A Modular Learning Model with Addition and Integration of Modules

Masahiro Ono, Mamoru Sasaki and Atsushi Iwata

Graduate School of Advanced Sciences of Matter, Hiroshima University,
1-3-1 Kagamiyama, Higashi-Hiroshima-shi 739-8527, Japan
Phone and Fax: +81-824-22-7358, Email: {masa, sasaki, iwa}@dsl.hiroshima-u.ac.jp

1. Introduction

Recently, the progress of robots is remarkable in the field of high speed operation, humanoid, imitation of behavior and entertainment. Nevertheless, few autonomous robots having behavior-learning capability are developed. The main reason is that learning is insecure to environmental changes and the learning speed is slow.

Jordan et al. [1] proposed "modular learning model" for solving this problem. Doya et al. [2] proposed a module based learning model "MMRL", which improves Jordan's model in flexibility to environmental changes. It has many modules, each of which is composed of a situation predictor and a learning controller. The output of each module is an action. The situation predictor estimates the current environment dynamics and outputs the expected situation of next time. The learning controller selects and the suitable action and learns a value function for the policy of actions, based on the "responsibility signal", which stands for accuracy of the estimation of the predictor. The output of the model is calculated by the summation of the output of each module weighted by the responsibility signal.

However, MMRL cannot add and integrate (or delete) some modules even if the number of the kinds of environments increases or decreases. Thus, it is limited with adaptation to complex dynamical environments. Besides, the model has a problem about power dissipation that is important for autonomous robot applications because multiple modules can be active simultaneously by selecting multiple modules. Then, we propose the advanced learning model, which selects only one module and has functions of addition and integration of modules. We show the effectiveness of the proposed model in simulations by comparing MMRL.

2. Learning Model

We propose a new modular learning model in Figure 1. It consists of three elements: expert modules, an environmental observer (EO) and a policy observer (PO). We define the dynamic environment as a set of multiple stationary tasks. The expert module has an environmental model (EM) of the stationary task and a policy for the task. A policy is set of actions to a task. EO is defined as a set of probabilities $p_i(s(t+1) \mid s(t), a(t))$ (i: index of module, s(t+1), s(t): situation at the next time and the present time, respectively, a(t): action at the present time). The policy is defined as $a(t) = f_i(s(t))$. The structure of the environmental observer and the policy observer is also same as the EM and the policy (expressed as p_o and f_o), respectively. The value of the probability of EO and EM is either 0 or 1.

Next, we explain the processing procedure.

0. The current situation (s(t)) is observed.
1. The model ($p_o(s(t) \mid s(t-1), a(t-1)) = 1$) in the current environment is memorized by the EO.
2. Only the effective policy ($f_o(s(t-1))$) with which robots can achieve to the goal in the current environment is memorized by the PO.
3. The environmental change (change of the task) is detected by the EO and environmental change signal is generated. When the environment (subtask) changes, p_o changes from 1 to 0 or from 0 to 1. By using those information, the environmental change is found.

3-1. If the environmental change is not detected, the most suitable module with the smallest error between the EM and the EO for the current environment is selected.

3-2. If the environmental change is detected, it is decided whether the EO and the PO is copied as the EM and the policy of the new module (Addition) or unified with one of the existing modules (Integration). Then, EO and PO are reset.

The criterion for addition and integration is as follows.

- Addition: $Po \neq Pi$ for all the existing modules
- Integration: $Po=Pi$ (if the criterion is satisfied, the integration processing is performed: $Po \cup Pi$, $fo \cup fi$)

3. Experiment

3.1. Task Setting

We choose maze problems that are often used for the test of behavior-learning. We prepare the six mazes shown in Figure 3. We change the kind of the maze in constant time in order to add dynamic elements. The degree of resemblance between the structures of the mazes is shown in Figure 4. The robot selects one of four possible actions: {north, east, south, west}. We define selecting an action as "one step". A trial is terminated and the robot is given a positive reward, when the robot reaches to the goal. Then, the new trial begins at the start point.

3.2. Experimental Results

Figure 5 shows the difference of the performance between the proposed model and MMRL in case that two mazes: {(a): A3-B3, (b):A1, A2} are switched in 100 steps. The vertical axis is the number of steps required in a trial. The horizontal axis is the number of trials. While, in case of MMRL, the number of steps is larger as the degree of resemblance is larger, the number of steps does not change much in case of the proposed model (the smallest number of steps are realized within 200 trials.).

