

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

メタデータ	言語: eng 出版者: 公開日: 2011-09-29 キーワード (Ja): キーワード (En): 作成者: ZHANG, Kun, MAEDA, Yoichiro, TAKAHASHI, Yasutake メールアドレス: 所属:
URL	<a href="http://hdl.handle.net/10098/4141">http://hdl.handle.net/10098/4141</a>

Paper:

# 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 910-8507, Japan

E-mail: {kzhang, maeda, yasutake}@ir.his.u-fukui.ac.jp

[Received March 5, 2011; accepted May 9, 2011]

**Research on multi-agent systems, in which autonomous agents are able to learn cooperative behavior, has been the subject of rising expectations in recent years. We have aimed at the group behavior generation of the multi-agents who have high levels of autonomous learning ability, like that of human beings, through social interaction between agents to acquire cooperative behavior. The sharing of environment states can improve cooperative ability, and the changing state of the environment in the information shared by agents will improve agents' cooperative ability. On this basis, we use reward redistribution among agents to reinforce group behavior, and we propose a method of constructing a multi-agent system with an autonomous group creation ability. This is able to strengthen the cooperative behavior of the group as social agents.**

**Keywords:** group behavior learning, multi-agent systems, reinforcement learning, state communication, social interaction

## 1. Introduction

In Multi-Agent Systems (MAS), it is necessary that plural autonomous agents cooperate to achieve the objective. At the same time, it is very important for agents to have the ability to learn because the tasks are generally complex [1]. Reinforcement Learning (RL) is a machine learning method by which a machine learns the behavior suitable to solving the problem autonomously through repeated trials and errors without teaching signals [2]. It is easy for RL to adjust to dynamic environmental transformations flexibly. It is expected that a MAS will be constructed that is able to learn group behavior and group strategy by the application of RL to autonomous agents [3].

When RL is used for a MAS in that way, problems such as perceptual aliasing problems, concurrent learning problems, and reward distribution problems will be generated [4]. In a cooperative social environment, it is effective for agents to learn by sharing sensing infor-

mation, episodes, and learned policies, or by exchanging advice [5]. The sharing of environment states can improve cooperative ability. However, one of the main problems concerned with learning in a MAS environment is how agents can benefit from mutual interaction during the learning process [6]. Some researchers deal with the influence of interaction by handling the loss of stationarity based on game theory, such as friend-or-foe Q-learning in general-sum games [7] and correlated Q-learning [8]. Nevertheless, these methods may fail to learn best-response policies even against simple, non-learning opponents. Besides, best-response learners learn best-responses by adjusting self-strategies or by modeling the strategies of opponents while still considering whether the resulting algorithm converges in some form [9].

In this paper, agents can benefit from the communication of state interaction. In the state communication model, when an agent gets an environment state that other agents do not get due to their restricted perception, if it cannot achieve the task by itself, it will change the environment state to the special signal information among agents in order to tell other agents. Under the hint of special signal information, the overall performance of agents will gradually increase. This state communication is different from the exchanging of information that is unrelated to the target because special signal information includes some information about task performance.

In addition, in MAS, although the purpose of each agent is to obtain more desired rewards, it does not mean that the whole system can obtain more desired rewards. If the reward cannot be distributed appropriately, it not only has a bad effect on the behavior of each agent but also reduces the overall performance of the system [10]. In our proposed method, it is possible for the exchanging of rewards among agents to reinforce cooperative behavior autonomously when the expected actions have appeared. Therefore, all agents not only get the reward from the environment but also exchange the reward among themselves based on signal information hints and their actions. Agents are able to exhibit excellent cooperative group behavior by taking rewards as a social interaction strategy, as people do in society. In order to evaluate the interaction ability of agents, a method of evaluation is also proposed based on task mining and cooperative support abilities.

We perform prey chase simulations to prove the effectiveness of the proposed methods.

## 2. 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 made up of an enormous number of autonomous agents. On the other hand, Reinforcement Learning (RL) is an attractive method, as it allows the agent to learn a behavior on the basis of sparse, delayed reward signals provided only when the agent attains 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].

### 2.1. Multi-Agent Systems

A Multi-Agent System (MAS) is composed of multiple autonomous agents that solve problems by working with each other autonomously. Communication and cooperation among intelligent agents can be used to deal with incomplete and uncertain problems [12]. The actions of agents are not restricted by others in MAS, but the actions of any one agent will have a significant impact on other agents, not only on the environment. Therefore, the interaction of agents becomes especially important in joint tasks that cannot be completed by a single agent.

### 2.2. Reinforcement Learning in MAS

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 cooperative learning of multi-agents becomes possible by using RL in MAS. In the reinforcement learning of MAS, the state of the environment will become a new environment through the influence of multiple agents, not through that of single agents as in normal RL. The learning of agents becomes more difficult than in a single environment because the actions of single agents have unexpected effects on other agents and the environment [14].

