Journal of Robotics, Networking and Artificial Life, Vol. 1, No. 3 (December 2014), 231-236

# An Action Selection Method Using Degree of Cooperation in a Multi-agent Reinforcement Learning System

**Masanori Kawamura**
*Aichi Prefectural University, 1522-3 Ibaragabasama*
*Nagakute, Aichi 480-1198, Japan*

**Kunikazu Kobayashi**
*Aichi Prefectural University, 1522-3 Ibaragabasama*
*Nagakute, Aichi 480-1198, Japan*
*E-mail: kobayashi@ist.aichi-pu.ac.jp*
*http://www.ist.aichi-pu.ac.jp/~koba/*

**Abstract**

In recent years, a concept of a dividual is proposed by a Japanese novelist to interact properly with another person. To construct a model of the dividual, the degree of cooperation is assigned to the corresponding dividual. By introducing the degree of cooperation into a multi-agent system, we evaluate what kind of changes appears in the agent behavior. In addition, we propose an action selection method by introducing the degree of cooperation into the soft-max method in a multi-agent system. Using the proposed method, we confirm whether the cooperative action is promoted or suppressed through computer simulations.

*Keywords: Multi-agent system, Reinforcement leaning, Cooperative action, Dividual, Degree of cooperation.*

## 1. Introduction

In recent years, robots such as a cleaning robot and a nursing care robot become something familiar to us. In the near future, a convivial society is supposed that the robots can communicate with each other like a person and also they can smoothly cooperate with persons. It is supposed that smooth communication between a person and a robot can realize by emulating communication between persons.

Nagayuki et al. presented a policy estimation method which can estimate the other's action to be taken based on the observed information about the other's action sequence[1,2]. They successfully applied it to the reinforcement Q-learning method[3] and showed to get effective the other's policy. Meanwhile, Yokoyama et al. proposed an approach to model action decision based on the other's intention according to atypical situation such as human-machine interaction[4,5]. They presented three estimation levels of the other's intention and presented a computational model of action decision process to solve cooperative tasks through a psychological approach. In this context, Kobayashi et al. successfully presented an adaptive approach for automatically switching the above three estimation levels depending on the situation[6].

In the human society, a person act cooperatively by taking some kinds of communication such as gesture, language, and eye contact. Recently, a concept of a dividual is proposed by Hirano to interact properly with another person[7]. At present, by introducing the above concept into a multi-agent system[8,9,10], we construct a model of the dividual to realize cooperative behavior.

In the present paper, we treat a difference of how to interact with another person, which characterizes the dividual model. When a self-agent recognized the other agent, a dividual is formed in the self-agent. At the same time, the degree of cooperation proposed in the present paper is assigned to the corresponding dividual.

By introducing the measure into a multi-agent system, we evaluate what kind of changes appears in the agent behavior. In addition, in the present paper, we

Published by Atlantis Press

propose an action selection method based on the degree of cooperation in a multi-agent reinforcement learning system. Using the proposed method, we confirm whether the cooperative action is promoted or suppressed through computer simulations.

In section 2, an action selection method using degree of cooperation is proposed. In section 3, the performance of the proposed method is evaluated through computer simulations. In section 4, we give a summary of the present paper and describe future problems.

## 2. Proposed Method

First of all, we define a problem to be treated in the present paper (section 2.1). Then, we explain a concept of a dividual (section 2.2) and the degree of cooperation (section 2.3). After that, we proposed an action selection method using the degree of cooperation (section 2.4).

### 2.1. *Problem Definition*

We treat a problem that plural agents arrive the same goal in a grid world in Fig.1.

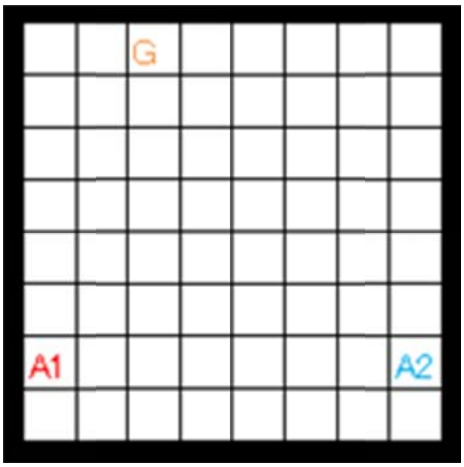


Fig. 1: An example of field.

