ICIC Express Letters
Volume **13**, Number **1**, January **2019**

 ISSN 1881-803X
pp. **59–64**

CONSIDERATION ON A SOCIAL DILEMMA PROBLEM USING AGENTS WITH UNCERTAINTY OF INFORMATION

Junko Shibata[1], Koji Okuhara[2], Shintaro Mohri[1] and Shogo Shiode[3]

[1]Faculty of Economics
Kobe Gakuin University
518 Arise, Ikawadani, Nishi-ku, Kobe, Hyogo 651-2180, Japan
shibata@eb.kobegakuin.ac.jp

[2]Graduate School of Engineering
Toyama Prefectural University
5180 Kurokawa, Imizu, Toyama 939-0398, Japan

[3]Faculty of Business Administration
Kobe Gakuin University
1-1-3 Minatojima, Chuo-ku, Kobe, Hyogo 650-8586, Japan

Received June 2018; accepted September 2018

Abstract. *When the individual who composes chooses an egoistic act independently with others in a society system, the reward obtained sometimes decreases. This situation is called a social dilemma. In the environment, after an individual chose irrational behavior, he gets a higher reward. The information which an agent uses for the selection of actions is supposed to be accurate in most of past researches though it seems more practical that it includes uncertainty and delay. In this paper, focusing on Hogg-Huberman model originally composed of only competitive agents with the uncertain and delayed information, we investigate the total reward of the whole system in a tragedy game of the common ground. And we discuss the degree of information uncertainty on the total reward of the whole system. As a result, the computer experiment shows that the total reward using more accurate information becomes high.*
Keywords: Multi-agent system, Hogg-Huberman model, Tragedy game of commons

1. **Introduction.** In recent years, each of the constituent elements of the system has been studied to analyze the complex behavior of the system [1,2]. The component is called an agent, and the system constituted by many agents is called multi-agent system. This system is effective in a virtual space (artificial society) on a computer social phenomenon that it is difficult to perform experiments. And it is used to reproduce artificial society, to facilitate data collection and analysis, to analyze social phenomena in detail, and to deepen their understanding.

An agent constituting a multi-agent system targeting social phenomena is usually a model abstracting people constituting society. To realize a better society, research on cooperative behavior of agents is attracting attention [3-5]. The agent that is an autonomous entity selects an individual rational act that optimizes its purpose. Upon this selection, the agent uses external information. Since there are old information, unnecessary information and fake information in the real world, it is necessary to consider the uncertainty and time delay of the information used by the agent in selecting an action that matches the purpose. However, few researches deal with these. From the viewpoint of computational ecology, Hogg and Huberman proposed an agent model that iteratively selects and uses resources that best improve its objective from multiple resources [6,7]. The agent in this model uses uncertain and time delayed information. We call this model

DOI: 10.24507/icicel.13.01.59

Hogg-Huberman model. In addition, they classify agents into multiple types due to bias, and report the behavior of the agent at that time.

In this research, we focus on the problem of social dilemma in multi-agent environment. Under the social dilemma, as we pursue profits here, the benefits obtained decrease, and sometimes acting irrationally is required. Therefore, we deal with social dilemma problem using agents who make decisions according to Hogg-Huberman model. Information used by the agent for his or her decision influences the behavioral policy of the agent. We think that collaborative behavior can be realized using the average of total reward obtained from the entire system. By numerical experiments, we discuss the influence of the degree of uncertainty of information on the social dilemma problem. This paper is organized as follows. Section 2 explains the environment of the agent system and agent rules based on Q-learning. Section 3 describes Hogg-Huberman model with three resources. By computer simulations, we compare the agent who does the decision making based on the Hogg-Huberman model, and discuss the results. Finally, our conclusion is given in Section 5.

2. Tragedy Game of the Commons.

2.1. The environment of the agent system.

The environment is classified into the following in a relation between each agent and system which consists of that [8].