Compared with normal RL, the RL of MAS keeps the previous problems in normal RL and also has its unique problems, which are identified 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 perceived information is imperfect. 2) When two or more agents learn independently, it is difficult to judge to which the result of learning should be credited. 3) Although the task is completed, how to distribute the reward, etc., is problematic.

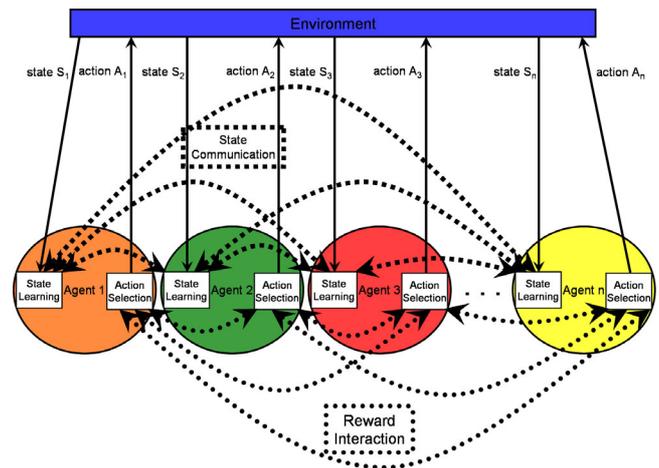


Fig. 1. RL with interaction model in MAS.

## 3. Reinforcement Learning Based on Social Interaction in MAS

In this paper, the agents learn cooperative behaviors through the interaction of state communication and group behavior learning, as shown in Fig. 1. In normal RL, an individual agent can only use trial and error repeatedly through interaction with the environment. In our proposed method, all agents can learn cooperative behaviors through trial and error attempts among autonomous agents through the interaction information, not only with the environment.

We have aimed at the formation of group behavior among multi-agents who not only act based on their own profit but also based on a consideration of joint profit with other agents through the exchange of the reward as social agents [15].

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

In a MAS environment, if the agent selects its action based on the environment state as in normal RL, it will have some problems, such as the perceptual aliasing problem identified by Arai [4]. In a cooperative social environment, the agents learn not only by trial and error but also through cooperation by sharing information instantaneously [5]. Therefore, we verify this idea by sharing the state of the environment (which we call RL-SES). When an agent experiences a new state that can not be perceived by others, the new environment state will be shared by other agents through  $S_t^s$ , which means the state related by the agent who can perceive the new state. In the RL method, the environmental state is described by  $S_t$ , not including  $S_t^s$ . In the RL-SES method, the Q-value is described as follows:

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

where,

- $S_t$ : environment state in current moment  $t$ ,
- $S_t^s$ : shared state communicated by another agent in the current moment  $t$ ,
- $a_t$ : current action in current moment  $t$ ,
- $\alpha$ : learning rate,
- $\gamma$ : discount rate,
- $r_t$ : immediate reward received after performing action of action  $a_t$  at current state  $s_t$ .

The action evaluation value  $w(s, b)$  becomes similar to that in Eq. (2) in the profit sharing method.

$$w(s_t \vee s_t^s, b) \leftarrow w(s_t \vee s_t^s, b) + f(h), \dots \dots (2)$$

$w(s_t \vee s_t^s, b)$  is the action evaluation value when the state is  $s_t \vee s_t^s$  and the action is  $b$  in current moment  $t$ . Reinforcement function  $f(h)$  is the same as in the normal profit sharing method. Efficient cooperative behaviors will be expected under the communication of this new signal hint among agents.

### 3.2. State Communication Reinforcement Learning (SC-RL)

For agents in MAS, it is necessary to achieve a task through autonomous cooperation when the task cannot be achieved only by one agent. Therefore, in our paper, once a target has been discovered, each agent sends special signal information (called Help) to other agents if the task cannot be achieved by itself alone in a dynamic environment. However, the other agents are uncertain of the meaning of this Help or special signal information at first and record it as a special state, similar to environment changing information by  $S_t^h$  in order to communicate. In this way, we proposed a State Communication Reinforcement Learning (SC-RL) method by which agents can perceive the change in an environment state though interaction with other agents. In addition, this perception of environment changing information is also related to target discovery or task performance.

When the target is discovered by an agent but the agent cannot achieve it, it will send Help information to all others, even they are outside the range of its own perception. At the same time, other agents receive the coordinates of the generated Help and compute the relative direction from its current position to Help. The emergence of Help and the relative direction will be recorded as the environment change information by others. All agents will perceive the emergence of a task through the interaction of the agents. Therefore, this SC-RL method based on the emergence of Help and the relative direction can solve the problem of perceptual aliasing in MAS. The Q-value of SC-RL is changed as follows:

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

The action evaluation value  $w(s, b)$  becomes similar to that in Eq. (4) in the profit sharing method.