### 2.2. *Dividual*

A person usually communicate by using how to interact other person properly. For example, you face your boss, I bet you communicate politely with him/her. But you face a good friend, I bet you communicate friendly with him/her. Although we give a simple example, a person may change a way to interact with other person according to sex, nationality, relationship with him/her. This concept is named as a dividual by Hirano[7].

Dividual is roughly divided into three types. The first one is a social dividual. This is a standard dividual to interact with a stranger or an unfamiliar person. The second one is a group-oriented dividual. This is a dividual for a specific group such as a school class or a tennis club. The third one is an individual-oriented dividual. This is a dividual for a specific person such as family members or a bully. In the present paper, the last dividual, i.e. individual-oriented dividual is treated.

A different dividual interacts with the other per-son, i.e. families, friends, and acquaintances. The number of dividual therefore corresponds to that of other persons who interact with. When you communicate with person A and become A's acquaintance, A's dividual is constructed in yourself. Similarly, when you communicate with person B and become B's acquaintance, B's dividual is also constructed in yourself. It is suggested that a set of your dividuals may characterize a human personality. In the present paper, when a dividual is created, its degree of cooperation is defined so as to cooperate with other person.

### 2.3. *Degree of Cooperation*

In the present paper, the degree of cooperation c which corresponds to its dividual is defined. The c takes a scalar value and fall within the range of $0 \leq c \leq 1$. When a dividual is firstly created, a social dividual is defined as a default dividual and grown by interacting each other. The degree of cooperation for the social dividual is defined as 0.5. If $c>0.5$ and $c<0.5$, the degree of cooperation is regarded as high and low cooperation, respectively. The high and low degrees of cooperation promotes and suppress the cooperation.

### 2.4. *Action Selection Method Using Degree of Cooperation*

In this section, we propose an action selection method using the degree of cooperation in order to cooperate with others. In the proposed method, we identify the direction of the other agent and realize the cooperation by reflecting the degree of cooperation toward its direction. This is explained using Fig.2.

In Fig.2, there is a goal in the top direction and a cooperative partner in the right direction. Let us assume

that the self-agent can select an action out of moving the right, left, top, or bottom. The proposed method realizes that the self-agent tends to select an action toward the other agent as $a_0$ out of available actions by reflecting the degree of cooperation on the probability of selecting $a_0$.

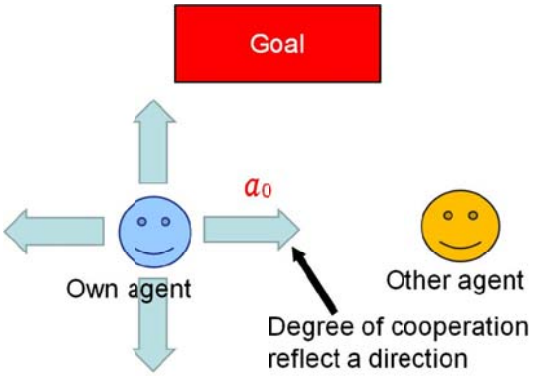


Fig. 2: Relationship between two agents.

In the proposed method, we use the Q-leaning method[3] which is one of the representative reinforcement learning methods[11,12] and soft-max action selection method. But, our approach apply to ε-greedy method. The proposed action selection method is defined as Eq.(1).

$$p(a|s) = \frac{\exp(Q(s,a)g(c))}{\sum_{b\in A}\exp(Q(s,b)g(c))} \quad (1)$$

where A is a set of available actions. The degree of cooperation is introduced as a function g(c). In proposed method, g(c) is defined as Eq.(2).

$$\mathrm{g(c)} = \begin{cases} 2c & (a = a_0) \\ 1 & (otherwise) \end{cases} \quad (2)$$

We assume that the degree of cooperation for the social dividual is set to 0.5. The value of function g(c) is calculated as g(0.5) = 1 so that the probability for the social dividual does not change at all. This corresponds to not considering cooperation with a stranger. On the other hand, when the degree of cooperation is larger than 0.5, it becomes g(c)>1 and the cooperation is promoted. When the degree of cooperation is smaller than 0.5, it becomes g(c)<1 and the cooperation is suppressed. Therefore, function g(c) is set as Eq.(2).

