Effect of Patten Coding on Pattern Classification Neural Networks

Tomohiro Tanno, Kazumasa Horie, Takaaki Kobayashi, and Masahiko Morita

***Abstract*—A recent practical research reported that a layered neural network when applies pattern coding, which is a method to convert analogue input value into a multidimensional binary pattern, has a high capability of pattern classification. However, it does not conduct sufficient basic analysis; thus, the effect of pattern coding is unclear. The present study examines the effectiveness of pattern coding in pattern classification. Numerical experiments on two-dimensional two-class classification problems show that a multilayer perceptron can learn complex decision boundaries when pattern coding is applied. The results also indicate that pattern coding is also effective for a simple perceptron if selective desensitization is applied jointly.**

***Index Terms*—Neural networks, pattern classification, pattern coding, selective desensitization.**

I. Introduction

Simple perceptron (SP), the simplest layered neural network, is a linear pattern classifier whose decision boundary is a hyperplane [1]. When a problem cannot be linearly separated, the multilayer perceptron (MLP), which contains hidden layer(s), is used in general [2]. MLP is known to have high expression ability and can construct almost any kind of boundaries between classes. However, in the case of complex boundaries, learning by MLP often takes a very long time or does not converge. MLP also requires high efforts for parameter tuning because of its high parameter dependency.

Pattern coding is a method for improving the performance of neural networks for pattern classification. This method converts the analogue input value into a multidimensional binary pattern. The effect of pattern coding is less dependent on parameters and can thus be used in most networks effortlessly. A recent practical study showed that the performance of SP for pattern classification improved significantly when both pattern coding and selective desensitization, which is a method to enhance the effect of pattern coding, were applied together [3].

However, the study did not provide sufficient basic analysis on the effect of pattern coding. Thus, the effectiveness of pattern coding on each neural network and its applications, i.e., the types of problems to which it is appropriately applicable, are unclear.

Accordingly, in the present study, we examined the effects of pattern coding when applied to SP and MLP through numerical experiments on several types of two-dimensional two-class classification problems. We also compared the results with the case applying selective desensitization instead of adding hidden layers.

II. Pattern Classifiers

We used two types of neural networks, SP and MLP. For each network, the ordinary types with analogue inputs were written as SP-A and MLP-A, and the pattern coding-applied were written as SP-P and MLP-P. The selective desensitization-applied SP was written as SP-SD. These five neural networks (i.e., SP-A, MLP-A, SP-P, MLP-P, SP-SD) are the pattern classifiers we used in this study for investigating the effects of pattern coding (Fig. 1).

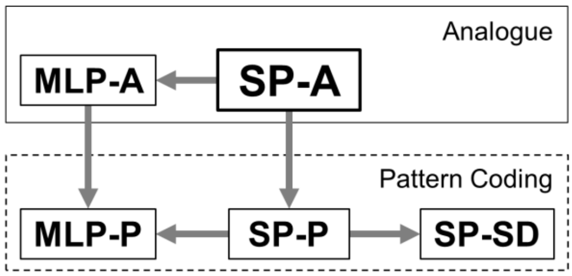


Fig. 1. Layered neural networks.

A. Simple Perceptron

When SP consists of two input units (x, y) and one output unit z, then z is calculated by

$$z = g(w_1 x + w_2 y + h) \tag{1}$$

where w_i is the synaptic weight, h is a threshold, and $g(u)$ is an activation function. SP classifies two classes by using the Heaviside function whose value is 1 when $u > 0$ and 0 otherwise. The error correction learning algorithm is used for learning.

Studies have reported that SP is a linear classifier and cannot solve the XOR problem [4]. In other words, SP cannot separate classes such as $z(0,0) = z(1,1) = 0$; $z(0,1) = z(1,0) = 1$. SP can construct only a straight line as a decision boundary; therefore, its classification ability is limited.

B. Multilayer Perceptron

The three-layered MLP is a layered neural network that contains a hidden layer between the input and the output layers. It can be regarded as combining SP to realize

Manuscript received October 14, 2014; revised February 25, 2015. This work was supported partly by JSPS KAKENHI, Grant Numbers 22300079 and 26590173.

T. Tanno, K. Horie, and T. Kobayashi are with the Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan (e-mail: tanno@bcl.esys.tsukuba.ac.jp, horie@bcl.esys.tsukuba.ac.jp, takaaki@bcl.esys.tsukuba.ac.jp).

