Modeling Sequential Bargaining Game Agents towards Human-like Behaviors: Comparing Experimental and Simulation Results

Tetsuro Kawai, Yuhsuke Koyama, and Keiki Takadama
Tokyo Institute of Technology
4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8502 Japan
kawai@cas.dis.titech.ac.jp
koyama@dis.titech.ac.jp
keiki@dis.titech.ac.jp

ABSTRACT

The objective of this paper is to clarify whether the two types of agent-modeling (i.e. Roth's three parameter RE learning agent and Q-learning agent) are valid in terms of the reproduction of human-like behaviors or not. Specifically, we discuss it from both superficial viewpoint (i.e. simulation results) and internal viewpoint (i.e. simulation modeling). Concretely, we conducted subject experiments of sequential bargaining game and compared two types of simulation results (agents employ two types of learning mechanisms) with experimental results. An intensive comparison of experimental results and simulation results has revealed the following implications: (1-a) from superficial viewpoint, Q-learning agents reproduce human-like behaviors well against Roth's learning agent; (1-b) even Q-learning agents cannot reproduce the decreasing trend of negotiation size shown in subject experiments; (2-a) from internal viewpoint, the combination with Q-learning and boltzmann distribution selection has the possibility to reproduce human-like behavior which predicts the intention of the opponent; (2-b) even this combination cannot reproduce the decreasing trend of the negotiation size; and (2-c) the maximum limit of Q-values makes the differences of the converged Q-values for all actions in the same state smaller, which contributes to the acquisition of sequential negotiations.

Keywords: subject experiments, agent-based social simulation, sequential bargaining game, reinforcement learning

1. INTRODUCTION

Agent-based social simulation (ABSS) should be able to reproduce human-like behaviors more proficiently for its higher reliability, but it is really complex problem. Here, the comparisons of experimental results and simulation results are really important to find out the tips for modeling simulations with higher reproduction capabilities.

As the first step towards above challenge, we investigated the sensitivity of agent-modeling toward simulation results by switching both learning mechanisms and action selections in an agent [1]. Concretely, the simulation results of Roth's-learning agents (the representative model in experimental economics) [2], [3] and Q-learning agents (the representative model in computer science) with three types of action selections in a sequential bargaining game [4] are compared, and the superiority of Q-learning mechanism was revealed in terms of that: (1) Q-learning mechanism enables agents to produce two different kinds of results by changing action selections and its parameters; (2) Q-learning mechanism enables agents to acquire sequential negotiation; and (3) Roth's learning agents generate different simulation results for every executions of simulation. Here, such a conclusion was not compared with experimental results.

In this paper, we conduct subject experiments on a sequential bargaining game under the objectives: (1) to investigate when and what kind of payoff is accepted by subjects; (2) to observe human psychologies in the bargaining game; and (3) to validate our hypotheses that subjects continue the negotiation to acquire biggar payoff. Then we weigh and discuss the reproduction capabilities of Q-learning agents and Roth's three-parameter RE learning agents [2], [3] about the experimental results from both superficial viewpoint and internal viewpoint. Through such comparisons and discussions, we explore the agent-modeling which is able to reproduce human-like behaviors well. This attempt would directly contribute to the development of agent modeling methodology.

2. BARGAINING GAME

The sequential bargaining game [4] is a well-known example in the context of social science. In this game, two agents decide the dividing-ratio of money R through negotiations as shown in the Figure 1. MAX_STEPS which means the maximum number of negotiations (i.e. offering count) in a game is fixed and

both agents know this information as a common knowledge. This paper focuses on the sequential bargaining game where MAX_STEPS $>$ 1, while the bargaining game where MAX_STEP = 1 is called ultimatum game. In our simulations, we set MAX_STEPS to 6. We think this number is appropriate for our simulation because it is not too small and not too big number as sequential negotiations. It would make our simulations simple.