1) Incoherency situation: Society as a whole gets best when all agents pursue profits.

2) Quagmire situation: Society as a whole is best when all agents do not pursue profits.

3) Competitive situation: Society as a whole is best when some agents pursue profits and the rest do not pursue profits.

In this paper, we discuss the tragedy game of the commons which is a game modeling the n prisoners' dilemma [9]. Therefore, it handled the Quagmire situation. Prisoner's dilemma is a situation where individual rational behavior causes the deterioration of society as a whole and individual profits are also decreasing.

2.2. Agent rules based on Q-learning.

Q-learning is a learning method of updating the value $Q(s,a)$ associated with the combination of the state s and the action a or each learning cycle and selecting the action using this value [10]. At time t $(t = 1, 2, \ldots, T)$, agent i $(i = 1, 2, \ldots, I)$ gains the reward r_{t+1} after acting.

However, when the agent i acts, the cost c_i is generated for the agent society as a whole. Each agent acts according to the following algorithm.

Step 0: Set initial Q values.

Step 1: Agent i selects one action out of three egoistic, cooperative and altruistic behaviors. The egoistic Selfish behavior is $+\Delta r_c$, the cooperative behavior is 0, and the altruistic behavior is $-\Delta r_c$.

Step 2: Agent i gets a reward according to his selected action. The value of the reward is $3 - r_c$ if it is egoistic, $1 - r_c$ if it is collaborative and $-3 - r_c$ if it is altruistic. Here, r_c is a common cost and is represented by $r_c = \sum_i c_i$.

Step 3: Update $Q(s,a)$ according to the following equations.

$$Q_t(s,a) = Q_{t-1}(s,a) \text{ if } s \neq s_t \text{ or } a \neq a_t$$

$$Q_t(s_t,a_t) = \left(1-\alpha^Q\right) Q_{t-1}(s_t,a_t) + \alpha^Q \left(r_{t+1} + \gamma \max_b Q_{t-1}(s_t,b)\right) \tag{1}$$

where α^Q is the learning rate and γ is the discount factor.

Step 4: The agent i selects the next action stochastically in the roulette system according to the probability obtained by the Boltzmann selection

$$p(a) = \frac{\exp\left(Q_{t-1}(s_t,a)/T^B\right)}{\sum_b \exp\left(Q_{t-1}(s_t,b)/T^B\right)} \tag{2}$$

obtained from the Q values, where T^B is the temperature parameter.

Step 5: If the time t is last time, stop. Otherwise, update time and process. Go to Step 1.

3. **Hogg-Huberman Model with Three Resources.** The Hogg-Huberman model assumes that the information from the environment has uncertainty and time delay, and the agent evaluates the resources to gain more profit for the two resources. This model shows various phenomena such as periodic behavior and chaotic behavior by changing the value of the parameter. We explain the Hogg-Huberman model in discrete time is extended to three resources.

The symbol f_t^r is the fraction of agents using resource r $(r = 1, 2, 3)$ at discrete time t. In this system, α^H is the rate of agents which reevaluate the choice of resources at time and decide the resource used at the next time $(t+1)$. The rest of agents keep using the same resource at time t. Then, reevaluating agents select resource r used at the next time $(t+1)$ with the probability ρ_t^r that an agent will prefer resource r to the other. The dynamics of the system with two resources is then governed by following equation [11].

$$f_{t+1}^r = f_t^r + \alpha^H\{\rho_t^r - f_t^r\} \tag{3}$$

The probability ρ_t^r that the re-evaluated agent prefers resource r is given by the following equation.

$$\rho_t^r = \frac{P_t^r}{\sum_{u=1}^3 P_t^u} \tag{4}$$

where

$$P_t^r = \frac{1}{2}\left[1 + erf\left(\frac{G_{t-\tau}^r - G_{t-\tau}^0}{\sqrt{2}\sigma}\right)\right] \tag{5}$$

