Emergence of Individuality and Sociality by Reinforcement Learning in Multi-Agent Systems

Katsunari Shibata, Masahide Ueda, & Koji Ito
Dept. of Computational Intelligence & Systems Science, Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan. shibata@ito.dis.titech.ac.jp

Abstract

In this paper, a new viewpoint, "individuality" and "sociality", is introduced for analyzing the multi-agent system's behavior. A hypothesis is set up that the both behavioral characters emerge to generate different actions from the other agent to increase each agent's benefit. "Individuality" is defined as the difference of actions between agents based on the difference of its internal information processing. While "sociality (rule)" is defined as the difference of actions based on the difference of its sensory inputs.

Next, a model is proposed in which "individuality" and "sociality" are obtained by reinforcement learning. It is also mentioned that there exist some factors like asymmetry of the environment, which influence the differentiation into one of the two characters. Finally through some simulations of conflict avoidance problems among passengers getting on and off a train, it is examined that the differentiation is adaptive to some of the above factors appropriately, and the rule that the passengers getting off have a priority to go.

1 Introduction

When we see our human society, we make and obey many rules, while each of us have a variety of individualities. Too many rules reduce the effectiveness of the society like a signal at the cross with less traffic. On the other hand, excessive individualities break the order of the society. Both are the opposed characters with each other, and necessary but must be moderate.

In some works in multi-agent(robot) systems and game theory, some individualities are given and the ratio of the agents having each individuality changes adaptively like the Hawk-Dove model[1] and so on [2]. Ota et al. proposed a distributed strategy-making using the instant reinforcement signals[3]. A kind of rule that was changed adaptively according to the given environment, emerged among the agents based on each agent's learning by only its own reward. However, they did not mention "individuality" and "sociality" at all, and the given task in their work is too simple in the following meanings. The reinforcement signal is given for each decision, and the number of the states and the number of actions for each agent are only two. Furthermore no serious conflict is assumed. The serious conflict means that in order to avoid the conflict, an agent has to take a strategy which is bad when the opponent agent does not exist.

In this paper, the existence of individuality and society is not dealt with as a premise, but the emergence of the both characters itself is also a subject. At first, "individuality" and "sociality" are introduced, and the factors of the differentiation into one of the two are considered. Then some simulations are done with more realistic environment.

2 Individuality and Sociality

2.1 Hypothesis and Definition

A hypothesis set up here is that both "individuality" and "sociality" are not inherited, but are emerged through each agent's learning to increase its benefit. The latter is the same standpoint as Adam Smith's[4].

Both behavioral characters are defined between two agents at first. Fig. 1 shows the definitions. "Individuality" is defined that each agent generates a different action from the other agent's even if the sensory inputs are the same. While, "sociality" is defined that two agents generate the same action when the sensory inputs are the same, but they generate different actions when the sensory inputs are different. If each agent's action is observed to be exchanged with each other when their states were exchanged, it is called that there exists "sociality" between them. If the action is the same, it is called that there is "individuality" between them.

In multi-agent case, it is not easy to define each agent's behavioral character. Here, for the present, it is judged by observing all the relations to the other agents. If "individuality" is the relation with more agents than "sociality", it is defined that "individuality" is stronger in the agent. In the contrary case, the agent is defined to have strong "sociality".

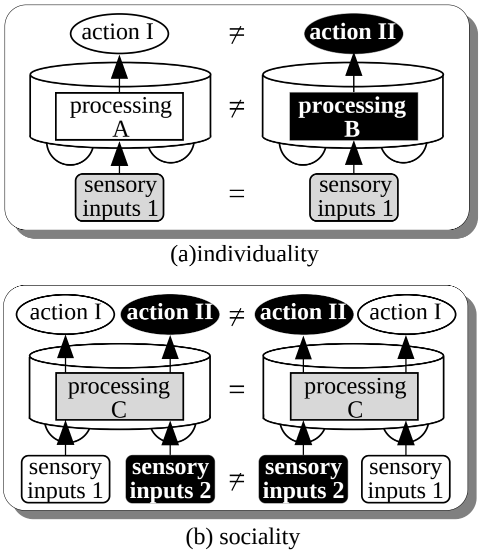


Figure 1: Definition of "identification" and "sociality" between two agents.

