Engineering

Industrial & Management Engineering fields

Okayama University *Year* 2000

An incremental state-segmentation method for reinforcement learning using ART neural network

Hisashi Handa
Okayama University

Akira Ninomiya
Okayama University

Tadashi Horiuchi
Osaka University

Tadataka Konishi
Okayama University

Mitsuru Baba
Okayama University

An Incremental State-Segmantation Method for Reinforcement Learning Using ART Neural Network

H. Handa* **A. Ninomiya*** **T. Horiuchi**** **T. Konishi*** **M. Baba***

*Dept. of Information Technology
Faculty of Engr., Okayama Univ.
Tsushima-naka 3-1-1, Okayama, JAPAN
handa@sdc.it.okayama-u.ac.jp

**Dept. of Advanced Reasoning
I.S.I.R., Osaka Univ.
Mihogaoka 8-1, Ibaraki, JAPAN
horiuchi@ar.sanken.osaka-u.ac.jp

Abstract

In this paper, we propose a new incremental state segmentation method by utilizing information of agents' state transition table which consists of tuple of (state, action, state) in order to reduce the effort of designers and which is generated by ART Neural Network. In the proposed method, if inconsistent situation in the state transition table is observed, agents refine their map from perceptual inputs to states such that such inconsistency is resolved. We introduce two kinds of inconsistency, i.e., "Different Results Caused by the Same States and the Same Actions" and "Contradiction due to Ambiguous States." Several computational simulations on cart-pole problems confirm us the effectiveness of the proposed method.

1 Introduction

Reinforcement learning is one of the most active research areas in intelligent systems. In this approach to machine learning, an agent tries to maximize the total amount of reward it receives when interacting with a complex and uncertain environment. The object of reinforcement learning agents is to discover effective policy how agents decide actions against any perceptual inputs in order to receive the most reward via their trying. Many reinforcement learning algorithms, i.e., Q-Learning, TD (λ), SALSA, Profit-Sharing, and so on, have been developed by many researchers and have been confirmed the effectiveness of them experimentally or theoretically so far [1, 2, 3]. In usual cases to apply such reinforcement learning algorithms to certain practical problem, designer who wants to apply reinforcement learning to such problems has to define the constitution of the states, namely, perception-state maps, in advance. This definement is quite important since it causes various learning results of agents. That is, if grained-scale state segmentation is given to agents, agents would have to necessitate much learning time to acquire moderate perception-action rules. On the other hand, i.e., in the case of coarse-scale state segmentation, agents may not distinguish certain states such that these must be aware as the different ones due to yield the different results from the same action, that is, it may be caused to a famous problem, called perceptual alias problem, in reinforcement learning community. Therefore, in this paper, we propose a new incremental state segmentation method by utilizing information of agents' state transition table which consists of tuple of (*state*, *action*, *state*) in order to reduce the effort of designers. In our approach, if inconsistent situation in the state transition table is observed, agents refine their map from perceptual inputs to states such that such inconsistency is resolved. We discuss on two kinds of inconsistency in this paper: "Different Results Caused by the Same States and the Same Actions" and "Contradiction due to Ambiguous States."

2 Related works

Incremental state segmentation has been studied by many researchers [4, 5, 6]. Many of them used reinforcement signals to delineate state segmentation more precisely. In the proposed method, we adopt state transition information to acquire adequate segmentation. Dubrawski and Reignier proposed perceptual state categorization method using Fuzzy-ART Neural Network. Our method utilizes modified ART Neural Network based on distance between input vectors and refines state segmentation by using the notion of contradiction. The notion of contradiction used in this

0-7803-6456-2/00/$10.00 ©2000 IEEE

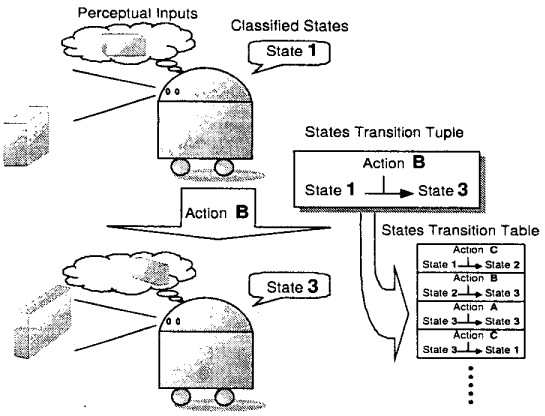

Figure 1: A framework of the proposed method

work is inspired by Piaget's one [7]. In his work, the notion of contradiction is classified into three categories by seeing into children' behavior:

1. Contradictions such that it looks that the same actions yield the different results.
2. Contradictions characterized by incomplete disagreement among certain classes
3. Contradictions caused by incorrect reasoning, especially incorrect implication.

In his work, he concluded that contradictions are emerged from inconsistent complements. Two kinds of contradictions introduced in this paper, i.e., "Different Results Caused by the Same States and the Same Actions" and "Contradiction due to Ambiguous States," are belonging to category 1. and 2., respectively.

3 Proposed state segmentation method

3.1 Overview

The diagram of our approach is depicted in Figure 1. As depicted in this figure, we assume continuous inputs from environments, such like sensors, cameras and so on. For the purpose of using traditional reinforcement algorithms, discrete state of agents is decided by the continuous inputs. In this paper, we adopt a kind of Adaptive Resonance Theory (ART) originally proposed by Grossberg as a map from such continuous inputs to discrete states [8]. Then, agents carry out proper action associated to such state based on his action selection mechanism, and recognize new perceptual inputs from the environments again. In

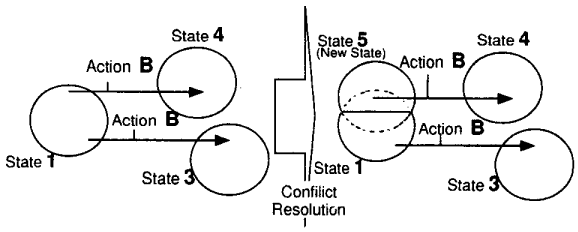

Figure 2: A depiction of contradiction such that different results are caused by the same states and the same actions

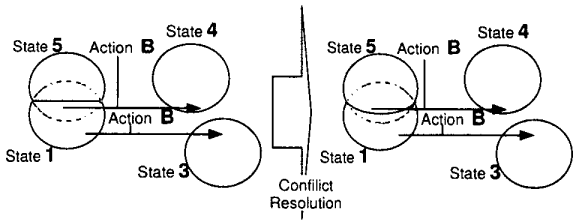

Figure 3: A depiction of contradiction due to ambiguous states

this paper, tuples $(state, action, state)$ which indicates state transition are recorded into a table called state-transition tables. Moreover, if inconsistent state transition against the state-transition tables is acquired, we define such inconsistent state transition as **contradiction** and introduce two manners with the aim of solving such contradiction.

3.2 Classification of states from perceptual inputs by ART

In this paper, we adopt ART to realize the map from perceptual inputs to corresponding state. ART consists of two levels of neurons: F_1 and F_2. The neurons in the level F_1 and F_2 are corresponding to a particular combination of sensory features and recognition code which represents states in the case of this paper, respectively. In ART, given inputs are classified into the most resonant code that is decided by referring to a selection strength and vigilance criterion. If there are no resonant codes against certain inputs to classify, namely, there is no selection strength associated to code which is greater than the vigilance criterion, a new recognition code is added to the level F_2 by adopting the input vector as the sample vector to the added recognition code.

Detailed description is shown in followings: Let $\boldsymbol{x}$ and $\boldsymbol{w_i}$ be a perceptual input vector for ART and sample vectors linked from all neurons in the level F_1 to a