$erf()$ is the error function. And $G_{t-\tau}^r$ represents the profit obtained by selecting resource r at discrete time $(t-\tau)$. $G_{t-\tau}^0$ is the average profit obtained from three resources.

$$G_{t-\tau}^0 = \frac{1}{3}\sum_{u=1}^{3} G_{t-\tau}^u \tag{6}$$

In this model, in order to consider the uncertainty σ and the delay τ of information, each of the agents is assumed to select resource r with the probability ρ_t^r based on $G_{t-\tau}^r$ which is newest information available at time t. The parameter σ means the degree of uncertainty about information. When $0 < \sigma \ll 1$, information is sure and reliable. As σ becomes larger, information becomes more ambiguous, and unreliable. Then, the agent reevaluates the resources based on the benefits before τ time.

Humans may act irrationally according to circumstances. In other words, we do not necessarily choose resources that can get higher profits at the next time. Therefore, the agent i stochastically performs the roulette method using the probability

$$P_r\left(x_{t+1}^{i,r} = 1\right) = \max_r f_{t+1}^r \tag{7}$$

that selects the resource r at the next time $\left(x_{t+1}^{i,r} = 1\right)$.

4. **Computer Simulations.** In this paper, we confirm what action selection when the agent who does the decision making based on the Hogg-Huberman model plays the tragedy game of the commons. We assume that the profit $G_{t-\tau}^r$ obtained by selecting the resource r is the value of the reward according to the action selected by the user in Step 2 of 2.2. Also, it is assumed an agent system consisting only of homogeneous 10 agents.

For comparison, we use the three types of the agent: (1) random agent, (2) normal Q-learning agent, (3) neighborhood reward agent. The agent of (1) randomly selects actions. The agent of (2) performs ordinary reinforcement learning by the behavior rule shown in

2.2. The agent of (3) uses the reward r_{t+1} plus λ_{t+1} as learning reward r'_{t+1} instead of r_{t+1} [8].

$$r'_{t+1} = r_{t+1} + \lambda_{t+1} \tag{8}$$

where

$$\lambda_{t+1} = \sum_{A_k \in N_i \backslash A_i} r_{k,t+1} \tag{9}$$

$N_i \backslash A_i$ represents a set N_i consisting of neighbors of agents A_i excluding himself. N_i is defined by the following equation.

$$N_i = \{A_k | k = (i + j) \bmod 10,\ j = 0, 1, 2, 3\} \tag{10}$$

Next, Table 1 shows a list of parameter values used in this simulation.

TABLE 1. List of parameter values

The tragedy game of the commons	$I = 10$	$T = 30000$	$\Delta r_c = 1$
Q-learning	$\alpha^Q = 0.5$	$\gamma = 0.5$	$T^B = 1$
Hogg-Huberman model	$\alpha^H = 0.85$	$\tau = 0$	

In the agent system of Table 1, when all agents act altruistically, the total reward is 70. Conversely, if all agents behave egoistically, it is -70.

Figure 1 and Figure 2 show the transition of the total reward relative to the number of cycles. This graph approximates the average value every 100 cycles after calculating the average of 5 experiments. Therefore, the value on the horizontal axis is scaled to 1/100. In Figure 1, the black line shows the results of (1) random agent, the gray line (2) normal Q-learning agent, and the light gray line (3) neighborhood reward agent. And in Figure 2, the black line shows the results for $\sigma = 0.01$, the gray line for $\sigma = 0.25$, the light gray line for $\sigma = 1.00$, and the black dotted line for $\sigma = 10.00$.

From Figure 1, we can see that the total reward is the highest in the case of (1) random agent. Next, the normal Q-learning agent (2) is high, and the neighbor reward agent (3) is the lowest. This result is different from [8]. Figure 2 shows the results of agents based on the Hogg-Huberman model. When the information is more accurate, it is seen that higher rewards are obtained. Also, if the information is uncertain, it is almost the same as the random agent (1) in Figure 1.

