Learning Intialized by Topologically Correct Representation

Pitoyo Hartono
School of System Information Science
Future University-Hakodate
Hakodate, Japan
hartono@fun.ac.jp

Thomas Trappenberg
Faculty of Computer Science
Dalhousie University
Halifax, Canada
tt@cs.dal.ca

Abstract—In this research, we proposed a model of a hierarchical three-layered perceptron, in which the middle layer contains a two dimensional map where the topological relationship of the high dimensional input data (external world) are internally represented. The proposed model executes a two-phase learning algorithm where the supervised learning of the output layer is proceeded by a self-organization unsupervised learning of the hidden layer. The objective of this study is to build a simple neural network model which is more biologically realistic than the standard Multilayer Perceptron model and that can form an internal representation that supports its learning potential. The characteristics of the proposed model are demonstrated using several benchmark classification problems.

Index Terms—Self-Organizing Map (SOM), Supervised Learning, Unsupervised Learning, Topological Representation, Perceptron

I. INTRODUCTION

In this study, we proposed a hierarchical neural network model, namely Map-Initialized Perceptron (MIP) that has an internal representation of external data (world) and use the representation to support a supervised learning strategy. The internal representation in the proposed MIP is a two dimensional Self-Organizing Map (SOM) [1], which organizes high dimensional data into a topologically correct low dimensional representation. Unlike the conventional Multilayer Perceptron (MLP) [2], the low dimensional map provides a more intuitive understanding on the internal representation. We believe that this advantage is important for developing and analyzing a more biologically realistic learning model.

In this paper we argue that the formation of the internal representation as a map in the hidden layer, significantly helps MIP to learn better, and we support this argument with several computational experiments. This argument also has a significant relevance with biological findings on cortical maps [3] [4]. We consider that the proposed model has more biological plausibility compared to MLP. Although in the experiments the task of MIP is to learn several benchmark classification problems, our focus is not in building a classifier superiors than MLP in the sense of classification performance and learning convergence. Our objective is to build an alternative neural network model which is more plausible in relating internal representation with the learning ability and more consistent with biological systems.

Some previous work have utilized internal maps for further learning. For example, in [5] a standard Q-learning algorithm is built on top of internal representation. Although closely related, this work differs from ours in that it did not focus on the correlation between the fidelity of the internal representation and the inherent learning potential. Self-Organizing Relationship (SOR) Network proposed in [7] [8] and Vector Quantized Temporal Associative Memory (VQTAM) in [9] are new models of SOM that executes supervised learning mechanism. These studies differ from our in objective and also in the learning method, in that we execute a two-step learning method where an unsupervised self-organization precedes a supervised learning.

The reminder of this paper is organized as follows. First, we will provide an explanation on the structure and learning strategy of the proposed MIP. The section that follows provides details on the learning experiments and brief analysis of the relationship between internal representation and the learning ability. Conclusion and future works will be discuss in the final section.

II. MAP-INTIALIZED PERCEPTRON

In this research, we built a three-layered neural network, namely the Map-initialized Perceptron (MIP) shown in Fig.1. The structure of this neural network is similar to that of MLP, in that it has an input layer to receive external input, a hidden layer for internal representation of external world, and an output layer that produces signals into the external world. MIP differs from MLP in that the hidden layer is a two-dimensional SOM [1], and consequently also differs in the learning method.

The training process of MIP consists of two phases. The first one is the initialization process in which SOM layer is trained to self-organize a topologically correct map with regard to the high dimensional external input. The second phase is the supervised learning phase, in which the connection weights between the SOM layer and the output layer are trained to produced desired signals [10].

The dynamics of MIP can be explained as follows.

