IEICE TRANS. FUNDAMENTALS, VOL.E94–A, NO.9 SEPTEMBER 2011

LETTER

# Multi-Layer Perceptron with Glial Network for Solving Two-Spiral Problem

**Chihiro IKUTA**[†a)], ***Nonmember***, **Yoko UWATE**[†b)], ***and*** **Yoshifumi NISHIO**[†c)], ***Members***

**SUMMARY** In this study, we propose a multi-layer perceptron with a glial network which is inspired from the features of glias in the brain. All glias in the proposed network generate independent oscillations, and the oscillations propagate through the glial network with attenuation. We apply the proposed network to the two-spiral problem. Computer simulations show that the proposed network gains a better performance than the conventional multi-layer perceptron.

***key words:*** *multi-layer perceptron, back propagation, glia, glial network, two-spiral problem*

## 1. Introduction

A human's brain is made by many nervous cells; neurons and glias. The neurons have been extensively studied by many researchers, because we can see that the neurons actively work in the brain. They transmit electrical signals to other neurons in the network. The glias had not been noticed for a long time compared with the neurons, because we could not observe the activity of the glias clearly. Currently, we can research the activity of the glias by checking ions in the brain. Some researchers discovered that the glias have some ion receptors [1], and that the glias transmit signals to other glias and neurons using by waves of ions; calcium ion ($Ca^{2+}$), glutamic acid (Glu), adenosine triphosphate (ATP), and so on [2], [3]. These ions are considered to be important for activity of a nerve in the brain. Especially, $Ca^{2+}$ density propagates to wide range of the brain and affects neurons' membrane potentials. We consider that the biological behavior of the glias is applicable to artificial neural networks to enhance their abilities.

Artificial neural networks are inspired by the features of the biological neurons and their networks. Now, various artificial neural networks are proposed for different purposes and some of them have been utilized in real engineering systems. The Multi-Layer Perceptron (MLP) is the most famous feed-forward neural network. The MLP can be applied to solve many different kinds of tasks, for example, a pattern recognition, a pattern classification, a data mining, and so on. In general, the MLP is trained by the Back Propagation (BP) algorithm [4]. In the BP algorithm, we calculate the error between the teacher value and the output value, after that, this error is propagated backward in the network to modify the weights between neurons. However, the BP algorithm often falls into a local minimum, because the BP uses the steepest decent method. The MLP should escape out from the local minimum to achieve a good performance. In order to overcome this problem, some researchers proposed a method giving noise to the MLP [5], [6]. We notice that the same effect as noise can be realized by the glia's feature in the brain. We also consider that the artificial glial network can make the position dependency of the neurons in the MLP, which may give a novel feature to the MLP.

In this study, we propose the MLP with the glial network. The artificial glial networks are designed to shake the threshold values of the neurons by random but correlated oscillation signals. Computer simulation results show that the proposed networks gain a better performance to solve the Two-Spiral Problem (TSP) than the conventional MLP.

## 2. Multi-Layer Perceptron with Glial Network

In this study, we consider the MLP with four layers (one input layer, two hidden layers, and one output layer) and we assume that the glial network is connected to the second hidden layer only. In the proposed network, each glia is connected to one neuron and its neighbor glias.

The standard updating rule of the neuron in the MLP is given by Eq. (1).

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

where $y$ is the output of the neuron, $x$ is the input of the neuron, $\omega$ is the weight between neurons, $n$ is the number of the neurons in the prior layer, and $f$ is the output function.

For the neurons in the second layer, we assume that the updating rule can be given by Eq. (2).

$$y_i(t) = f\left(\sum_{j=1}^{n} \omega_{ij} x_j(t) - \theta_i(t) + \alpha\Psi_i(t)\right), \tag{2}$$

where $\Psi$ is the output of the glia and $\alpha$ controls the strength of the effect from the glia. We assume that the glias influence the thresholds of their corresponding neurons, because the glias in the brain are reported to affect to neurons' membrane potentials by $Ca^{2+}$ densities.

We use the sigmoidal function as the output function $f$.

Manuscript received February 25, 2011.
Manuscript revised May 24, 2011.