In the task (environment) in which two agents cannot achieve their purposes or can get less reward when they do the same action, they are expected to generate the different actions and avoid such situations by applying reinforcement learning to each agent. A model is proposed that the both characters emerge through such experiences and learnings

2.2 Differentiation and Its Factors

As in the previous section, there are two ways for an agent to generate a different action from the other agent. The reinforcement learning can decide which behavioral character is appropriate in the relation to the other agents. The factors which affect the differentiation can be picked up as follows. (1) Asymmetry of the environment, (2) Identification of the other agents, (3) Physical and mental characters of each agent, and (4) Communication with the other agents. If the environment is so asymmetric that it influences the agent's performance, some rules emerge, but the environment is symmetric, the agent can utilize only "individuality" to generate the different action. If the other agents can be identified, the agent can change its strategy according to the other, and "individuality" is apt to emerge. If an agent's physical or mental character is different from the others, it is easy for the agent to generate the different actions based on "individuality". The communication with the other agent helps to identify the other, and that results in the emergence of "individuality" [5]. In this paper, it is examined if some of those factors influence the differentiation appropriately.

3 Simulation

In this paper, the problem in which passengers are getting on and off a train as shown in Fig. 2 is taken as examples. In this problem, when both the passengers getting on and getting off are going to go straight, conflicts may happen. It is well-known that the rule is effective that the passengers getting off have priority to go.

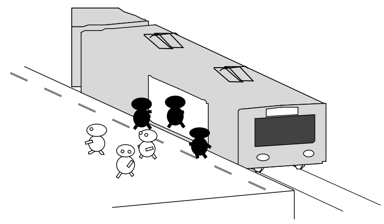


Figure 2: Conflict avoidance problem of passengers getting on and off a train.

3.1 Experimental Setup

Whole the space is represented as a grid world as shown in Fig. 3, and the boxes are divided into two regions, "in the train" and "out of the train". Some agents are initially located randomly in the stated places in the both regions. Only one agent (passenger) can exist in a box, and the agent can move only to the next box. If the number of whole the agents is larger than that in one simulation run, the agents for each run are chosen randomly. The order that the agents action depends on the initial locations.

The number of sensory inputs for each agent is 8. The first input is the role of the agent which represents whether the agent is "getting on" or "getting off". That depends on the initial location. The second input represents whether the doorway exists "in front or at the back", "right hand side", or "left hand side" of the agent. The states of the next boxes are indicated by the following four inputs. The state can be one of the four states, "empty", "occupied by the same group agent", "occupied by the opponent", and "wall exists between the boxes". Finally the other two inputs represent the state of the front box from the view of the right-next and left-next boxes respectively.

Each agent can select one of 5 actions, "move to each of the four next boxes", and "does not move". If there exists some agent already in the destination box, it does not move. If each agent arrives at its goal, it stops to learn and move randomly while restricted to move only in the goal area. The process until all the agents arrive at their goals, is defined as one trial, and one cycle means that all the agents move once.

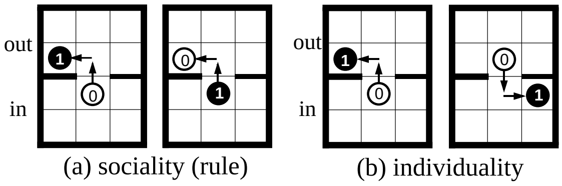


Figure 3: Example of "sociality" and "individuality" between two agents in the passengers problem.

In this simulation, one of the agents facing each other has to give its way to the other to avoid the conflict. It is observed which agent gives its way, and after the place of one agent is exchanged to the other's, it is observed again. Then if the different agents gives their ways as shown in Fig. 3 (a), there exists a rule and "sociality" exists between them. If the same agent gives its way as shown in Fig. 3 (b), the relation is judged as "individuality". The agent who gives its way has "individuality" of seeming "altruistic", while the other has "individuality" of "selfish".