$$w(s_t \vee s_t^h, b) \leftarrow w(s_t \vee s_t^h, b) + f(h), \dots \dots (4)$$

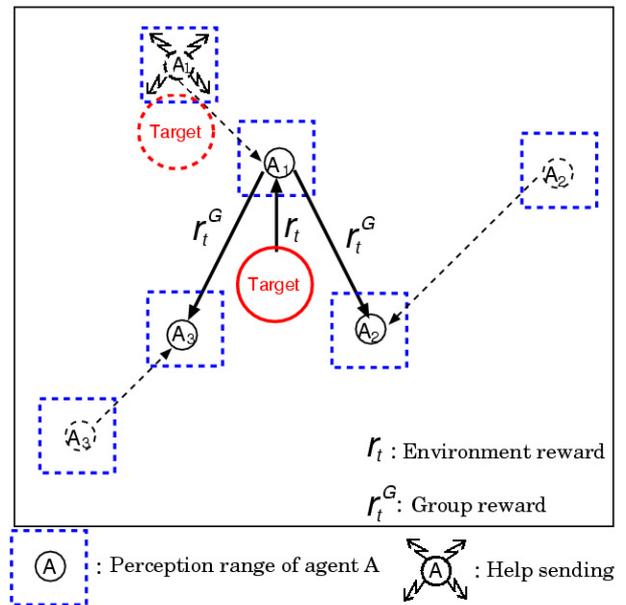


Fig. 2. GB-RL based on group reward.

where  $S_t^h$  is the special information state of environmental change called the Help state, as communicated by other agents in current moment  $t$ . It includes the relative direction from one's own position to the Help position. Efficient group behaviors will be expected due to the influence of state communication among agents.

### 3.3. Group Behavior Reinforcement Learning (GB-RL)

In the SC-RL method, interaction among agents is possible through the state communication about the emergence of Help. This is never known to agents; they just receive partial information from each other without any evaluation regarding action in response to this information. In our human society, if person A likes the behavior of person B, then A will praise B. There is a possibility that that perfect behavior on the part of B will be repeated again for that praise [16]. Based on this idea, we believe that if social agents want to have the strong cooperative ability of human beings, agents should be praised if their actions are perfect. Through the power of praise, we believe that agents also acquire group behavior learning ability.

In the Group Behavior Reinforcement Learning (GB-RL) method, good behavior will be praised through a group reward, as illustrated in Fig. 2. In this approach, at a certain time  $t$ , if agent  $A_1$  discovers a target that it cannot achieve by itself, it then sends Help information to the other agents.  $A_1$  then achieves the target with the help of other agents  $A_2$  and  $A_3$ . Finally,  $A_1$  should distribute a part of reward it received from the 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  $A_1$ , as agent  $A_1$  sent Help information, but not from the agents that did no. Because the cooperative behavior of agents  $A_1, A_2,$

and  $A_3$  was strengthened in the learning process, when Help information appears the next time, the possibility of selecting group cooperative behavior will rise.

In the SC-RL method, each agent interacts with other agents through state communication. Therefore, agents can have better interactive communication than through the normal RL method, but they learn cooperative behavior only by themselves. In the GB-RL method, each agent not only interacts with other agents through state communication but also reinforces the cooperative behavior with others by exchanging rewards, as in Fig. 2, when the expected action has appeared.

In the GB-RL approach, when the state is  $s_t$  in current moment  $t$ , if the state changes to  $s_{t+1}$  for the selection of action  $a_t$ , then the Q-value becomes as shown in Eq. (5) in the Q-learning method.

$$Q(S_t \vee S_t^h, a_t) \leftarrow Q(S_t \vee S_t^h, a_t) + \alpha[r_t + \beta \cdot r_t^G + \gamma \max_{b \in A} Q(S_{t+1} \vee S_{t+1}^h, b) - Q(S_t \vee S_t^h, a_t)] \quad (5)$$

The action evaluation value  $w(s, b)$  is the same as in Eq. (4) in the profit sharing method, but reinforcement function  $f(h)$  becomes as follows in Eq. (6).

$$f(h) = \alpha \cdot \gamma^h (r_t + \beta \cdot r_t^G) \dots \dots \dots (6)$$

Each agent that gives help will obtain a group reward  $r_t^G$ , as shown in Eq. (7), from the agent who sent the signal.

$$r_t^G = \frac{\theta \cdot r_t}{n}, \dots \dots \dots (7)$$

where,

- $\beta$ : rate of social interaction regarding the group reward,
- $r_t^G$ : group reward in current moment  $t$ ,
- $\theta$ : percentage of reward given compared to  $r_t$  ( $\theta \leq \theta_{lim}$ ),
- $n$ : number of agents cooperating on help-sending agent ( $n \leq n_{lim}$ ).

For the agent sending new signal information, it is necessary to supply reward  $n \cdot r_t^G$  to be given to cooperating agents.