[†]The authors are with Tokushima University, Tokushima-shi, 770-8506 Japan.
a) E-mail: ikuta@ee.tokushima-u.ac.jp
b) E-mail: uwate@ee.tokushima-u.ac.jp
c) E-mail: nishio@ee.tokushima-u.ac.jp

DOI: 10.1587/transfun.E94.A.1864

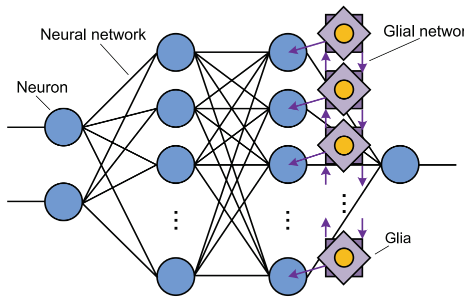


**Fig. 1** MLP with glial network.

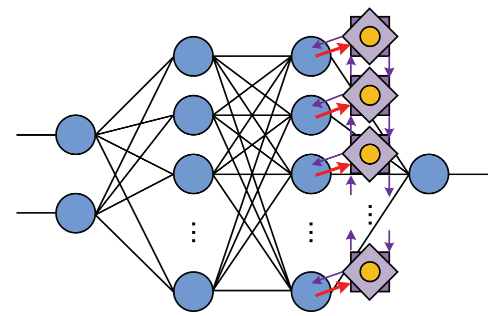

**Fig. 2** MLP with talking glial network.

## 3. Structure of Glial Networks

In this study, we propose two kinds of structures of the glial networks; the glial network and the talking glial network. For both networks, we assume that the glias produce independent random oscillations and the oscillations propagate through the glial network with attenuation. We also assume that the boundary condition of the glial network is a ring condition, namely the one end of the network is connected to the other end of the network.

### 3.1 Glial Network

In the first model, the output of the glial network is defined as Eq. (3).

$$\Psi_i(t) = \sum_{k=-m}^{m} \beta^{|k|}\psi_{i+k}(t-|k|), \tag{3}$$

where $\psi$ is the random oscillation in the range $[-1, 1]$ generated by each glia, $\beta$ is the attenuation parameter, and $m$ is the propagating range of the oscillations.

Figure 1 shows the structure of the MLP with the glial network. In this model, the MLP does not affect to the glial network.

### 3.2 Talking Glial Network

In the second model, the output of the glial network is defined as Eq. (4).

$$\Psi_i(t) = \{y_i(t-1) - 0.5\} \sum_{k=-m}^{m} \beta^{|k|}\psi_{i+k}(t-|k|), \tag{4}$$

where $y_i$ is the output of the neuron. Hence, in the talking glial network, the output of the glial network is controlled by the output of the MLP. This means that the neurons and the glias influence stronger and more dynamic in the talking glial network during the learning process than the case of the above simpler glial network. We consider that this model is reflected better the relationship between the neurons and the glias in the real brain.

Figure 2 shows the structure of the MLP with the talking glial network.

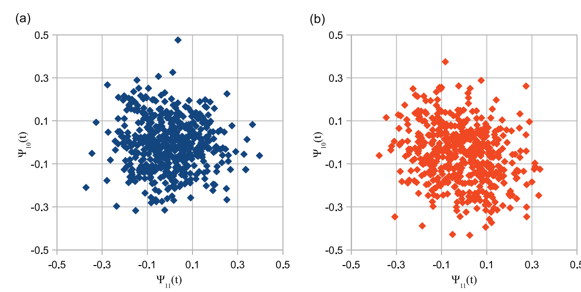

**Fig. 3** Relationship between two different glias of (a) MLP with glial network and (b) MLP with talking glial network.

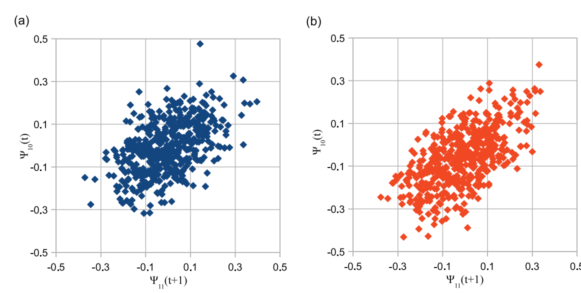

**Fig. 4** Relationship between two different glias at different time steps of (a) MLP with glial network and (b) MLP with talking glial network.

## 4. Output of Glial Networks

Because the random oscillations generated in the glias propagate through the glial network, the output of the glial network has some correlation characteristics.

We investigated the relationship between the different positions of the outputs of the glial networks. The outputs are recorded during the BP learning with the following network parameters; the numbers of the neurons of the MLP are 2-20-20-1, the number of the glias is 20, the attenuation parameter is $\beta = 0.8$, and the propagating range is $m = 9$.

Figure 3 shows the relationship between the outputs of the glial network at the positions of the 11th glia and the 10th glia. They are not correlated for the both types of MLPs, because the oscillations propagate through the glial network with time delay.

