A region selecting method which performs observation and action in the multi-resolution environment.

Toshihiro MATSUI, Hiroshi MATSUO and Akira IWATA

Department of Electrical & Computer Engineering, Nagoya Institute of Technology,
Gokiso-cho, Showa-ku, NAGOYA, 466-8555, JAPAN

Abstract. We propose a method for selecting the characteristic region in the environment based on the occurrence probability of the pattern. If the occurrence probability of the pattern is unknown in initial stage, estimation of the distribution of the pattern and selection of the characteristic region must be done simultaneously. We noticed that a method for exploration of the state-space in reinforcement learning was similar to such task. Then, we propose a method for selecting the characteristic region by repeating observation and action in the environment. In the observation using only one resolution, the position in the environment can not be decided. The multi-resolution concept is introduced in order to solve this problem. The experimental result shows that the characteristic region is selected from the environment.

1 Introduction.

In the tasks such as pattern recognition, the part of information observed for the environment is selected as a important region at one time. A criteria to select such region is the occurrence probability of the partial pattern. Because the characteristic area is correspondent to the pattern of which the occurrence probability is small, it is appropriate to select mainly such pattern in order to decide the important region. If the region is selected in proportion to the occurrence probability of the pattern, the frequency of the observation for each pattern becomes equal. When the occurrence probability of the pattern is unknown in the initial condition, it is reasonable to correct the polarization of the observation frequency for each pattern by each point of time in the selection of the region. It requires large calculation cost to select next region from all regions. And, when it is selected in the absolute position, the error of the position becomes a problem. Therefore, it is efficient to relatively select the region from the neighborhood defined for the current position.

In reinforcement learning [1] which is one of the unsupervised learning methods, the agent starts the learning from the condition without knowledge on the environment, and it acquires the rule. The rule reflects the series of the action which obtains the reward from the environment. The agent repeats observation and action in the environment, and it constitutes the state-space. In the environment identification type learning, it is necessary to select the action that explores all states in order to constitute the state-space which reflects the environment. Miyazaki et al. proposed the algorithm for exploring all state-space[2]. In this algorithm, the agent explores environments by selecting all actions at least $k = 1$ time. Next, the agent repeats the similar action, after the value of k is increased for 1. This method is the selection of the action which averages the trial frequency of all actions in all known states. In the environment, the distribution of the pattern observed as an identical state will influence the distribution of the position of the agent. In reinforcement learning, if an identical state is observed at different positions in the environment, it causes a mismatching between the state-space of the agent and the environment. However, we discuss how the position in the environment of the agent is distributed in the repetition of action and observation. It is not possible to classify the identical pattern observed in the different position, if the agent has no information about the position in the environment. In such case, the agent will explore around local area. For the solution of this problem, the multi-resolution is introduced into the environment. We define the multi-resolution for observation and action in the environment. Then, we propose a framework which aim to equalize frequency of each action on each state. By this method the agent mainly exists at the characteristic position in the environment.

The concept of peripheral vision and multi-resolution is used for visual attention in the computer vision. The multi-resolution concept is used in the modeling of the saccade phenomenon[6]. B. Takacs et al. proposed a method using a dynamic and multi-resolution model[7]. C. Bandera et al. applies reinforcement learning for the visual attention. Its purpose is the model based target recognition[5]. The visual field of the multi-resolution is used even in this method.

In the following, we describe our technique, and it is evaluated by the experiment.

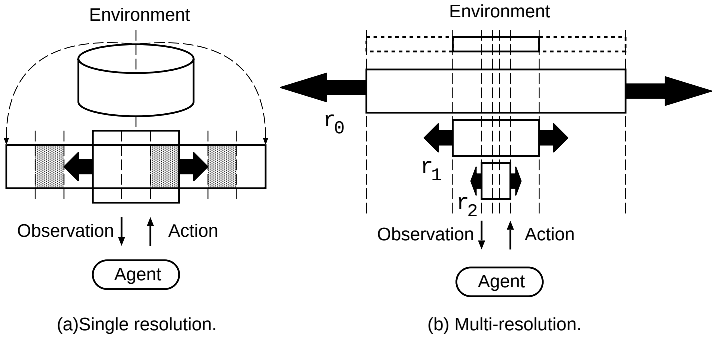
2 The modeling of the problem.

We describe environment, observation and action in order to simplify the problem. We assume that the environment is a n-dimensional torus. Reasons for assuming the n-dimensional torus are as follows: (1) It is possible to limit the whole of the space in the observation. (2) There is no contradiction in our definition on the neighborhood.