Next, Figure 3 shows the number of times when the total reward reaches 70, -70 in all simulation steps (30000 cycles $\times$ 5 times).

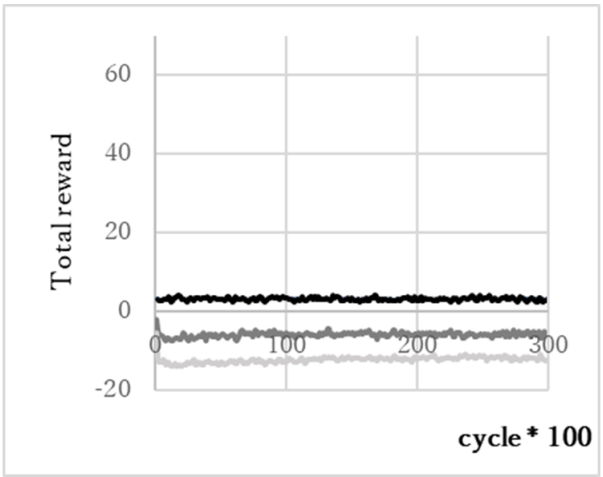

FIGURE 1. Results of the three types of the agents (1), (2) and (3)

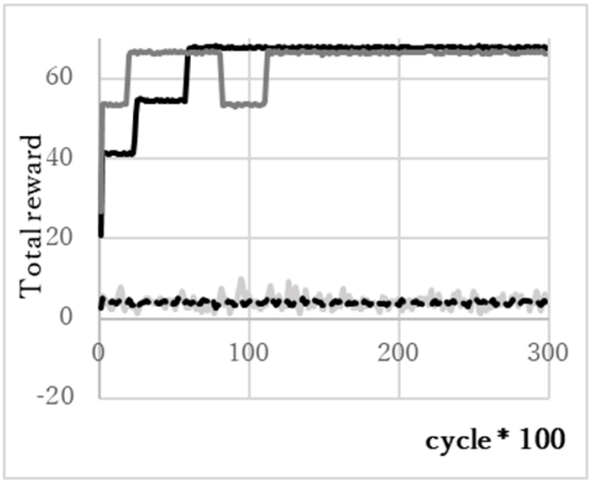

FIGURE 2. Results of the agent based on Hogg-Huberman model

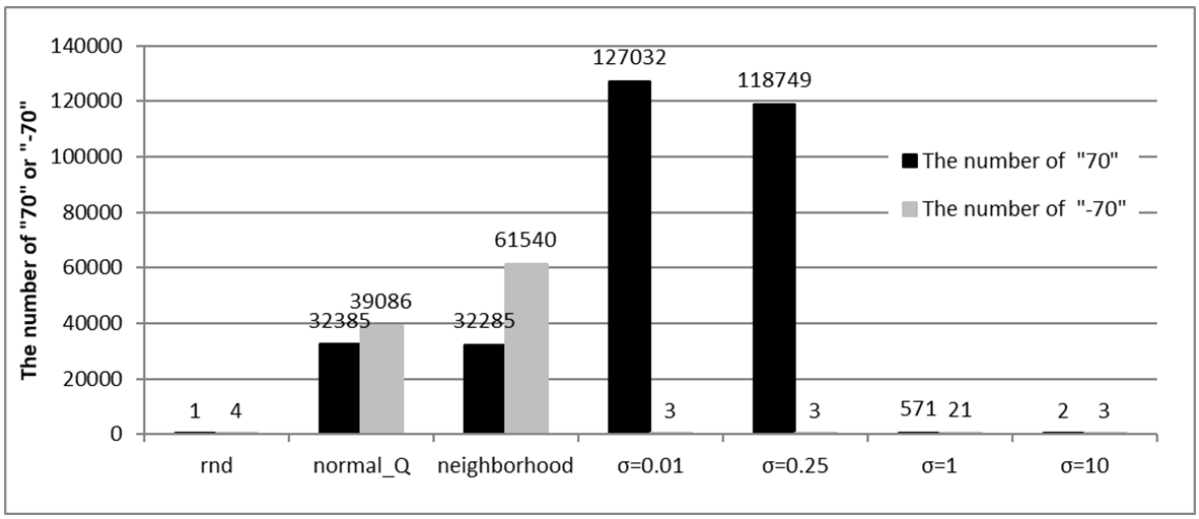

FIGURE 3. Results of the number of "70" and "−70"

