

Multi-Layer Perceptron Including Glial Pulse and Switching between Learning and Non-Learning

Chihiro Ikuta[†], Yoko Uwate[†], Yoshifumi Nishio[†], and Guoan Yang[‡]

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

Dept. of Automation Science and Technology, School of Electronic and Information Eng.,
Xi'an Jiaotong University, No. 28 Xianning West Road, Xi'an City, Shaanxi Province, China
Email: gayang@mail.xjtu.deu.cn

Abstract—A glia is a nervous cell which is existing in a brain. This cell changes a Ca^{2+} concentration. This ion affects a neuron membrane potential and it is propagated to the neighboring glia. Moreover, the Ca^{2+} directly affects the human memory by increasing of a D-serine. From these functions, we propose a Multi-Layer Perceptron (MLP) including glial pulse and switching between a learning and non-learning. In this method, the neurons in the hidden-layer received the pulse from connected glias. The pulse is generated depending on the neuron outputs and it is propagated to the neighboring glias and neurons. Moreover, the neurons are separated to some groups. Each group periodically switches a learning term and a non-learning term. Each group starts the learning term having a small lag each other. We consider that a performance of the MLP improves by two different methods influencing each other. By two simulations, we confirm that the MLP obtains the high solving ability by using our methods.

I. INTRODUCTION

The functions of human brain are composed of nervous cells which are a neuron and a glia. The neuron transmits signals to other neurons by using an electrical stimulus. However, the glia did not attract attention, because many researchers considered that the glial function is only a support of the neurons. Recently, some researchers discovered the novel glial functions [1][2]. The glia uses various ions concentrations which become a transporter of the signals [3][4]. We notice the Ca^{2+} , because the Ca^{2+} transmits to a wide range in the brain and relates closely to the neuron works. We have proposed the several applications of the glial functions to Multi-Layer Perceptron (MLP).

We know that Ca^{2+} to be generated by the glia affects the state of the neighboring glias and the neuron membrane potential [5][6]. From this function, we proposed the MLP with positive and negative pulse glial chain [7]. In this method, the neurons in the hidden-layer receive the pulse from the connected glia. This pulse affects neuron threshold and it excites the neighboring glias. The excited glia also generates the pulse, thus the pulse is propagated to the wide range. In the MLP with positive and negative pulse glial chain, this network can solve the tasks with a high nonlinearity, however the pulse disrupts the solution convergence in easy tasks.

On the other hand, Takatas reported that a Ca^{2+} affects the synaptic Long-Term Potentiation (LTP) by an experiment on living animals [8]. In that research, they used two mice. One

mouse has a normal glia. The other mouse has the genetically-modified glia which cannot increase the Ca^{2+} concentration. In the mouse with genetically-modified glia, the neuron could not observe the LTP. In the normal mouse, the researchers observed the LTP from the increase of the response. Moreover, the increase of the D-serine was observed in increasing the Ca^{2+} concentration. The D-serine is an important ion for LTP. Thereby, we can say that the glia relates closely to the LTP of the synapse. We considered that this experimental result can be applied to an artificial neural network. We proposed the MLP with the learning and non-learning terms [9]. In this method, the neurons in the hidden-layer are separated to some groups and have the learning term and the non-learning term. Two terms are periodically switched and the start times of the learning term are different in each group. On the other hand, the MLP with learning and non-learning terms has a better performance than the standard MLP in the easy and difficult tasks. However, this method basically depends on Back Propagation (BP) algorithm [10]. Thus, this MLP does not have the superiority for the tasks with high nonlinearity.

In this study, we propose a Multi-Layer Perceptron (MLP) including the glial pulse and switching between the learning and the non-learning. We combine two glial application models which have mutually different characteristics. When the pulse is input to the MLP, the amount of updates of weights are increased. Thus the MLP can obtain the energy for escaping out from the local minimum, however the learning becomes unstable by this pulse. In the proposed MLP, only a part of the weights is updated, because the weights are not updated during the non-learning term. The learning and non-learning terms inhibit that the learning becomes unstable. We believe that the MLP obtains both characteristics by combining two glial methods. By the simulations of two different tasks, we confirm the performance of the proposed MLP.

II. PROPOSED METHOD

The MLP is a famous feed forward neural network. It is applied to an approximation of function, a classification task, a data mining, and so on. For the MLP learning, the BP algorithm is proposed by D.E. Rumelhart [10]. In this algorithm, updated weights are calculated for decreasing an error from a partial differentiation of error of the MLP.

However the solution of the MLP is often falls into a local minimum. Because the BP algorithm uses the steepest decent method. For this problem, we use a noise. We hope that the noise gives the energy for escaping out from the local minimum. The noise is efficient to difficult tasks, because a difficult task has many local minima. However, the noise encumbers the learning when the MLP solves the easy task. In this study, we propose the MLP including the glial pulse and switching between the learning and non-learning which is inspired from the glial works in the brain. We give two glial works to the MLP. The construction of the proposed MLP is presented in Fig. 1.