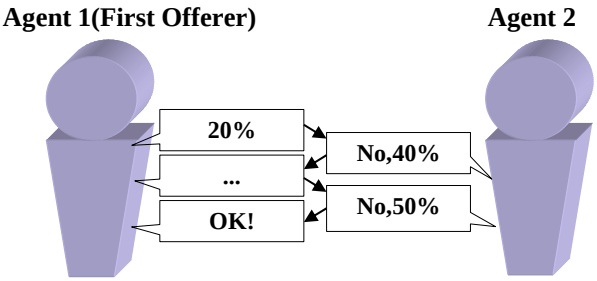

Fig.1 Sequential bargaining game

The concrete flow of the sequential bargaining game is explained below.

1. Agent 1 (first offerer) offers the value r_1 $(0 \langle r_1 \langle R)$ which means the dividing-ratio of R for Agent 2.
2. Agent 2 decides to accept the above r_1 or make a counter-offer r_2 $(0 \langle r_2 \langle R)$.
3. If agent 2 accepts the offer from agent 1, agent 1 acquires $R - r_1$ as the reward, while agent 2 acquires r_1. If agent 2 makes a counter-offer, both agents change their position each other and go to step 2.

This negotiation process is repeated up to MAX_STEPS. Here, if acceptance is not selected until the last of negotiation, both agents acquire no reward.

For example, focusing on the case that offering-values are integers from 1 to 9, the equilibrium payoff of theoretical approach is 9:1 whatever the MAX_STEPS is. Offered-value 1 was offered by the advantageous agent who can make the last offer to maximize his reward in this case, and the other agent accepts the offer because he prefers bigger reward (1 is larger than 0). If he does not accept the offer, he would acquire no reward. Thus the advantageous agent obtains 9, while the other obtains 1 as the reward [5]. The experimental results of ultimatum game are 5:5, 6:4 and so on (they are different by countries where the experiments were conducted) [2]. For sequential bargaining game, since there was no experimental result, we conducted subject experiments to collect the experimental results of 6-negotiation bargaining game.

3. MODELING AGENTS

This section explained the modeling agents for the framework of a sequential bargaining game described in the previous section.

3.1 Architecture of agents

An agent in a sequential bargaining game is illustrated in Figure 2. An agent consists of five elements: (1) *detector* for receiving information from the environment; (2) *effecter* that acts toward the environment; (3) *memory/knowledge* that stores information; (4) *learning mechanism* for updating the memory; and (5) *action selection* to select an action from the memory and input information which indicate the current state.

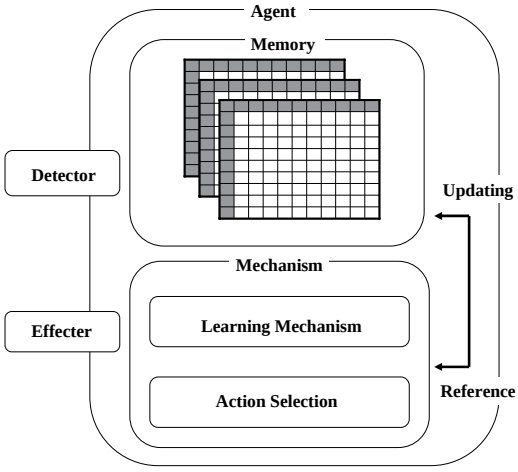

Fig.2 Architecture of Agents

3.2 State-space and Q-table

This subsection focuses on both Q-table which corresponds to the agent's memory and the state-space which is necessary to constitute Q-table.

Our sequential bargaining game simulations employ reinforcement learning as learning mechanisms of an agent, which needs appropriate design of the state-space. As agent's current state, the pair of the count of elapsed negotiations in the bargaining game and the offered payoff is applied. Agent's action is any of acceptance or offering payoff. In our framework of simulations, payoff r is a positive integer in the range of $1 \le r \le 9$, and $R = 10$. Discrete value from 1 to 9 is selected as an offer-value. Q-table (shown in Fig.3) is matrix of

Q-values which represents probabilities[1] for each action in all states. In this research, MAX_STEPS of the bargaining game is set to 6, thus agents have four or three Q-tables. Concretely, both agents are offered r three times and facilitate different Q-table for each r to select an action if the negotiation goes to the last stage. And first offerer needs one more Q-table to select the first action besides those three Q-tables.

