

Characteristics of Multi-Layer Perceptron Switching between Learning and Non-Learning Neurons

Chihiro Ikuta

Dept. of Electrical and Electronic Eng., Dept. of Electrical and Electronic Eng., Dept. of Electrical and Electronic Eng.,
Tokushima University Tokushima University Tokushima University
2-1 Minami-Josanjima, 2-1 Minami-Josanjima, 2-1 Minami-Josanjima,
Tokushima Japan Tokushima Japan Tokushima Japan
Email: ikuta@ee.tokushima-u.ac.jp Email: uwate@ee.tokushima-u.ac.jp Email: nishio@ee.tokushima-u.ac.jp

Yoko Uwate

Yoshifumi Nishio

Abstract—A glia has important functions for a higher brain functions. Among them, we notice that the glia influences the synaptic Long-Term Potentiation (LTP). The fired glia increase a Ca^{2+} concentration. The Ca^{2+} generates a d-serine. The d-serine happens the synaptic LTP. In previous study, we proposed the Multi-Layer Perceptron (MLP) with switching learning and non-learning neurons. In this method, neurons in the hidden-layer are connected with glia and separated to some groups according to the connected glia. The glia are regularly fired. The neurons are switched the learning and non-learning terms according to the glia firing. We consider that switching between two terms is efficient to the MLP performance. By the simulations, we confirm that the change of the number of neurons is how to influence to the learning performance.

standard MLP is often falls into local minimum, because the BP uses the steepest decent method. The proposed MLP has learning neurons and non-learning neurons during iterations. We consider that the non-learning neurons gives an energy to the MLP for escaping out from the local minimum. The non-learning neurons periodically change to the learning neurons with time. The hidden-layer neurons repeat the switching of the learning term and the non-learning term. By the simulations, we confirm the performance and characteristics of the proposed MLP.

I. INTRODUCTION

II. PROPOSED METHOD

Glia had not been investigated, because it was considered to the support cell of a neuron. However, some researchers discovered that the glia transmits signal by using various ions [1][2]. For example, the ions are an adenosine triphosphate, a glutamate acid, a calcium, and so on [3][4]. They are important for the brain works. Actually, they are often used in the synapse. Among them, we have noticed the Ca^{2+} . Takatas reported that the Ca^{2+} affects the synaptic Long-Term Potentiation (LTP) by an experiment on living animals [5]. In this research, the increase of the Ca^{2+} concentration induces the d-serine at the periphery of the fired synapse. The increase of the Ca^{2+} concentration is generated by the glia. If the glia does not increase the Ca^{2+} , the d-serine is not generated. The d-serine relates closely the LTP, thus the LTP is not observed when the increase of the Ca^{2+} is not happened.

In previous study, we proposed a Multi-Layer Perceptron (MLP) with switching between learning and non-learning neurons which is presented in NOLTA'12 [6]. In the biological system, the neurons are learned in the increase of the Ca^{2+} concentration, moreover the glial effect is propagated. We attract these features and apply these features to the MLP. We connected the glia with the neurons in the hidden-layer and the neurons are separated to some groups. The glia are regularly fired. The neurons in a same group learn at same time. Thus the neurons are switched between learning and non-learning term according to the firing of glia. The MLP learns by a Back Propagation (BP) algorithm [7]. However, the

The MLP is a famous feed forward neural network. This network is composed of layers of neurons and the outputs are tuned by changing the weights of connections. In general, it is learned by the BP which was proposed by D.E. Rumelhart [7]. The MLP can be applied to a pattern recognition, a pattern classification, a data mining, and so on. However, the MLP has the local minimum problem, because the BP uses the steepest decent method. For the local minimum problem, we often add a noise to any part of the MLP.

The MLP with switching between learning and non-learning neurons was proposed in NOLTA'12. The proposed MLP is inspired from relationships the neurons and the glia in the biological system. The glia is one of nervous cells in the brain. This cell can change the Ca^{2+} concentration which is propagated to neighboring glia [8][9]. The Ca^{2+} is important for the LTP of the synapse. In order to increase the response amplitudes of the synapse by the learning, the glia must increase the Ca^{2+} concentration [5]. We apply these features to the MLP. In the proposed MLP, the hidden-layer neurons are separated some learning groups. The neurons in the same group learn by the BP algorithm at same time. The groups periodically switch between the learning term and the non-learning term. During the learning term, the hidden-layer neurons are learned. During the non-learning term the weights of connections between the hidden-layer and the output layer are not updated.

