Back Propagation Learning of Neural Networks with Chaotically-Selected Affordable Neurons

Yoko Uwate and Yoshifumi Nishio
Dept. of Electrical and Electronics Engineering, Tokushima University
2-1 Minami-Josanjima, Tokushima, Japan
Email: {uwate, nishio}@ee.tokushima-u.ac.jp

Abstract— **Cell assembly is one of explanations of information processing in the brain, in which an information is represented by a firing space pattern of a group of plural neurons. On the other hand, effectiveness of neural network has been confirmed in pattern recognition, system control, signal processing, and so on, since the back propagation learning was proposed. In this study, we propose a new network structure with chaotically-selected affordable neurons in the hidden layer of the feedforward neural network. Computer simulated results show that the proposed network exhibits a good performance for the back propagation learning.**

I. Introduction

Recently, studies on the brain have been carried out actively on various levels. On the system level, many researchers investigate how the brain operates to realize higher functions such as learning, memory, emotion, and so on. Cell assembly [1]-[3] is one of explanations of information processing in the brain, in which an information is represented by a firing space pattern of a group of plural neurons. This mechanism has been proposed by Hebb who is a psychologist in 1949. Cell assembly could achieve multiple information processing carried out at the same time and parallel dispersion processing of the brain by forming a functional circuit with some neurons according to the necessary processing. Although this concept is impressive, we consider that more complex mechanism is realized in the brain.

On the other hand, Back Propagation (BP) learning [4] is one of engineering applications of neural networks. The BP learning operates with a feedforward neural network which is composed of an input layer, a hidden layer and an output layer, and the effectiveness of the BP learning has been confirmed in pattern recognition, system control, signal processing, and so on [5]-[7]. When the BP learning is applied, the hidden layer plays the essential role to decide the performance of the neural network. Hence, for example, there have been a lot of reports on the number of the neurons in the hidden layer. However, there have not been many studies on changing the structure of the hidden layer to improve the performance on convergence speed or learning efficiency.

In this study, we propose a new network model which is a feedforward neural network with chaotically-selected affordable neurons in the hidden layer for more efficient BP learning. In this network, we prepare some extra neurons in the hidden layer. When the network executes the BP learning, all of the neurons in the hidden layer are not used at every updating. Namely, some of the neurons are selected for the learning and the rest of the neurons are deactivated. Further, the selected set of the neurons is changed in a chaotic manner at every updating. Computer simulated results show that the proposed network exhibits a good performance for the BP learning on both convergence speed and learning efficiency.

II. Network Model with Chaotically-Selected Affordable Neurons

Cell assembly is one of explanations of information processing in the brain, in which an information is represented by a firing space pattern of a group of plural neurons. This mechanism could achieve multiple information processing carried out at the same time and parallel dispersion processing of the brain by forming a functional circuit with some neurons according to the necessary processing. Conceptual diagram of cell assembly is shown in Fig. 1. Informations A and B are expressed by the corresponding functional connections between plural neurons. It is a special characteristics of cell assembly to exist overlapping neurons for the informations A and B. Although this concept is impressive, we consider that more complex mechanism is realized in the brain. For example, in the case of cell assembly, an information A is expressed by a specific group of the neurons. However, we consider that the same information A can be expressed by several different groups of neurons in the brain.

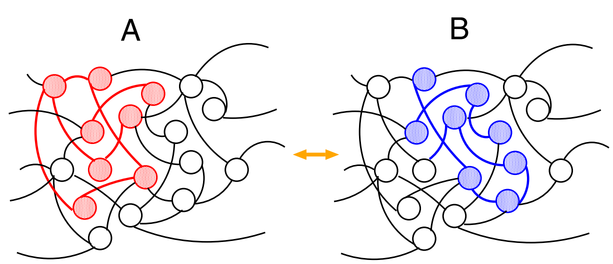


Fig. 1. Conceptual diagram of cell assembly.

A. *Network Model with Affordable Neurons*

We introduce affordable neurons in the hidden layer of the feedforward neural network to reflect a function of the brain. The extra neurons in the hidden layer are prepared in advance. During the BP learning, all of the neurons in the hidden layer are not used at every updating. Namely, some of the neurons are selected for the learning and the rest of the neurons are deactivated. The network model with the affordable neurons is shown in Fig. 2.