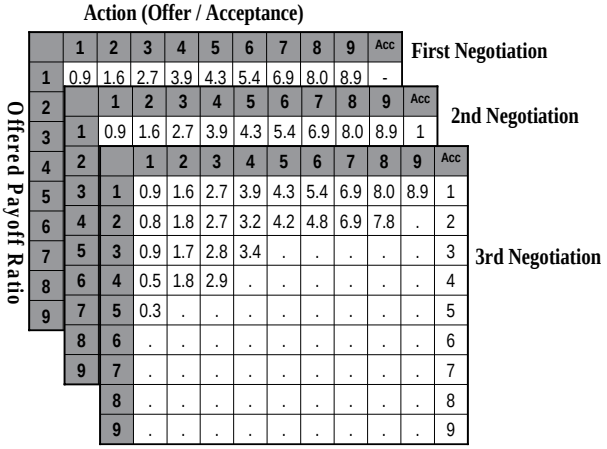

Fig.3 Q-Table

3.3 Learning mechanisms

Two types of reinforcement learning mechanisms are employed in our comparisons.

(1) Q-learning

Q-learning is one of famous reinforcement learning methods. In Q-learning, Q-values are updated by the following equation (1).

$$Q(s,a) \leftarrow Q(s,a) + \alpha\{r + \gamma \max_{a' \in A(s')} Q(s',a') - Q(s,a)\} \quad (1)$$

In equation (1), s, s', a, a' and $Q(s,a)$ indicate the current state, the next state where an agent will actually move, the current action, the next action in the next state, and Q-value which represents the probability that an agent takes action a at state s, respectively. r is the reward. The parameter $\alpha(0 < \alpha \le 1)$ is the learning rate which changes the learning speed. The parameter $\gamma(0 \le \gamma \le 1)$ is the discount rate which is the propagation rate of the reward. $A(s)$ is the set of actions that can be selected at a current state s.

In Q-learning, it is known that all Q-values converge as shown in the following equation (2), when setting appropriate parameters.

$$\lim_{t \to \infty} Q(s,a) = r + \gamma \max_{a' \in A(s')} Q(s',a') \quad (2)$$

In equation (2), t means the number of learning (defined as learning iteration). In our bargaining game simulations, $\lim_{t \to \infty} Q(s,a) = 9$ for all $Q(s,a)$.

(2) Roth's three-parameter RE learning [2], [3]

This learning method is an extended version of Roth's basic reinforcement learning [2], [3] by adding the experimentation/generalization parameter λ[2] and the forgetting parameter ϕ. In this learning, Q-values are updated by the following three equations (3).

$$\begin{aligned} &Q(s,a) \leftarrow (1-\phi)Q(s,a) + r(1-\lambda) \\ &Q(s,a \pm 1) \leftarrow (1-\phi)Q(s,a \pm 1) + r(\lambda/2) \\ &other\ \ Q \leftarrow (1-\phi)Q + 0 \end{aligned} \quad (3)$$

In this learning, not only the Q-value of the selected action but also one or two actions which have liner relation with the selected action are updated. For example, when an agent offers 4 and acquires reward $r = 6$ with $\lambda = 0.05$, three of Q-values are updated by $Q(s,4) \leftarrow Q(s,4) + 6(1-0.05)$, $Q(s,3) \leftarrow Q(s,3) + 6(0.05/2)$ and $Q(s,5) \leftarrow Q(s,5) + 6(0.05/2)$. If selected action is 1, two Q-values are updated by $Q(s,1) \leftarrow Q(s,1) + 9(1-0.05/2)$ and $Q(s,2) \leftarrow Q(s,2) + 9(0.05/2)$. In the case of 9, two Q-values are updated like this. Here, the parameter ϕ decreases Q-value along with the increment of learning iteration.