### 3.4. Evaluation of Interaction Ability

In SC-RL, when an agent finds a target, it will send Help information to others if it cannot achieve the task by itself, and such state communication causes all agents to become aware that a task has emerged. The frequency of sending Help information can be treated as task mining ability in some learning trials. After Help information has been sent out, others can cooperate better through group rewards as GB-RL. Therefore, the social interaction creates relationships between agents, giving them the ability to interact with different characters. This basic interaction ability should include task mining and cooperative support abilities.

In our proposal,  $R_{i \rightarrow j}$  represents the reward transferred from agent  $i$  to agent  $j$ , and  $\sum_{i=1}^m \sum_{j=1}^m R_{i \rightarrow j}$  is the total rewards transferred among  $m$  agents. Interaction ability  $I_{i \rightarrow j}$

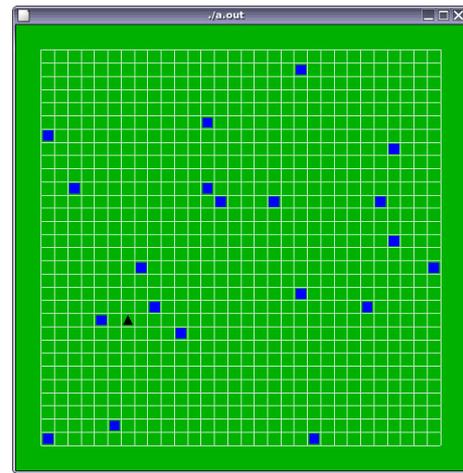


Fig. 3. Prey chase simulator without obstacles.

of agents  $i$  and  $j$ , agent  $i$ 's task mining ability  $T_i$ , and agent  $j$ 's cooperative support ability  $C_j$  are defined in Eqs. (8), (9), and (10), respectively, through transferred rewards.

$$I_{i \rightarrow j} = \frac{R_{i \rightarrow j}}{(\sum_{i=1}^m \sum_{j=1}^m R_{i \rightarrow j})/m^2}, \dots \dots \dots (8)$$

$$T_i = \frac{\sum_{j=1}^m R_{i \rightarrow j}}{(\sum_{i=1}^m \sum_{j=1}^m R_{i \rightarrow j})/m}, \dots \dots \dots (9)$$

$$C_j = \frac{\sum_{i=1}^m R_{i \rightarrow j}}{(\sum_{i=1}^m \sum_{j=1}^m R_{i \rightarrow j})/m}, \dots \dots \dots (10)$$

$I_{i \rightarrow j}$  describes the relative frequency of interaction when agent  $i$  is in trouble with a joint task and agent  $j$  gives help.  $T_i$  and  $C_j$  describe the overall performance of task mining and cooperative support abilities with all  $m$  agents. The performance among individuals can be estimated by  $I_{i \rightarrow j}$ , and agents' overall performance as a group can be estimated by  $T_i$  and  $C_j$ .

## 4. Simulation Experiments

We performed a prey chase simulation experiment to prove the effectiveness of the proposed methods. Because the profit sharing method has been proven more effective than Q-learning in the dynamic environment, all experiments are performed using the profit sharing method in this paper. This simulation is programmed in the C language with the OpenGL graphics library and executed on a Linux system. The simulator is shown in Fig. 3.

### 4.1. Simulation Preconditions

The preconditions of this simulation are as follows:

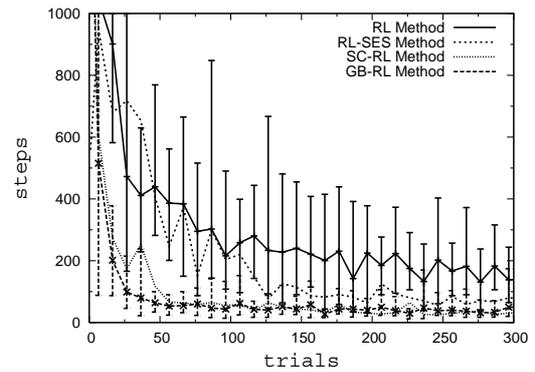
- The field is a  $30 \times 30$  grid in a two-dimensional environment.
- 20 hunter agents are set and represented as a square on the grid. All agents have equal action and perception capabilities.
- Both the hunter and prey, the prey represented by a triangle, can move in eight directions: up, down, left, right, up and to the right, up and to the left, down and to the right, and down and to the left. They can also stay at their current positions.
- The perception range is set to  $7 \times 7$  grid squares in all directions.
- The default positions of all agents are arranged randomly, and a prey agent always moves randomly.
- All hunter agents can obtain the information other agents have perceived through interactive communications.
- When the prey agent becomes adjacent to three or more hunter agents, the task of pursuit is achieved. All agents then return to the initial state.
- The task of hunter agents is to capture the prey agent as quickly as possible. When the prey is captured, they will obtain the basic reward  $r_t$ , assumed to be 100.

