Design of a Strategy Learning Model for Robot Brain and LSI Implementation of the Model

Masahiro Ono (Graduate School of Advanced Sciences of Matter, D1),
Mamoru Sasaki (Associate Prof., Graduate School of Advanced Sciences of Matter),
Atsushi Iwata (Prof., Graduate School of Advanced Sciences of Matter),

1. Research Target

Recently, the progress of robots is remarkable. It is expected that the role which robots will bear in society from now on is expanded more, and the demand of robots will increase. The research of the robots which interact with persons such as welfare robots and communication robots, is mentioned as one of the fields from which the success is especially expected. The environments and the situations (person's characteristics) which the robots recognize tend to change very quickly because of the persons' learning. Under that condition, it is necessary for the robots to have more robust brain (We call it "Robot Brain") in order to accomplish the tasks. It can recognize person's characteristics and choose the most suitable action according to memorized experiences obtained by learning. We have two research targets. The first target is to propose a model for the Robot Brain which accomplishes the above process and to evaluate the effectiveness quantitatively using a numerical simulation. The second target is to realize the model by the custom LSI which is very compact and has low power consumptions.

2. Research Results

The proposed model is shown in Fig.1. It consists of the three sections: 1. selection of a strategy, 2. construction of a new strategy by learning, 3. addition of it or elimination of some strategies. At first, the opponent's behavior is observed in order to estimate his characteristics. Secondly, using the estimation result, the optimum strategy for the current opponent is selected from many candidates stored in the memory. This is carried out in the first section. The many strategies have been obtained by past learning. The selected strategy is copied into working memory. The robot acts based on the copied one. Thirdly, actions are decided according to the strategy in the working memory, and simultaneously the strategy is tuned in order to obtain better one. This is carried out in the second section. Finally, the tuned one is estimated with a criterion. It is added in the memory, if the estimation judges that it is needed. And, some needless strategies are eliminated to restrain memory overflow. This is carried out in the third section. We have devised only the first and second sections, and evaluated their effectiveness.

We briefly explain the algorism of the first section and the second section. We start the explanation of the second section, to easily understand the model. The algorism in this section is Reinforcement Learning (RL) [1]. We define a strategy. It consists of a set of ``situation" and ``action" pairs. By another expression, it consists of a set of if-then rules. Here, Q-function, $Q_x(\mathbf{s}_i, \mathbf{a}_j)$ has very important role. x is an index expressing each strategy. It expresses the quality of an action $\mathbf{a}_j$ in a situation $\mathbf{s}_i$ (larger Q_x means that an action is better). In the situation $\mathbf{s}_i$, an action is decided according to $Q_x(\mathbf{s}_i, \mathbf{a}_j)$. The robot is given a positive reward (or a negative reward) in the case of achieving its goal (or not), after the robot's taking the action. $Q_x(\mathbf{s}_i, \mathbf{a}_j)$ is updated by the reward. The strategy learning means updating $Q_x(\mathbf{s}_i, \mathbf{a}_j)$ by using the rule. Next, we explain the first section in Fig1. Q-function can be also employed in order to evaluate the opponent's characteristics. So, we prepare $Q_{obs}(\mathbf{s}_i, \mathbf{a}_j)$ which is Q-function, for evaluating the opponent's characteristics. As well as Q-function in the second section, Q_{obs} is also updated during a sequence. The strategy with the closest Q_x to Q_{obs} is selected.

In order to confirm the ability of the model, we applied it to "air hockey game" as the example of the task which needs persons. In this experiment, an opponent is the same program with a simple strategy. The strategy is fixed and is not tuned by learning. We had many experiments by varying the opponent's characteristics. Fig.2 shows one of the experiments. The opponent's characteristic has been varied by stick position as shown in Fig2. Fig3 shows the result of the experiment. The stick position of the opponent changes from A to E (see Fig.2) every set. It is one set in 20 points. The number of the stored strategies is four. Before the experiment, the four strategies have been obtained by learning in cases that the stick positions of opponent are fixed at A, B, C and D, respectively. The strategies are called S_A, S_B, S_C and S_D. Fig.3 shows that proper strategy except B was quickly selected. Here, we consider the case of the position B. It is the second set. Fig.4 shows that the robot overwhelmingly won even by using the Sc. So, the model has also selected the proper strategy in this case. Next, let us consider the case of the position E. Note that no strategy has been previously prepared for the position E. In order to confirm the selection of Sc for the position E is appropriate, we had another experiment. In the experiment, we applied the strategies from S_A to S_D to the position E. Fig.5 shows the result: The Sc is the best strategy because the robot overwhelmingly won as comparison with the others. From the above discussions, the model's effectiveness has been confirmed apparent.

3. Relation between COE program and this research result

A target of COE program is "Realization of integrated systems with high-level recognition and learning capabilities by innovative circuits and architectures".

Therefore, we can provide COE program with this research result, “a strategy learning model” as the learning function of the integrated systems by integration technologies.

We will apply this model to the 3-demension integration system which is the goal of COE program. At first, we will realize Sec.1, Sec.2, and Sec.3 in Fig.1 on a custom LSI chip. Secondary, several LSI chips is used as the memory of strategies in Fig.1. The function of this model will be realized by communicating between the chips with wireless integration technology. We will be able to contribute to the accomplishment of COE program target by embedding high-level learning capabilities into 3-demension integration system.