3.4 Action selections

(1) Boltzmann distribution selection

This method selects the action by the probabilities shown in the following equation (4).

$$P(a \mid s) = e^{Q(s,a)/T} \Big/ \sum_{a_i \in A(S)} e^{Q(s,a_i)/T} \quad (4)$$

In the equation (4), T is the temperature parameter which adjusts randomness of action selection. Agents

[1] Strictly speaking, those probabilities are led by Q-values and action selections (explained in this Section).

[2] In Roth's original models, this parameter is represented by ε.

frequently make random actions when T is high, while agents selects actions greedily when T is low.

4. SUBJECT EXPERIMENTS

4.1 Objective

Our major objectives for the subject experiments are summalized as follows: (1) to investigate when and what kind of payoff is accepted by subjects; (2) to observe human psychologies in the bargaining game; and (3) to validate our hypotheses that subjects continue the negotiation to acquire biggar payoff.

4.2 Outline of our experiments

The major setting of our subject experiments are shown in Table 1. The maximum number of negotiation (MAX_STEPS) is 6. Twenty subjects (10 conbinations) are graduate students who are not familiar with bargaining game and they are not informed how many times the bargaining game iterated (actually, 20 times). Ten games are conducted by 10 combinations of subjects, and each game are iterated 20 times. Subjects can accept the offer from the opponent or make counter-offer in each negotiation. In each game, subjects decide the dividing ratio of 3000 Japanese Yen through some negotiations. The rewards are calculated according to the game results in every games, but the rewards paid to subjects are decided randomly from the 20 games. For objective (2), questionaires are distributed to subjects. Major questions are as follows: (1) did you feel advantage or disadvantage in the game? (2) were you going to continue the negotiations? (3) what is your strategy for bargaining game? (4) how did you feel about your opponent? and (5) for what objective did you conduct this game, e.g. the maximization of your own payoff, acquiring bigger payoff than your opponent, or other objective?.

What is important to be noticed is that we do not apply time-discount for the payoff in this game, thus the maximum payoff (3000 Japanese Yen) never becomes smaller. As shown in the objective of our subject experiments, our hypotheses is that subjects continue the negotiation to acquire biggar payoff. Since it is clear that the time-discount factor works to decrease negotiation size, we do not apply time-discount to observe that subjects continue the negotiations in order to acquire bigger payoff.

Table1 Setting on our subject experiments

Maximum number of negotiation (MAX_STEPS)	6
Number of subjects	20
Number of games	10 (2 human players make conbination)
Number of iterations in each game	20 (subjects do not know this information)
Action of human players	1, 2, …, 9 and ACCEPT

4.3 Results of our experiments

We eliminated the data of non-completion of the deal from 10 results by 10 combinations, then averaged out the 10 results and made graphs of moving averages about payoff (Fig.4(a)) and negotiation size (Fig.4(b)). Each point of moving averages is calculated from 3 points (itself and the nearest 2 points). From these results, in most of cases, the payoff converged into 5:5, while the negotiation size converged into more than 2 (sequential negotiations), but it decreases as iteration increases.

Questionnaires have revealed subjects' feelings in the game. Most of subjects learned that the subject who can make the last offer has the advantage to acquire bigger payoff. Most of subjects tried to predict the intentions and characters of their opponents, and tried to imply their own intentions to their opponents through negotiations. The objectives of 20 subjects in the game are: (1) the maximization of their own payoff (8 persons); (2) an acquisition of bigger payoff than their opponent (7 persons); and (3) equality (5 persons). Finally, several subjects had feeling of anger to their opponent.

