfect action of B will be repeated in next time for the praising force [15]. Based on this idea, we believe that if social agents want to have strong cooperative ability like human, the agent should be praised if its action was perfect.

Therefore, in social interactive reinforcement learning (SI-RL) method, the perfect action will be praised through a group reward like Fig.2. In this approach, at a certain time  $t$  if the agent  $A_1$  discovered the target but it couldn't achieve the task with itself power, then  $A_1$  sent the Help information to the other agents. And  $A_1$  achieved the target with other agents  $A_2$  and  $A_3$  lastly, then  $A_1$  should distribute a part of reward got from environment to  $A_2$  and  $A_3$  as a group reward to praise them for their help.  $A_2$  and  $A_3$  would get the reward from the agent  $A_1$  who sent Help information but not from the agent who didn't send Help information. Because the cooperative action of agent  $A_1$  and agent  $A_2$ ,  $A_3$  was strengthened as the learning process, when the Help information appears in the next time, the possibility of selecting the group cooperative

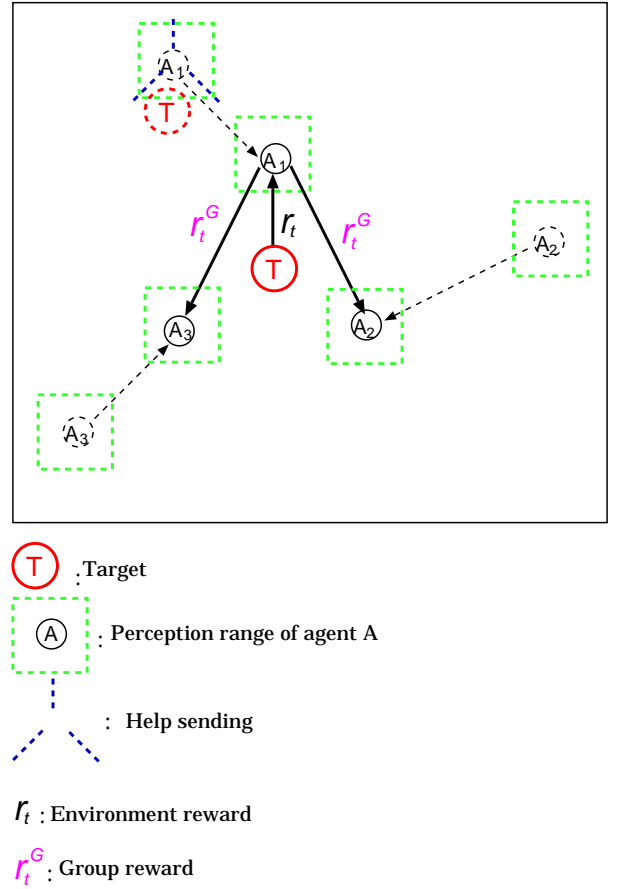


Fig. 2. RL of cooperative behavior based on group reward

behavior will rise. Although the range of perception is limited to each agent for the state space, the task is widely perceivable through the interaction of sending Help information, and the group behavior with cooperation is also strengthened.

In the SI-RL, when the state is  $s_t$  in current moment  $t$ , if the state changed to  $s_{t+1}$  for selecting the action  $a_t$ , then the Q-value becomes like in Eq.(3) used Q-learning in case of static environment. The action evaluation value  $w_{sb}$  becomes like in Eq.(4) used Profit Sharing method in case of dynamic environment.

In case of static environment:

$$Q(S_t \vee S_t^h, a_t) \leftarrow (1 - \alpha) \cdot Q(S_t \vee S_t^h, a_t) + \alpha(r_t + \beta \cdot r_t^G + \gamma \max_{a \in A} Q(S_{t+1} \vee S_{t+1}^h, b)), \quad (3)$$

In case of dynamic environment:

$$w_{sb} \leftarrow w_{sb} + \alpha \cdot \gamma^h (r_t + \beta \cdot r_t^G), \quad (4)$$

For the agent of sending Help information, it's necessary to supply a part of reward to the agents giving help and  $r_t^G$  of Eq.(3) and (4) is defined as follows:

$$r_t^G = -\theta \cdot r_t, \quad (5)$$

For the agents of giving help, each agent will obtain a part of reward from the agent who sent Help information as group reward. Therefore,  $r_t^G$  is defined as follows:

$$r_t^G = \frac{\theta \cdot r_t}{n}, \quad (6)$$

where,

$w_{sb}$ : the action evaluation value when the state is  $s_t$  and the action is  $b$ ,

$\beta$ : the percentage of social interaction about group reward,

$r_t^G$ : the group reward in current moment  $t$ ,

$n$ : the number of agents giving cooperative behavior to Help sending agent ( $n \leq n_{lim}$ ),

$\theta$ : the percentage of given reward as compared with  $r_t$  ( $\theta \leq \theta_{lim}$ ).