The function $f(x) = \{0, 1\}$ is defined for the all position $\{x\}$ in the environment. For the observation, the whole of $\{x\}$ is equally divided into the unit interval $u_i (i = 0, \cdots, n)$ in proportion to the resolution(Fig.1(a)). In the observation, the unit interval u_i is selected, and $\max_{u_i} f(x)$ is obtained as the value by the observation. In the observation, the values of the unit intervals, which are selected as the center and adjoining it, are obtained.

There is no constraint on the movement in the environment. However,the unit in the transfer is supposed to be identical with the unit interval in the observation. In addition, the moving range in each iteration is limited to unit intervals selected as the center and adjoining it.

By the above, the state-space is constituted. The agent repeats observation and action using this state-space. This framework is similar to general reinforcement learning. However, our purpose is that the distribution of the position of the agent is made to adapt to the distribution of the pattern in the environment. For this purpose, we combine two approaches. One approach aims to equalize the frequency of the observation of all states. Another approach is to introduce the multi-resolution into the state-space.



(a)Single resolution.

(b) Multi-resolution.

Fig. 1. Resolution for the action and the observation.

3 An algorithm which equalize the number of trials of all action on all observed state.

In our framework, the agent repeats observation and action as well as general reinforcement learning. In this framework, each state does not have information of the position in the environment. Therefore, the agent can fail to local exploration. We describe this problem in the next section.

The observation frequency of the pattern follows the occurrence probability, if the agent equally scans the environment. Reversely, the distribution of the position of the agent is dependent on the occurrence probability of the pattern, if the frequency of the observation of all states is equalized. Since each state does not have information about the position, it is impossible to decide the action which equalizes the frequency of the observation of each state directly. Then, we use the method for equalizing the frequency of the selection of each action instead of each state. In reinforcement learning, it is not desirable that the agent detects the different position in the environment as an identical state. But, our purpose is not the acquisition of the rule by reinforcement learning.

Miyazaki et al. proposed the k-Certainty Exploration Method as an algorithm for the environment identification type learning. The purpose of this method is that the agent explores all state-space. In this algorithm, the agent explores all environments by selecting all actions to over $k = 1$ time. Next, the agent repeats the similar action, after the value of k is increased for 1. The algorithm is as follows (We arranged the notation).

```
procedure k-Certainty Exploration Method
begin

if the k-Uncertainty rule exists for current state then k:=1.
else if all known rules are k-Certainty then k:=k+1;

if the k-Uncertainty rule exists for current state then
    One of the such state is randomly selected.
else
    begin

    set flags for all known states.
    for all known states except for present condition do

       if the k-Uncertainty rule exists or
       the rule which makes transition to the state
       that the flag is off, is exist then
       reset flag of the state.

    while the state that the flag was newly reset exists;

    One of the rules which can make transition
    to the state that the flag is off from present state,is randomly selected.

    end;

end.
```

In this method, only the action which does not satisfy the k value becomes a candidate for the selection. This is a very simple method. However, it is appropriate as the constraint which brings the distribution close to the desired value, because it can be understood as a weighting based on upper part accumulation probability of normal distribution (k,0). And, the calculation cost is small, since it does not need to calculate the distribution. In this algorithm, the exceptional processing is executed when all action in present state has already been selected over k time. Such mechanisms are necessary in order to prevent the local exploration.

4 The region selecting method using multi-resolution to correct the scale mismatch between the environment and the state space.

We have assumed the state-space based on the single resolution. However, an identical pattern in different positions can not be distinguished. In such case, the frequency of the action on each state is equalized, while the agent explore around local area. Then, we introduce multi-resolution $r_j (j = 0 \cdots n)$ in order to solve this problem(Fig.1(b)). The whole of $\{x\}$ is equally divided into the unit interval $u_i^n (i = 0, \cdots, 3^n)$ in proportion to the highest resolution r_n. At the resolution r_n, any of the unit interval u_i^n is selected. In the observation, the values of the unit intervals which are selected as the center and adjoining it, are obtained. At other resolution r_j, the visual field at the resolution r_{j+1} is selected as a center. In the observation, the values of the unit intervals, which are selected as the center u_i^j and adjoining it, are obtained. The whole environment is fixed, because the lowest resolution wraps the environment.