From the results of questionnaires, it is inferred that subjects tried to continue the negotiations to meet above objectives (i.e. the predictions of their opponents' intentions and the implications of their own intentions) in the early iterations, but they have no incentive to continue the negotiation after their objectives were met, so the decreasing trend of negotiation size is observed. This consideration motivates us to investigate the model which predicts the intention of the opponent such as Roth's fictitious play model [3].

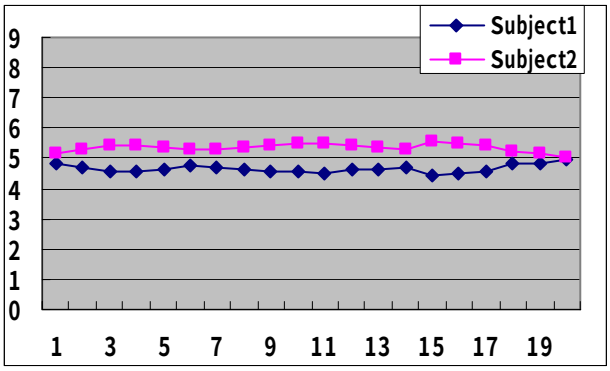

Fig.4(a) Payoff of subject experiments

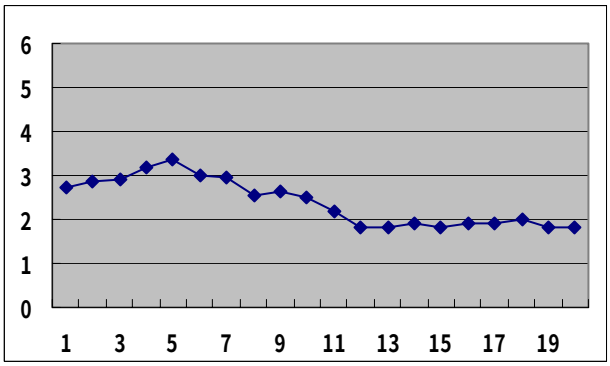

Fig.4(b) Negotiation Size of subject experiments

5. COMPARISON

The experimental results and simulation results are compared on sequential bargaining game. In our comparisons, two types of agent-modeling, Q-learning agents and Roth's three-parameter RE learning agents are explored for the reproduction capabilities of human-like behaviors. In our simulations, several combinations of learning mechanisms and action selections have already been investigated[3,4][1].

5. 1. Comparisons of results

Graph (a) in Fig. 5 and 6 indicates the payoff, while graph (b) indicates how many times the negotiation are continued until the bargaining game is concluded. The vertical axis in the figures indicates these two cases, while the horizontal axis indicates the iterations of each simulation. In graph (a), the gray line indicates the payoff of agent which can make the last offer (theoretically advantageous agent), while the black line indicates that of other agent. Note that these graphs are calculated by the moving averages of simulation results.

We conducted 10 simulations with 10 different random seeds for each case. Each graph of simulation results shows average of the 10 results.

The graphs of Roth's three parameter RE learning agents indicate: (1) the payoff converges to approximately 6:4; and (2) the negotiation size converges to 1. On the other hand, the graphs of Q-learning agents indicate: (1) the payoff converges to approximately 5:5; and (2) the negotiation size converges to approximately 2.

Roth's three-parameter RE learning agents

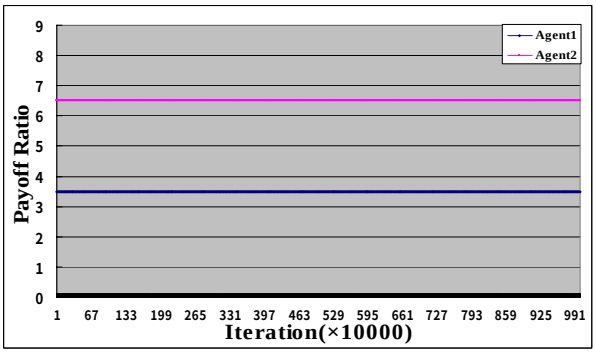

Fig.5(a) Payoff (Boltzmann selection T=0.5)