M. Morita is with the Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan (e-mail: mor@bcl.esys.tsukuba.ac.jp).

DOI: 10.7763/IJMLC.2015.V5.531

nonlinear separation. The sigmoid function is used as the activation function of the hidden and output layers. The backpropagation algorithm, one of the gradient methods, is used for learning.

Theoretically, the three-layered MLP can express almost any complex boundaries between classes [5], [6]. This is realized by an extremely large number of hidden units, in which each unit takes charge of separating a small area. However, generalization ability cannot be expected by this way of separating classes. Furthermore, learning a finite number of training samples can result in over-fitting.

Moreover, the time required for learning is another disadvantage of MLP. In MLP, learning often requires time because many learning cycles are required to train the hidden units. Also, learning often does not converge by falling into local minimums because of backpropagation algorithm. Therefore, several trials of learning with different parameters are necessary. For these reasons, it is not yet suitable for practical use.

C. Pattern Coding

Pattern coding is a method to express an analogue variable value by a corresponding binary pattern. The input range is divided into q intervals, and each interval is assigned with individual n-dimensional binary pattern, $P_i (i = 1 \dots q)$. An element of each pattern has a value of either $+1$ or -1. Furthermore, it is necessary to fulfill the first two conditions below. The remaining two conditions are also desirable to be fulfilled.

1) Information about the variable value disperses widely in the whole pattern, and the pattern has enough redundancy as the whole. In other words, the variable value cannot be identified with one or a few elements of a pattern but rather with about a half of the elements.
2) The pattern changes gradually as the variable value changes successively, resulting in a high correlation between the patterns when the values are close and a low correlation when the values are far apart.
3) Patterns have the same number of $+1$ and -1 elements.
4) Correlation between patterns of the values that are far apart is 0.

Several methods can be used to create patterns according to the above conditions. This study considers a simple method that can be used to create the patterns easily. This simple method can be applied when n is large enough.

First, the pattern P_1 is randomly made so that the number of $+1$ and -1 elements can be the same (therefore, n is even). P_2 is made by inverting the sign of r elements of $+1$ and r elements of -1 of P_1 which are selected randomly. Likewise, P_k is made by inverting the sign of $2r$ elements of P_{k-1}. An example of pattern coding $(n = 8, r = 1)$ is shown in Fig. 2.

When pattern coding is applied to layered neural networks, the analogue input values get converted into binary patterns. Moreover, it does not interfere with the output or the calculation of the network, and can, thus, be applied together with the hidden layer(s). The structure of an ordinary MLP using analogue input (MLP-A) is shown in Fig. 3. The structure of pattern coding-applied MLP (MLP-P) is shown in Fig. 4.

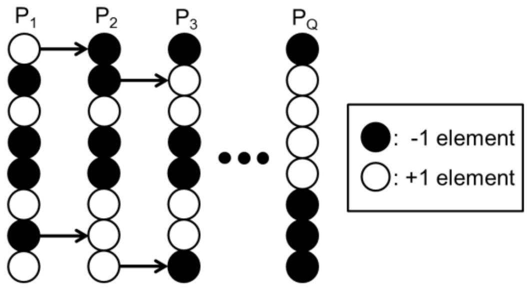


Fig. 2. Example of pattern coding (n=8, r=1).

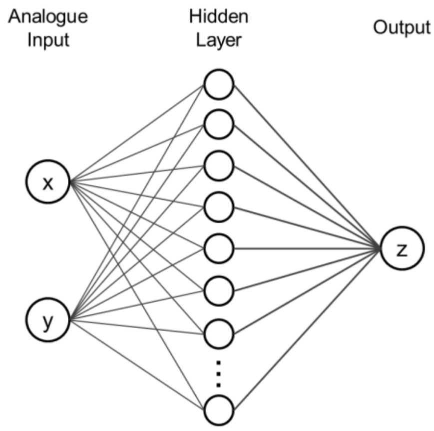


Fig. 3. Structure of an ordinary MLP that uses analogue input (MLP-A).