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

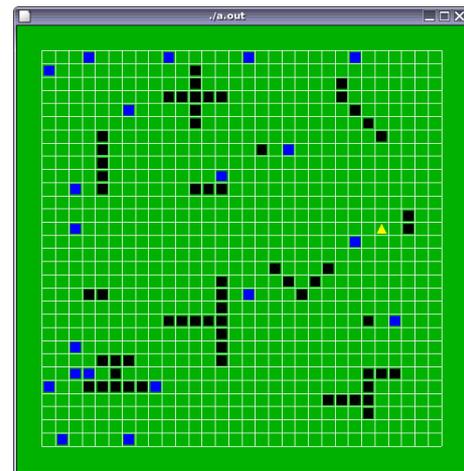
### 4.2. Simulation Results

All of the methods were performed five times in this simulation; the average and deviation range values of simulation results are shown in the graphs of the results. The vertical axis shows the steps required for the prey to be captured, and the horizontal axis represents the number of trials taken for learning.

The results of the four methods compared in the simulation experiment without obstacles are shown in **Fig. 4**. From the average results in **Fig. 4**, we have confirmed that the Reinforcement Learning Sharing Environment (RL-SES) method took less time to capture the prey agent than did the normal Reinforcement Learning (RL) method; the State Communication Reinforcement Learning (SC-RL) method obtained better results than the RL-SES method; and the Group Behavior Reinforcement Learning (GB-RL) method obtained the best learning results of all. From the deviation range value results in



**Fig. 4.** Results of experiment without obstacles.



**Fig. 5.** Prey chase simulator with obstacles.

**Fig. 4**, we have also confirmed that the stability of GB-RL methods is better than that of the normal RL method. The interaction among agents based on the emergence of Help and the relative direction information produces efficient group cooperative behavior. Furthermore, it can be seen that the GB-RL method obtained better results than the SC-RL method in the average and the deviation range value. That shows the group rewards among agents can serve to reinforce cooperative behaviors, thus promoting efficient task achievement.

In addition, we also performed the comparison experiments with obstacles on the board, as seen in **Fig. 5**. The comparison experiment results of the four methods are shown in **Fig. 6**. These results confirm that the performance of SC-RL and GB-RL are better than those of RL and RL-SES, even when the board has obstacles.

It is necessary for social agents to determine the best decision even with several targets in a complex environment. We performed the simulation experiment with five targets, as shown in **Fig. 7**. The comparison experiment results of the four methods are shown in **Fig. 8**. These results are similar to those obtained in experiments with only one target, but the degree of convergence was not very good because it is naturally more difficult to capture five preys than only one.

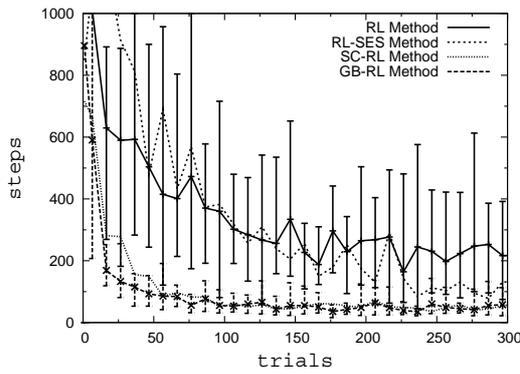


Fig. 6. Results of experiment with obstacles.

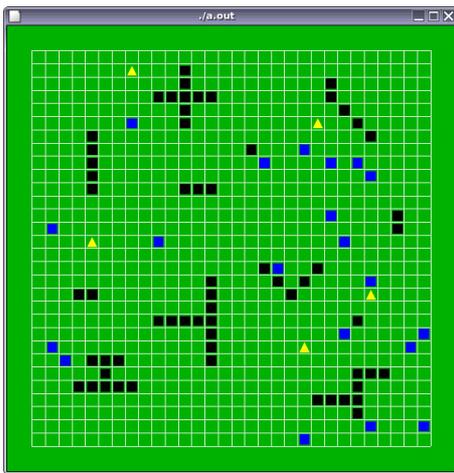


Fig. 7. Prey chase simulator with 5 targets.

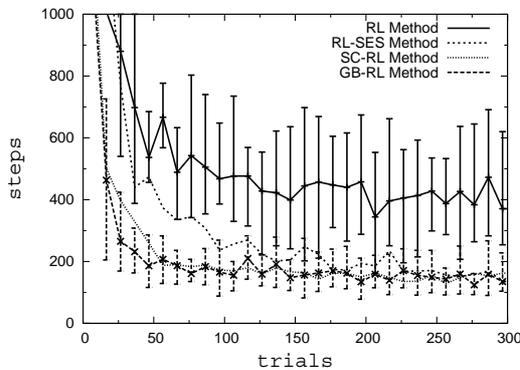


Fig. 8. Results of experiment with 5 targets.

Table 1. Total steps in target catching over 300 trials.

Methods	RL	RL-SES	SC-RL	GB-RL
no obstacle	89566	66102	26699	21779
obstacles	104340	106366	30310	25406
5 targets	163146	106533	68068	65214

The total number of steps taken to catch the target by using four methods are shown in Table 1. These data clearly show that our proposed methods are more effective than normal RL and RL-SES methods.