4. Conclusion and Schedule

To realize Robot brain, which can execute tasks interacting with persons, we have proposed a model with RL. The model can select the most suitable strategy so quickly and construct a new strategy by learning. It was confirmed by the simulation experiment. In present study, we use the virtual opponent. As the next step, we’ll have the experiment using a person as the opponent, and we’ll design the circuit which has functions of this model.

References

[1] Richard S.Sutton and Andrew G. Barto, Reinforcement Learning, MIT Press, 1998.

3. Published Papers and Patents

Proceedings

1. M. Ono, M. Shiozaki, M. Sasaki and A. Iwata, "A Strategy Learning Model for Robot Brain, " 2nd Hiroshima International Workshop on Nanoelectronics for Tera-Bit Information Processing (2004), pp. 114-116

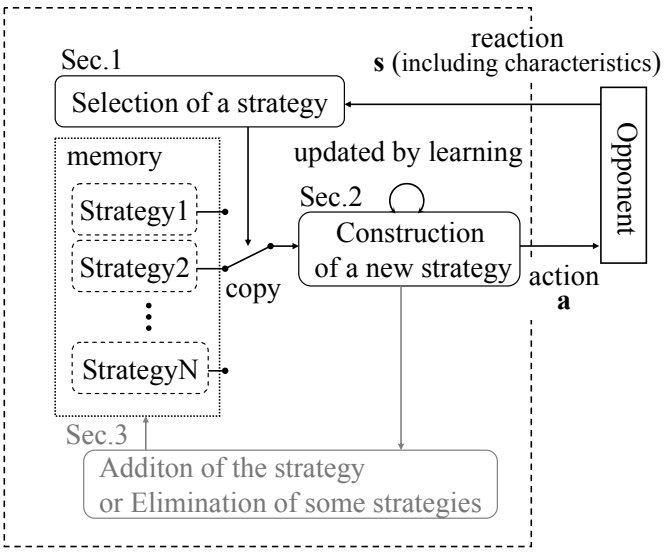

Fig. 1: A proposed model for Robot brain

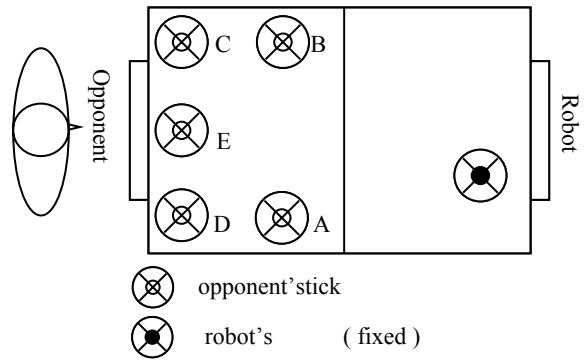

Fig. 2: An example of the opponent’s characteristics

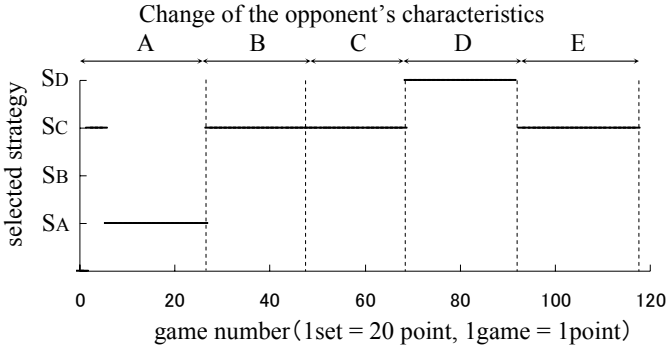

Fig. 3: The result of the experiment where the opponent’s stick position changes A from E by one set(= twenty points)

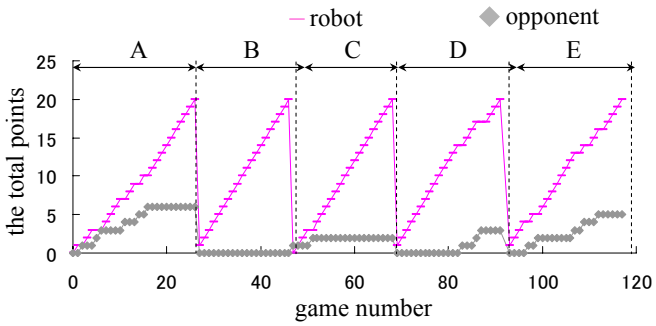

Fig. 4: The total points that the robot and the opponent got

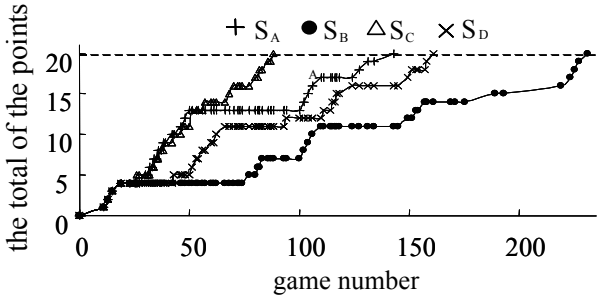

Fig. 5: The total points given to the robot by using each strategy during position E