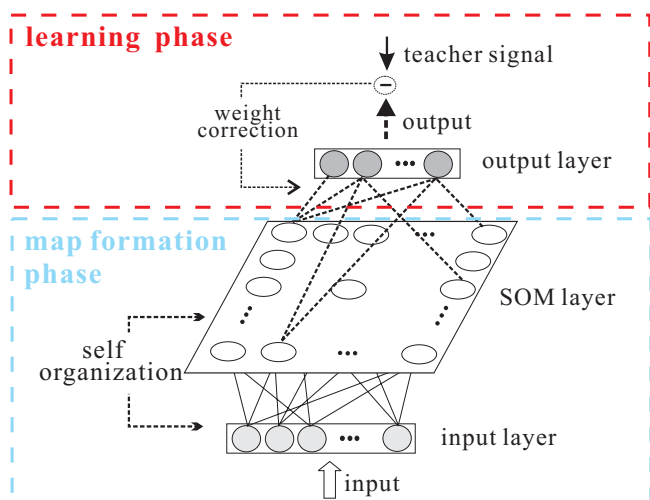The output of the j-th neuron in the output layer at time t, $O_j(t)$ is defined as

Fig. 1.   Outline of MIP

$$
\begin{aligned}
O_j(t) &= f(\sum_{i}^{K} v_{ij}(t)M_j(t) - \theta_j(t)) \\
M_j(t) &= \frac{1}{D_j(t) + \epsilon} \\
D_j(t) &= \frac{1}{d_{in}}|W_j(t) - X(t)|,
\end{aligned}
\qquad (1)
$$

where, $v_{ij}(t)$ and $\theta_j(t)$ denote the connection weight from the i-th neuron in SOM layer to the j-th output neuron and the potential of the j-th output neuron at time t, respectively. $W_j(t)$, $X(t)$ and K are the connection weight vector leading from the input layer to the j-th neuron in the SOM layer, the input vector , and the number of neuron in the SOM layer, respectively. $D_j(t)$ and $M_j(t)$ are the distance between the input vector and the j-th neuron in SOM layer, M_j is the output of the of that neuron, respectively. ϵ is an empirically decided positive constant, while d_{in} is the dimension of the input vector. Equation 1 shows that a neuron in SOM layer generates a large output value in reaction for an input vector with a small distance to it, and oppositely produces a small output value..

In Eq.1, $f(x)$ is a sigmoidal function, where α is a positive constant.

$$
f(x) = \frac{1}{1 + e^{-\alpha x}}
\qquad (2)
$$

A. Initial Learning Phase: Map Formation

The supervised learning of MIP is preceded by an initial learning, in which an internal representation of input data are self-organized in the form of a topologically correct map in the hidden layer. The conventional unsupervised learning method of SOM are utilized. In this initial learning phase, for every presentation of sample, $X(t)$, a neuron in the SOM layer which has the minimum distance is designated as a winner as in Eq.3

$$
win = arg \min_{j} |W_j(t) - X(t)|,
\qquad (3)
$$

The connection weight vectors are randomly initialized at the start of the learning phase.

A topologically correct map is formed by correcting the weight vectors as follows.

$$
W_i(t+1) = W_i(t) + \eta_s(t)\Lambda(dist(win, i))(X(t) - W_i(t))
\qquad (4)
$$

In Eq.4, η_s denotes the learning rate, and $dist(win, i)$ is the distance between the winner neuron and neuron i in SOM layer. Λ denotes a neighborhood function, such that $\Lambda(0) = 1$ and decreases along with the distance.

The unsupervised learning in the SOM layer is executed until a specified learning epoch is reached. This learning process generates a topologically correct representation of the data in the hidden layer of MIP to precede the supervised learning for the next layer.

B. Supervised Learning Phase

The task of the supervised learning stage is to learn a mapping function F that maps an input vector, $X \in R^n$ into its designated output vector $T \in R^m$.

$$
F : X \to T
\qquad (5)
$$

The learning process is executed according to the delta rule [10] as follows, in which the learning error E_{learn} at time t is defined as follows.

$$
E_{learn}(t) = \frac{1}{m}|O(t) - T(t)|,
\qquad (6)
$$

in which, $T(t)$ is the teacher signal associated the input vector, $X(t)$.

During this learning stage, the connection weights between the input layer and the hidden SOM layer are fixed, while the connection weights between the hidden layer and the output layer are corrected as follows.

$$
\begin{aligned}
V(t+1) &= V(t) - \eta_{lrn}\frac{\partial E_{lrn}(t)}{\partial V(t)} \\
\Theta(t+1) &= \Theta(t) + \eta_{lrn}\frac{\partial E_{lrn}(t)}{\partial \Theta(t)},
\end{aligned}
\qquad (7)
$$

where V and Θ are the connection weight matrix from the hidden layer to the output layer and the threshold vector of the output neurons, respectively, while η_{lrn} denotes the learning rate.