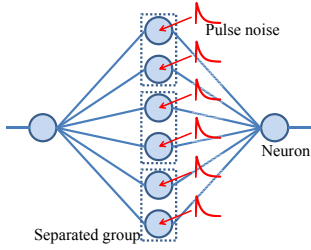


Fig. 1. Proposed MLP.

A. The method of pulse doping

We proposed the efficient method of doping pulse for the MLP [7]. It is inspired from the biological glia. The pulse is generated depending on the neuron output. If the neuron has a large output, the positive pulse is generated. If the neuron has a small output, the negative pulse is generated. Moreover, the pulses are propagated to the neighboring neurons. We confirmed that the pulses which have position relationships improve the learning performance. The glia has two different states which are the positive response and the negative response. We define the output function as the positive response of the glia in Eq. (1).

$$\psi_i(t+1) = \begin{cases} 1, & \{(\theta_n < y_i \cup \psi_{i+1,i-1}(t-i*D) = 1) \\ & \cap (\theta_g > |\psi_i(t)|)\} \\ \gamma\psi_i(t), & \text{else,} \end{cases} \quad (1)$$

where ψ is an output of a glia, γ is an attenuated parameter, y is an output of a connecting neuron, θ_n is a glia threshold of excitation, θ_g is a period of inactivity, and D is a delay time of a glial effect. Moreover, we define the output function for the negative response of the glia in Eq. (2).

$$\psi_i(t+1) = \begin{cases} -1, & \{(1-\theta_n > y_i \cup \psi_{i+1,i-1}(t-i*D) = -1) \\ & \cap (\theta_g > |\psi_i(t)|)\} \\ \gamma\psi_i(t), & \text{else,} \end{cases} \quad (2)$$

The neuron has multi-inputs and a single output. We can change the neuron output by tuning of the weights of connections. The standard updating rule of the neuron is defined by Eq. (3).

$$y_i(t+1) = f \left(\sum_{j=1}^n w_{ij}(t)x_j(t) - \theta_i(t) \right), \quad (3)$$

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. Next, we show a proposed updating rule of the neuron. We add the glial effect to the threshold of neuron, because the glial effect influences to the membrane potential in the biological system. This updating rule is used to neurons in the hidden layer. It is described by Eq. (4).

$$y_i(t+1) = f \left(\sum_{j=1}^n w_{ij}(t)x_j(t) - \theta_i(t) + \alpha\psi_i(t) \right), \quad (4)$$

where α is a weight of the glial effect. We can change the glial effect by tuning α . In this equation, the weight of connection and the threshold are changed by BP algorithm as same as the standard updating rule of the neuron. However, the glial effect is not changed. It is updated by Eq. (1).

B. The method of the switching between learning term and the non-learning term

The neurons in the hidden-layer are separated to some groups. Each group has the learning term and the non-learning term. Two terms are periodically changed. The weights of connections are tuned according to the standard BP algorithm. However, the weights between the input-layer and the hidden-layer are only updated during the learning term. Thus, during the non-learning term, the MLP charges the error, because the other weights are changed. By this method, the MLP obtains a better performance in easy and difficult tasks than the standard MLP. This method is useful to some tasks, however, it cannot escape out from the deep local minimum. Because this method basically depends on the standard BP algorithm.

C. Time flow of the proposed method

We show the flow of the learning of the proposed MLP in Fig. 2. The pulse is generated by the connected glia depending on the neuron output. By this pulse, we believe that the MLP escapes out from the local minimum. The generated pulse is propagated to the neighboring neurons with the iterations. 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 periodically changed. During the learning term, the neurons are learned by BP algorithm. During the non-learning term, the weights of connections between the hidden-layer neuron and the input-layer neurons are not updated. First group goes into the learning term. Other groups sojourn to the non-learning term. Second group goes into the learning term, then the first group remains the learning term. In the middle of the learning term of the second group, the first group goes into the non-learning term. After that, every group repeats changing between the learning term and the non-learning term. During the iterations, the MLP receives the pulse from the glia.

The pulse doping is switching between the learning and non-learning terms have different characteristics. The pulse locally

gives a large energy to the MLP. When the pulse is input to the MLP, the updated weights are increased. Thus the learning becomes unstable by this pulse. By combining of two kinds of methods, only a part weights are updated, moreover the MLP can obtain the energy from the pulse. In the proposed method, the MLP receives the pulse and this pulse affects only a part of weights. We consider that the MLP can search good solutions in the wide range by including these methods.