## 3. Computer Simulation

We conducted computer simulations to evaluate the proposed action selection method based on the degree of cooperation. At first, we describe problem setting (section 3.1) and parameter setting (section 3.2). Then, we give simulation results and discuss them (section 3.3).

### 3.1. *Problem Setting*

The proposed method is evaluate using the 10 × 10 field in Fig.1. We prepare for two agents and one goal, two agents act in a discrete grid world. In the field, the black surrounding is wall, there are two agents $A_1$ and $A_2$ and one goal G. The agents can select an action out of moving the right, left, top, or bottom and their aim is moving towards the goal.

### 3.2. *Parameter Setting*

The number of episodes is set as 10,000 and the maximum number of steps is limited to 100. The performance is evaluated by the distance between two agents at every step. To evaluate the degree of cooperation, we prepare two levels of the degree. In case of a cooperative agent, the degree of cooperation is set as 0.8 and in case of a non-cooperative agent, it is set as 0.2. In this simulation, agents can only move four directions, i.e. up, down, right, and left. The minimum numbers of steps toward the goal for agents $A_1$ and $A_2$ are 8 and 11 steps, respectively.

### 3.3. *Simulation Result*

Firstly, the transition of the distance between two agents whose degrees of cooperation are both high is shown in Fig.3. The image of an action sequence in the field is also illustrated in Fig.4. In these figures, the degree of cooperation from $A_1$ to $A_2$ is denoted by $C(A_1, A_2)$.

As a result, two agents approach each other and move toward the goal while keeping close distance. Therefore, it followed that they cooperated when each other's degree of cooperation was high. In addition, considered minimum step of agent $A_2$ was 11 steps, it is thought that they arrived at a goal after they drop in a little by they give priority to cooperate.

Secondly, we show the transition of the distance between two agents when the degree of cooperation from $A_1$ to $A_2$ is high but that from $A_2$ to $A_1$ is low in

Fig.5. The image of an action sequence in the field is also illustrated in Fig.6.

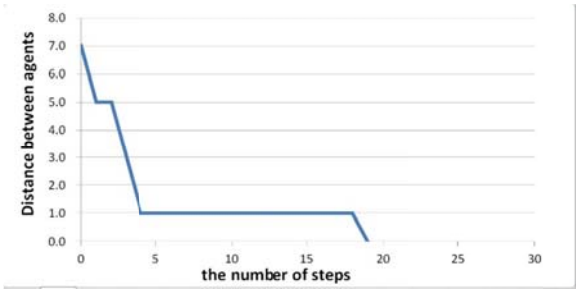


Fig. 3: Transition of distance between agents when $C(A_1, A_2) = 0.8$ and $C(A_2, A_1) = 0.8$.

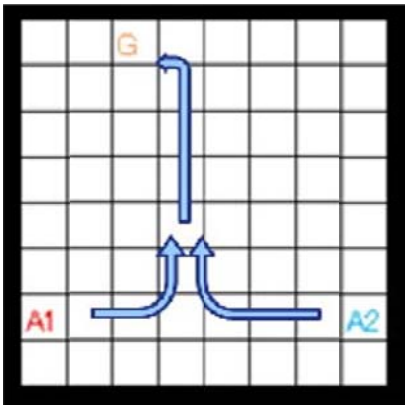


Fig. 4: An action image when both $C(A_1, A_2)$ and $C(A_2, A_1)$ are high.

As a result, the distance between agents does not decrease rapidly but become gradually close toward the goal. Therefore, when $C(A_1, A_2)$ is high and $C(A_2, A_1)$ is low, cooperative action is slightly suppressed. In addition, as considering the number of steps for $A_2$, it takes 23 steps until it arrives at the goal. This shows that $A_1$ gives first priority to cooperate with $A_2$ than to arrive at the goal and $A_2$ gives first priority not to cooperate with $A_1$ than to arrive at the goal.

Thirdly, we show the transition of the distance between two agents when the degree of cooperation from $A_1$ to $A_2$ is low but that from $A_2$ to $A_1$ is high in Fig.7. The image of an action sequence in the field is also illustrated in Fig.8.