III. EXPERIMENTS

We run experiments using several benchmark problems from UCI [11]. The objective of these experiments is to observe about the effect of the initial self-organization of the input data to the learning ability of MIP in the supervised-learning stage. For the clarity sake of this observation, we limit the size of SOM in the hidden layer to 4 by 4. In this study, we observed the supervised learning ability of MIP based on four degrees of initial data organizations in its hidden layer. The first one is the execution of the supervised learning

without SOM layered training initialization. It is obvious that this learning mechanism is equivalent to that of three-layered Perceptron in which the connection weights between the first and the second layers are fixed. In this learning mechanism, the supervised learning stage of MIP is not preceded by the formation of internal data organization, thus the supervised learning is executed over a random representation. The second and the third supervised learnings are initialized by 10 epochs and 100 epochs of SOM unsupervised learning. It is clear that in these cases, the supervised learning is supported by the topologically correct data representation in the hidden layer. In the fourth case, the unsupervised learning of SOM runs simultaneously with the supervised learning. In this case, the supervised learning is executed while at the same time MIP also tries to form a topological internal representation of the data. All the experiment results for MIP learning are averaged over 30 runs. The parameter settings are $\alpha = 0.1$, $\epsilon = 0.1$, $\eta_s = 0.1$ and $\eta_{lrn} = 0.1$ for all the experiments.

TABLE I
BENCHMARK DATA

data	input dimension	number of classes	data size
iris	4	3	150
balance	4	3	625
breast cancer	9	2	100
wine	10	3	178
thyroid	5	3	215

Figure 2 shows the error curve during the learning process of SOM, calculated according to Eq.8, with regard to five UCI benchmark problems, whose characteristics are shown in Table.I. From this figure, it is clear that the topological representation of the input data in SOM improves with the increase of the learning epoch.

$$E_{som}(t) = \frac{1}{P} \sum_{j=1}^{P} |W_{win_j(t)} - X_j|, \qquad (8)$$

where $X_j(t)$ and $win_j(t)$ are the j-th sample of the data and the winner neuron in SOM layer associated with it at the t-th learning epoch, respectively, while P is the size of the data.

Figure 3 shows the learning curve of the supervised learning phase of MIP with regard to the iris classification problem. In this graph, ”percept”, ”SOM10”, ”SOM100” and ”sim” show the supervised learning not preceded by data organization (thus a Perceptron), supervised learning preceded by 10 epochs of SOM learning, supervised learning preceded 100 epoch of SOM learning and simultaneous learning of SOM and the supervised learning.

The internal representations in the middle layer of MIP can be visualized as two dimensional maps as shown in Fig. 4. The left map in Fig. 4 shows the topology of the input data when the training process of SOM is not executed. The middle map shows the data topology after 10 learning epochs of SOM, while the right map shows the data topology after 100 learning epochs. The circles show the winner neurons while their colors
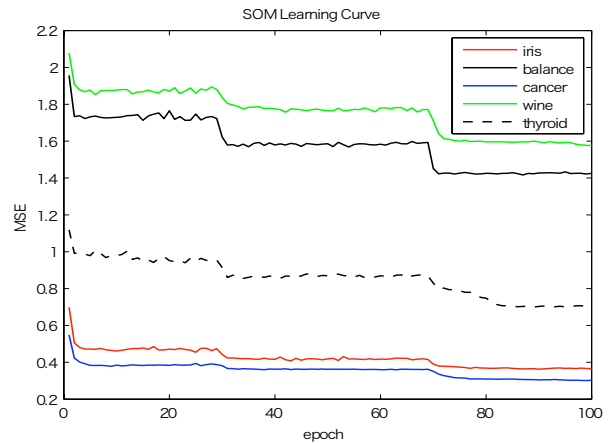

