A Strategy Learning Model for Robot Brain

Masahiro Ono, Mamoru Sasaki and Atsushi Iwata

Graduate school of Advanced Sciences of Matter, Hiroshima University

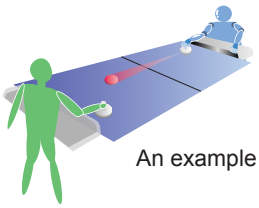
Introduction

Robots interacting with human being
are expected. (welfare, entertainment robots etc...)

In order to realize such robots,
"Robot Brain", that can recognize human characteristics and
select the best strategy and situations is needed.

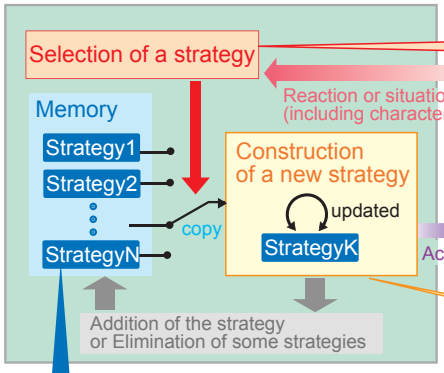
Purpose of Research

Proposition and evaluation of a model (Strategy Learning Model)
which flexibly adapts to the situations by learning.



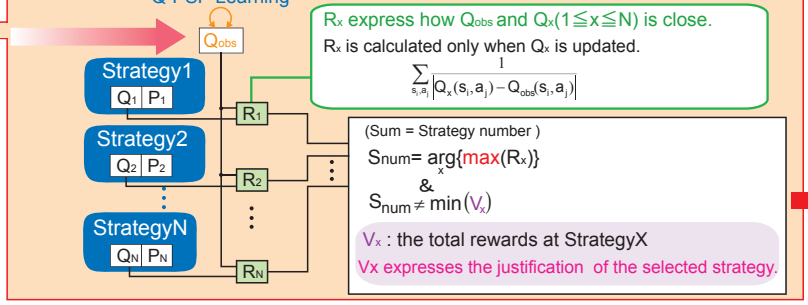
An example of tasks : air hockey game

A Strategy Leaning Model



◎ The strategy with the closest Q to Q_{obs} is selected.

(Q_{obs} is Q-function which estimates opponent's characteristics.)



Strategy

Q(s, a) : Quality function (strongly affected by opponent's characteristics)

s : situation vector
(ex. where the opponent returned a puck?)

a : action vector
(ex. where a robot will return a puck ?)

Q(s,a) : larger
at the situation s
the action a : better

The way of updating Q

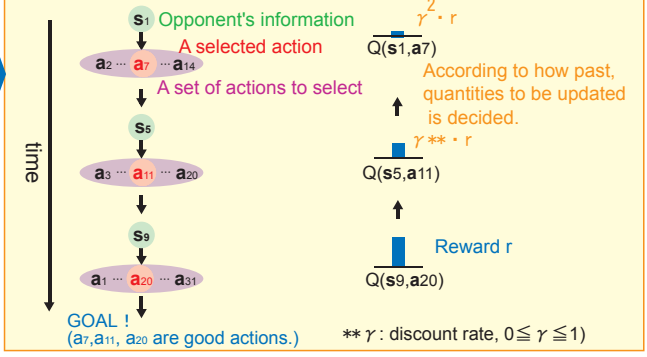
P(Q) : Probability about selection of an action

$$P(\mathbf{s}_m, \mathbf{a}_k) = \frac{\exp(Q(\mathbf{s}_m, \mathbf{a}_k)/T)}{\sum_{\mathbf{a}'} \exp(Q(\mathbf{s}_m, \mathbf{a}')/T)}$$

T : thermal parameter
small → Selection of an action strongly depends on Q.
large → An action is randomly selected.

Q-PSP Learning (one of Reinforcement Learning (RL))

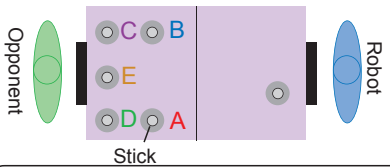
Good action → ▪ Q : up Bad → Q : down



Numerical Simulation

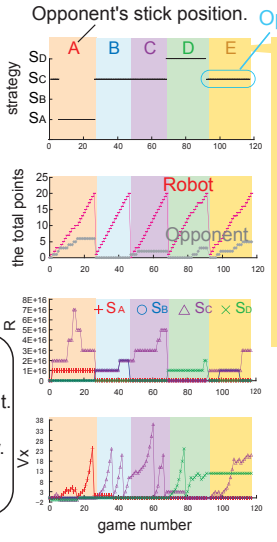
In the games, the opponent changes its stick position to defense from A to E by 1 set (20 points). (Robot's position is fixed.)

Opponent's characteristics = stick positions



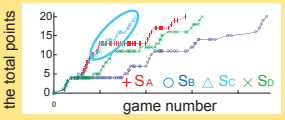
Assumption :

- The robot has already had the strategies for A, B, C and D by learning before the experiment. (called S_A, S_B, S_C and S_D, respectively)
- The opponent has only a simple fixed strategy.
- The performance of the robot is the same as that of the opponent.



Evaluation the selection during E

The total points given to the robot by using each strategy during E.



SC turned out the best strategy.

Selection is correct.

Conclusions

We confirmed the effectiveness of this model with RL by the simulation experiment.

- Adaptation to the change of 5 patterns
- Selection and construction of the optimum strategy

The future schedule

- Proposition of a method for addition of the strategy or elimination of some strategies
- Experiment with a person as the opponent