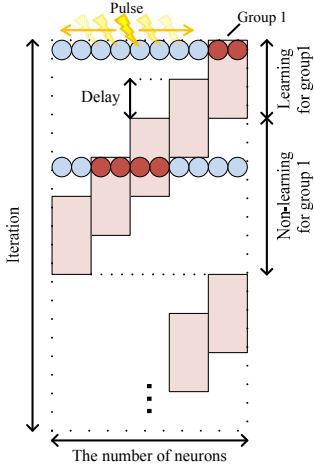


Fig. 2. Time flow of the learning term and non-learning term.

III. SIMULATIONS

In this section, we show two simulation results. We give two kinds of tasks to the MLP. Firstly, the MLP learns the logistic map. This task shows the performance of the function approximation. The logistic map is described by one-dimensional map, thus, we presume that it is easily solved by the MLP. Secondly, the MLP computes the Two-Spiral Problem (TSP). This task has high nonlinearity. From two simulations, we confirm the solving performance of the proposed MLP. We compare the performance of the proposed MLP to the standard MLP, MLP with positive and negative pulse glial chain (previous 1), and MLP with learning and non-learning terms (previous 2). In the MLP with positive and negative pulse glial chain, the neurons in the hidden-layer receive the pulses from connected glias. The MLP with learning and non-learning terms has the learning term and the non-learning term. For the index of the error, we use the Mean Square Error (MSE) for both simulation results. The MSE is described by Eq. (5).

$$MSE = \frac{1}{P} \sum_{n=1}^P (T_n - O_n)^2, \quad (5)$$

where P is the number of the pairs of the input and the supervised value, T is a supervised value, O is an output of the MLP.

A. Simulation results of learning logistic map

The logistic map is described by Eq. (6).

$$\chi(t+1) = a\chi(t)\{1 - \chi(t)\}, \quad (6)$$

where a is control parameter of chaos. We use $a = 4$, then, the logistic map generates fully-developed-chaos. The MLP receives $\chi(t)$ as input and learns $\chi(t+1)$. In this simulation, the neurons are arranged in 1-12-1 and the iterations are 10000 times. The simulation results are shown in Fig. 3. We show three curves which are the proposed MLP, the standard MLP, and the previous 1. In the proposed MLP (the noise amplitude is zero), this MLP is same as the previous 2, because the influence of pulse is removed. We change the pulse amplitude which is input to the neurons threshold of the hidden-layer. From this result, the learning term and the non-learning term are efficient to the MLP performance. Because the proposed MLP is better than the standard MLP without reference to the pulse amplitude. The previous 1 has the pulse in the thresholds of the hidden-layer neurons. In general, the noise doping improves the performance because it give the energy for escaping out from the local minimum. In the easy task, the standard MLP is not trapped into the local minimum and rapidly finds a good solution. The average error of the previous method increases according to the increase of the pulse amplitude. We consider that the pulse disrupts the MLP learning. Although the proposed MLP and the previous MLP have the same generation mechanism of the pulse in the hidden-layer, the performances of the MLPs are different. Moreover, the performance of the proposed MLP is the best of all when the noise amplitude is 0.02. From this result, we can say that the two glial effects improves the MLP performance by the mutual influence.

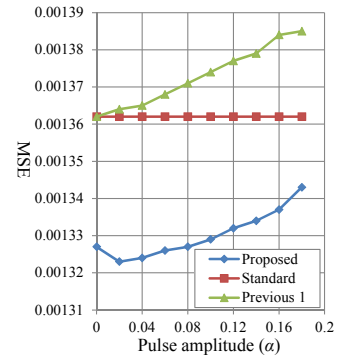


Fig. 3. Results of the MLPs for learning the logistic map.

B. Simulation results of learning TSP

The TSP is a famous task for the artificial neural network [11]. It is a linearly-inseparable problem, thereby it has a high nonlinearity [12]. The MLP receives the coordinates of the spirals for the inputs and learns the corresponding classifications. In this simulation, we increase the number of neurons in the hidden-layer. The MLP is composed of the neurons (constructed 2-40-1). The iterations are 1000000 times. Figure 4 shows the two spirals for the MLP. In this figure, the spirals are composed of 130 points.

The simulation result is summarized in Table I. The average error of the standard MLP is the worst of all. Because the TSP has high nonlinearity, thus the standard MLP is trapped into

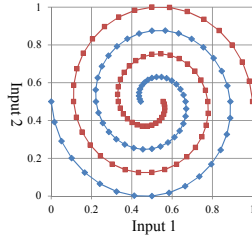


Fig. 4. Two-spiral Problem.

the local minimum. The simulation result of the previous 2 is worse than the previous 1. The previous 2 has the learning term and the non-learning term. During the non-learning term, the weights are not updated, thus it affects the MLP learning. However, it basically depends on the BP algorithm. This network cannot escape out from the deep local minimum. The previous 1 is better than the proposed MLP and the previous 2. The pulse locally gives a large energy to the MLP. The pulse helps the MLP for escaping out from the local minimum. In the proposed MLP, the pulse and the learning and non-learning terms improves the learning performance. The pulse helps MLP learning for escaping out from a deep local minimum. Moreover, the non-learning term inhibits the learning curve becoming oscillatory. Because the weights are not updated during the non-learning term. The MLP has the large error by the pulse, however only a part of weights is updated.