At each resolution, the algorithm similar to the k-Certainty Exploration Method is prepared. It is necessary to select one action at a time, because the agent has only one body. It means that only one resolution must be selected to decide the action. For this selection, we use the method which is similar to the k-Certainty Exploration Method. The m value is introduced for k_j value at each resolution r_j. One of the resolution with k_j value under the m value is selected. However, the resolution in which the frequency of each action in present state has achieved k_j is removed from the selection candidate. This expansion is based on the assumption that the increase of k_j value at each resolution reflects the number of the observed state. Our algorithms using the multi-resolution are as follows.

```
procedure
begin

for all resolution $r_j (j = 0, \cdots, n)$
   begin
   if the unknown state is detected then
      begin
      All $k_j$:=1.
      $m$:=1.
      end
   else if all known rules are executed over $k_j$ times then $k_j$:=$k_j$+1;
   end;

if all $k_j > m$ then $m$:=$m$+1;

if some of $k_j \leq m$ and one of rules of current state are not executed
more than $k_j$ times then
select one of such resolution randomly.
```

```
else select a resolution randomly from all resolutions;

if for selected resolution, rules of current state are not executed more
than $k_j$ times
then select one of such rule randomly.
else select a rule randomly form all rules of selected resolution;

execute the rule.

end.
```

5 Experiment.

We show the experimental result using proposed method. We apply our method in a 1-dimensional torus environment including the characteristic pattern. And, we also show the example in a two-dimensional torus environment.

5.1 The result in 1-dimensional torus.

The result in the 1-dimensional torus environment is shown. In this example, the number of resolutions is 5. Therefore, the environment is divided into $3^{5-1} = 81$ in the highest resolution. The accumulation of the agent of the position in the environment where the characteristic pattern exists is shown in the figure 2. The m value is 1000. It takes 411520 iterations. In the result, the characteristic pattern has mainly been selected. The change of the accumulation from start point of time is shown in the figure 3.

The result as the pattern contains the simple texture is shown in the figure 4. The distribution increases, while the characteristic pattern occurrence probability is reflected.

5.2 The result in 2-dimensional torus.

The result in the 2-dimensional torus environment is shown(Fig.5). In this example, the number of resolutions is 4, and the environment is divided into 27×27. The m value is also 1000,but it takes 10837593 iterations. This result reflects that the occurrence probability of the pattern of the vertices are small.

6 Summary.

In the proposed method, the agent repeats observation and action in the environment using the multi-resolution. The agent equalizes the trial frequency of each action, and as the result, the characteristic region in the environment is mainly explored. It is indicated that the exploration which adapted to the environment is a filter processing for the spatial characteristic of the environment.

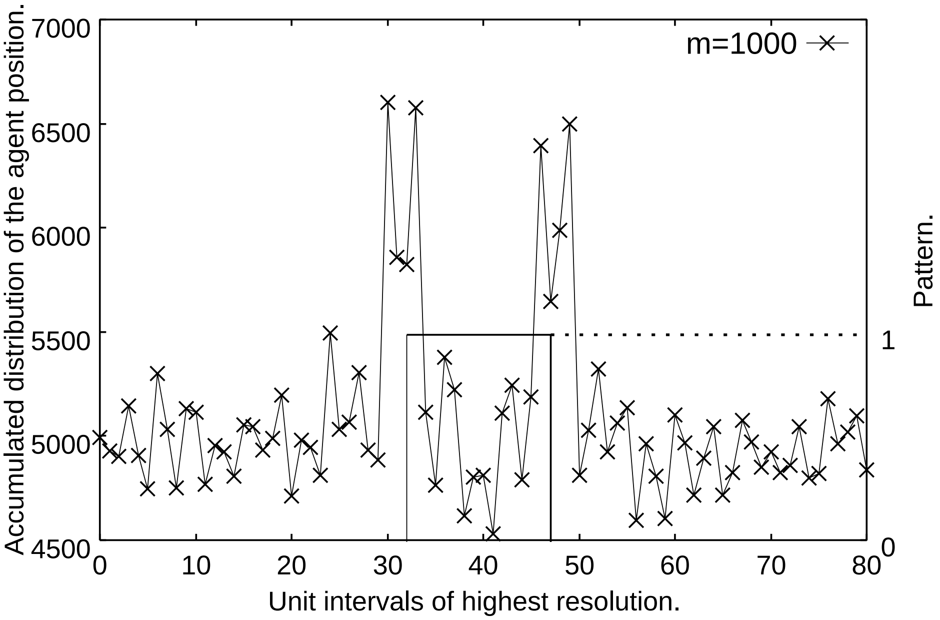


Fig. 2. The accumulation of the agent position (m=1000).