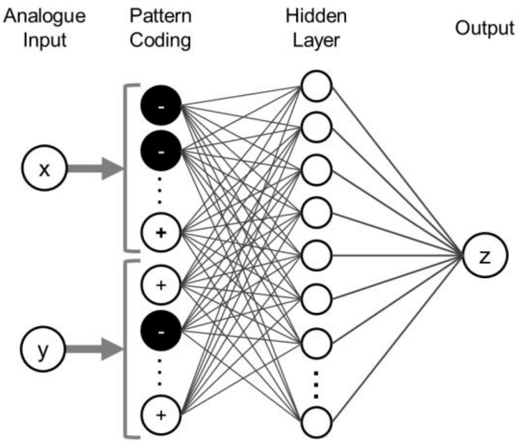


Fig. 4. Structure of pattern coding-applied MLP (MLP-P).

D. Selective Desensitization

Selective desensitization is a method to enhance the effect of pattern coding by integrating the information represented by two binary patterns into one three-value pattern. When there are two n-dimensional binary patterns, $S = (s_1 \dots s_n)$ and $C = (c_1 \dots c_n)$, half of the elements of S are selected according to C and converted into 0. This newly obtained three-value pattern is called "S modified by C" and is written as $S(C) = (x_1 \dots x_n)$. The value of x_i is calculated by

$$x_i = \frac{\left(1 + c_{\sigma(i)}\right)}{2} s_i \quad (2)$$

where σ denotes a fixed permutation function that is used to avoid the harmful influence of correlation between S and C. This creates a pattern in which half of the elements are 0 and the rest are the same with S (either $+1$ or -1).

When selective desensitization is applied to SP that consists of two input and one output units, the two input values are first converted into binary patterns by pattern coding. Then, each pattern is modified by the other. The structure of SP-SD is shown in Fig. 5.

Recently, pattern coding and selective desensitization are also used for function approximation by neural networks [7], [8]. However, basic analysis on these methods has not been done; thus, their effects are unclear.

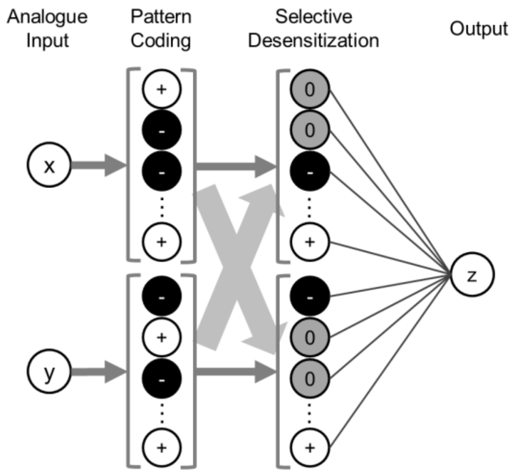


Fig. 5. Structure of selective desensitization applied SP (SP-SD).

III. Numerical Experiment

A. Classification Problems

We conducted numerical experiments on two-dimensional two-class classification problems to examine the effect of pattern coding on pattern classification. Each problem is a binary (black and white) image of size 101×101 (10201 points in total). Black corresponds to class-0 and white corresponds to class-1. For each image, 2000 points are randomly selected as training samples.

We conducted experiments on four problems to determine the kind of decision boundaries that each classifier can learn and construct. The problems are shown in Fig. 6.

Problem A is a linearly separable problem. Problem B is not linearly separable but is relatively simple, and it does not contain any set of points in an XOR relationship. Problem C is a complex problem and contains XOR. Problem D is the most complex problem.

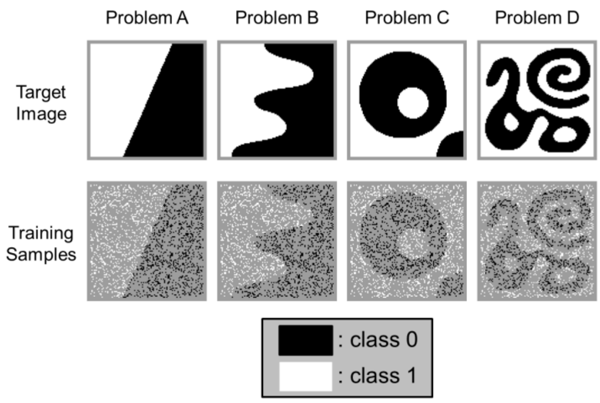


Fig. 6. Classification problems.

B. Methods

First, each classifier learned the set of training samples for a sufficient number of times. Each classifier then predicted the class for all the other points. We examined if each classifier is able to construct the boundary of the problems. We also compared the decision boundaries, classification errors, generalization errors, and the time required for learning between the classifiers. Classification error is the percentage of the training samples that were not classified correctly and generalization error is the percentage of unlearned data points that were not classified correctly.