A. Updating rule of neuron

The neuron has multi-inputs and a single output. We can change the neuron output by tuning weights of connections. The updating rule of the neuron is defined by Eq. (1).

$$y_i(t+1) = f\left(\sum_{j=1}^n w_{ij}(t)x_j(t) - \theta_i(t)\right), \quad (1)$$

where y is an output of the neuron, w is a weight of connection, x is an input of the neuron, and θ is a threshold of neuron. In this equation, the weights of the connections and the thresholds of neurons are learned by BP algorithm. We used sigmoidal function to an activating function which is described by Eq. (2).

$$f(a) = \frac{1}{1 + e^{-a}} \quad (2)$$

where a is an inner state.

B. Flow of the proposed MLP

We show the flow of the learning of the proposed MLP in Fig. 1. In this example, one group is composed of two neurons. Every group has same time length of the learning term and the non-learning term. The learning term and the non-learning term are changed with time. During the learning term, the neurons are learned by BP algorithm. On the other hand, the weights of connections between the hidden-layer neuron and the output-layer neuron are not updated. First group goes into the learning term. Other groups are the non-learning term. Second group goes into the learning term, then the first group remain the learning term. In the middle of the learning term of the second group, the first group finishes the learning term and start the non-learning term. After that, every group repeats changing the learning term and the non-learning term.

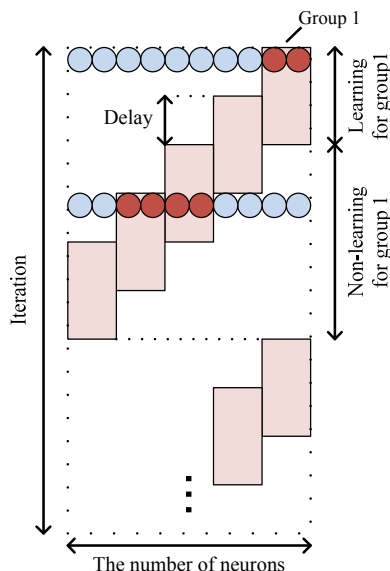


Fig. 1. Flow of the learning of the proposed MLP.

III. SIMULATIONS

In this simulation, we use the Two-Spiral Problem (TSP) for a learning task. The TSP is a famous task for the artificial neural network and it has a high nonlinearity. In this task, the coordinates of the spirals are input to the MLP and the MLP learns the correlating classifications. The input spirals are shown as Fig. 2.

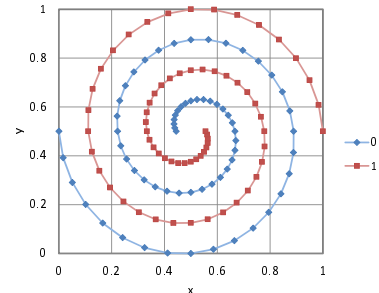


Fig. 2. Two-Spiral Problem.

The average of error is obtained from 100 trials and each trial has 100000 iterations. Moreover, we fix the delay of the learning term to 5. We use a Mean Square Error (MSE) to an error function. The MSE is described by Eq. (3).

$$MSE = \frac{1}{N} \sum_{n=1}^N (T_n - O_n)^2, \quad (3)$$

A. Approximation performance

We compare approximation performances the proposed MLP and the standard MLP. The simulation result is shown Table I. From this table, the proposed MLP has better performance than the standard MLP. The TSP is difficult task, thus the standard MLP is trapped into local minimum. From comparison result, the proposed MLP can escape out from the local minimum. Moreover, the proposed MLP has a high solution stability.

TABLE I
LEARNING PERFORMANCE.

	Average	Minimum	Maximum	Std. Dev.
Standard	0.1030	0.0010	0.2237	0.0549
Proposed	0.0358	0.0002	0.1694	0.0322

B. Convergence performance

We show the convergence performance of the proposed MLP shown as Fig. 3. The decrease of error is fixed about 100000 iterations in the both MLPs. However, in the proposed MLP, the error is rapidly decreased from 50000 iterations to 100000 iterations. We can see that the influence of the proposed method is slowly started. Because the proposed MLP has the non-learning term. During the non-learning term, the weight updating is skipped, thus the decrease of the error is slow.