#### IV. SIMULATION EXPERIMENTS

To prove the effectiveness of the proposed method we performed the prey chase simulation experiment using Profit Sharing method. This simulation is programmed by C language with the graphic library of OpenGL executed in Linux system. The view of simulator is shown in Fig.3.

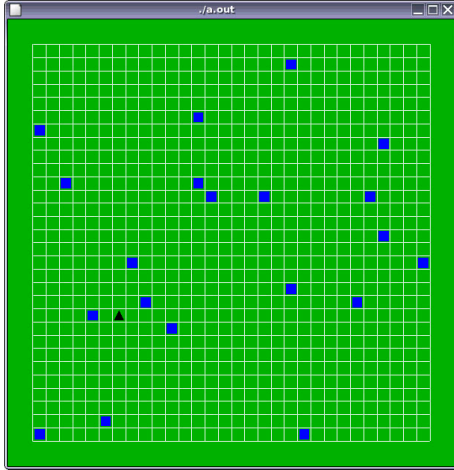


Fig. 3. The view of prey chase simulator without obstacles

##### A. Simulation Preconditions

The preconditions of this simulation are set as follows:

- The field is composed by  $30 \times 30$  grids in two dimensional environment.
- 20 hunter agents are set and represented as a square on the grid. The action and the perception ability of all agents is equal.
- Both the hunter and a prey represented as triangle agents can move in eight directions: top, down, left, right, top-right, top-left, down-right, down-left. They can stay at their current positions.
- The perception range is set  $7 \times 7$  grids around itself.
- The default positions of all agents are arranged randomly and a prey agent moves always randomly.

- All hunter agents can obtain the other agents' perception information according to the interaction communication.
- When the prey agent becomes adjacent by three or more hunter agents, the task of chase is achieved. All the agents will return to the initial state after that.
- Hunter agent's task is to capture the prey agent as soon as possible. When the prey capturing is achieved, they will obtain the basic reward  $R$  assumed to be 100.

The parameters in this simulation are set as follows:  $\gamma = 0.8$ ,  $\lambda = 0.8$ ,  $\alpha = 0.8$ ,  $\beta = 1$ ,  $\theta = 0.2$ . To confirm the effectiveness of the proposed method, we performed comparison experiments with normal Profit Sharing reinforcement learning method. When the action is selected in Profit Sharing, the action is generally decided by the roulette selection of the action evaluation value added to the rule.

##### B. Simulation Results

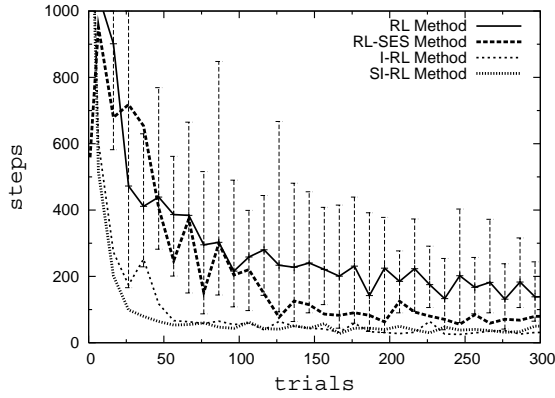
All of the methods are performed 5 times in this simulation, the average and deviation of simulation results are shown in Fig.4. Vertical axis shows the steps required for capturing the prey and horizontal axis shows the trials of learning.