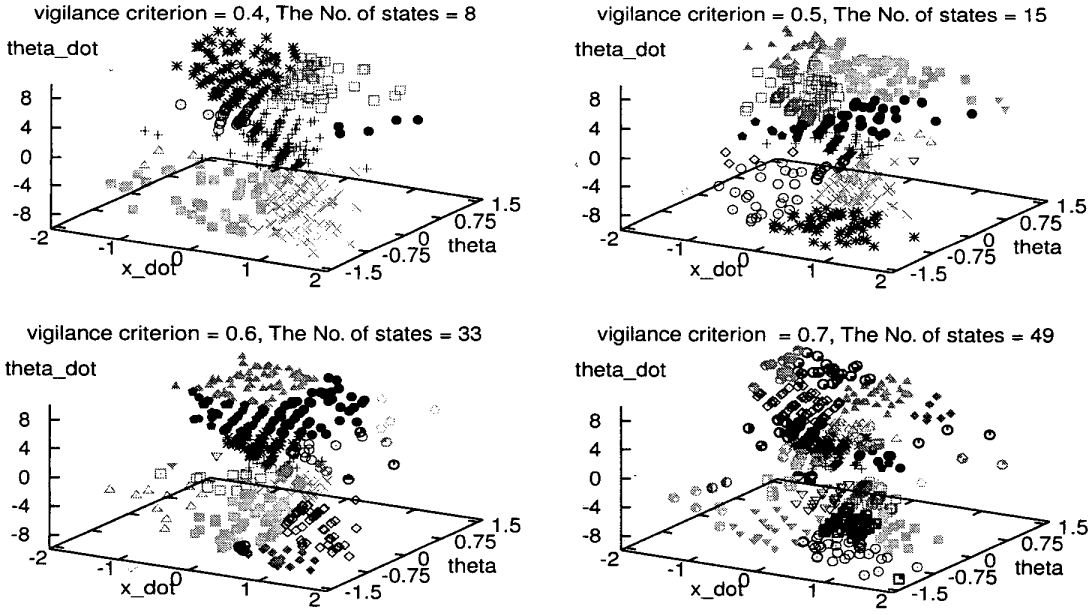

Figure 4: The distribution of segmented states with ART without weight change for each value of the vigilance criterion of ART

neuron i (recognition code) in the level F_2. We adopt following selection strength T_i for state i (recognition code i) based on the distance between the input vector and the sample vector:

$$T_i = \frac{1}{\epsilon_\alpha + \frac{|\boldsymbol{x}-\boldsymbol{w}_i|^2}{\epsilon_\gamma+|\boldsymbol{x}|^2}}$$

where, ϵ_α and ϵ_γ indicate small positive constant values fixed in advance. For each state i, this selection strength T_i is calculated, and if the most resonant state i^*, i.e., a state which has the highest selection strength, is greater than the vigilance criterion σ, such state i^* is chosen as a state corresponding to the perceptual inputs. Otherwise, a new state j whose sample vector is the same as the perceptual inputs $\boldsymbol{x}$ is added into a set of sample vectors. That is,

$$\boldsymbol{w_j} = \boldsymbol{x}.$$

Also, in traditional ART, activated sample vector is updated for following to a current perceptual input vector described as follows:

$$\boldsymbol{w}'_{i^*} = \beta\boldsymbol{x} + (1-\beta)\boldsymbol{w}_{i^*}.$$

However, in the proposed method, state transition information is utilized vigorously so that such improvement of an activated sample vector is not carried out.

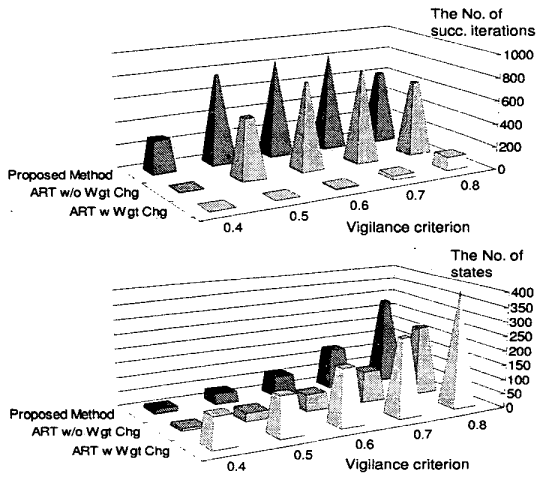

Figure 5: The number of success iterations (UPPER) and segmented states (LOWER)

3.3 Description of State Transition Table

In the proposed method, a state transition table is recorded in order to detect inconsistent transition. How to record the transition table is as follows: First, suppose that a state *current* corresponding to a perceptual input $x_{current}$ is classified by the means of previous section at a current time step. Moreover,