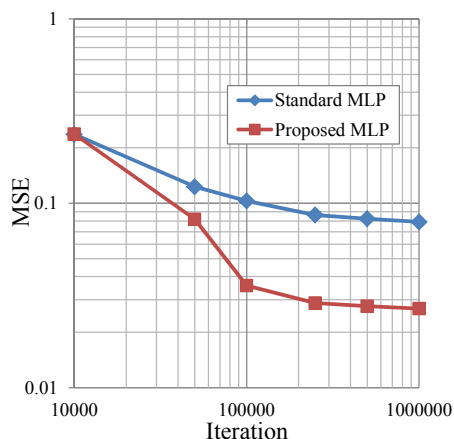


Fig. 3. The convergence performance of the proposed MLP.

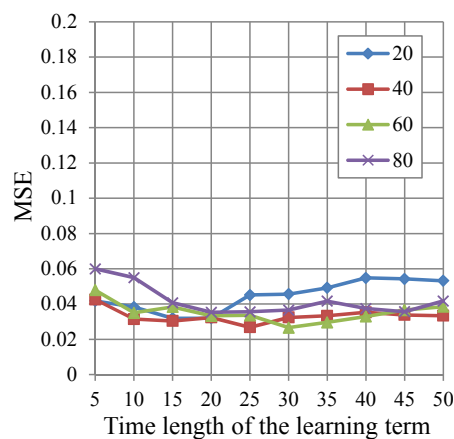


Fig. 5. The performance of different number of neurons (5 neurons in the same group).

C. Number of neurons in the hidden-layer

Figure 1 shows the performances of the proposed MLP when we change the number of neurons in the hidden-layer. The proposed MLP which has 40 neurons is the best of all. In general, the MLP obtains a better performance if we use the larger number of neurons, however, the MLP which has large number on neurons needs a high simulation cost and a longer iterations. In this case, the proposed MLP which has 80 neurons is the worst performance of all. We consider that the proposed MLP which has 80 neurons cannot converge, thus it obtains the worst result.

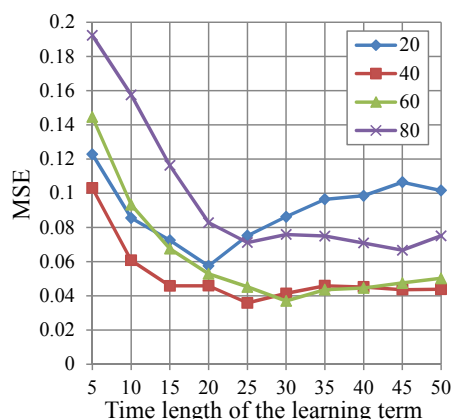


Fig. 4. The performance of different number of neurons (5 neurons in the same group).

We increase the iterations to 1000000 and obtain the result. Figure 5 shows the simulation results. In this case, every result is similar to each other. From this result, the proposed MLP needs the longer iterations when we use the large number of neurons. Moreover the proposed MLP which has 40 neurons can balance an approximation performance with a convergence performance.

Next, we change the number of neurons in the same group. In this simulation, we use 100000 iterations for the MLPs. Figure 6 simulations result when we use 10 neurons to one

learning group. We confirmed that all MLP can obtain the similar performances to Fig. 5 if they can converge. In this simulation, the proposed MLPs which have 60 and 80 neurons improves the performances than Fig. 4. On the other hand, the performance of proposed MLP which has 40 neurons worth than Fig. 4 when we use the large time length of the learning term. We consider that this MLP becomes similar to the standard MLP, because we increase the number of neurons in the same group and the time length of the learning term. Thus, the proposed MLP which has 40 neurons is trapped into local minimum similar to the standard MLP.

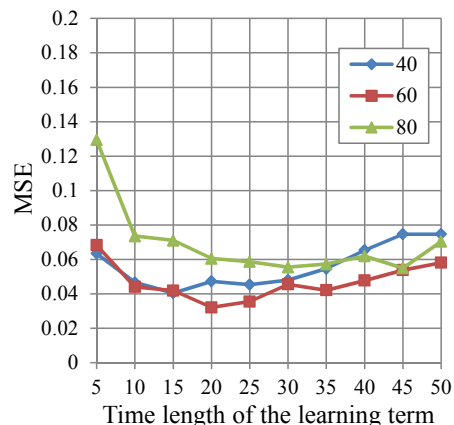


Fig. 6. The performance of different number of neurons (10 neurons in the same group).