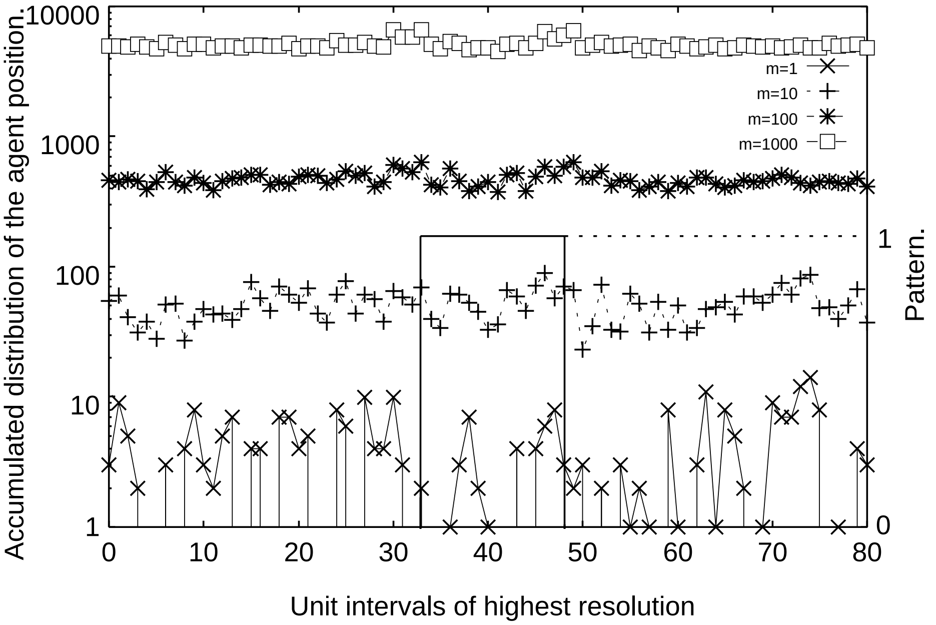


Fig. 3. The accumulation of the agent position (m=1,10,100 and 1000).

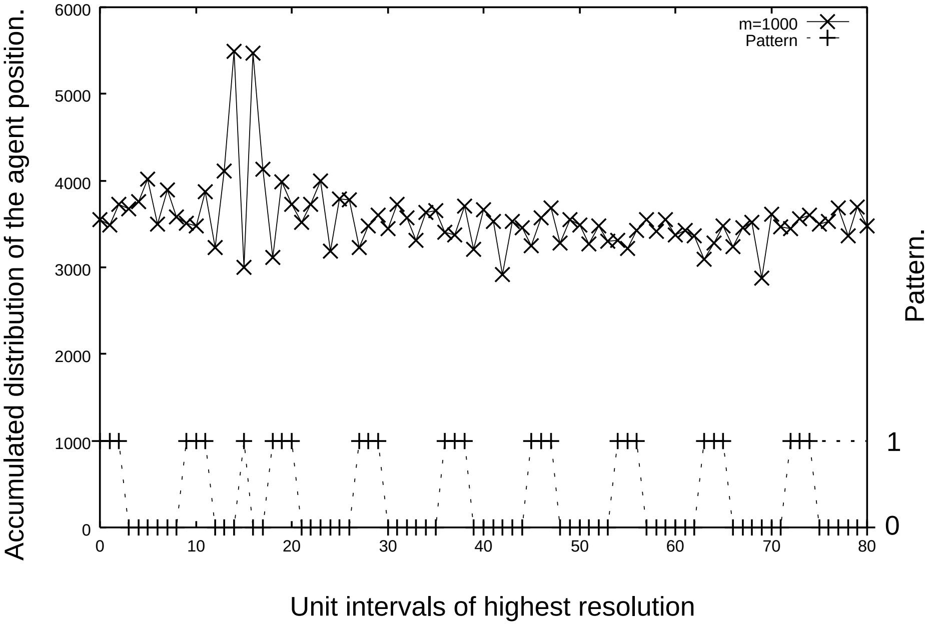


Fig. 4. The accumulation of the agent position in the texture.(m=1000).