Receiver Sender	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	$\sum_i R_{ij}$
1	0	60	140	40	60	80	40	60	80	30	60	40	60	40	40	60	20	90	40		1100
2	40	0	250	120	230	200	150	60	220	100	180	170	120	110	210	160	190	210	120	10	2850
3	60	100	0	80	120	330	100	30	250	50	130	160	100	120	180	150	170	140	130	40	2440
4	0	90	80	0	110	170	110	30	160	70	140	110	70	50	90	160	230	110	80	60	1920
5	60	130	90	60	0	210	120	50	200	90	180	120	80	110	300	210	110	100	80	60	2360
6	90	130	190	80	150	0	210	70	390	150	210	230	140	190	210	310	250	250	140	90	3480
7	60	60	30	40	70	130	0	50	160	70	100	110	100	110	130	140	100	70	60	40	1630
8	20	30	20	10	60	40	30	0	70	30	60	40	20	110	50	50	80	70	70	60	920
9	40	80	130	30	140	160	90	50	0	130	160	190	80	280	140	150	370	180	70	40	2510
10	20	80	30	40	40	60	40	0	70	0	20	80	100	70	40	100	120	110	70	40	1130
11	30	30	90	60	100	120	100	10	50	20	0	60	90	80	150	220	110	110	120	70	1620
12	30	70	80	40	140	160	40	10	80	20	40	0	110	150	220	190	190	170	50	50	1960
13	10	50	30	20	20	30	30	20	60	60	20	60	0	120	70	110	100	150	20	30	1010
14	30	20	90	30	100	90	40	30	130	90	60	60	140	0	160	100	230	270	120	40	1830
15	10	70	20	20	60	70	80	10	40	0	120	70	30	40	0	220	60	50	110	50	1130
16	0	40	50	20	110	20	20	0	90	50	40	30	20	10	110	0	190	100	90	60	1050
17	10	10	40	20	30	90	20	10	80	20	40	20	30	30	20	0	120	70	60	800	
18	20	30	10	10	20	10	30	40	60	40	30	20	20	20	10	50	30	0	40	30	520
19	20	10	20	10	20	20	10	0	10	0	30	0	0	10	10	20	10	0	0	40	240
20	120	70	60	20	110	130	130	60	90	80	20	90	80	40	50	40	60	40	40	0	1330
$\sum_j R_{ij}$	670	160	1450	750	690	2120	1390	590	2290	1160	620	700	1360	710	2200	2440	2660	2700	1690	10	31830

Fig. 9. Total transferred reward  $R_{ij}$  among agents.

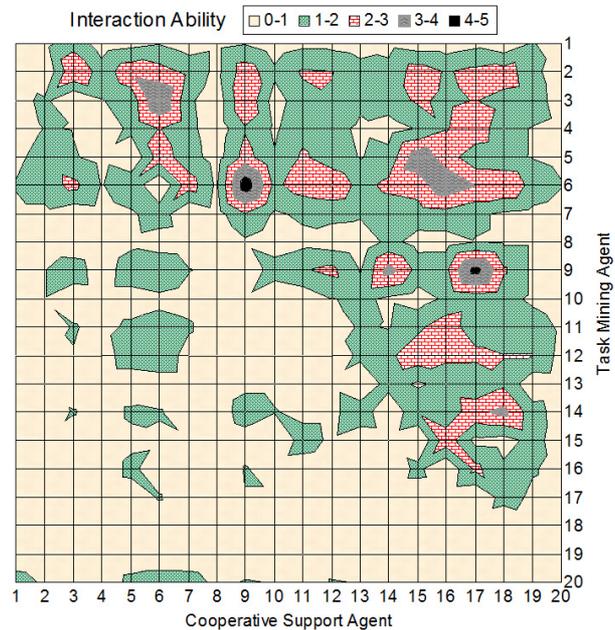


Fig. 10. Interaction ability.

### 4.3. Evaluation of Interaction Ability

We have aimed at the generation of group behavior and recorded the interaction information among agents, as shown in Fig. 9. The numbers down the left and across the top represent twenty agents, and the other numbers represent reward exchanging among agents. The agents on the left first found a target and sent Help signals to the agents across the top. The agents across the top received the signal and selected their actions. Therefore, reward is passed from the agents on the left to those across the top. Based on Eq. (8), the ability for interaction among individuals is shown in Fig. 10. When a task-mining agent is in trouble with a joint task, a cooperative support agent gives help. Different interaction abilities among agents become obvious when compared with the average value

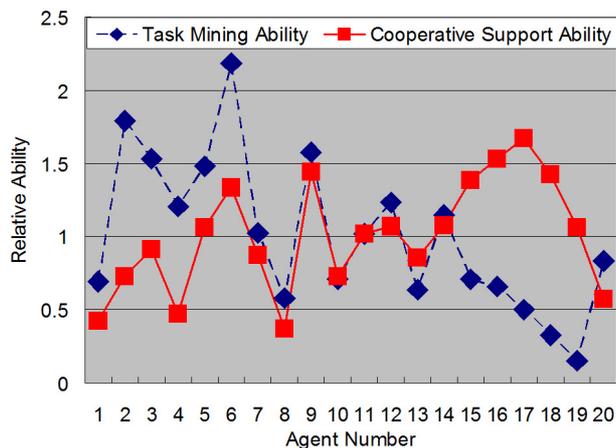


Fig. 11. Task mining and cooperative support ability.

of 1.

