Fast and Stable Learning in Direct-Vision-Based Reinforcement Learning

Katsunari Shibata*, Masanori Sugisaka* & Koji Ito**

**Dept. of Electrical & Electronics Engineering, Oita University, 700 Dannoharu, Oita 870-1192, Japan.

***Dept. of Computational Intelligence & Systems Science, Tokyo Institute of Technology

4259 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan. shibata@cc.oita-u.ac.jp

Abstract

Direct-Vision-Based Reinforcement Learning has been proposed not only for the motion planning but for the learning of the whole process from sensors to motors in robots, including recognition, attention and so on. In this learning, raw visual sensory signals are put into a layered neural network directly, and the network is trained by the training signals generated based on reinforcement learning. On the other hand, it has been pointed out that the combination of neural network and TD-type reinforcement learning sometimes leads to instability of learning.

In this paper, it is shown that each visual sensory cell makes a role of localization of our continuous 3-dimensional space and it helps the learning to be fast and stable. Further by processing the localized input signals in the layered neural network, a global representation is reconstructed adaptively in the hidden layer through learning as shown in the previous papers.

1 Introduction

Reinforcement learning has been focused recently for its autonomous and adaptive learning ability. However, in general, only the mapping from the state space to the action space is formed by reinforcement learning. The state space is generated from sensory signals through some pre-processings, and is usually designed by a designer. Some of the authors have proposed Direct-Vision-Based Reinforcement Leaning [1]. In this learning, raw visual sensory signals are put into a neural network directly without any pre-processings and the outputs of the network are utilized as the motion commands. By this, the continuous and adaptive state space is formed in the hidden layer through learning. Furthermore, the reinforcement learning is extended to the total learning for the whole process from sensors to motors, including recognition, memory, sensory integration, and so on[2] [3].

Layered neural networks have been already utilized widely in TD(Temporal Difference)-type reinforcement learning [4][5][6]. However, Boyan et al. has been pointed out that the combination of TD-type reinforcement learning and layered neural network is not robust and may produce an entirely wrong policy[7]. The problem occurs when the discontinuous motion or value function is required such as "2-D puddle world" and "car-on-the-hill" task in [7]. The reason might be that the sigmoid function in the neural network is a smooth and monotonically incremental function, in other words, a semi-linear function and has weak non-linearity.

For this problem, it has been shown that localization of the continuous input space such as CMAC, k-nearest-neightbor and RBF (Radial Basis Function) is effective [8][9][10]. However, some of the authors have pointed out that since they are close to the table-look-up approach and have no hidden units, they cannot reconstruct the adaptive continuous state space from localized signals[11]. This means that whenever a robot with RBF or CMAC learns a new task, it cannot utilize the knowledge obtained from the learning of the previous tasks and has to learn from scratch. Therefore it is not suitable as a brain of intelligent robots.

Some of the authors have proposed Gauss-Sigmoid neural network[11]. In this network, continuous input signals are put into a RBF network at first, and then the outputs of the RBF network are put into a sigmoid-based neural network. This network has the both advantages of the RBF-based network and the sigmoid-based neural network. In this paper, it is shown that each visual sensory cell makes a role of localization of our 3-dimensional space like the RBF unit in Gauss-Sigmoid neural network, and it helps the learning to be fast and stable.

2 Localization and Visual Sensor

2.1 Localization

Localization means to represent the spatial information using multiple signals each of which mainly shows the information of a local area in the whole continuous space. All of table-look-up, RBF, and CMAC are the localization-based function approximators.

By using local signals, the learning effects only at the local area, and strong non-linear functions from the continuous input space can be approximated easily as the sum of the weighted local signals. So as to realize better approximation, the continuous space has to be localized more finely and many localized signals are required. However, there is a dilemma that the