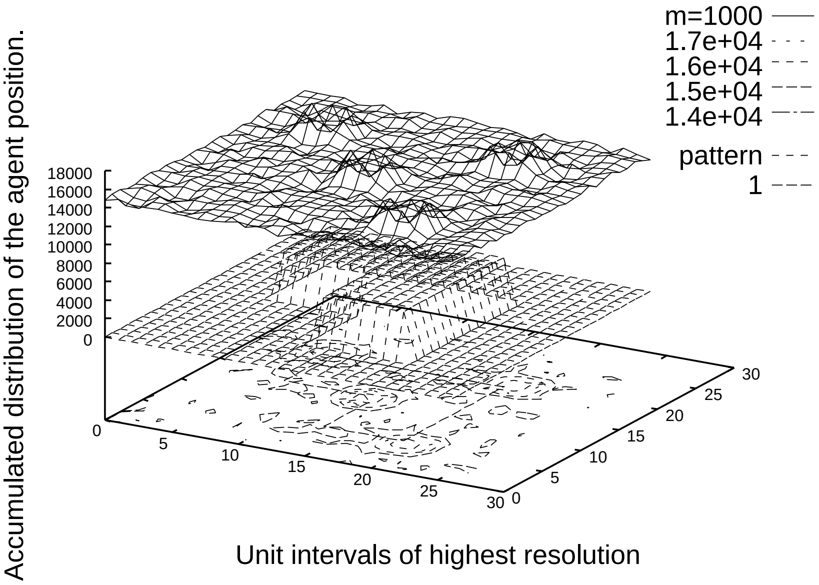


Fig. 5. The accumulation of the agent position in 2d torus.(m=1000).

Information about the position in the environment is necessary so that this distribution may be correspondent to the whole environment. The introduction of the multi-resolution enables the decision of the relative positional relation. In the experiment using this method, the distribution in proportion to the characteristic region in the environment was obtained.

There is another type of implementation the multi-resolution, that places the visual field of the logarithm scale in the single layer. This implementation seems to be essentially equivalent to the proposed method. However, the dimension of observed vector increases. The state-space exponentially increases for the dimension of the state vector observed. Therefore, we currently divided the state-space at each resolution.

The agent view whole environment in the lowest resolution. Observed pattern in the lowest resolution environment is not changed while the agent explores. However, the visual field can be limited in the practical environment. In such case, the agent will fail to local exploration.

How to abstract the environment for multi-resolution is also important problem. To classify the observed pattern with high accuracy, another multi-resolution model is needed.

The serious problem of this method is to require many trials. In this method, the frequency of the selection of all actions is equalized. Therefore, many trials are necessary for the case in which there are many types of patterns in the environment in order to try all actions. And, the global perturbation by the low resolution does not supplement the local perturbation at the high resolution, because the weight of the selection of each resolution is simply equalized. Exploration of the whole environment and selection of the specific region are trade-off. Both balance must be set in proportion to actual purpose.

It is a future problem to introduce some pattern recognition mechanism in order to apply it to the incremental feature extraction processing. Some theoretical consideration on this problem is also necessary.

References

1. L. P. Kaelbling, M. L. Littman, Andrew W. Moore: Reinforcement Learning: A Survey; Journal of Artificial Intelligence Research 4, pp.237-285 (1996).
2. K. Miyazaki, M. Yamamura and S. Kobayashi: k-Certainty Exploration Method: An Action Selector on Reinforcement Learning to Identify the Environment; Journal of Artificial Intelligence, Vol.91, pp.155-171 (1997).
3. K. Miyazaki, M. Yamamura and S. Kobayashi: l-Certainty Exploration Method: An Action Selector to Identify the Identify the Environment – An Extension of k-Certainty Exploration Method to Stochastic MDPs –; Journal of Japanese Society for Artificial Intelligence, Vol.11, pp.804-808 (1996).
4. K.Miyazaki, M. Yamamura and S. Kobayashi: MarcoPolo-A Reinforcement Learning System Considering Tradeoff Exploration and Exploitation under Markovian Environment; Proc. of 4th Int. Conf. on Soft Computing, pp.561-564 (1996).
5. Bandera, C., Vico, F. J., Bravo, J. M., Harmon, M. E., and Baird, L. C.: Residual Q-learning applied to visual attention; Proceedings of the Thirteenth International Conference on Machine Learning, Bari, Italy, 3-6 July, pp. 20-27. (1996).

6. Moddeling Saccadic Targeting in Visual Search; R. P. N. Rao, G. J. Zelinsky, M. M. Hayhoe, D. H. Ballard.: Advances in Neural Information Processing Systems 8, D. S. Touretzky, M. C. Mozer, M. E. Hasselmo, eds., MIT Press (1996).
7. B. Takacs, H. Wechsler: A Dynamic and Multiresolution Model of Visual Attention and Its Application to Facial Landmark Detection; Computer Vision and Image Understanding Vol.70, pp.63-73 (1998).
8. I. Marsic: Data-Driven Shifts of Attention in Wavelet Scale Space; CAIP-TR-166, CAIP Center, Rutgers University, September (1993).