Along the horizontal axis of Fig. 9, we add the reward, and the sum means the reward transferred to other agents from each agent on the left. For the twenty agents on the left, the higher the sum of the reward, the better the task mining ability to search out a target. Based on such sum-total rewards along the horizontal axis and Eq. (9), we can get the relative percentage of task mining ability, shown in Fig. 11. The better and worse performances are obvious when compared with the average value of 1.

Using the same method, we also got cooperative support ability along the vertical axis of Fig. 9, as shown in Fig. 11. For each agent, it shows the percentage of transferred reward from other agents, and it represents the cooperative support ability of agents across the top in Fig. 9 because they received rewards from others only when their behavior was cooperative.

In addition, the total performance of twenty hunter agents to achieve tasks, based on the catch times over 300 trials, is shown in Fig. 12. It can be seen that agents 6 and 17 have excellent performance in achieving the task because they have many catches after 300 trials. In Fig. 11, agents 6 and 17 have the best task mining ability and best cooperative support ability, respectively. On the other hand, agent 9 also has excellent performance, although not the best, in task mining ability and cooperative support ability.

## 5. Conclusions

We have aimed at the formation of group behavior among multi-agents through the interaction of state communication and group behavior learning. In this paper, we have proposed SC-RL, GB-RL for multi-agent reinforcement learning and did simulation experiments to prove its effectiveness. From the results of the comparison simulation, we also found that the SC-RL method, where environment state information is shared among agents, obtains better cooperative ability than the RL-SES and RL

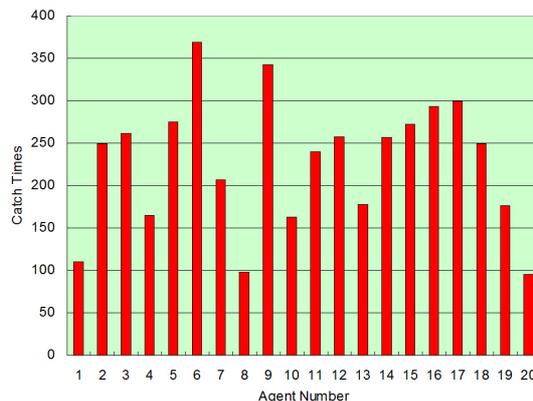


Fig. 12. Performance of 20 hunters based on catch times.

methods. Agents can obtain some environment states that are outside their range of perception and therefore cannot be obtained directly through interaction with the other agents.

In addition, we have proposed the GB-RL method, which can provide some evaluation to strengthen group behavior. Finally, the interaction among agents can be changed to behavioral strategy communication, similar to that in human society, and the autonomous cooperative group behavior of the GBRL method has better performance than do the other methods. From the performance of twenty hunter agents, we also found the importance of reward exchange among agents. Social interaction relationships represent the interactive information among agents through interaction ability, task mining ability, and cooperative support ability.

In future studies, we plan to add more communication models to agents and give social agents robust autonomous cooperative learning abilities. As a social agent, it is necessary to understand more unknown signals, especially in complex environments.

## References:

- [1] Y. Maeda, "Evolutionary Simulation for Co-Operative Behavior Learning on Multi-Agent Robots," J. 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 and 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," J. of The Japanese Society for Artificial Intelligence, Vol.16, No.4, pp. 476-481, 2001.
- [5] M. Tan, "Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents," Proc. of Tenth International Conference on Machine Learning, pp. 330-337, 1993.
- [6] L. Nunes and 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, "Friend-or-foe Q-learning in general-sum games," Proc. of Eighteenth International Conference on Machine Learning, pp. 322-328, 2001.
- [8] A. Greenwald, K. Hall, and R. Serrano, "Correlated Q-learning," Proc. of Twentieth Int. Conf. on Machine Learning, pp. 242-249, 2003.