Figure 6 shows the difference of the total rewards between the proposed model and MMRL when we change six mazes are switched in T steps. T is a parameter which means the interval from a maze to another maze. In view of the total rewards, it is confirmed that the proposed model learns the expert modules about 5 - 8 times faster than MMRL.

4. Conclusion

We have proposed a new modular learning model. The model can not only adapt to environmental changes faster than MMRL by the simulation experiments, but also change the number of modules flexibly. Therefore, it is expected that the power dissipation is less than conventional models.

References

[1] R.A. Jacobs, M.I. Jordan, S.J. Nowlan, and G.E. Hinton, “Adaptive mixture of experts”, Neural Computation, vol.3, pp.79-87, 1991

[2] Kenji Doya, Kazuyuki Samejima, Ken-ichi Katagiri and Mitsuo Kawato, “Mutiple Model-Based Reinforcement Learning”, Neural Computation, vol.14, pp1347-1369, 2002.

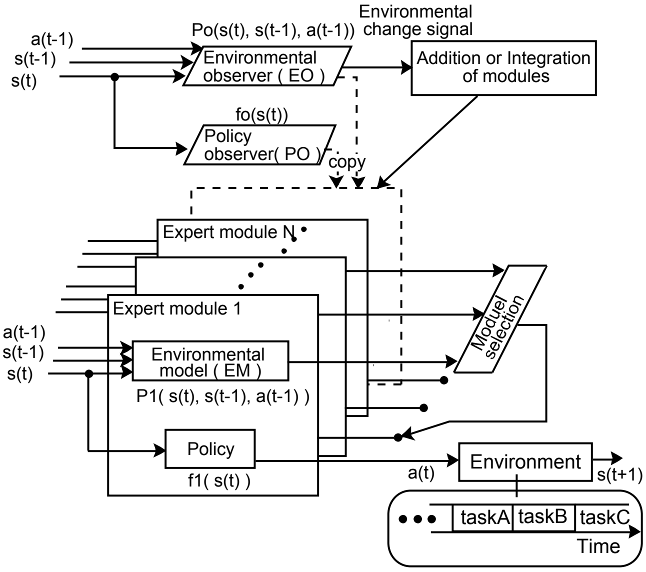
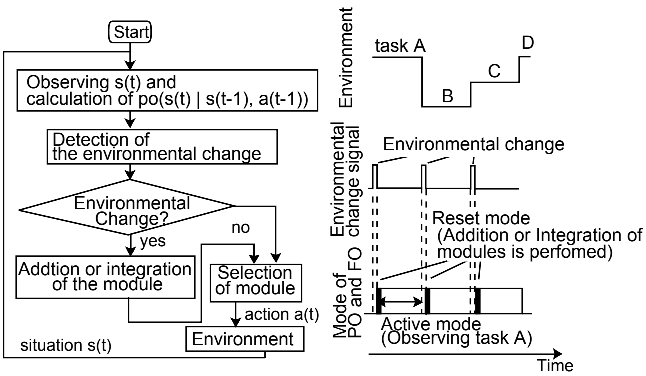


Figure 1: Schematic diagram of the proposed model.



(a) Processing procedure of the proposed model

(b) Timing chart of processing

Figure 2: Processing procedure.

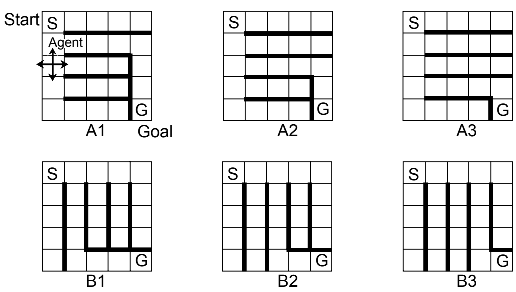


Figure 3: Maze problem.

(%)

	A1	A2	A3
B1	25	21	17
B2	21	17	13
B3	17	13	8

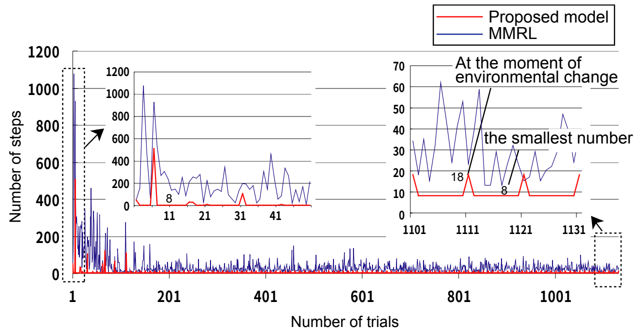
(a) between different types

(%)