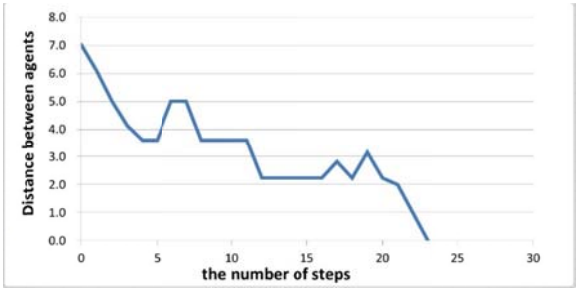


Fig. 5: Transition of distance between agents when $C(A_1, A_2) = 0.8$ and $C(A_2, A_1) = 0.2$.

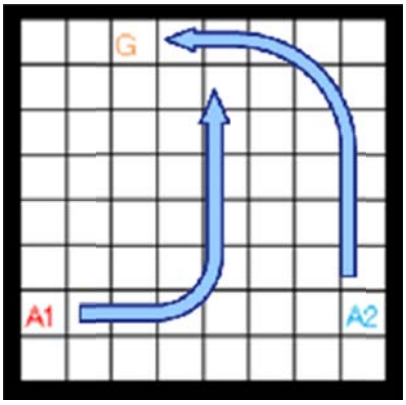


Figure 6: An action image when $C(A_1, A_2)$ is high and $C(A_2, A_1)$ is low.

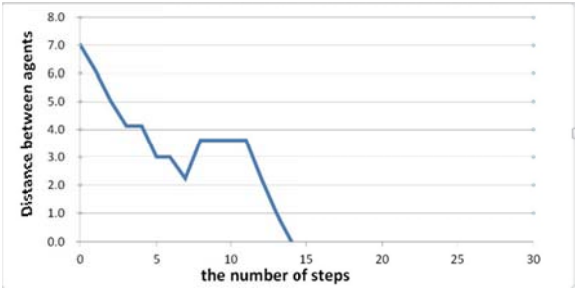


Fig. 7: Transition of distance between agents when $C(A_1, A_2) = 0.2$ and $C(A_2, A_1) = 0.8$.

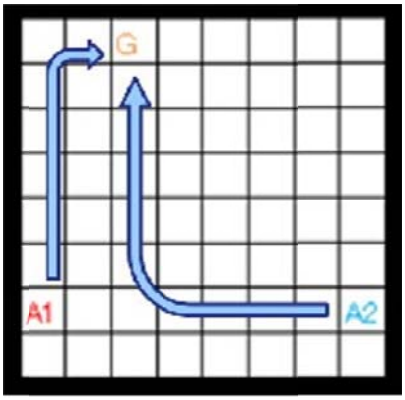


Fig. 8: An action image when $C(A_1, A_2)$ is low and $C(A_2, A_1)$ is high.

As a result, the distance between agents does not also decrease rapidly but become gradually close toward the goal. Therefore, when $C(A_1, A_2)$ is low and $C(A_2, A_1)$ is high, cooperative action is slightly suppressed. In addition, as considering the number of steps for $A_2$, it takes 14 steps until it arrives at the goal. This number of steps is a little bit smaller than the above case, i.e. $C(A_1, A_2) = 0.8$ and $C(A_2, A_1) = 0.2$ because the goal position is closer to $A_1$ than $A_2$. This shows that $A_2$ gives first priority not to cooperate with $A_1$ than to arrive at the goal and $A_1$ gives first priority to cooperate with $A_2$ than to arrive at the goal.

Finally, we show that the transition of the distance between two agents whose degrees of cooperation are both low is shown in Fig.9. The image of an action sequence in the field is also illustrated in Fig.10.

As a result, the distance between agents does not decrease at all. This shows that both agents give first priority not to cooperate with each other than to arrive at the goal.

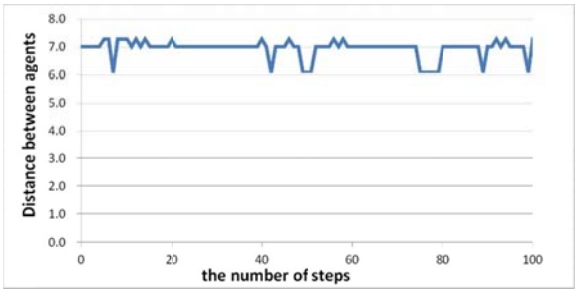


Fig. 9: Transition of distance between agents when $C(A_1, A_2) = 0.2$ and $C(A_2, A_1) = 0.2$.