- [9] M. Bowling, "Convergence and no-regret in multiagent learning," Proc. of the Annual Conf. 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 and H. Matsuo, "A Theory of Profit Sharing in Dynamic Environment," PRICAI 2000, LNAI 1886 pp. 115-124, 2000.
- [12] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems," AAAI/IAAI, pp. 746-752, 1998.
- [13] R. Ribeiro, A. P. Borges, and F. Enembreck, "Interaction Models for Multiagent Reinforcement Learning," Computational Intelligence for Modelling, Control and Automation, Int. Conf., pp. 464-469, 2008.
- [14] K. Zhang and Y. Maeda, "Multi Agent Reinforcement Learning Based on Contribution Degree of Individual and Group Evaluation," The 27TH Annual Conf. of the Robotics Society of Japan, CD-ROM, RSJ2009AC1F1-03, 2009.
- [15] K. Zhang, Y. Maeda, and Y. Takahashi, "Group Behavior Learning in Multi-Agent Systems Based on Social Interaction among Agents," Joint 5th Int. Conf. on Soft Computing and Intelligent Systems and 11th Int. Symposium on advanced Intelligent Systems, TH-B3-1, pp. 193-198, 2010.
- [16] D. Barrios-Aranibar and L. M. G. Goncalves, "Learning Coordination in Multi-Agent Systems using Influence Value Reinforcement Learning," 7th Int. Conf. on Intelligent Systems Design and Applications (ISDA07), pp. 471-478, 2007.



**Name:**  
Kun Zhang

**Affiliation:**  
Department of System Design Engineering,  
Graduate School of Engineering, University of  
Fukui

**Address:**

3-9-1 Bunkyo, Fukui 910-8507, Japan

**Brief Biographical History:**

2003- Major of Flight Vehicle Design and Engineering, Faculty of Aeronautic and Mechanical Engineering, Nanchang Hangkong University  
2008- Dept. of Human and Artificial Intelligent Systems, Graduate School of Engineering, University of Fukui  
2010- Dept. of System Design Engineering, Graduate School of Engineering, University of Fukui

**Membership in Academic Societies:**

- The Robotics Society of Japan (RSJ)
- Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)



**Name:**  
Yoichiro Maeda

**Affiliation:**  
Department of Human and Artificial Intelligent  
Systems, Graduate School of Engineering, Uni-  
versity of Fukui

**Address:**

3-9-1 Bunkyo, Fukui 910-8507, Japan

**Brief Biographical History:**

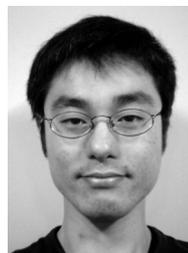
1983- Researcher, Central Research Lab., Mitsubishi Electric Corp.  
1989-1992 Senior Researcher, Laboratory for International Fuzzy  
Engineering Research (LIFE)  
1995- Associate Prof., Osaka Electro-Communication University  
1999-2000 Visiting Researcher, University of British Columbia (UBC),  
Canada  
2002- Associate Prof., Faculty of Engineering, University of Fukui  
2007- Prof., Graduate School of Engineering, University of Fukui

**Main Works:**

- "Behavior-Decision Fuzzy Algorithm for Autonomous Mobile Robots," Information Sciences, Vol.71, No.1, pp. 145-168, 1993.
- "Emotional Generation Model for Autonomous Mobile Robot," KANSEI Engineering Int., Vol.1, No.1, pp. 59-66, 1999.
- "Personal Preference Analysis for Emotional Behavior Response of Autonomous Robot in Interactive Emotion Communication," J. of Advanced Computational Intelligence and Intelligent Informatics (JACIII), Vol.14, No.7, pp. 852-859, 2010.

**Membership in Academic Societies:**

- The Society of Instrument and Control Engineers (SICE)
- The Robotics Society of Japan (RSJ)
- Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
- The Japanese Society for Artificial Intelligence (JSAI)
- Japan Society of Kansei Engineering (JSKE)



**Name:**  
Yasutake Takahashi

**Affiliation:**  
Department of Human and Artificial Intelligent  
Systems, Graduate School of Engineering, Uni-  
versity of Fukui

**Address:**

3-9-1 Bunkyo, Fukui 910-8507, Japan

**Brief Biographical History:**

2000-2009 Assistant Professor, Department of Adaptive Machine Systems,  
Graduate School of Engineering, Osaka University  
2003-2009 Member of Exec Committee for RoboCup middle size league  
2006-2007 Visiting Researcher, the Fraunhofer IAIS  
2009- Senior Assistant Professor, Department of Human and Artificial  
Intelligent Systems, Graduate School of Engineering, University of Fukui

**Main Works:**

- "Efficient Behavior Learning based on State Value Estimation of Self and Others," Advanced Robotics, Vol.22, No.12, pp. 1379-1395, 2008.
- "Emulation and behavior understanding through shared values," Robotics and Autonomous Systems, Vol.58, No.7, pp. 855-865, 2010.
- "Observed Body Clustering for Imitation Based on Value System," J. of Advanced Computational Intelligence and Intelligent Informatics, Vol.14, No.7, pp. 802-812, 2010.

**Membership in Academic Societies:**

- The Robotics Society of Japan (RSJ)
- Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
- The Japanese Society for Artificial Intelligence (JSAI)
- The Institute of Electrical and Electronics Engineers (IEEE)