Experiments were conducted with five classifiers. SP-A and MLP-A are the ordinary neural networks that used the analogue input values. SP-P, MLP-P and SP-SD are the neural networks that applied pattern coding. The parameters and other settings used for each classifier are provided below.

SP: The error correction learning algorithm is used here. Learning ends when all the training samples are correctly classified or the number of times the sample set is learnt reaches maximum (50000 for SP-A and 2000 for SP-P and SP-SD). The parameters of pattern coding for SP-P and SP-SD are set as $n = 10000, r = 250$, and $q = 101$.

MLP: The backpropagation learning algorithm is used here. The number of hidden units is 50. Learning ends when the square error reaches a certain value or the number of times the sample set is learnt reaches maximum (50000 for MLP-A and 2000 for MLP-P). Several trials with different default values of synaptic weight were considered because of the presence of local minimums, and the one with the best result was chosen. Parameters of pattern coding for MLP-P are set as $n = 200, r = 5$, and $q = 101$.

C. Result

The decision boundaries constructed by classifiers are shown in Fig. 7. Classification error, generalization error, and learning time are shown in Table I-Table III.

TABLE I: Classification Error

Classifier	Problem A	Problem B	Problem C	Problem D
SP-A	0.00 %	20.10 %	50.20 %	50.35 %
SP-P	0.00 %	0.00 %	26.85 %	34.00 %
MLP-A	0.15 %	0.15 %	7.75 %	19.75 %
MLP-P	0.05 %	0.00 %	0.00 %	0.00 %
SP-SD	0.00 %	0.00 %	0.00 %	0.00 %

TABLE II: Generalization Error

Classifier	Problem A	Problem B	Problem C	Problem D
SP-A	0.01 %	17.83 %	54.96 %	52.18 %
SP-P	1.67 %	2.34 %	28.97 %	35.34 %
MLP-A	0.06 %	0.88 %	7.94 %	21.46 %
MLP-P	1.68 %	2.19 %	2.45 %	6.05 %
SP-SD	0.68 %	1.48 %	2.89 %	6.73 %

TABLE III: Time Required for Learning

Classifier	Problem A	Problem B	Problem C	Problem D
SP-A	0.45 s	3.53 s	3.58 s	3.44 s
SP-P	10.39 s	10.16 s	543.45 s	542.87 s
MLP-A	15.38 s	158.29 s	824.52 s	832.46 s
MLP-P	17.50 s	14.61 s	17.44 s	28.98 s
SP-SD	8.81 s	8.68 s	11.44 s	28.72 s

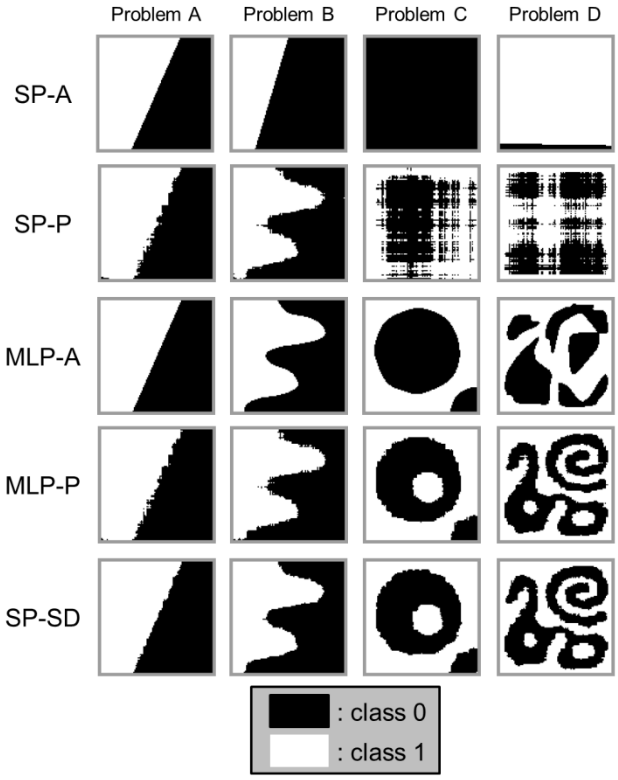


Fig. 7. Boundaries constructed by classifiers after training.