Fig. 2.   SOM Learning Curve

indicates the class labels of the data. A black rectangle in the maps is an ”ambiguous” neuron, which is a neuron which represent two or more data belonging to different classes. This kind of neuron represents ”border” between different classes.
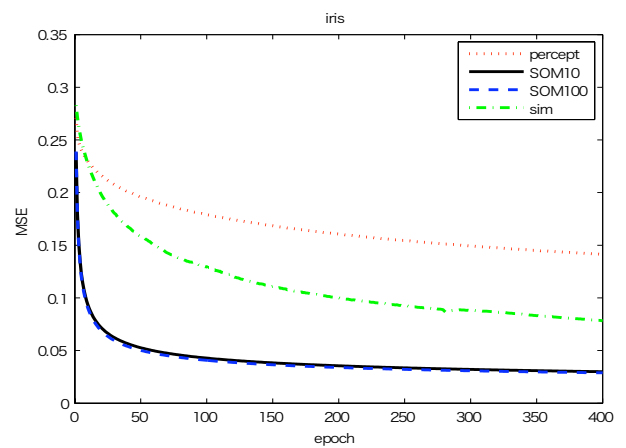

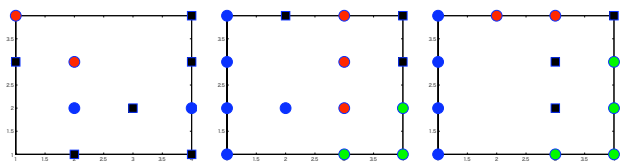Fig. 3.   Learning Curve (Iris)

Fig. 4.   SOM (Iris)

From Figs.2, 3, 4, we can clearly observe that topological organization in the hidden layer contributes significantly with the ability of MIP to perform in the supervised learning phase. The simultaneous learning of SOM and supervised training show a better performance than a supervised learning unsupported by internal representation. In this learning mechanism, the supervised learning is executed while the internal representation of the data is changing, which evidently, is a

hard task for MIP, in that MIP failed to reach its better potential in the supervised stage. These figures shows that some degree of internal organization significantly helps in achieving a good learning performance in the supervised learning phase. We can observe that the learning performances of MIP after 10 and 100 epochs of SOM learning are almost identical, which is consistent with similar topological representations shown in Fig. 4.
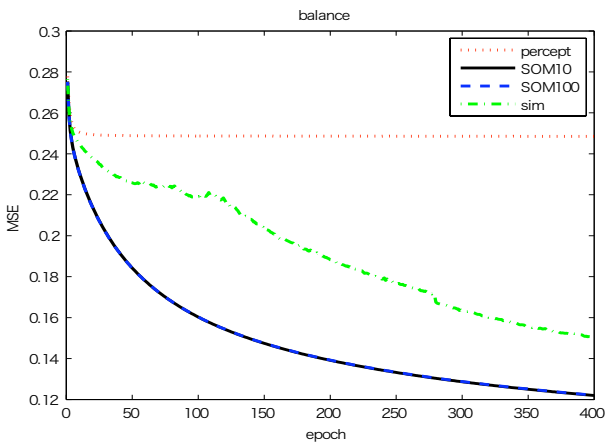
The results our experiments with other benchmark problems are as follows.

Fig. 5.   Learning Curve (Balance)
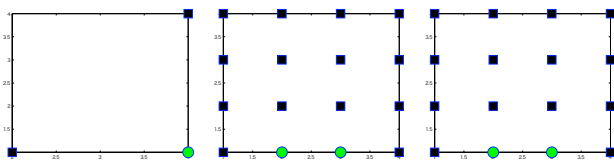
Fig. 6.   SOM (Balance)
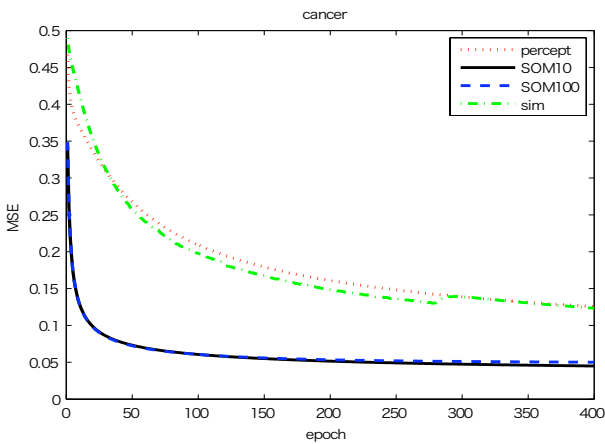
Fig. 7.   Learning Curve (Cancer)