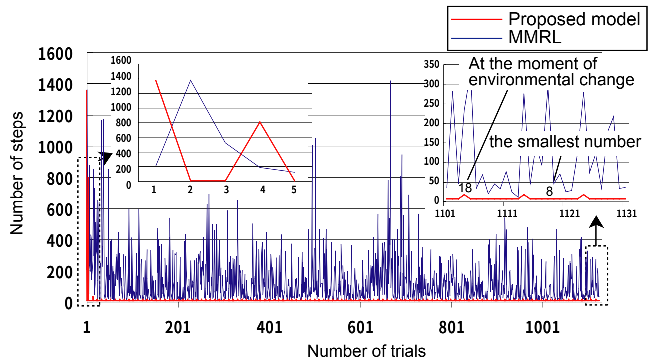
	A1(B1)	A2(B2)	A3(B3)
A1(B1)		64	56
A2(B2)	64		64
A3(B3)	56	64	

(b) between similar types

Figure 4: Degree of resemblance between mazes.



(a) : A3, B3(the lowest degree of resemblance)



(b) : A3, B3(the highest degree of resemblance)

Figure 5: Experimental result of change of the number of steps (2 mazes, interval T= 100, (a): A3, B3, (b): A1, A2).

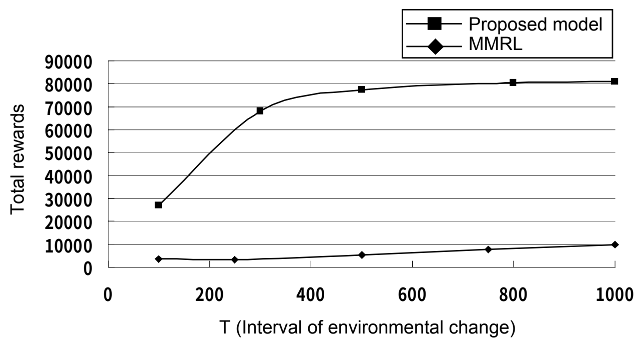


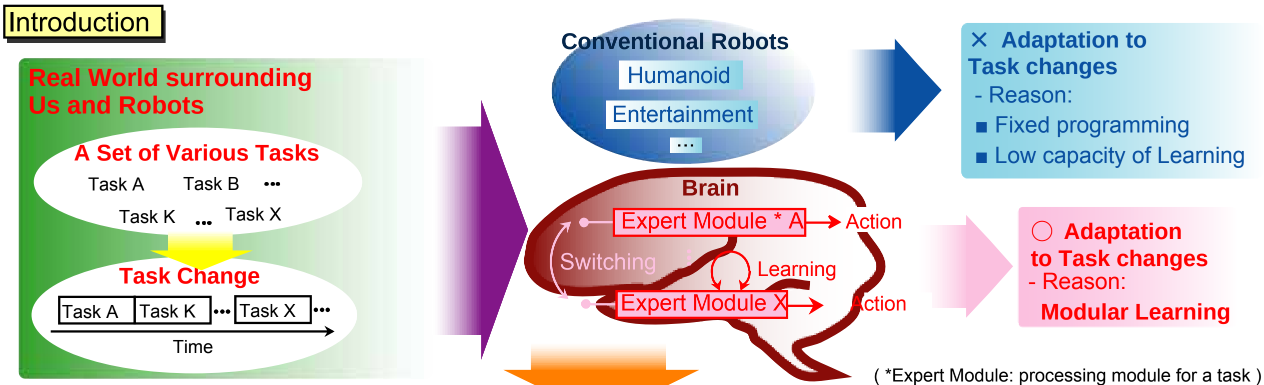
Figure 6: Experimental result of total rewards for interval T (6 mazes, during 10^5 steps).