All the five classifiers succeeded in constructing the linear boundary of problem A. The generalization error of SP-A and MLP-A are significantly small.

SP-A is a linear classifier and thus, could not construct the boundaries of the other three problems. The other four classifiers constructed the nonlinear boundary of problem B. The boundary of MLP-A is particularly smooth and has the lowest generalization error, but the time it required for learning was the longest in the five classifiers.

MLP-P and SP-SD were the only classifiers that were able to construct the complex boundaries of problems C and D. The time they required was also relatively short.

IV. Discussion

With the results of the experiment, we discuss the effect of pattern coding and selective desensitization.

A. *Effect of Pattern Coding*

Although SP-A is a linear classifier and can only construct linear boundaries, SP-P was able to construct nonlinear boundaries. This indicates that pattern coding enables nonlinear separation.

However, SP-P could not construct the boundaries of problems C and D. We conducted several experiments with a larger number of elements for pattern coding or a larger number of times for learning; however, no significant difference in the results was observed. This is because SP-P cannot solve the XOR problem. Therefore, it can be said that the expression ability of SP-P is limited. Pattern coding alone is insufficient to improve the performance of SP; the method to solve the XOR problem is necessary.

Adding hidden layer(s) to SP is a generally used method to solve the XOR problem. MLP-A, which contains one hidden layer, constructed the nonlinear boundary of problem B. However, learning of MLP-A took a long time for problem B and did not converge for problems C and D. We conducted several experiments with larger number of hidden units or larger number of times for learning, but none of them could successfully construct the boundaries completely. It is confirmed that ordinary MLP with analogue input values has difficulties in learning complex boundaries. Even though the expression ability of MLP is high, it is not guaranteed that the learning is appropriate in practical use.

On the other hand, MLP-P was able to learn and construct complex boundaries of problems C and D in a relatively short time. It can be said that pattern coding solves the defects about learning of MLP, and that the insufficiency of pattern coding's expression ability is compensated by training hidden units. As a result, the performance of SP for pattern classification improves greatly when both pattern coding and a hidden layer are added together.

Another difference between ordinary networks and pattern coding-applied networks is the smoothness of the decision boundary. The boundaries constructed by SP-A and MLP-A are smooth compared to the rough boundaries constructed by SP-P, MLP-P and SP-SD. Therefore, although training samples are mostly correctly classified, generalization errors occurred in the areas near the boundary. It can be said that pattern coding is especially suitable for complex problems such as C and D.

B. *Effect of Selective Desensitization*

SP-SD was able to construct boundaries for all the problems, suggesting that selective desensitization enhances the expression ability of SP-P and enables it to solve the XOR problem.

The decision boundary, learning error and generalization error are all at the same level with MLP-P. This implies that selective desensitization enhances the performance of SP-P as much as adding hidden layer(s).

The time required for learning of SP-SD is shorter than that of MLP-P for all of the problems. One of the reasons is the difference in learning algorithms. The backpropagation algorithm used in the learning of MLP requires complicated calculations including the use of a differential of an error function. MLP learns all the training samples in every learning cycle. In contrast, error correction used in the learning of SP is a simple calculation of incrementing or decrementing the synaptic weight only when the sample was classified incorrectly.

Moreover, backpropagation is highly dependent on parameters such as learning rate, learning completion conditions, and default values of synaptic weight. Learning rate decides the amount of influence of one learning. It takes time for the square error to decrease if the learning rate is too small. If the learning rate is too large, the value of synaptic weight oscillates around the solution and it takes time to converge. Learning completion conditions are concerned with the accuracy or learning time. It must be strict enough to keep the accuracy, but it takes a long time for learning to complete if the condition is too strict. Over-fitting is also more likely to occur in such a case. The default values of synaptic weight are decided randomly and should not be 0. Local minimums occur depending on the default values of synaptic weight; therefore, several trials of experiments are necessary. It takes effort to set appropriate values for these parameters. We conducted several experiments with different values to search for appropriate values.

Parameter tuning is unnecessary for error correction learning used in SP. The result is less dependent on default values of synaptic weight, and they can also be 0. Learning rate is unnecessary for error correction learning. It can be said that the learning of SP-SD is much easier than that of MLP-P.

C. *Parameters of Pattern Coding*

There are three parameters of pattern coding; q, n and r. Parameter q concerns with input space discretization. Input space must be discretized in order to apply pattern coding.