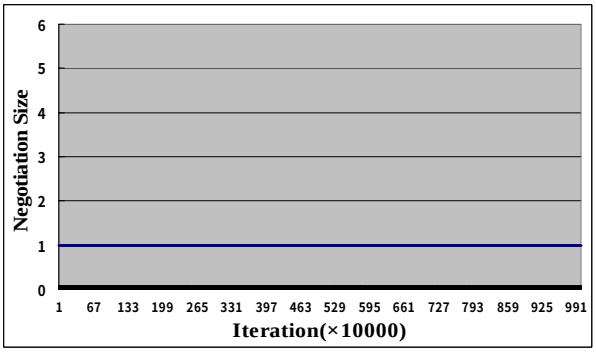

Fig.5(b) Negotiation Size (Boltzmann selection T=0.5)

[3] The details of our agent-modeling and simulations are described in [1].

[4] For Roth's learning agents, all parameter settings followed Roth & Erev's papers [2], [3].

Q-learning agents

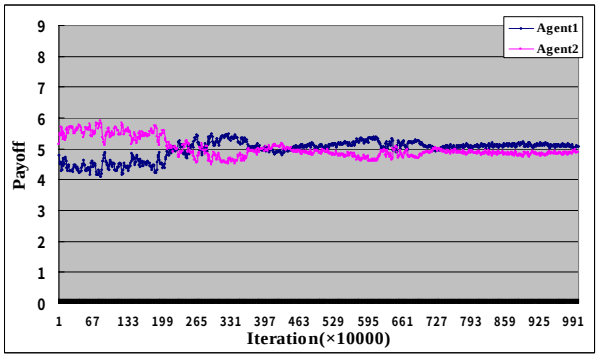

Fig.6(a) Payoff (Boltzmann selection T=0.5)

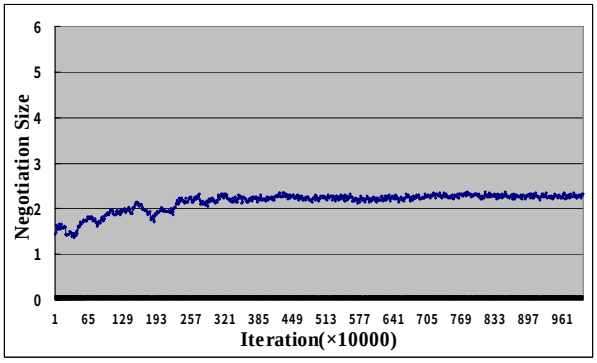

Fig.6(b) Negotiation Size (Boltzmann selection T=0.5)

5. 2. Discussion

5.2.1 Discussion from superficial viewpoint

First, we discuss the 2 types of agent-modeling from the superficial viewpoint (simulation results). In Fig.5(a), the payoffs of Roth's learning agents look like converging to approximately 6:4. But, this is the average of 10 results with 10 random seeds, and each 10 results are not equal at all (shown in Table2). This table shows the payoffs of each 10 simulations. For example, the payoff on the seed 1 indicates that the dividing ratio between agent 1 and agent 2 is 5.20 : 4.80. Thus, in our simulation, Roth's learning agents cannot conduct consistent learning. Sequential negotiations (more than one time negotiation) have not reproduced by Roth's learning agents.

Table2 Each 10 results of Roth's 3-parameter RE learning

	Seed 1	Seed 2	Seed 3	Seed 4	Seed 5	…
Payoff	5.20 : 4.80	7.78 : 2.22	7.74 : 2.26	7.05 : 2.95	2.00 : 8.00	…

On the other hand, Q-learning agents reproduced human-like payoffs well, and 10 results with 10 random seeds have strong consistency. Additionally, Q-learning agents reproduced sequential negotiations which are conducted by human players to a certain extent. Thus, Q-learning agents reproduced human-like behaviors well against Roth's learning agents in our comparisons. Even Q-learning agents, however, cannot reproduce the decreasing trend of negotiation size.