In addition, we increase the number of neurons in the same learning group to 20. The simulation results are shown as Fig. 7. We can see that the performances improve as the MLP having the large number of neurons. Moreover, we can say that the proposed MLP has a negative correlation between the number of neurons in the same group and the time length of the learning term.

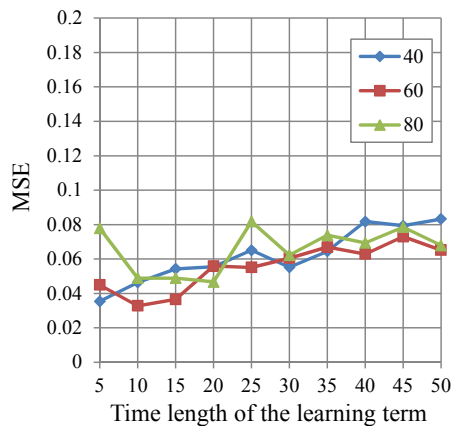


Fig. 7. The performance of different number of neurons (10 neurons in the same group).

IV. CONCLUSIONS

In this study, we have investigated about the MLP with switching between learning and non-learning neurons which is inspired from the biological characteristics of the glia. In this model, the hidden-layer neurons are connected with the glia and are separated to some groups. The neurons in the same group is learned at same time. The glia are regularly fired. Every learning group is switched between learning and non-learning term according to the firing of glia. By the simulation, we showed that the proposed method is efficient to the MLP learning. Moreover, we confirmed the characteristic of the proposed MLP.

ACKNOWLEDGMENT

This work was partly supported by MEXT/JSPS Grant-in-Aid for JSPS Fellows (24-10018).

REFERENCES

- [1] P.G. Haydon, "Glia: Listening and Talking to the Synapse," *Nature Reviews Neuroscience*, vol. 2, pp. 844-847, 2001.
- [2] S. Koizumi, M. Tsuda, Y. Shigemoto-Nogami and K. Inoue, "Dynamic Inhibition of Excitatory Synaptic Transmission by Astrocyte-Derived ATP in Hippocampal Cultures," *Proc. National Academy of Science of U.S.A.*, vol. 100, pp. 11023-11028, Mar. 2003.
- [3] S. Ozawa, "Role of Glutamate Transporters in Excitatory Synapses in Cerebellar Purkinje Cells," *Brain and Nerve*, vol. 59, pp. 669-676, 2007.
- [4] G. Perea and A. Araque, "Glial Calcium Signaling and Neuro-Glia Communication," *Cell Calcium*, vol. 38, pp. 375-382, 2005.
- [5] N. Takata, T. Mishima, C. Hisatsune, T. Nagai, E. Ebisui, K. Mikoshiba, and H. Hirase, "Astrocyte Calcium Signaling Transforms Cholinergic Modulation to Cortical Plasticity In Vivo," *The Journal of Neuroscience*, vol. 31, pp. 18155-18165, Dec. 2011.
- [6] C. Ikuta, Y. Uwate, Y. Nishio, and Guoan Yang, "Multi-Layer Perceptron Decided Learning Neurons by Regular Output Glia," *Proc. NOLTA'12*, pp. 719-722, Oct. 2012.
- [7] D.E. Rumelhart, G.E. Hinton and R.J. Williams, "Learning Representations by Back-Propagating Errors," *Nature*, vol. 323-9, pp. 533-536, 1986.
- [8] S. Kriegler and S.Y. Chiu, "Calcium Signaling of Glial Cells along Mammalian Axons," *The Journal of Neuroscience*, vol. 13, pp. 4229-4245, 1993.
- [9] M.P. Mattoson and S.L. Chan, "Neuronal and Glial Calcium Signaling in Alzheimer's Disease," *Cell Calcium*, vol. 34, pp. 385-397, 2003.

- [10] J.R. Alvarez-Sanchez, "Injecting knowledge into the Solution of the Two-Spiral Problem," *Neural Computing & Applications*, vol. 8, pp. 265-272, 1999.
- [11] H. Sasaki, T. Shiraishi and S. Morishita, "High precision learning for neural networks by dynamic modification of their network structure," *Dynamics & Design Conference*, pp. 411-1-411-6, 2004.