3.2 Learning

Here, one-step Q-learning is employed as a reinforcement learning algorithm. The agent can get a reward 1.0 when it arrives at its goal. The goal is the upper end for the agent going upward and the bottom for the agent going downward. It gets no reward by the other agent's goal. That assumption is based on the selfish agent hypothesis as in the subsection 2.1. As each agent's strategy of action selection, Boltzmann Selection is employed, and the temperature is reduced exponentially from 1.0 to 0.01 during learning. The learning rate is 0.01 and the discount factor γ is 0.92.

3.3 Results

At first, the adaptive differentiation according to the specification of the other agent and the asymmetry of the environment is examined in the two-agent case. In this case, two agents are chosen randomly among some agents, and the choice and trial are repeated 40000 times. The grid world consists of 4x5 boxes, that is the same as Fig. 4 (a). In the case that the agents identify the other agents, the state of the next box can be not only opponent or wall, but can be agent1..6 except for the agent itself. This means that if the opponent is different from the previous trial, the state is perfectly different, and the previous learning does not affect the present learning at all. When the number of whole the agents is 2, the identification does not have any meaning because the number of the opponent is only one. In the symmetrical environment,

Table 1: Acquired behavioral character after learning in the case of two-agent. 5 simulation runs are done for each case. (ind: all the agents acquire "individuality", soc: all the agents acquire "sociality", mix: some agents acquire "individuality" and the others acquire "sociality", and fail: the task fails for some combinations of the agents)

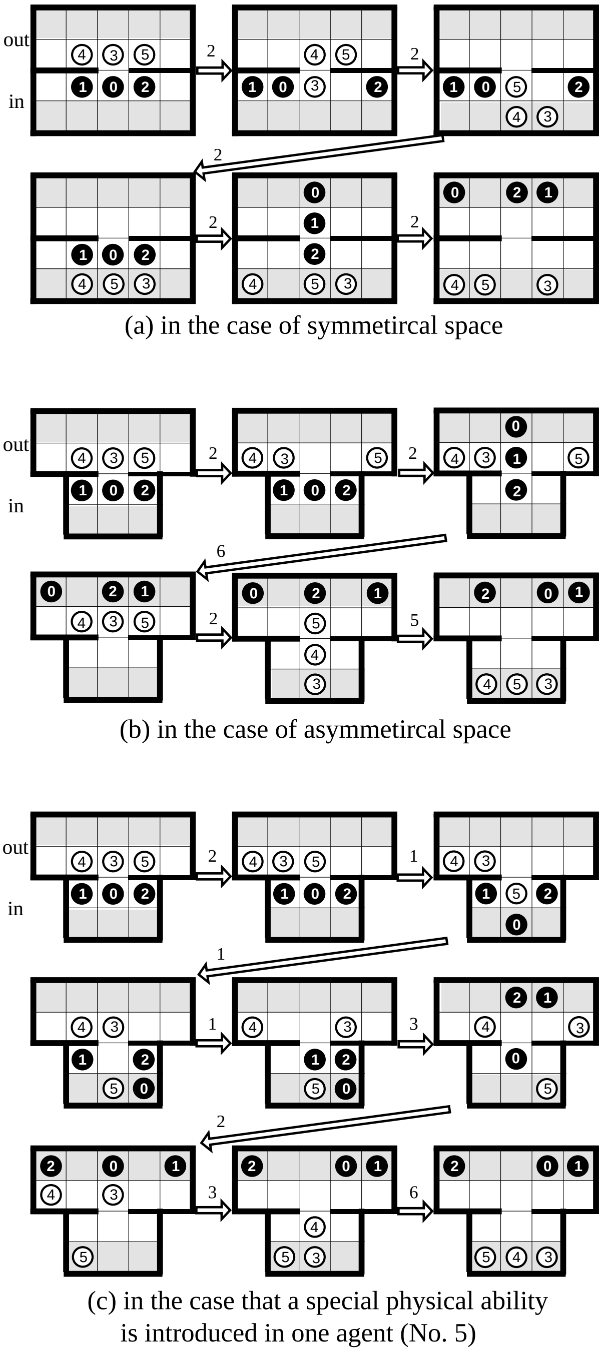
	number of whole agents	identify	non-identify
symmetrical inputs	2	ind 5	
	3	ind 5	mix 3 fail 2
	4	ind 5	fail 5
asymmetrical inputs	2	ind 3 soc 2	
	3	ind 1 mix 4	soc 5
	4	mix 5	soc 5

the agents are assumed that they cannot know whether they are getting on the train or getting off the train.