From Figure 3, the highest black bar is the result of Hogg-Huberman model assuming accurate information ($\sigma = 0.01$). In this model, it means that there are many agents acting altruistically. Also, as the information becomes uncertain, the number of times that all agents choose the altruistic action decreases. However, the number of times that all agents choose the egoistic action is not influenced by the uncertainty of information.

5. **Conclusion.** In this paper, we focused on Hogg-Huberman model which is one of models of multi-agent system considering information uncertainty and time delay. We considered the result of the tragedy game of the commons by agents who select resources according to this model. Several numerical experiments discussed the influence of the degree of uncertainty of information on the total reward. In addition, by comparing with three models, we showed effectiveness on agent resource selection in Hogg-Huberman model. In [8], they have discussed that the agent tends to altruistic behavior by adding neighboring reward compensation. In this study, we showed that cooperative behavior can be realized by considering the average reward of the whole system, not the reward of a specific agent. The future work is to investigate the influence on the learning speed by the change of the number of agents constituting the system and finally adapt to the actual social problem.

REFERENCES

[1] T. Ito and T. Shintani, Implementation technologies for multiagent systems and their applications, *Journal of Japanese Society for Artificial Intelligence*, vol.16, pp.469-475, 2001.

[2] K. Kurumatani, Perspective of multiagent social simulation, *Systems, Control and Information*, vol.46, pp.518-523, 2002.

[3] P. A. M. Van Lange, J. Joireman, C. D. Parks and E. V. Dijk, The psychology of social dilemmas: A review, *Organizational Behavior and Human Decision Processes*, vol.120, no.2, pp.125-141, 2013.

[4] J. Z. Leibo, V. Zambaldi, M. Lanctot, J. Marecki and T. Graepel, Multi-agent reinforcement learning in sequential social dilemmas, *Proc. of the 16th Conference on Autonomous Agents and Multiagent Systems*, pp.464-473, 2017.

[5] F. Uwano, N. Tatebe, Y. Tajima, M. Nakata, T. Kobacs and K. Takadama, Multi-agent cooperation based on reinforcement learning with internal reward in maze problem, *SICE Journal of Control, Measurement, and System Integration*, vol.11, no.4, pp.321-330, 2018.

[6] T. Hogg and B. A. Huberman, Controlling chaos in distributed systems, *IEEE Trans. Systems, Man, and Cybernestics*, vol.21, pp.1325-1332, 1991.

[7] B. A. Huberman and T. Hogg, Behavior of computational ecologies, in *The Ecology of Computation*, B. A. Huberman (ed.), North-Holland, Amsterdam, 1988.

[8] K. Moriyama and M. Numao, Construction of a learning agent handling its rewards according to environmental situations, *The 1st International Joint Conference on Autonomous Agents and Multiagent Systems*, 2002.

[9] S. Mikami and Y. Kakazu, Co-operation of multiple agents through filtering payoff, *Proc. of the 1st European Workshop on Reinforcement Learning*, pp.97-107, 1994.

[10] C. J. C. H. Watkins and P. Dayan, Technical note: Q-learning, *Machine Learning*, vol.8, pp.279-292, 1992.

[11] M. Inoue, T. Tanaka, N. Takagi and J. Shibata, Dynamical behavior of a discrete time Hogg-Huberman model with three resources, *Physica A: Statistical Mechanics and Its Applications*, vol.312, pp.627-635, 2002.