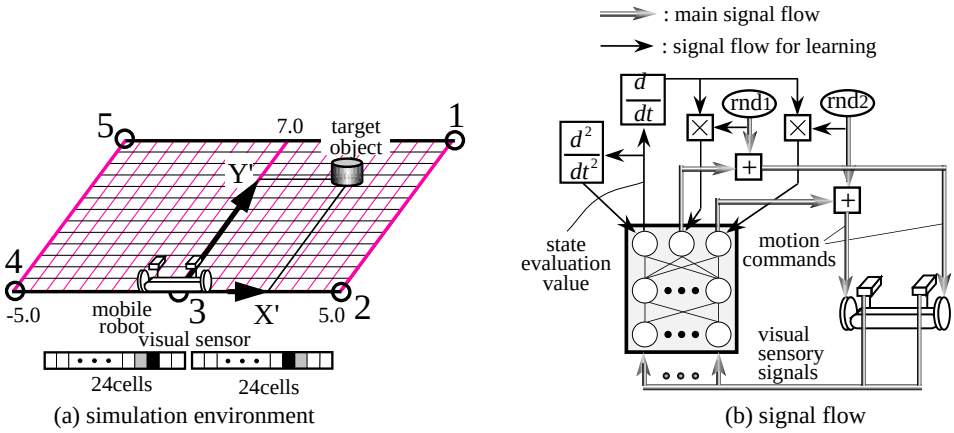

Figure 1: (a)Environment of the basic task. A mobile robot with two wheels has two visual sensors and it can obtain a reward when it arrives at the target. (b)Signal flow of the learning. Visual sensory signals are put into a layered neural network directly, and the neural network is trained by the reinforcement learning based on Temporal Smoothing Learning.

generalization ability becomes less and less when the number of localized signals more. That is the same as table-look-up approach.

2.2 Visual Sensor and Receptive Field

A visual sensor usually consists of many visual cells. Each visual cell has a local receptive field, and catches only a part of the whole visual field. It can be said that the visual sensor makes a role of localization of the continuous 3-dimensional space where we live.

So it can be thought that when the visual sensory signals are put into a neural network, the learning is faster than when the continuous spatial signals, such as object location in the sensor, is put into the neural network. Further, in order to generate the continuous spatial signals from the localized signals, pre-processing is necessary. The pre-processing is usually designed by a designer, and it is difficult to modify the pre-processing flexibly through the learning. It has been already shown that the localized visual signals can be integrated in the hidden layer of the neural network very adaptively through reinforcement learning according to the given task and the motion character of the robot[2].

3 Comparison of Learning Speed and Stability

3.1 Task

Here let's consider the task that a mobile robot with two wheels and two visual sensors obtains an target object as shown in Fig.1(a). Each visual sensor has 24 visual cells arranged in a row. The receptive field of each sensory cell spread out radially, and the sensor has a total of 180 degree of visual field. The robot obtains a reward only when it reaches the target object. Concretely, when the center of the target goes through the robot, the state evaluation output is trained to be 0.4. When the robot misses the target, i.e., the target disappears out of the visual field, it is trained to be -0.4 as a penalty. The diameter of the target is 1.0 and length of the robot is 2.0. Figure 1 (b) shows the signal flow of this simulation. Before learning, the input-hidden connection weights are small random numbers, and all the hidden-output connection weights are 0.0. Accordingly all the outputs of the neural network are

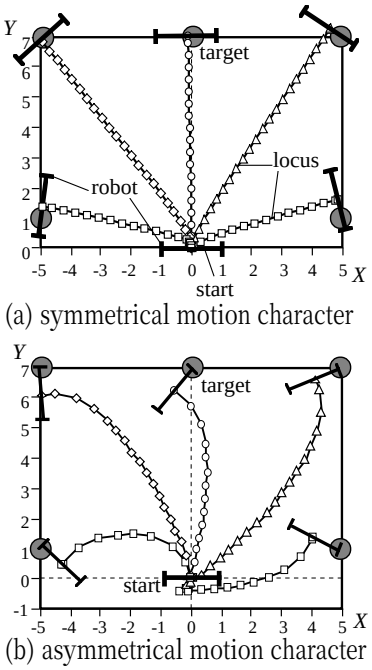

Figure 2: The robot loci after learning.

0.0 before learning. The number of outputs for motion signals is two, and the robot rotates its wheels according to the outputs. In the early stage, since the robot moves only by random numbers, the target is located within the range that is close to the robot. According to the progress of the learning, the range of the target location becomes wider gradually until $-5 \leq x \leq 5, 0 \leq y \leq 7$. After the robot reaches or misses the target, the target is located at another place chosen randomly. One trial indicates a sequence from the initial state to the target state. Fig. 2(a) shows the robot loci after learning. It is seen that the robot rotates until it catches the target around the center of the visual sensor and then it moves forward.