For comparisons, we also trained a 3-layered MLP with 16 hidden neurons with the same problems and show the result
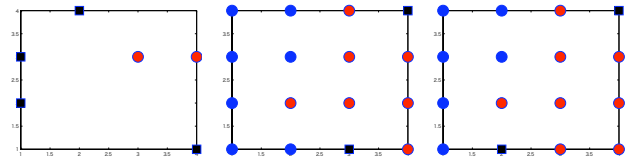

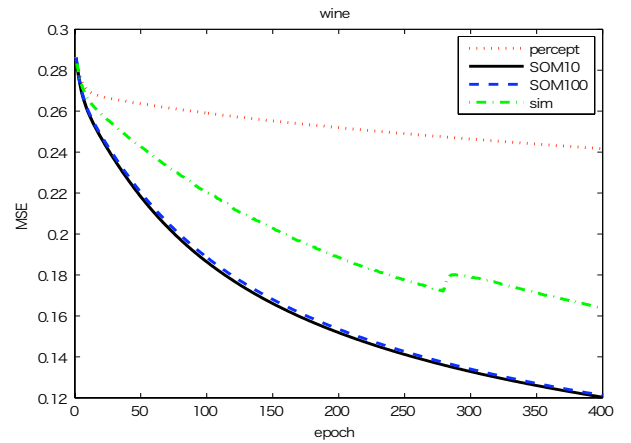Fig. 8.   SOM (Cancer)

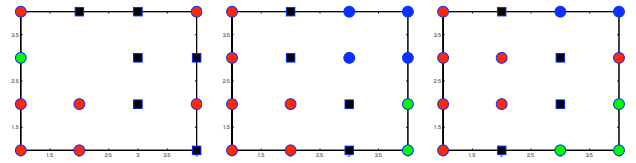Fig. 9.   Learning Curve (Wine)

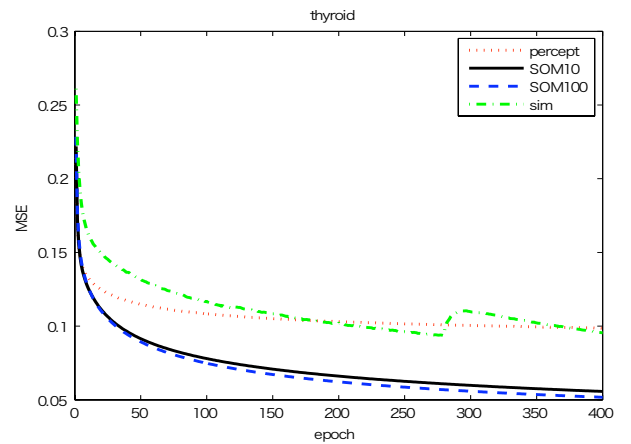Fig. 10.   SOM (Wine)
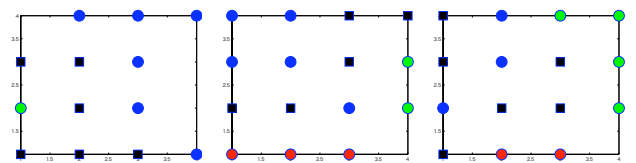

Fig. 11.   Learning Curve (Thyroid)

Fig. 12.   SOM (Thyroid)