Table 1 shows the number of the acquired characters after 5 simulation runs for each case. It can be seen that in the case of non-identification and asymmetrical environment, "sociality" is apt to emerge, while "individuality" is apt to emerge in the case of identification and symmetrical environment. In the case of non-identification and symmetrical environment, some agents fail to arrive at its goal. The reason is that if both selfish and altruistic agents exist, the third agent has no effective solutions. Further, if they made a rule that the agent who moves first gives its way, the system becomes unstable. It is always profitable for an agent not to obey the rule and to stay at the same location, because the opponent agent always gives its way at the next time step.

Next, it is examined in six-agent simulations whether the rule emerges that the passengers getting off the train have a priority to go. The experimental setting is almost the same as the previous simulation, but the number of trials is 100000. At first, the case of a symmetrical space is examined as Fig. 4 (a). Since each agent is assumed to be able to know its role, the inputs are asymmetric. In this case, the rule that the agent who moves first gives its way, emerged in all the simulation runs in 100 runs. However, in the case of asymmetrical space, the expected rule emerged as shown in Fig. 4 (b).

Finally one of the agents is assumed to have a special physical ability, i.e., "strong power". When the opponent exists in front of the agent, no one exists on



(a) in the case of symmetircal space

(b) in the case of asymmetircal space

(c) in the case that a special physical ability is introduced in one agent (No. 5)

Figure 4: Results of the simulations in the case of six agents. It is examined whether the rule emerges that the passengers getting off a train have a priority. The number on the arrows indicates the number of cycles for the transitions. The number in each circle indicates the order of action in one cycle.

the box beyond the front opponent, and no wall exists at the back of the front agent, it can move forward and makes the front opponent go back. Fig. 4 (c) shows the result of the simulation. It can be seen that the agent with "strong power" got on the train at first.

4 Discussion

The authors expect that this model can explain "individuality" and "sociality" in our human society on the micro-level, i.e., individual learning level. We think that the evolution also makes an important role for the emergence. However, the both behavioral characters are defined by observing their actions, and we don't have enough strategy of action when we were born. Then we have dealt with the evolution as one of the factors which influences the learning, and represented it by "physical and mental character".

5 Conclusion

"Individuality" and "sociality" has been introduced, and the model has been proposed in which the both behavioral characters can be obtained by reinforcement learning. It is examined that the model is feasible in some simulations.

Acknowledgments

A part of this research was supported by the Scientific Research Foundation of the Ministry of Education, Science, Sports and Culture of Japan (#10450165) and by Research for the Future Program by The Japan Society for the Promotion of Science (JSPS-RFTF96I00105).

References

[1] Maynard Smith J (1982), Evolution and the Theory of Games, Cambridge Univ. Press.

[2] Zamora J, Millan JR, Murciano A (1998), Learning and Stabilization of Altruistic Behaviors in Multi-Agent Systems by Reciprocity, *Biological Cybernetics*, Vol. 78, No. 3, pp.197-205, 1998

[3] Ota J, Yokogawa Y, Arai T (1996), Study on Asymptotic Strategy-Making in Multiple Autonomous Robot System, *Journal of Robotic Society Japan*, Vol. 14, No. 3, pp. 379-385 (in Japanese)

[4] Smith A (1776), An Inquiry into the Nature and Caufs of the Wealth of Nations

[5] Shibata K, Ito K (1999), Emergence of Communication for Negotiation by a Recurrent Neural Network, *Proc. of ISADS(Int'l Sympo. on Autonomous Decentralized Sys.)'99*, pp.294-301