3.2 Integration of Visual Sensory Signals

Here, at first, it is introduced that the neural network has an ability to represent the spatial information in the hidden layer by integrating the localized visual signals[1]. When the 5-layered sand-glass neural network with two hidden units in the middle layer as shown in Fig. 3 is utilized in the reinforcement learning, the representation of the target location in the middle layer is as shown in Fig. 4(b). It can be noticed that the two neurons represents the target location that is the global and continuous information. The representation is not uniform, and is magnified around the boundary area whether the robot gets the target or miss the target and whether the robot has to change its motion from rotation to forward movement.

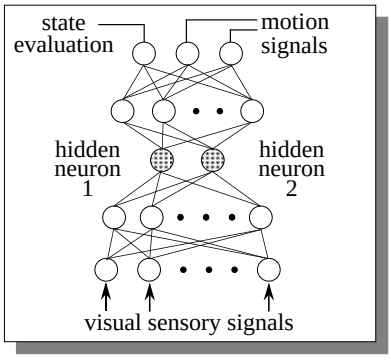

Figure 3: Five-layered sand-glass neural network. It is also used to examine the representation of hidden neurons in the next chapter.

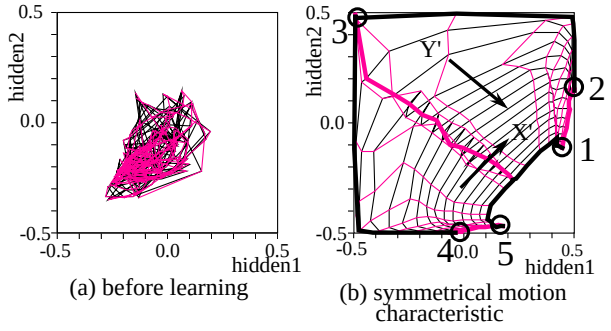

Figure 4: Representation of the target location in hidden neurons' space. Circles in (b) are typical target locations which correspond to the circles in Fig. 1 (a).

Next, the hidden representation was also examined for the case of the regular 3-layered neural network. After the reinforcement learning, one output neuron was added to the neural network with 0.0 connection weight and the output is trained only at six locations by the training signals as black and white circles in Fig. 5. Then the output distribution for all the target locations is as shown in Fig. 5(b). Fig. 5(a) shows the distribution when the reinforcement learning was not applied. The output distribution in the case of no reinforcement learning is spread radially. The distribution after reinforcement learning is divided into two regions, left and right. From this result, the hidden representation becomes to represent the global information that is required in the learning.

3.3 Comparison

Here the comparison of the learning performance is made between two input forms, i.e., visual sensory signals input and two dimensional relative location input from the robot to the target. Here three combinations of the input forms and network structures are employed as (1)visual sensory inputs & three-layered network, (2)visual sensory inputs & five-layered sand-glass network, (3)relative location inputs & three-layered neural network. The sand-glass network is the same as the previous subsection that is shown in Fig.3. The number of the hidden neurons are 20 for three-layered network, and 20-2-20 for the sand-glass network. The sand-glass network is employed to map the input information into two dimensional space that is the same as the relative distance inputs. The given task is the target reaching task as mentioned before, but the more difficult task is employed by introducing asymmetry of the wheel motion as follows. The

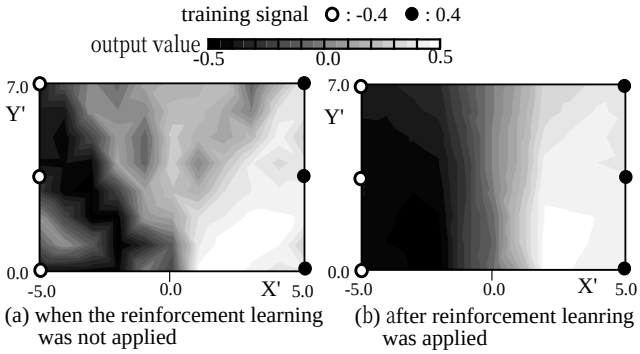

Figure 5: Comparison of the output distribution after the supervised learning in which the six target locations are used as training data s et.