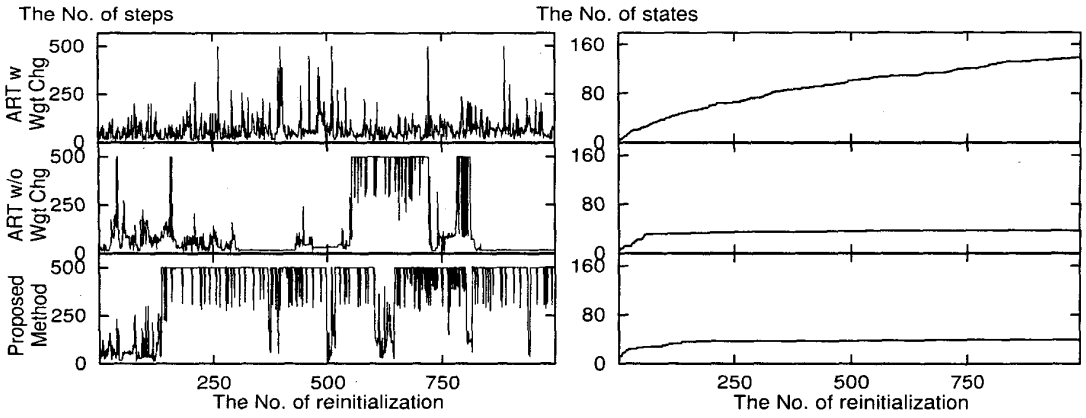

Figure 6: The changes of the number of steps until tumbling pole (LEFT SIDE) and the number of segmented states (RIGHT SIDE); ART with weight change (UPPER), ART without weight change (MIDDLE), and the proposed method (LOWER)

suppose that, at the time step, the agent behaves an action act_i and, as a consequence of the action, next perceptual input x_{next} is received and classified into a state *next* at the next time step. Such process is recorded that

$$f_{act_i}(current) = next.$$

Note that state transitions with respect to ambiguous states which mean that selection strengths for several states exceed the vigilance criterion is not recorded into the state transition table.

3.4 State Segmentation by Using the Notion of Contradiction

In this paper, we adopt two kinds of the notions of contradiction to constitute state segmentation: "different results caused by the same states and the same actions" and "contradiction due to ambiguous states." Following subsections introduce them.

Different Results Caused by the Same States and the Same Actions

Suppose that there is a record in the state transition table such that an action *act* in a state A brought out a state B:

$$f_{act}(A) = B.$$

Moreover, now, the same action *act* in the same state A causes the different state C:

$$f_{act}(A) = C.$$

Above equations are inconsistent each other. In this case, it is possible that inadequate mapping from perceptual inputs to states is carried out by ART. Hence, a new sample vector $\boldsymbol{w_{j+1}}$ which indicates a new state D is added to the ART by using a perceptual input $\boldsymbol{x_D}$ which causes above contradiction, i.e., $\boldsymbol{w_D} = \boldsymbol{x_D}$. Therefore,

$$f_{act}(D) = C.$$

There is no contradiction in a state transition table as depicted in Figure 2. Because the addition of the new state D affects to not only the definition of the state A but also neighbor states around the new state D, all records in the state transition table are destroyed at the time step. By adopting such destruction of the state transition table, meaningless detections of further contradiction are prevented.

Contradiction due to Ambiguous States

In the case that several states are resonant simultaneously called ambiguous states in this paper, even if the contradiction in the mean of last subsection is occurring, other resonant state might be consistent with the transition table. If such consistent resonant state is found, a new state by the means of the former subsection is not generated. Instead, following process is carried out: In the ambiguous state, instead of selection strength T_i described in section 2.2, biased selection strength T_i' for state i is used to decide a state for certain perceptual input. That is,

$$T_i' = T_i + \sum_{j \in Ambiguous(i)} b(i,j),$$

where $Ambiguous(i)$ and $b(i,j)$ denote states which are in ambiguous state with state i for current percep-

tion and a bias term of state i against j, respectively. $b(i,j)$ is positive value and is updated as follows:

$$\Delta b(i,j) = \delta,$$

if state i is consistent for state transition table, where δ is constant value fixed in advance. By introducing this mechanism, an excess of state generation is avoided.