Parameter n is the number of elements in each binary pattern and is desirable to be larger than q. High expression ability realizes when n is large. However, it is insignificant to make n too large because the expression ability of SP-P is limited. This can also be said for MLP-P because each hidden unit functions as SP-P. When selective desensitization is applied, the three-value pattern itself solves the XOR problem. Therefore, a large value for n enhances the performance of SP-SD. If n is sufficiently large, there is no significant difference in the performance, thus it does not need to be set precisely.

Parameter r is the number of inverted elements between P_i and P_{i+1}. It should be decided according to the value of n and q. If the required conditions of pattern coding is fulfilled, the performance depends little on the value of r. Therefore, it does not need to be decided precisely.

V. Conclusion

We examined the effects of pattern coding and selective desensitization when applied to layered neural networks through numerical experiments on various types of two-dimensional two-class classification problems. The results indicate that, although only a limited effect can be obtained by applying pattern coding alone on SP, a complicated decision boundary can be easily learnt by introducing a hidden layer in addition to pattern coding. It can be said that pattern coding enhances the performance of MLP, a generally used neural network for nonlinear problems. Pattern coding-applied MLP can easily construct complex boundaries compared to ordinary MLP.

The results also indicate that selective desensitization enhances the classification ability of SP as much as multilayering. Learning is much easier because it is unnecessary to train hidden units using selective desensitization. However, no differences in boundaries or errors were observed between the pattern coding-applied MLP and selective desensitization-applied SP. We plan to conduct experiments with other types of problems to identify the differences between their classification abilities.

In the future, we plan to carry out an experiment on higher dimension problems to verify the generality of this result. It is also an upcoming task to compare pattern coding applied neural networks with other pattern classifiers such as support vector machines.

Acknowledgment

This work was supported partly by JSPS KAKENHI Grant Numbers 22300079 and 26590173.

References

[1] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, no. 6, pp. 386-408, Nov. 1958.

[2] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D. E. Rumelhart, J. L. McClelland and the PDP research group, Eds. Cambridge, MA: MIT Press, vol. 1, pp. 318-362, 1986.

[3] H. Kawata, F. Tanaka, A. Suemitsu, and M. Morita, "Practical surface EMG pattern classification by using a selective desensitization neural network," *Neural Information Processing (Part II), Lect. Notes Comput. Sci.*, vol. 6444, pp. 42-49, 2010.

[4] M. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry*, MIT Press, 1969.

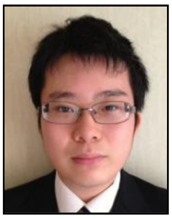
[5] B. Irie and S. Miyake, "Capabilities of three-layered perceptrons," in *Proc. the IEEE International Conference on Neural Networks*, 1988, vol. 1, pp. 641-648.

[6] K. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks*, vol. 2, pp. 183-192, 1989.

[7] K. Horie, A. Suemitsu, and M. Morita, "Direct estimation of hand motion speed from surface electromyograms using a selective desensitization neural network," *Journal of Signal Processing*, vol. 18, no. 4, pp. 225-228, 2015.

[8] T. Kobayashi, T. Shibuya, and M. Morita, "Q-learning in continuous state-action space with redundant dimensions using a selective desensitization neural network," in *Proc. the SCIS&ISIS 2014*, Dec. 2014, pp. 801-806.

Tomohiro Tanno received B. Eng degree from University of Tsukuba, Japan, in 2013. He is currently a student at the Graduate School of Systems and Information Engineering, University of Tsukuba. His research interests are neural networks and machine learning.



Kazumasa Horie received B. Eng. and M. Eng. degrees from University of Tsukuba, Japan, in 2012 and 2014, respectively. He is currently a student at the Graduate School of Systems and Information Engineering, University of Tsukuba. His research interests are biological signal processing, neural networks, and human computer interaction.

Takaaki Kobayashi received his B. Eng. and M. Eng. degrees from University of Tsukuba, Japan, in 2010 and 2012, respectively. He is currently a student at the Graduate School of Systems and Information Engineering, University of Tsukuba. His research interests are neural networks and reinforcement learning.

Masahiko Morita is currently a professor at the Faculty of Engineering, Information and Systems, University of Tsukuba, Japan. He received his Ph. D. degree from the University of Tokyo, Japan, in 1991. He has been engaged in research on biological information processing and neural computation.