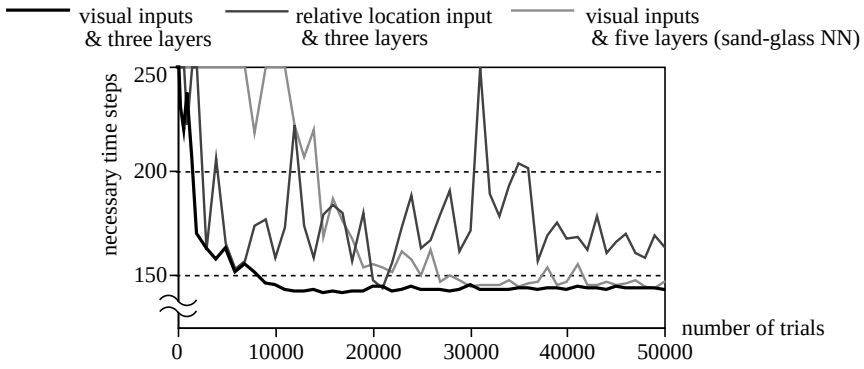

Figure 6: Comparison of the learning performance with respect to the form of the input.

right wheel is rotated according to three times as the corresponding motion signal, while the left wheel is rotated according to the motion signal itself. If the robot rotates both wheels in the same way, the motion direction gradually changes to the left. The robot loci after learning in the case of regular 3-layered neural network are shown in Fig. 2(b).

Figure 6 shows the learning curves for the three cases. The vertical axis shows the average number of steps until reaching the target state over the various target locations. Each trial is stopped at 250 time step even if the robot cannot reach the target. It is known that the learning is the fastest in the combination of visual sensory signals and 3-layered neural network. On the other hand, in the case of the relative location input, the learning is slightly slow and very unstable. This result is similar to that of Boyan et al.[7]. The learning in the case of sand-glass network is very slow, but more stable than the relative location input case.

4 Conclusion

It was shown that the visual sensor makes a role of localizing the spatial information. The difference of the learning speed and stability with respect to the input form was examined between the global spatial signals and the visual sensory signals. Then it was known that the learning is fast and stable in the case of visual sensory signal inputs.

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] Shibata, K. & Okabe, Y. (1997) "Reinforcement Learning When the Visual Signals are Directly Given as Inputs", *Proc. of ICNN'97*, **3**, pp.1716-1720

[2] Shibata, K., Ito, K. & Okabe, Y. (1998) "Direct-Vision-Based Reinforcement Learning in "Going to an Target" Task with an Obstacle and with a Variety of Target Sizes," *Proc. of NEURAP'98*, pp. 95–102

[3] Shibata, K., Sugisaka, M. & Ito, K. (2000) "Hand Reaching Movement Acquired through Reinforcement Learning" , *Proc. of 2000 KACC*, 90rd (CD-ROM)

[4] Anderson, C. W. (1989) "Learning to Control an Inverted Pendulum Using Neural Networks", *IEEE Control System Magazine*, **9**, 31-37

[5] Williams, R. J. (1992) "Simple Statistical Gradient-Following Algorithm for Connectionist Reinforcement Learning", *Machine Learning*, **8**, 229-256

[6] Tesauro, G., (1992) "Practical Issues in Temporal Difference Learning", *Machine Learning*, **8**, 257-277.

[7] Boyan, J. A. & Moore, A. W. (1995) "Generalization in Reinforcement Learning: safely approximating the value function", in *Advances in Neural Information Processing Systems* (Vol.7),

[8] Sutton, R. S. (1996) "Generalization in Reinforcement Learning: Successful Examples Using Sparse Coarse Coding", *Advances in Neural Information Processing Systems*, **8**, pp. 1038–1044

[9] Gordon, G. J.(1995) "Stable Function Approximation in Dynamic Programming", *Proc. of ICML*, pp. 261–268

[10] Sutton, R.S. & Barto, A.G. (1998) "Reinforcement Learning", The MIT Press

[11] Shibata, K., Maehara, S., Sugisaka, M. & Ito, K. (2001) "Gauss-Sigmoid Neural Network", *Proc. of 13th SICE Sympo. on Decentralized Autonomous Systems* (to appear in Japanese)