In the simulation experiment without obstacles, the comparison experiment results of 4 methods are shown in Fig.4(a) ~ (d). From the average results in Fig.4(a) ~ (d), we confirmed that the reinforcement learning sharing environment state (RL-SES) method costed little time to capture the prey agent than normal reinforcement learning (RL) method, the interactive reinforcement learning (I-RL) method obtained better results than RL-SES method, the social interaction reinforcement learning (SI-RL) method obtained the best learning result than the other methods. From the deviation results in Fig.4(a) ~ (d), we also confirmed that the stability of I-RL and SI-RL methods are better than RL-SES and normal RL methods. The interaction among agents based on the emerging of Help and the relative direction information is efficient for the group cooperation behavior. Furthermore, we can find that SI-RL method obtained better results than I-RL method in the average and the deviation. That shows the group rewards among agents can reinforce the cooperative behaviors to achieve the task efficiently. From the interaction rewards of 20 hunter agents in SI-RL method shown in Fig.5, we can find agents have different characters. In case of agent 2, it has the best explore ability because it paid the most reward to other agents.

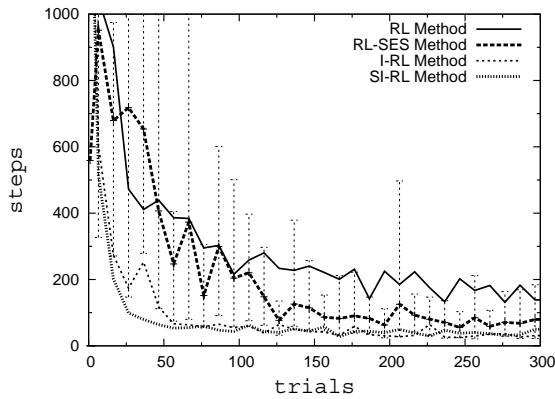
In addition, we also performed the comparison experiments with obstacles like Fig.6. The comparison experiment results of 4 methods are shown in Fig.7(a) ~ (d). The interaction rewards of 20 hunter agents in SI-RL method are shown in Fig.8 and agent 9 has the best cooperative ability because it got the most group reward. By these results, we could confirm that the performance of I-RL and SI-RL are better than that of RL and RL-SES in case of the simulation without obstacles.

#### V. CONCLUSIONS

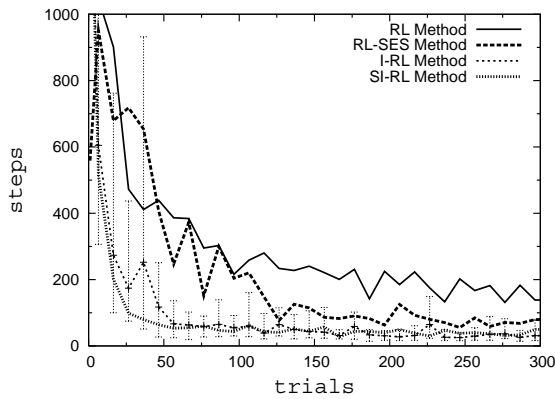
In this paper, we proposed I-RL, SI-RL of multi-agent reinforcement learning methods with an excellent cooperation ability through the social interaction of Help information and



