Self-Organization of Place Cells and Reward-Based Navigation for a Mobile Robot*

Takashi TAKAHASHI† Toshio TANAKA‡ Kenji NISHIDA‡
Takio KURITA‡

† Postdoctoral Research Fellow of the Japan Society for the Promotion of Science,
‡ Neuroscience Research Institute, Tsukuba Central 2,
The National Institute of Advanced Industrial Science and Technology,
Ibaraki 305–8568, Japan
Takashi@Takahashi.com

Abstract

We investigate a method to navigate a mobile robot by using self-organizing map and reinforcement learning. Modeling hippocampal place cells, the map consists of units activated at specified locations in an environment. In order to adapt the map to a real-world environment, preferred locations of these units are self-organized by Kohonen's algorithm using the robot's actual position data. Then an actor-critic network is provided the position information from the self-organized map and trained to acquire goal-directed behavior of the robot. It is shown by simulation that the network successfully achieves the navigation avoiding obstacles.

1 Introduction

It is reported that neurons in the dorsal hippocampus of the freely-moving rat are related with the animal's position in an environment[1, 2]. These neurons fire when the animal is located in a restricted portion of an environment (place field) but not in other parts[1]. These neurons are called "place cells". The place field of each place cell is determined by cues such as lights, sounds and feels, and is independent of distal cues fixed to the earth's axis such as geomagnetism[3].

Recently, Foster et al. proposed a model of how hippocampal place cells might be used for spatial navigation in watermaze tasks[4]. The model uses Temporal Difference (TD) learning[5], which is a local, incremental, and statistically efficient connectionist algorithm. A reward-based "actor-critic" network was applied to a navigation task, using place cells to provide information about state. By simulation experiments, it is shown that the actor-critic can learn the reference memory task in which the escape platform occupies a single location and rats gradually learn relatively direct paths to the goal over the course of days.

In this paper, we present preliminary experimental results on a mobile robot navigation using this model. Figure 1 shows the mobile robot used in the experiments. The robot has ultra-sonic sensors, and it can autonomously move around the floor avoiding obstacles. Although Foster et al. used a simplified model in which each place cell was systematically arranged on the grid, such map representation is inefficient for navigation in real-world system because many cells are necessary for fine map. Here the map is self-organized from actual position data obtained by the mobile robot. By this self-organization, we can obtain a space-variant map in which frequently visited places are represented with fine resolutions. Then an actor-critic model is trained to navigate the robot to a specific goal by using reinforcement learning on actor-critic model. The actor selects the next motion direction at each time step and the critic predicts the value function to evaluate the currently selected action. Both units receive the activities of place cells as input. By experiments on mobile robot navigation, it is confirmed that the network achieves the navigation avoiding obstacles.

*TR010808(accepted for ICONIP2001)

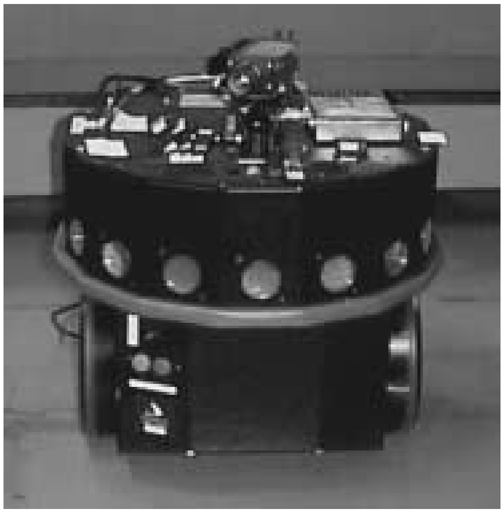


Figure 1: The mobile robot Nomad.

Self-organization of place map is explained in Section 2. Section 3 is for the navigation of the mobile robot using the self-organized place cells. Experimental results are shown in Section 4.

2 Self-organization of place map

2.1 Place cell and self-organizing map

From physiological studies, it is known that the activities of hippocampal place cells of rat can be represented as Gaussian functions of location in space[6]. If a rat is at position $\boldsymbol{p}$, then the activity of place cell $i = 1, \ldots, n$ is given by

$$f_i(\boldsymbol{p}) = \exp\left(-\frac{\|\boldsymbol{p} - \boldsymbol{s}_i\|^2}{2\sigma^2}\right) \tag{1}$$

where $\boldsymbol{s}_i$ is the preferred location of the i-th cell, and σ is the parameter which determines the activity tuning of the cell. Foster et al. used such model to explain how hippocampal place cells might be used for spatial navigation in watermaze tasks[4]. In this study, each $\boldsymbol{s}_i$ was systematically arranged on a grid and fixed beforehand. However, such map representation is inefficient for navigation in real-world since vast number of place cells are required to organize a fine map. It is necessary to organize the map which has fine representation for frequently visited place but coarse representation for others. Therefore, we investigate a method to self-organize the map from actual position data obtained by the mobile robot. The preferred locations of place cells $\boldsymbol{s}_i$ are self-organized by using Kohonen's algorithm[7]:

$$\Delta \boldsymbol{s}_i = \eta h(i, i^*)(\boldsymbol{p} - \boldsymbol{s}_i) \tag{2}$$

where $h(i, j)$ is the neighborhood function which determines the topology of the map, i^* denotes the index of the winner cell determined as

$$i^* = \arg\min_i \|\boldsymbol{p} - \boldsymbol{s}_i\| \tag{3}$$

and η is the learning rate.

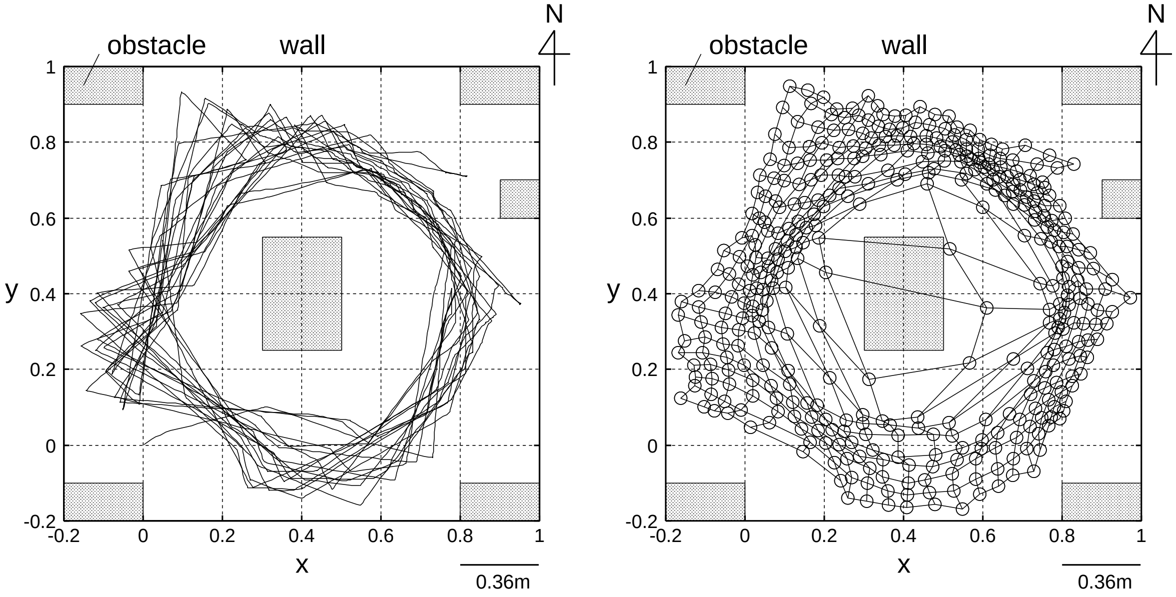
2.2 Experiment

We have done an experiment using the mobile robot shown in Figure 1. Figure 2 (a) shows the experimental field and the motion trajectory of the robot. At the beginning of the experiment, the robot was located at the origin heading the east. Then it wandered in the field recording position information by dead-reckoning. The robot was equipped with ultra-sonic sensors, and it could autonomously move around the floor avoiding obstacles. From this experiment, we obtained 5,327 position data. They were used to organize the place cell map by Kohonen's algorithm. The number of place cells n were set to 20×20 composing two-dimensional neighborhood relation. Figure 2 (b) shows the result.

3 Reward-based navigation

3.1 Actor-critic model

This section describes the navigation module which determines the robot's goal directed behavior. The module consists of two networks, the actor and the critic. The actor selects the next action among eight possible actions $\{A_1, \ldots A_8\}$, each of which corresponds to one motion direction(e.g., north, northeast, east), at each time step. The critic predicts the value function, the discounted total future reward that is expected, to evaluate the currently selected action. Both units receive the activities of place cells as input, and they are trained using Temporal Difference(TD) learning[5]. If the robot reaches the goal at time t, it receives the reward $R(t) = 1$; otherwise $R(t) = 0$. In addition, $R(t) = -1$ is given if the robot makes an invasion to the prohibited area(walls and obstacles).



(a) The trajectory of the robot.

(b) The self-organized map.

Figure 2: Experimental environment and the self-organized map of the place cells.