in Fig. 13. From this figure, it is obvious that MLP learns the given examples better in the sense of convergence speed compared to the proposed MIP. This is expected since the network with an arbitrary hidden topology represents a model with a larger degree of freedom, or, in other words, the SOM has additional constrains (the topography) relative to the MLP version with the same number of connections. Thus, a more sensible comparison must take other criteria into account, such as the generalization ability of the networks, their scalability, or the biological plausibility, which will be the subject of our future research. Thus, the purpose of this paper is not (at this point) to build a better classifier than MLP (although our architecture has the potential of outperforming the MLP and other machine learning methods), but to outline a comparable alternative that is more biologically plausible and offers clearer information on the knowledge representation and potentially the knowledge acquisition during the learning process.
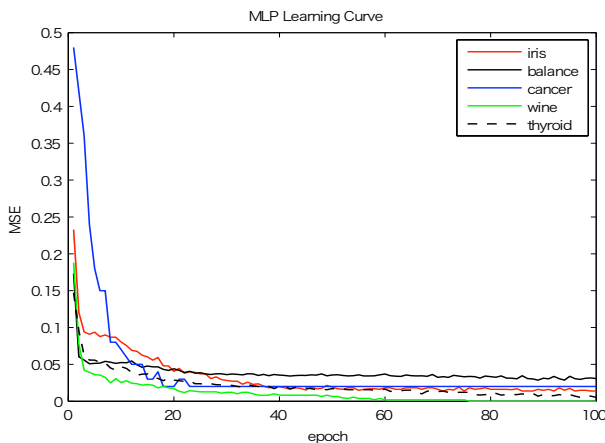

Fig. 13.   MLP Learning Curve

IV. CONCLUSION AND FUTURE WORKS

In this study we proposed a neural network model containing a map that self-organized high dimensional input data into a low dimensional internal representation. The proposed MIP is not better than the standard MLP in the sense of the learning speed. However, the proposed model does not only offer an alternative to MLP but also gives more plausible information in the form of topological map, on knowledge representation. We also believe that this model is more biologically realistic [12]. These two characteristics, potentially will allow us to expand our study in understanding the correlation between the degree of internal representation and the learning ability, not only in machine learning methods but also in their biology counterparts.

In this preliminary study we have built a simple hierarchical neural network model whose learning potential significantly correlates with the fineness of its internal representation. For the sake of clarity and simplicity we limit the size of SOM layer. It is of interest to investigate the behavior of the proposed model with a more flexible size of SOM layer.

We are also interested in expanding the model to include a learning feedback from the output layer to the formation of the representation in SOM layer. For example the effect of the environmental modification change the representation in the map. Application of the proposed model to the learning of autonomous physical robots is an immediate future interest for us.

REFERENCES

[1] T. Kohonen, "Self-Organized Formation of Topologically Correct Feature Maps", *Biological Cybernetics*, Vol.43, pp.59-69, 1982.
[2] D. Rumelhart. and J. McClelland, "Learning Internal Representation by Error Propagation", *Parallel Distributed Processing*, Vol.1 MIT Press pp.318-362, 1984.
[3] X. Zhou, and M. Merzenich, "Enduring Effects of Early Structured Noise Exposure on Temporal Modulation in the Primary Auditory Cortex", *Proc. Natl. Acad. Sci.*, Vol. 105, pp. 4423-4428, 2008.
[4] T. Trappenberg, P. Hartono, and D. Rasmusson, "Top-Down Control of Learning in Biological Self-Organizing Maps", *Proc. of Int. Workshop on Self-Organizing Maps (WSOM 2009)*, 2009 (accepted).
[5] Y. Choe and N. Smith, "Motion-Based Autonomous Grounding: Inferring External World Properties from Encoded Internal Sensory State Alone", *Proceedings of the 21st National Conference on Artificial Intelligence*, pp. 936-941, 2006.
[6] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, 1998.
[7] T. Koga, K. Horio and T. Yamakawa, "The Self-Organizing Relationship (SOR) Network Employing Fuzzy Inference Based Heuristic Evaluation", *Neural Networks*, Vol. 19, Nos.6-7, 2006.
[8] T. Yamakawa and T. Horio, "Self-Organizing Relationship (SOR) Network", *IEICE Trans. on Fundamentals*, Vol.E82-A, No.8, pp.1674-1677, 1999.
[9] G.A. Barreto, and A.F.R Araujo, "Identification and Control of Dynamical Systems Using the Self-Organizing Map", *IEEE Trans. on Neural Networks*, Vol.15, No. 5, pp.1244-1259, 2004.
[10] F.Rosenblat, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", *Psycological Review*, Vol. 65 pp. 386-408, 1958.
[11] UCI Repository, http://www.ics.uci.edu/ mlearn/MLRepository.html
[12] T. Trappenberg, *Fundamentals of Computational Neuroscience*, Oxford University Press, 2002.