A Robust Modular Learning Model with Addition and Integration of Modules

Masahiro Ono, Mamoru Sasaki and Atsushi Iwata

Graduate school of Advanced Sciences of Matter, Hiroshima University

Introduction

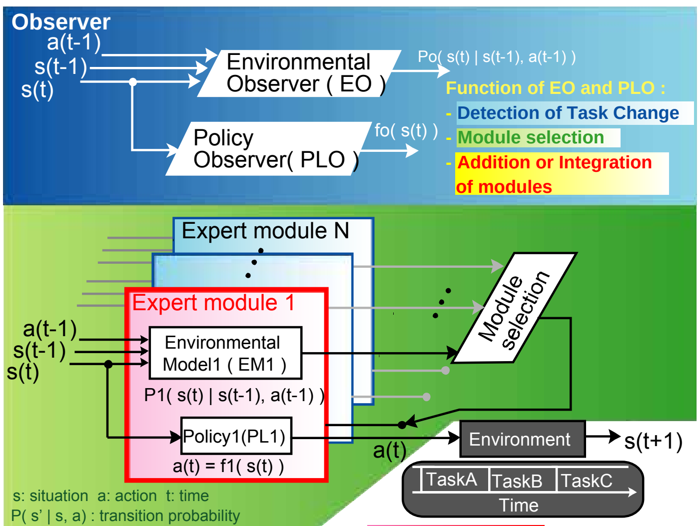


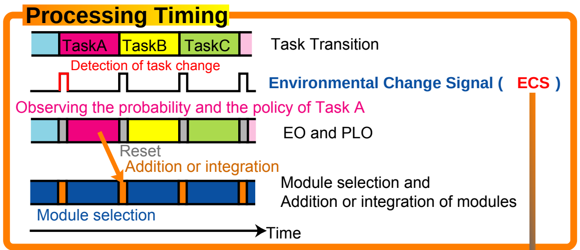
(*Expert Module: processing module for a task)

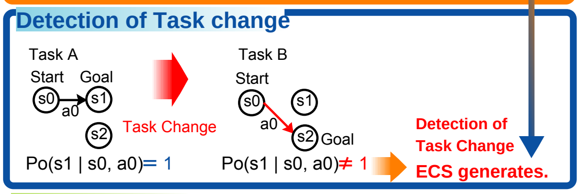
Conventional Modular Learning Model
- MMRL
(**M**ulti **M**odel Based **R**einforcement **L**earning**)
× Addition and integration of expert modules
(**: Doya et al. 2002)

Target of Research:
- **Proposition of a New Modular Learning Model**
with Addition and Integration of Modules
- Implementation of Modular Learning Model with LSI

Learning Model





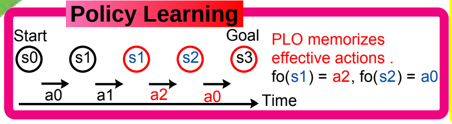


Module selection

Module number = arg max (Ri) (i = 0 ... N)

$R_i = \prod_{k=0}^{t\_start} P_i(t-k)$

(t_start: the time when ECS generates)



Addition or Integration of Modules

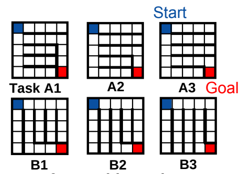
Condition of addition or integration: for all s, a, i;
- Addition: Po ≠ Pi
- Integration: (Po ⊆ Pi or Po ⊇ Pi) and (fo ⊇ fi)

Processing:
- Addition: P new ← Po fnew ← fo (Data Copy)
- Integration: Pi ← Po ∪ Pi, fi ← fo ∪ fi

Experiment

■ **Maze problem with Task changes**

- The robot selects four possible actions: {north, south, east, west}
- If the robot reaches to Goal, the new trial begins at Start

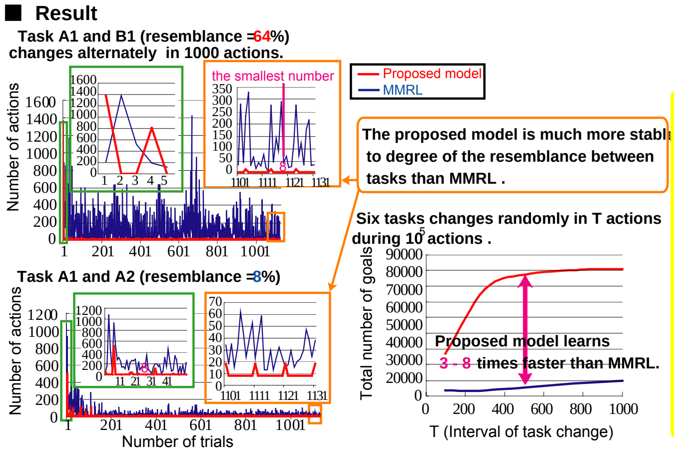


Degree of resemblance between tasks (%)

	A1	A2	A3
B1	25	21	17
B2	21	17	13
B3	17	13	8 Min

	A1(B1)	A2(B2)	A3(B3)
A1(B1)		64 Max	56
A2(B2)	64		64
A3(B3)	56	64	

■ **Result**



Conclusions

We proposed a new modular learning model with addition and integration of modules.

We confirmed the effectiveness of the model by maze problem with task changes.
- **Robust learning to the resemblance among tasks**
- **3 - 8 times faster learning than MMRL.**