3.2 Critic

We define an output of the critic when the robot is at position $\boldsymbol{p}$ as follows:

$$C(\boldsymbol{p}) = \sum_{i}^{n} v_i f_i(\boldsymbol{p}) \tag{4}$$

where v_i is the weight between the i-th place cell and the output cell. The critic learns the value function by TD learning, that is, the weights are updated so that the prediction error(TD signal) is reduced:

$$\delta(t) = R(t+1) + \gamma C(\boldsymbol{p}(t+1)) - C(\boldsymbol{p}(t)) \tag{5}$$

where γ denotes the discount factor. Then the updating rule of the weights v_i is given by

$$v_i(t+1) = v_i(t) + \alpha\delta(t) f_i(\boldsymbol{p}(t)) \tag{6}$$

where α denotes the learning rate.

3.3 Actor

The actor stochastically selects one action A_j according to the following probabilities:

$$P(A_j|\boldsymbol{p}) = \frac{\exp(2a_j)}{\sum_{k=1}^{n}\exp(2a_k)} \tag{7}$$

where a_j denotes the output of the j-th action cell corresponding to A_j. The output a_j is computed as

$$a_j(\boldsymbol{p}) = \sum_{i}^{n} w_{ji} f_i(\boldsymbol{p}) \tag{8}$$

where w_{ji} is the weight between the i-th place cell and the j-th action cell. Then the updating rule of the weights w_{ji} is given by

$$w_{ji}(t+1) = \begin{cases} w_{ji}(t) + \beta\delta(t) f_i(\boldsymbol{p}(t)) & \\ & \text{if } A_j \text{ was chosen,} \\ w_{ji}(t) & \text{otherwise.} \end{cases} \tag{9}$$

4 Simulation

In this section, we describe some simulation results of virtual robot navigation using the self-organized map and the trained actor-critic network. We simulated 1000 trials to train the actor-critic network by TD learning. The starting position of the robot was randomly chosen for each trial, and the goal was defined as the square region located around $x = 0.4, y = 0.8$. Trials were aborted if the robot

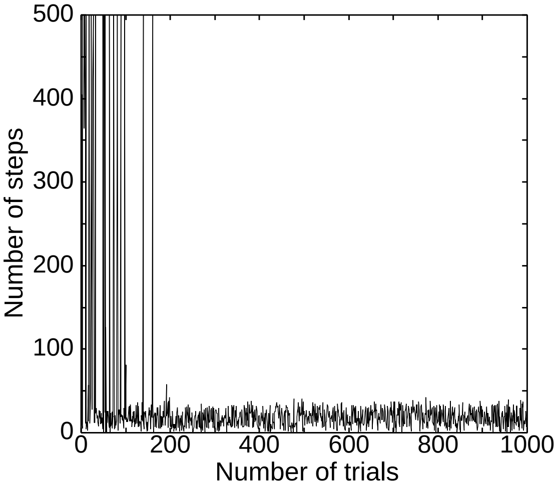


Figure 3: Learning Curve. The ordinate indicates the number of motion steps to reach the goal.

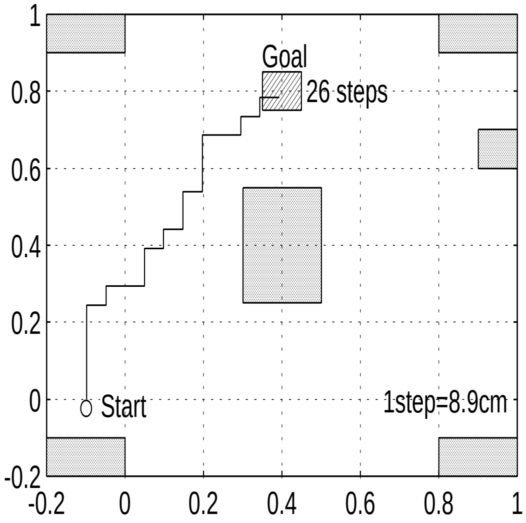


Figure 5: An example of the navigation.

failed to reach the goal within 500 steps. The distance of one motion step was set to 0.049, and the width of each place field σ was set to 0.041. The learning rates α and β were set to 0.1, and the discount factor γ was set to 0.99.

Figure 3 indicates the learning curve of TD learning. During these 1000 trials, the robot successfully reached the goal 980 times. Learning converged less than first 200 trials. After the convergence, the robot could reach the goal within at most 41 steps.

Figure 4 shows the map which depicts the motion direction selected by the actor at each position. It is shown that the actor achieves the goal directed navigation avoiding the obstacles. There is a divide around $x = 0.4, y = 0.0$, and paths are divided into two routes, eastern route and western one, depending on the starting position.

Figure 5 shows one example of navigation simulation. In this case, the robot reaches the goal at 26th step.