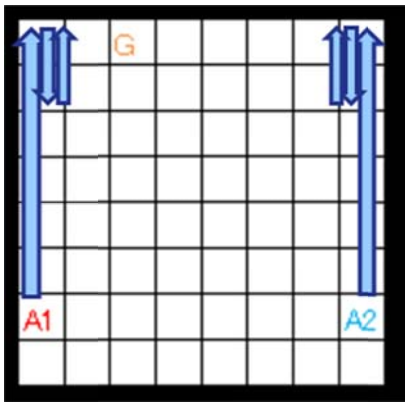


Fig. 10: An action image when both $C(A_1, A_2)$ and $C(A_2, A_1)$ are low.

## 4. Conclusion

In the present paper, we focused on the concept of the dividual and introduced the degree of cooperation. Then, we proposed the action selection method based on the soft-max method using the degree of cooperation. Through computer simulations, it was verified that two agents with high degree of cooperation approached each other and arrived at the goal. On the other hand, it was clear that two agents with low degree of cooperation did not approach each other and arrive at the goal neither. In addition, when one agent with high degree of

cooperation and the other agent with low degree of cooperation, it was shown that the distance between agents became gradually low and finally agents arrived at the goal.

In the present paper, the degree of cooperation is fixed throughout computer simulations, it is however feasible that it is adjustable in the real world. Although the degree of cooperation is depends on personal appearance and inner face, and also personal condition and impression, it is difficult how to adjust it.

It is supposed that a home robot introducing a concept of dividual benefits the elderly and children.

## Acknowledgements

This work was partly supported by JSPS KAKENHI Grant Number 23500181.

## References

1. Y. Nagayuki, S. Ishii, M. Ito, K. Shimohara, and K. Doya, "A Multi-Agent Reinforcement Learning Method with the Estimation of the Other Agent's Actions," *Proceedings of the Fifth International Symposium on Artificial Life and Robotics*, **1**, 255–259 (2000).
2. Y. Nagayuki and M. Ito, "Reinforcement Learning Method with the Inference of the Other Agent's Policy for 2-Player Stochastic Games," *Transactions on the Institute of Electronics, Information and Communication Engineers*, **J86-D-I**(11), pp.821–829 (2003) (in Japanese).
3. C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, **8**, pp.279-292 (1992).
4. A. Yokoyama, T. Omori, S. Ishikawa, and H. Okada, "Modeling of Action Decision Process Based on Intention Estimation," *Proceedings of Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on advanced Intelligent Systems*, No.TH-F3-1 (2008).
5. A. Yokoyama and T. Omori, "Model Based Analysis of Action Decision Process in Collaborative Task Based on Intention Estimation," *Transactions on the Institute of Electronics, Information and Communication Engineers*, **J92-A**, pp.734–742 (2009) (in Japanese).
6. K. Kobayashi, R. Kanehira, T. Kuremoto, and M. Obayashi, "An Action Selection Method Based on Estimation of Other's Intention in Time-Varying Multi-Agent Environments," *Lecture Notes in Computer Science*, **7064**, pp.76-85, Springer-Verlag (2011).
7. K. Hirano, "*Who am I?: From individual to dividual*," Kodansha shinsho (2012) (in Japanese).
8. P. Stone and M. Veloso, "Multiagent Systems: A Survey from a Machine Learning Perspective," *Autonomous Robots*, **8**, pp.345–383 (2000).
9. A. Ohuchi, M. Yamamoto, and H. Kawamura, "*Basics and Applications of Multi-agent Systems*," Corona Publishing (2003) (in Japanese).
10. K. Takadama, "*Multi-agent Learning*," Corona Publishing (2003) (in Japanese).
11. R. S. Sutton and A. G. Barto, "*Reinforcement Learning: An introduction*," MIT press (1998).
12. L. P. Kaelbling, M. L. Littman, and A. P. Moore, "Reinforcement Learning: A Survey," *Journal of Artificial Intelligence Research*, **4**, pp.237–285 (1996).