4 Computational Simulations

4.1 Simulated Environments

In this paper, we examine the proposed method with Q-Learning, which is one of the most famous reinforcement algorithms, on cart-pole problems. The agent receives the velocity $\dot{x}$ of the cart and the position θ and velocity $\dot{\theta}$ of the pole as the input. Hence, the dimension of a perceptual input vector in this examination is 3. Initial values of them are set to be randomly around 0. The agent can stress fixed force into left or right direction. The only negative reinforcement signal is given to the agent, provided that a pole is tumbling down, namely, $|\theta|$ is greater than constant value fixed in advance. When 500 steps are achieved without tumbling the pole or the pole is tumbling, simulated environment is reinitialized. One trial consists of 1000 iterations of such reinitializations. We examine three kinds of algorithms, i.e., State-Segmentation by ART, by ART without weight updating, and by the proposed method.

4.2 Experimental Results

First, we check the relevance between the magnitude of a vigilance criterion of ART and the number of segmented states as shown in Figure 4. The axises in these graphs in this figure denote the velocity of the cart, the angle of pole, and the angular speed of the pole, respectively. The points in these graphs denote the perceptual inputs for agent, and the same kinds of points means that such points are classified into the same states by the state segmentation method. These graphs in this figure are results for the same state segmentation method, i.e., the ART without weight change, and the same experiences to agents. The number of segmented states increases as the magnitude of the vigilance criterion enlarges.

Next, we investigate the effectiveness of the proposed method by referring to the number of success iterations and segmented states as delineated in Figure 6. These graphs denote the number of success iterations which indicates agents don't tumble until 500 steps (UPPER) and the number of segmented states when iterations are finished. Each row of prismoid in both graphs indicates the results for the proposed method (BACK), ART without weight change (MIDDLE), and ART with weight change. The front axis denotes the kinds of values of the vigilance criterion. These graphs are averaged results more than 30 trials. The proposed method outperforms the others in terms of the number of success iterations for various kinds of values of the vigilance criterion. However the proposed method incenses the states due to the detection of contradictory situations and such situations are frequently observed in each trial, the number of segmented states by the proposed method is roughly the same of the one by ART without weight change. We consider the reason why such phenomenons are appeared is that effective segmentation of states guides Q-Learning into the steady state.

Finally, we examine typical trial in which viglance criterion is 0.4, in order to confirm us the effectiveness of the proposed method. The graphs in Figure 6 indicate the number of steps until tumbling pole and the number of segmented states, resectively, for three state segmentation algorithms. In left graph of these, a stable situation in which the number of steps until tumbling pole is 500 means success iterations is repeated. As show in right graph, the number of segmented states in ART with weight change increases continuously. Furtermore, segmented states by these algorithms are illustrated in Figure 7. Axises and points in these graphs are the same sense as Figure 4. Despite of the obvious different reslut in the number of success iterations as shown in Figure 6 left, segmented states of ART without weight change and the proposed method look like the same. Thus, center area of these graphs which has the significant means for controlling pole are drawn in Figure 8. As shown in this figure, in the proposed method, perceputual input space is segmented near $\theta = 0$. It is excellent segmentation in center area since effective rules are easily constructed. That is the reason why the proposed method works well.

5 Conclusion

We proposed a new incremental state segmentation method by utilizing information of agents' state transition table which consists of tuple of $(state, action, state)$ in order to reduce the effort of designers in this paper. In the proposed method, if inconsistent situation in the state transition table is observed, agents refine their map from perceptual inputs to states such