In order to take the effect of the propagation delay into account, we investigated the similar relationship between

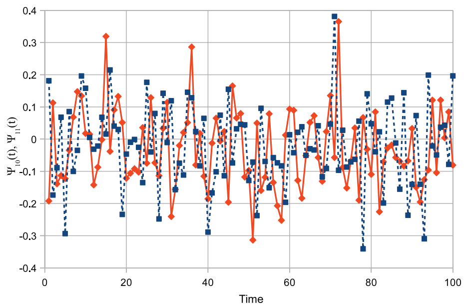


**Fig. 5** Time series of two different glias' positions of MLP with glial network.

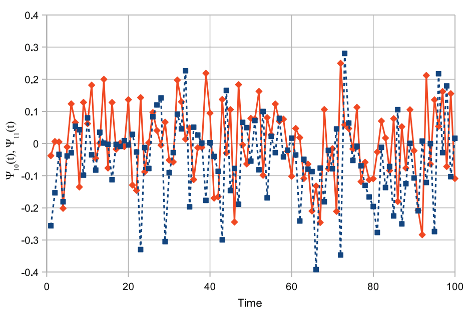


**Fig. 6** Time series of two different glias' positions of MLP with talking glial network.

the outputs but at the different time steps. Figure 4 shows the relationship between the outputs of the 11th position at $t$+1 and the 10th position at $t$. In this case, we can clearly observe the correlation of the data. This effect can be observed from the time series of the outputs of the glial network. Although it is not clearly visible, Figs. 5 and 6 show that the two time series take correlated values within ±1 time step.

Further, we confirmed that the distance of the positions corresponds to the propagation delay in the sense of the correlations. For example, the outputs of the 11th glia at $t + 2$ and the 9th glia at $t$ are correlated, though the relationship between them is smaller by the effect of the attenuation.

## 5. Simulation Result

We show the performance of the proposed network by solving the Two-Spiral Problem (TSP). The TSP is a famous benchmark of classification problems for the artificial neural networks [7], [8]. In this study, we prepare 130 points belonging to the two spirals as shown in Fig. 7 and train the MLPs by the BP learning to give the correct spiral to which a given point belongs, when the $x$ and $y$ coordinates of the point are inputted. This classification problem is known as a relatively difficult one and usually can not be solved with the standard three-layer MLP.

We compare the performance of four different MLPs; (a) the conventional MLP, (b) the MLP with the glial network, (c) the MLP with the talking glial network, and (d) the MLP with the completely independent random noise. The number of the trials is 200 and the MLPs learn 100000 times

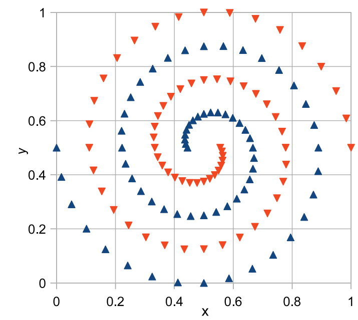


**Fig. 7** Two-spiral problems with 130 points.

**Table 1** Learning performance.

| | Average | Minimum | Maximum | Std. Dev. |
|---|---|---|---|---|
| (a) | 0.03346 | 0.00020 | 0.16970 | 0.03776 |
| (b) | 0.02410 | 0.00009 | 0.12997 | 0.02801 |
| (c) | 0.02035 | 0.00008 | 0.14149 | 0.02513 |
| (d) | 0.02809 | 0.00011 | 0.16140 | 0.03111 |

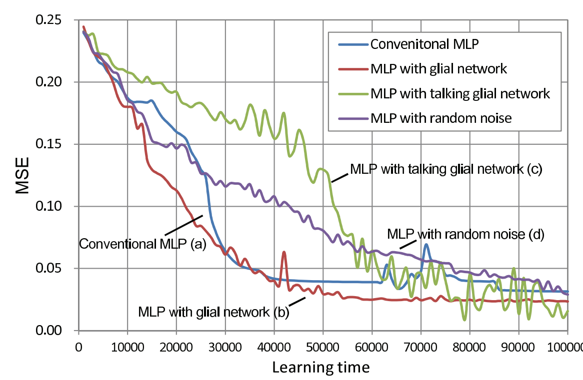


**Fig. 8** Example of learning curves of four MLPs.

during each trial. We calculate an average error, a minimum error, a maximum error, and the standard deviation (Std. Dev.). We use the Mean Square Error (MSE) for the error function given by Eq. (5).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(t_i - O_i)^2, \tag{5}$$

where $N$ is the number of the learning points and $t$ is the teacher value.

The simulated results are summarized in Table 1. In the table, (a), (b), (c), and (d) are the results of the conventional MLP, the MLP with the glial network, the MLP with talking glial network, and the MLP with random noise, respectively. We can see that the MLP with the talking glial network gives the best learning performance for the average error and the minimum error. The conventional MLP is the worst for all the measures.