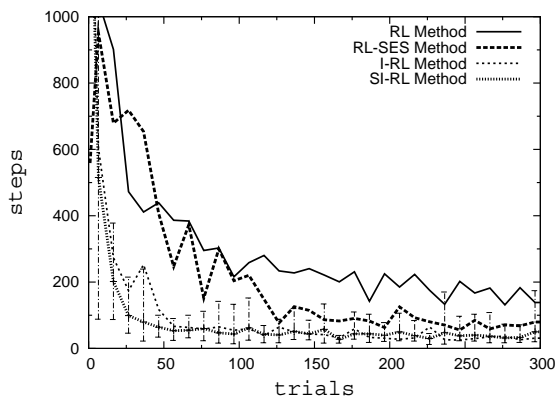
(a) with deviation of RL



(b) with deviation of RL-SES



(c) with deviation of I-RL



(d) with deviation of SI-RL

Fig. 4. Results of experiment without obstacles

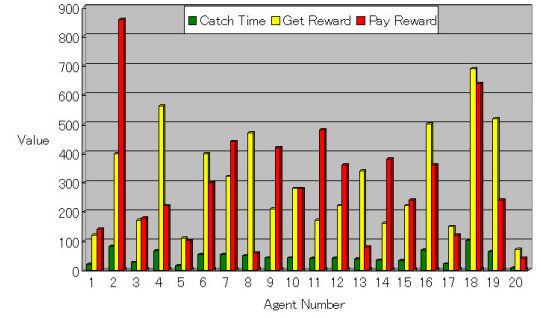


Fig. 5. The interaction rewards of 20 hunter agents without obstacles

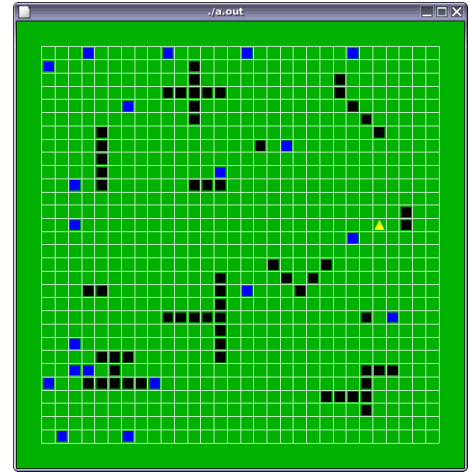


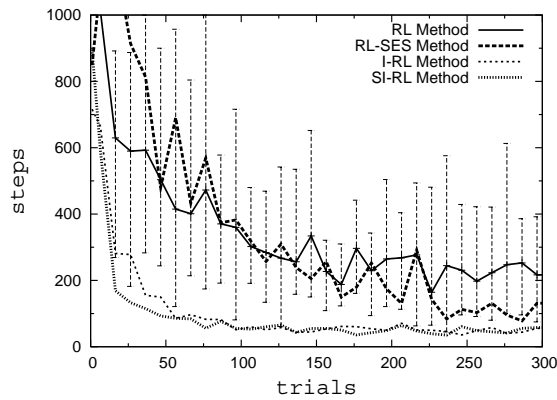
Fig. 6. The view of prey chase simulator with obstacles

the interactive reward which exchanged among autonomous agents in MAS.

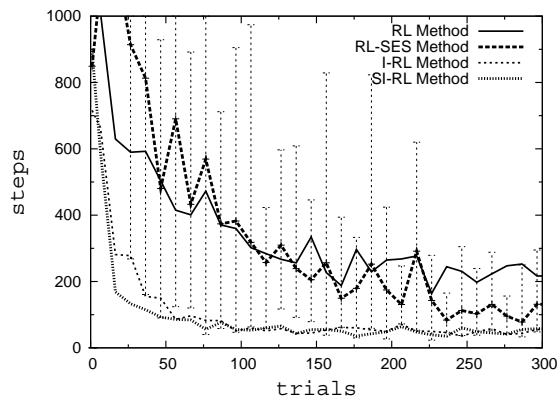
From the results of the comparison simulation between RL and RL-SES methods, we found that RL-SES can improve the overall performance in MAS. Moreover, we also found that the I-RL method obtained the environment state information among agents could obtain the better cooperative ability than RL-SES. Although outside of the perception range, the agent can obtain some environment states that can't be obtained from their restricted perception through the interaction with the other agents. The emergence of Help and the relative direction help the agent not only to perceive the environment state but also to get some informations about the target in the task.

Furthermore, we proposed SI-RL method that can give some evaluation to the action under the interaction of agents exchanging the partial information to each other like I-RL method. If the actions of agent A were good for agent B, then B would praise and reinforce these actions in the form of group reward. At last, the interaction among agents can be changed to the behavioral strategy communication like the human society and the autonomous cooperation group behavior of SI-RL method has the best performance than the other methods.