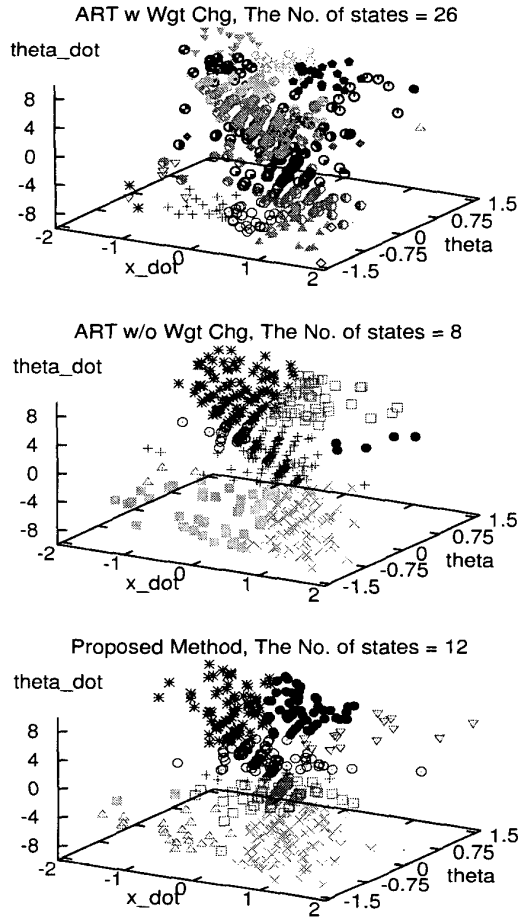

Figure 7: Segmented states in perceptual space by ART with weight change (UPPER), ART without weight change (MIDDLE), and the proposed method (LOWER), respectively.

that such inconsistency is resolved. We introduced two kinds of inconsistency in this paper, i.e., "Different Results Caused by the Same States and the Same Actions" and "Contradiction due to Ambiguous States." Several computational simulations on cart-pole problems carried out previous section confirmed us the effectiveness of the proposed method.

References

[1] R.Sutton and A.Barto, *Reinforcement Learning*, The MIT Press, 1999.

[2] C.Watkins and P.Dayan, "Q-learning", *Machine Learning*, Vol.8, pp.279-292, 1992.

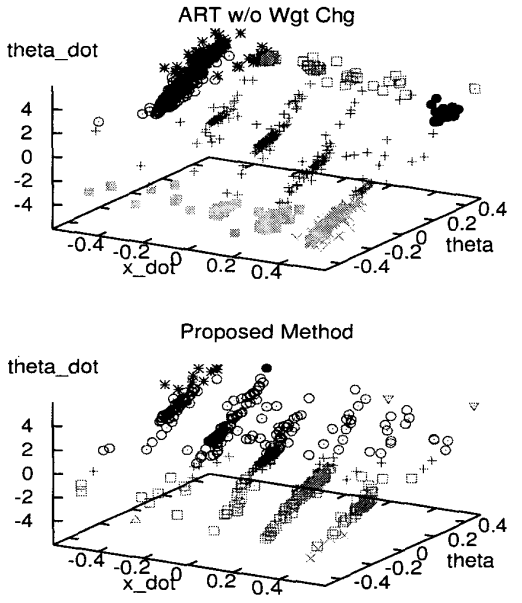

Figure 8: Detailed graphs in center area of graphs in Fig. 7, ART without weight change (UPPER), and the proposed method (LOWER), respectively.

[3] R.Sutton, "Learning to Predict by the Method of Temporal Differences", *Machine Learning*, Vol.3, pp.9-44, 1988.

[4] A.Dubrawski and P.Reignier, "Learning to Categorize Perceptual Space of a Mobile Robot Using Fuzzy-ART Neural Network", *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems IROS'94*, Vol.2, pp.1272-1277, 1994.

[5] H.Murao and S.Kitamura, "Incremental State Acquisition for Q-Learning by Adaptive Gaussian Soft-max Neural Network", *Proccedings of the 1998 IEEE ISIC/CIRA/ISAS Joint Conference*, Gaithersburg, pp.465-470, 1998.

[6] H.Murao and S.Kitamura, "Incremental Quantization of the Continuous Sensor Space for Leaning Agents", *Intelligent Autonomous Systems*, Y.Kakazu *et al.*(Eds.), IOS Press, pp.272-279, 1998.

[7] G.Drescher, *MADE UP MINDS*, The MIT Press, 1991.

[8] M.Snorrason and A.Caglayan, "Generalized ART2 Alogorithms", *World Congress on Neural Networks*, 1994.