Figure 8 shows an example of the learning curves by the MLPs. Before around 10000 times, all MLPs' learning curves are similar, however, the error curves are changed after that. The learning speed of the conventional MLP becomes very slow after around $MSE$ = 0.03 at 50000 times

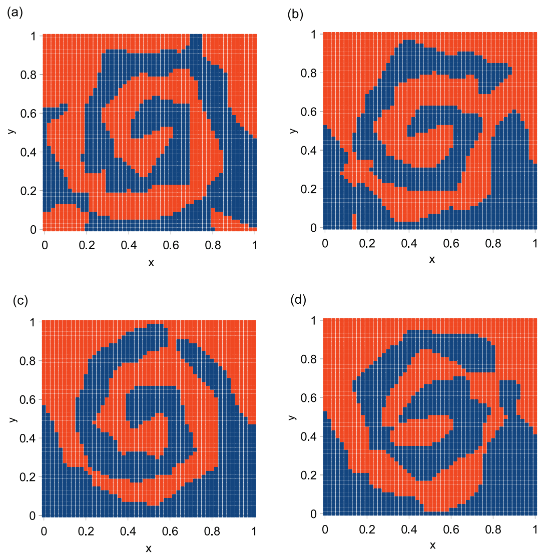


**Fig. 9** Classification results of TSP with 130 points. (a) Conventional MLP. (b) MLP with glial network. (c) MLP with talking glia network. (d) MLP with random noise.

and the ability escaping out from local minimums seems to be weak. The MLP with the random noise has better ability escaping out from local minimums but the learning speed is slow. The learning speed of the MLP with the glial network is very fast, but in this example this MLP does not learn further after around 60000 iterations. We should note that the learning curve of the MLP with the talking glial network cannot be expressed by a simple word. The curve has both slowly-learning part (10000-30000) and quickly-learning part (50000-60000), and especially many oscillating behaviors. We consider that this is because the MLP with the talking glial network has more complex structure than the others. This complex behavior on the learning curve seems to work better for learning.

Finally, we show the classification results of the TSP in Fig. 9. The generation of clear spirals also indicate the generalization capability of the MLPs. In this simulations, we give the coordinates ($x$ and $y$) to the trained MLPs and evaluate the outputs. We use the same initial conditions as Fig. 8 for learning of the MLPs in order to obtain similar error to the average $MSE$ listed in Table 1. We can see that the result of the MLP with the talking glial network is the best. Because, the spirals obtained by the MLP with the talking glial network has only one discontinuity. The others have two or more discontinuities in the spirals. Moreover, the spirals obtained by the MLP with the talking glial network have almost constant thickness compared with the others and are also smoother than the others.

From Figs. 8 and 9, the talking glial network gives a better learning performance and a better generalization capability to the MLP.

## 6. Conclusions

In this study, we have proposed the MLP with the glial network which was inspired from the features of glia. First type was the MLP with the glial network and the second one is the MLP with the talking glial network. All glias in the proposed network generated independent oscillations, and the oscillations propagated through the glial network with attenuation. We confirmed that the outputs of the glial network were correlated between the neighboring positions. We investigated the learning performance of the proposed networks by solving the TSP with 130 points. Computer simulated results showed that the proposed networks gained a better performance than the conventional MLP and the MLP with simple random noise. Further, we confirmed that the MLP with the talking glial network was better than the MLP with the simple glial network.

The important novel feature of the proposed networks is to give the MLPs the position dependency of the neurons, which is usually ignored in artificial neural networks because of the simplicity. It is true that more detailed investigation is needed in order to claim the effect of such a non-uniform feature on the learning performance. Real brains must be, of course, non-uniform and hence we believe that a kind of individuality will give wider chance to realize more intelligent artificial neural networks.

### 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, 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] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning representations by back-propagating errors," Nature, vol.323–9, pp.533–536, 1986.
[5] A. Guozhong, "The effects of adding noise during backpropagation training on a generalization performance," Neural Computation, vol.8, no.3, pp.643–674, April 2008.
[6] Y. Hayakawa, A. Marumoto, and Y. Sawada, "Effects of the chaotic noise on the performance of a neural network model for optimization problems," Physical Review E, vol.51, no.4, pp.2693–2696, April 1995.
[7] J.R. Alvarez-Sanchez, "Injecting knowledge into the solution of the two-spiral problem," Neural Computing & Applications, vol.8, pp.265–272, 1999.
[8] H. Sasaki, T. Shiraishi, and S. Morishita, "High precision learning for neural networks by dynamic modification of their network structure," Proc. Dynamics & Design Conference, pp.411-1–411-6, Sept. 2004.