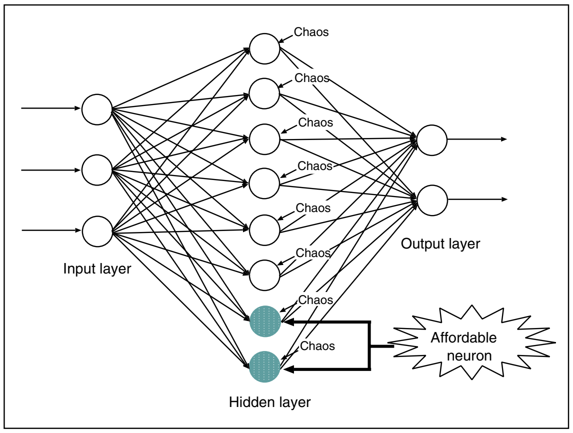


Fig. 2. Network model with affordable neurons.

B. Chaotic Selection

For the proposed network, some of the neurons have to be selected at every updating of the BP learning. The authors have investigated the performance of the Hopfield neural network solving combinatorial optimization problems when chaos is inputted to the neurons as noise [8]-[10]. By computer simulations, chaotic noise has been confirmed to gain better performance to escape out of local minima than random noise. Hence, we consider that various features of chaos are effective for neural networks.

In this study, we use the skew tent map as a simple chaotic map to realize chaotic selection of operating neurons in the hidden layer. We prepare the skew tent maps with different initial values, whose number is the same as the number of the neurons in the hidden layer, and each skew tent map is corresponded to each neuron. At the every updating of the BP learning, the skew tent maps are also updated and their values are referred. We select a certain number of the neurons in the order of the values of the skew tent maps. In this case, the selection is made in a chaotic manner. Note that chaos is not inputted to the neural network directly, but is used only for the selection of the operating neurons. The skew tent map is defined by the following equation and the map is shown in Fig. 3.

$$x_{n+1} = \begin{cases} \dfrac{2x_n + 1 - a}{1 + a} & (-1 \leq x_n \leq a) \\ \dfrac{-2x_n + 1 + a}{1 - a} & (a < x_n \leq 1) \end{cases} \quad (1)$$

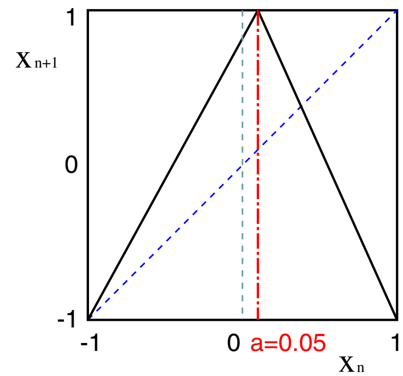


Fig. 3. Skew tent map.

III. BP Learning Algorithm

The standard BP learning algorithm was introduced in [4]. The BP is the most common learning algorithm for feedforward neural networks. In this study, we use the batch BP learning algorithm. The batch BP learning algorithm is expressed by similar formula of the standard BP learning algorithm. The difference lies in the timing of the update of the weight. The update of the standard BP is performed after each single input data, while the update of the batch BP is performed after all different input data.

The total error E of the network is defined as the following equation.

$$E = \sum_{p=1}^{P} E_p = \sum_{p=1}^{P} \left\{ \frac{1}{2} \sum_{i=1}^{N} (t_{pi} - o_{pi})^2) \right\} \quad (2)$$

where P is the number of the input data, N is the number of the neurons in the output layer, t_{pi} denotes the value of the desired target data for the pth input data, and o_{pi} denotes the value of the output data for the pth input data. The goal of the learning is to set weights between all layers of the network to minimize the total error E. In order to minimize E, the weights are adjusted according to the following equation:

$$\begin{aligned} w_{i,j}^{k-1,k}(m+1) &= w_{i,j}^{k-1,k}(m) + \sum_{p=1}^{P} \Delta_p w_{i,j}^{k-1,k}(m) \\ \Delta_p w_{i,j}^{k-1,k}(m) &= -\eta \frac{\partial E_p}{\partial w_{i,j}^{k-1,k}} \end{aligned} \quad (3)$$

where $w_{i,j}^{k-1,k}$ is the weight between the ith neuron of the layer $k-1$ and the jth neuron of the layer k, m is the learning time, and η is a proportionality factor known as the learning rate. In this study, we introduce inertia term in the 2nd term of the right-hand side of Eq. (4).

$$\Delta_p w_{i,j}^{k-1,k}(m) = -\eta \frac{\partial E_p}{\partial w_{i,j}^{k-1,k}} + \zeta \Delta_p w_{i,j}^{k-1,k}(m-1) \quad (4)$$

where ζ denotes the inertia rate.