5 Conclusion

This paper presented a network model for navigating a mobile robot. It was shown that the self-organizing map could acquire efficient representation for robot localization from the position data obtained in a real environment. Simulation studies also showed that the actor-critic network could learn how to navigate the robot utilizing such representation. Future work will include extensions of this approach for the case the other sensor information(e.g. vision, auditory) are available. It is also a problem to develop a method which incorporates the self-organization process of place cells with reinforcement learning.

Acknowledgment

The authors are grateful to Mr. Yukifumi Ikeda of Saitama University for his invaluable aid in the experiments. Takashi Takahashi was supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists.

References

[1] J. O'Keefe and J. Dostrovsky. The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely moving rat. *Brain Research*, Vol. 34, pp. 171–175, 1971.

[2] R. G. M. Morris, P. Garrud, J. N. P. Rawlins, and J. O'Keefe. Place navigation impaired in rats with hippocampal lesions. *Nature*, Vol. 297, pp. 681–683, 1982.

[3] J. O'Keefe and D. H. Conway. Hippocampal place units in the freely moving rat : Why they

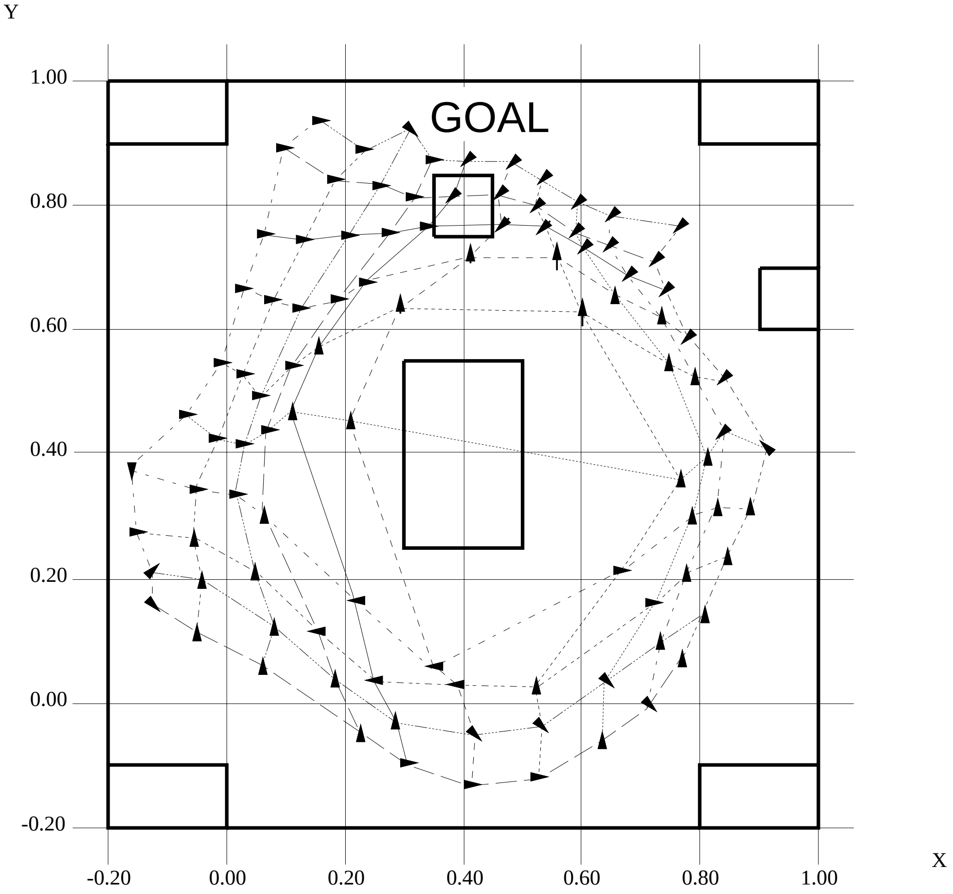


Figure 4: Action Map. Each small arrow indicates the motion direction which has the highest probability $P(A_j|\boldsymbol{p})$ at the center of each place field. The arrows are plotted alternately for conciseness.

fire where they fire. *Experimental Brain Research*, Vol. 31, pp. 573–590, 1978.

[4] D. J. Foster, R. G. M. Morris, and P. Dayan. A model of hippocampally dependent navigation using the temporal difference learning rule. *Hippocampus*, Vol. 10, pp. 1–16, 2000.

[5] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An introduction.* The MIT Press, 1998.

[6] J. O'Keefe and N. Burges. Geometrical determinants of the place fields of hippocampal neurons. *Nature*, Vol. 381, pp. 425–428, 1996.

[7] T. Kohonen. *Self-Organizing Maps.* Springer, third edition, 2001.