TABLE I
LEARNING PERFORMANCE.

	Average	Minimum	Maximum	Std. Dev.
Proposed	0.00904	0.00027	0.04685	0.01161
Standard	0.09604	0.00021	0.23087	0.05751
Previous 1	0.01175	0.00011	0.05417	0.01314
Previous 2	0.04830	0.00041	0.16941	0.03763

Next, we show an example of learning curves during the iterations. The learning curve of the standard MLP is converged earlier than the others and this curve becomes smooth. Because the solution of the MLP converges to the nearest local optimum solution from the initial condition. The MLP with learning and non-learning terms (previous 2) obtain a better solution than the standard MLP. However, this orbit is similar to the standard MLP. The oscillation of this curve is hardly observed. From this result, we can say that the learning basically depends on the BP algorithm. The curves of other two methods often oscillate during the iterations. We can see that the error is rapidly decreased after oscillation of the learning curve. From this behavior, the MLP is obtained the energy from the doping pulse, thus the MLP can escape out from the local minimum. In the proposed MLP, the learning curve becomes step-like after 60000 times. We consider that it happens by the influences both the pulse and the learning and non-learning terms. Because it has both characteristics of two previous method.

IV. CONCLUSION

We have proposed the MLP including the glial pulse and switching between learning and non-learning. In our method,

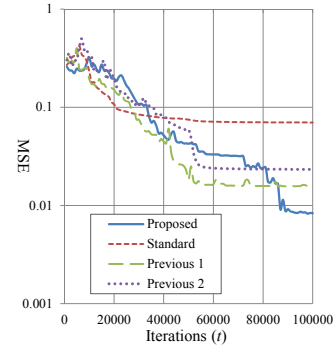


Fig. 5. Learning curves of three MLPs.

the neurons in the hidden-layer have two different methods. Firstly, these neurons receive the pulse to the thresholds. Secondly, these neurons are separated to some groups. Each group periodically switches the learning and non-learning terms. The neurons are only updated to the weights during learning terms. By two kinds of simulations, we showed the performance of the proposed MLP. From the results, we confirm that two kinds of methods are efficient to the MLP learning. Moreover, the MLP performance is improved by two methods corresponding each other.

ACKNOWLEDGMENT

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

REFERENCES

- [1] P.G. Haydon, "Glial: Listening and Talking to the Synapse," *Nature Reviews Neuroscience*, vol. 2, pp. 844-847, 2001.
- [2] S. Kriegler and S.Y. Chiu, "Calcium Signaling of Glial Cells along Mammalian Axons," *The Journal of Neuroscience*, vol. 13, pp. 4229-4245, 1993.
- [3] S. Ozawa, "Role of Glutamate Transporters in Excitatory Synapses in Cerebellar Purkinje Cells," *Brain and Nerve*, vol. 59, pp. 669-676, 2007.
- [4] M.P. Mattoson and S.L. Chan, "Neuronal and Glial Calcium Signaling in Alzheimer's Disease," *Cell Calcium*, vol. 34, pp. 385-397, 2003.
- [5] 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.
- [6] G. Perea and A. Araque, "Glial Calcium Signaling and Neuro-Glia Communication," *Cell Calcium*, vol. 38, pp. 375-382, 2005.
- [7] C. Ikuta, Y. Uwate, and Y.Nishio, "Multi-Layer Perceptron with Positive and Negative Pulse Glial Chain for Solving Two-Spirals Problem," *Proc. IJCNN'12*, pp. 2590-2595, Jun. 2012.
- [8] 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.
- [9] C. Ikuta, Y. Uwate, Y. Nishio, and G. Yang, "Multi-Layer Perceptron Decided Learning Neurons by Regular Output Glial," *Proc. NOLTA'12*, pp. 719-722, Oct. 2012.
- [10] D.E. Rumelhart, G.E. Hinton and R.J. Williams, "Learning Representations by Back-Propagating Errors," *Nature*, vol. 323-9, pp. 533-536, 1986.
- [11] J.R. Alvarez-Sanchez, "Injecting knowledge into the Solution of the Two-Spiral Problem," *Neural Computing & Applications*, vol. 8, pp. 265-272, 1999.
- [12] 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.