IV. Simulated Results

We consider the feedforward neural network producing outputs x^2 for inputs data x as one learning example. The sampling range of the input data is $[-1.0,\ 1.0]$ and the step size of the input data is set to be 0.01. We carried out the BP learning by using the following parameters. The parameter of the inertia rate is fixed as $\eta = 0.03$ and initial values of the weights are given between -1.0 and 1.0 at random. The learning time is set to 20000. We investigate the convergence speed and the learning efficiency as the total error between the output and the desired target, when the network structure of the hidden layer is changed as follows.

[Hidden:2] 2 neurons in the hidden layer.
[Hidden:6] 6 neurons in the hidden layer.
[Hidden:8] 8 neurons in the hidden layer.
[Hidden:8(6)] 8 neurons are prepared in the hidden layer and only 6 neurons are operated at every update time.

The simulation results are shown in Fig. 4. The horizontal axis is the learning time and the vertical axis is the total error. We can see both of the convergence speed and the learning efficiency from the figure. Figures 4(a), (b) and (c) show the cases of $\eta = 0.01, 0.1$ and 0.6, respectively. In each figure, four graphs representing the results of the four cases [Hidden:2], [Hidden:6], [Hidden:8] and [Hidden:8(6)] are shown.

When the learning rate is $\eta = 0.01$ (Fig. 4(a)), we confirmed that the all of the learning curves are similar. In this case, the network goes down to a local minimum corresponding to a stable state very slowly. Therefore, we can confirm that the convergence speed is slow and the learning efficiency is not good. (Note that the scales of the vertical axes are different in the figures.)

When the learning rate is $\eta = 0.1$ (Fig. 4(b)), we confirmed that the improvement of the learning efficiency is difficult by only increasing the number of the neurons in the hidden layer. However, the proposed network with chaotically-selected affordable neurons in the hidden layer [Hidden:8(6)] gains a good performance. We consider that the network with chaotically-selected affordable neurons has an advantage to escape from local minima corresponding to some stable states because of its chaotic behavior.

When the learning rate is $\eta = 0.6$ (Fig. 4(c)), we confirmed that the learning curves of the traditional network becomes oscillatory and can not converge. However, the learning curve of the proposed network with chaotically-selected affordable neurons [Hidden:8(6)] is not oscillatory and converges quickly.

From these results, we can say that the proposed network with chaotically-selected affordable neurons gains a good performance for the BP learning on both convergence speed and learning efficiency.

V. Other selection methods

In order to confirm the effectiveness of the chaotic selection of the affordable neurons, in this section, we investigate the performance of the BP learning of the network with affordable neurons selected other non-chaotic methods, namely regular method and random method. For the regular method, the operating neurons in the hidden layer are selected in order. For the random method, the operating neurons in the hidden layer are selected completely at random.

The simulation results are shown in Fig. 5. Figures 5(a), (b) and (c) show the cases of $\eta = 0.01, 0.1$ and 0.6, respectively. In each figure, three graphs representing the results of the three cases "chaos," "regular" and "random" methods are shown.

From the figures, we can confirm that the learning curve for the regular method becomes oscillatory and can not converge for any η. Namely, if the affordable neurons are selected in a regular way, the BP learning does not succeed. We consider that this is very interesting result stimulating our motivation to investigate the proposed network more deeply from the viewpoint of learning process of real brain.

The learning curves for both of the chaos and the random methods are not oscillatory and converge to certain error values. For small learning rate in Fig. 5(a), the two curves are similar and they converge to similar value. This is because the network goes down to a stable state very slowly. For larger learning rates in Figs. 5(b) and (c), the learning curve for the chaos method shows much better performance than the random method. Namely, the random method does not oscillate, but does not gain a good learning efficiency. We consider that the chaos method is better than the random method in the sense that chaotic behavior has an advantage over random behavior to escape from local minima corresponding to some stable states.

VI. Conclusion

In this study, we have proposed a new network model which was a feedforward neural network with chaotically-selected affordable neurons in the hidden layer for more efficient BP learning. Computer simulated results showed that the proposed network with chaotically-selected affordable neurons gained a good performance for the BP learning on both convergence speed and learning efficiency. Further, we confirmed that the chaotic selection was important to obtain the good performance.

References

[1] Y. Sakurai, "Dependence of Functional Synaptic Connections of Hippocampa and Neocortical Neurons on Types of Memory," *Neuroscience Letters*, vol. 158, pp. 181-184, 1993.