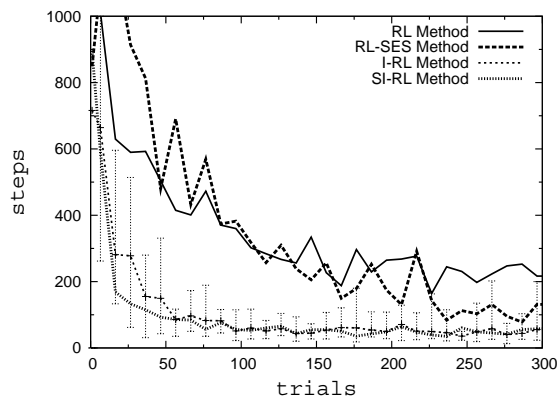




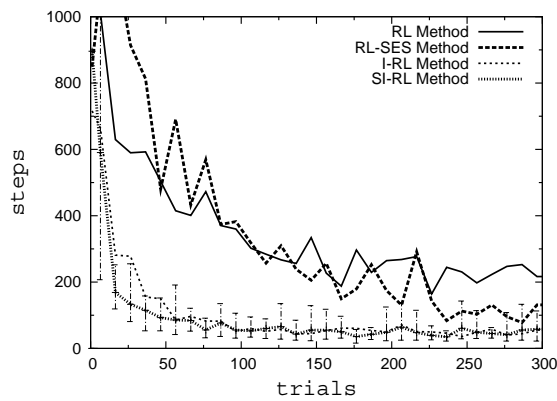
(a) with deviation of RL



(b) with deviation of RL-SES



(c) with deviation of I-RL



(d) with deviation of SI-RL

Fig. 7. Results of experiment with obstacles

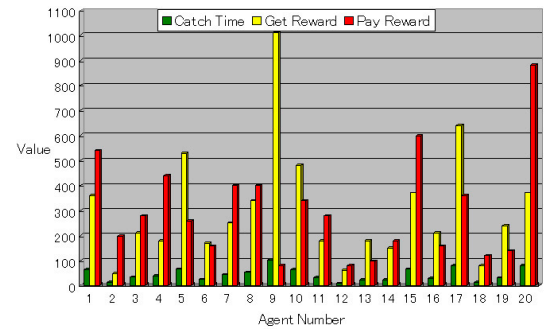


Fig. 8. The interaction rewards of 20 hunter agents with obstacles

In the further works, we have a plan to add more communication modes to agents and let social agents have robust autonomous cooperation learning ability. In addition, we'll perform the simulation with multiple targets. As a social agent, it's necessary to learn the best action selection even with several targets in the complex environment.

#### REFERENCES

- [1] Y.Maeda, " Evolutionary Simulation for Co-Operative Behavior Learning on Multi-Agent Robots, "Journal of Japan Society for Fuzzy Theory and intelligent informatics, Vol.13, No.3, pp.281-291, 2001. (In Japanese)
- [2] M.J.Wooldridge, An Introduction to MultiAgent Systems, John Wiley and Sons, Ltd. England, 2002.
- [3] T.Matsuura, Y.Maeda, " Deadlock Avoidance of a Multi-Agent Robot Based on a Network of Chaotic Elements, "Advanced Robotics, Vol.13, No.3, pp.249-251, 1999.
- [4] S.Arai, " Multiagent Reinforcement Learning Frameworks : Steps toward Practical Use, "Journal of The Japanese Society for Artificial Intelligence, Vol.16, No.4, pp.476-481, 2001.
- [5] M.Tan, " Multi-Agent Reinforcement Learning: Independent vs. Co-operative Agents, "Proceedings of Tenth International Conference on Machine Learning, pp.330-337, 1993.
- [6] L.Nunes, E.Oliveira, " Cooperative learning using advice exchange, " Adaptive Agents and Multiagent Systems, Lecture Notes in Computer Science, pp.33-48, 2003.
- [7] M.L.Littman, " Fierend-or-foe q-learning in general-sum games, "Proceedings of Eighteenth International Conference on Machine Learning, pp.322-328, 2001.
- [8] A.Greenwald, K.Hall, R.Serrano, " Correlated-q learning, "Proceedings of Twentieth International Conference on Machine Learning, pp.242-249, 2003.
- [9] M.Bowling, " Convergence and no-regret in multiagent learning. "Proceedings of the Annual Conference on Neural Information Processing Systems, pp.209-216, 2005.
- [10] G.Weiss, " Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence, "MIT Press, 1999.
- [11] S.Kato, H.Matsuo, " A Theory of Profit Sharing in Dynamic Environment, "PRICAI 2000, LNAI 1886 pp.115-124, 2000.
- [12] C.Claus, C.Boutilier, " The dynamics of reinforcement learning in cooperative multiagent systems, "AAAI/IAAI, pp.746-752, 1998.
- [13] R.Ribeiro, A.P.Borges, F.Enembreck, " Interaction Models for Multi-agent Reinforcement Learning, "Computational Intelligence for Modelling, Control and Automation, International Conference, pp.464-469, 2008.
- [14] K.Zhang, Y.Maeda, " Multi Agent Reinforcement Learning Based on Contribution Degree of Individual and Group Evaluation, "The 27TH Annual Conference of the Robotics Society of Japan, CD-ROM, RSJ2009AC1F1-03, 2009.
- [15] D.Barrios-Aranibar, L.M.G.Goncalves, " Learning Coordination in Multi-Agent Systems using Influence Value Reinforcement Learning, " 7th International Conference on Intelligent Systems Design and Applications (ISDA07), pp.471-478, 2007.