5.2.2 Discussion from internal viewpoint

Next, we discuss the 2 types of agent-modeling from the internal viewpoint (simulation modeling). Roth's learning agents cannot conduct consistent learning because of early convergence to a certain action which is selected randomly and acquired good reward in the early learning phase [1]. We tried to improve the performance of Roth's learning agents by changing parameter $S(1)$ which affects speed of learning. $S(1)$ is used to decide the initial probabilities of all actions, it is fairly divided among all actions at the same state s. Originally, $S(1)$ is set as $S(1)=10$ [2], and the same setting is used in our simulations shown in Fig.5. Here we conducted additional simulations with $S(1)=1000$ to delay the speed of the learning by avoiding that the differences of Q-values are generated rapidly, but an improvement for early convergence was not observed. Also, we have tried to improve this problem by adding a small change into forgetting effect, which enabled the negotiation size to become more than 1 (sequential negotiations) unexpectedly [1]. But, the negotiation size is too large against the experimental results.

On the other hand, Q-learning agents are able to acquire sequential negotiations because of the maximum limit of Q-value (see the equation (2)). This limit makes the differences of the converged Q-values for all actions in the same state smaller, which enables action selection has strong randomness permanently. High randomness contributes to the acquisition of sequential negotiations. This permanent randomness might represent the ambiguity of human. The combination with Q-learning and boltzmann distribution selection has the possibility to reproduce human-like behavior which decreases negotiation size after terminating to predict the intention of the opponent by changing the parameter T in the process of simulations. This change of the parameter in the process of simulations might be able to express the change of human feelings.

6. CONCLUSION

This paper showed our subject experiments on sequential bargaining game and comparisons of experimental results and simulation results. Specifically, we weighed reproduction capabilities of two types of agent-modeling for experimental results by both superficial and internal viewpoints. Our subject experiments have revealed: (1-a) from superficial viewpoint, Q-learning agents reproduce human-like behaviors well against Roth's learning agent; (1-b) even Q-learning agents cannot reproduce the decreasing trend of negotiation size shown in subject experiments; (2-a) from internal viewpoint, the combination with Q-learning and boltzmann distribution selection has the possibility to reproduce human-like behavior which predicts the intention of the opponent; (2-b) even this combination cannot reproduce the decreasing trend of the negotiation size; and (2-c) the maximum limit of Q-values makes the differences of the converged Q-values for all actions in the same state smaller, which contributes to the acquisition of sequential negotiations.

The following issues should be pursued in the near future: (1) reproduction of the decreasing trend; (2) comparisons of various agent-modeling from the viewpoint of the ultimatum bargaining game simulation; and (3) construction of the model which is able to predict the intention of the opponent.

ACKNOWLEDGEMENTS

The research reported here was supported in part by a Grant-in-Aid for Scientific Research (Young Scientists (B), No. 17700139) of Ministry of Education, Culture, Sports, Science and Technology (MEXT).

REFERENCES

[1] Kawai, T. and Takadama, K.: "Modeling Sequential-Bargaining-Game Agents by Switching Learning Mechanisms and Action Selections," *The fourth international workshop on Agent-based Approaches in Economic and Social Complex Systems (AESCS'05)*, pp. 129-140 (2005).

[2] Roth, A. E., and Erev, I.: "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, Vol.8, No.1, pp. 164-212 (1995).

[3] Erev, I. and Roth, A. E.: "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *The American Economic Review*, Vol.88, No. 4, pp. 848-881 (1998).

[4] Kagel, J. H. and Roth, A. E.: "The Handbook of Experimental Economics," *Princeton University Press*, pp. 253-342 (1995).

[5] Ståhl, I.: "Bargaining theory" *Economics Research Institute at the Stockholm School of Economics* (1972).