[2] E. Vaadia, I. Haalman, M. Abeles, H. Bergman, Y. Prut, H. Slovin and A. Aertsen, "Dynamics of Neural Interactions in Monkey Cortex in Relation to Behavioral Events," *Nature*, vol. 373, pp. 515-518, Feb. 1995.

[3] Y. Sakurai, "Hippocampa and Neocortical Cell Assemblies Encode Memory Processes for Different Types of Stimuli in the Rat," *Journal of Neuroscience*, vol. 16, pp. 2809-2819, 1996.

[4] D.E. Rumelhart, G.E. Hinton and R.J. Williams, "Learning Internal Representations by Error Propagation," *Parallel Distributed Processing*, vol. 1, MIT Press, MA, pp. 318-362, 1986.

[5] Y. LeCun, B. Boser, J.S. Denker, et al, "Handwritten Digit Recognition with a Back-Propagation Network," *Advances in Neural Information Processing Systems*, vol. 2, MIT Press, MA, 1990.

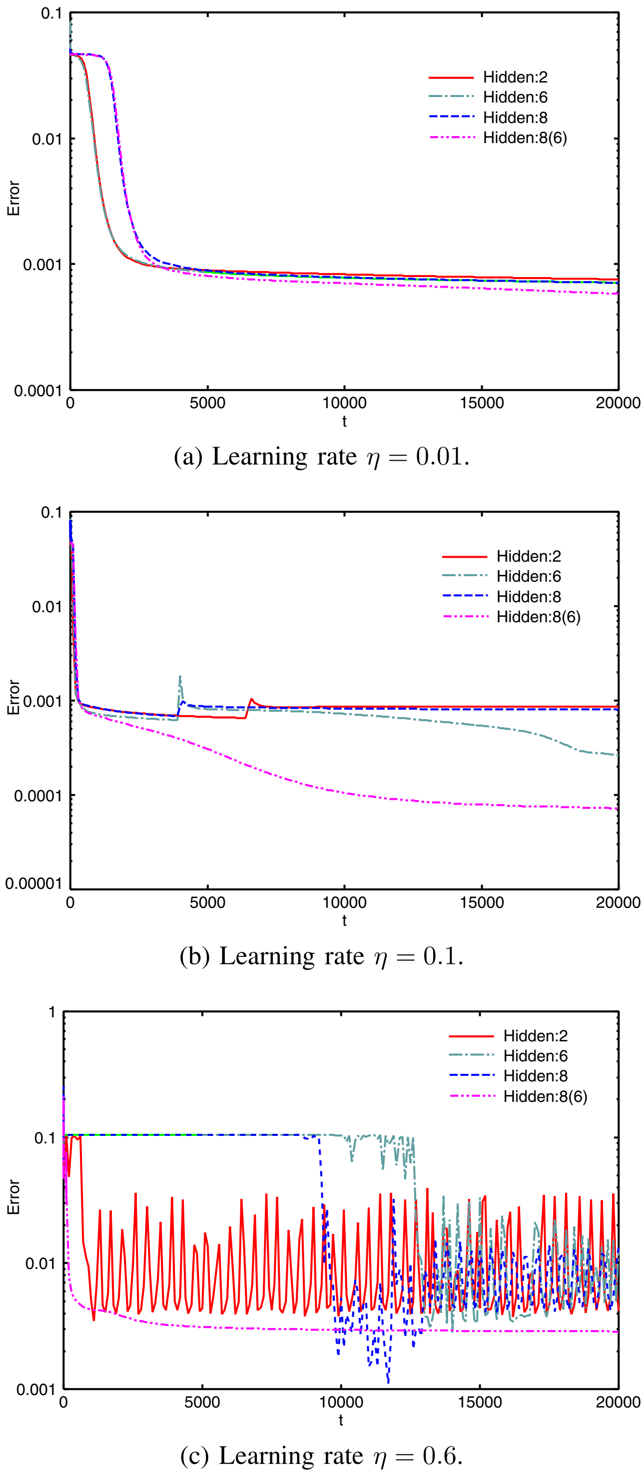
[6] M. Kawano, Y. Uno, M. Isobe and R. Suzuki, "A Hierarchical Neural Network Model for Voluntary Movement with Application to Robotics," *IEEE Control Systems Magazine*, vol. 8, pp. 8-16, 1988.

[7] T.J. Sejnowski and C.R. Rosenberg, "Parallel Networks that Learn to Pronounce English Text," *Complex Systems*, vol. 1, pp. 145-168, 1987.

[8] T. Ueta, Y. Nishio and T. Kawabe, "Comparison between Chaotic Noise and Burst Noise on Solving Ability of Hopfield Neural Networks," *Proc. NOLTA'97*, vol. 1, pp. 409-412, Nov. 1997.

[9] Y. Uwate, Y. Nishio and A. Ushida, "Markov Chain Modeling of Intermittency Chaos and its Application to Hopfield NN," *IEICE Transactions on Fundamentals*, vol. E87-A, no. 4, pp. 774-779, Apr. 2004.

[10] Y. Uwate, Y. Nishio, T. Ueta, T. Kawabe and T. Ikeguchi, "Performance of Chaos and Burst Noise Injected to the Hopfield NN for Quadratic Assignment Problems," *IEICE Transactions on Fundamentals*, vol. E87-A, no. 4, pp. 937-943, Apr. 2004.

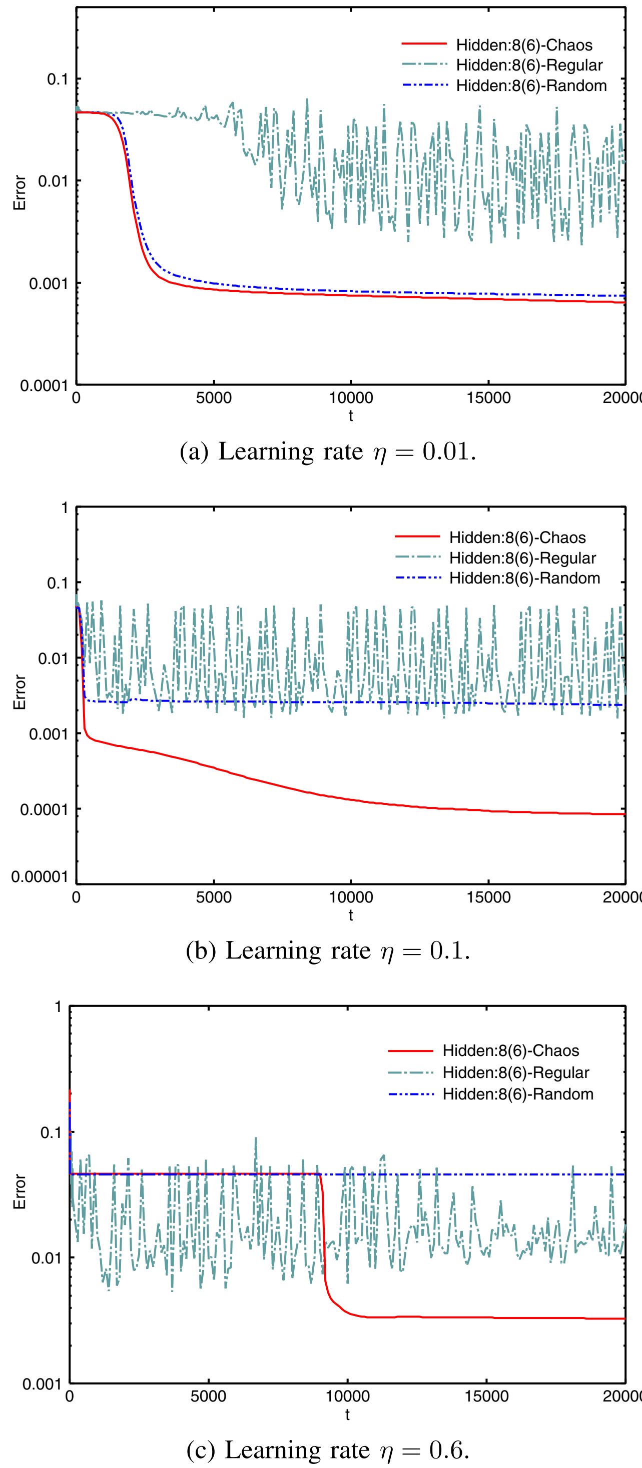


(a) Learning rate $\eta = 0.01$.

(b) Learning rate $\eta = 0.1$.

(c) Learning rate $\eta = 0.6$.

Fig. 4. Convergence of learning for different network structures.



(a) Learning rate $\eta = 0.01$.

(b) Learning rate $\eta = 0.1$.

(c) Learning rate $\eta = 0.6$.

Fig. 5. Convergence of learning for different selection methods of affordable neurons.