NOLTA 2006
11-14 september
Bologna - ITALY

Reunifying Self-Organizing Map

Haruna Matsushita† and Yoshifumi Nishio†

†Department of Electrical and Electronic Engineering, Tokushima University
Email: {haruna, nishio}@ee.tokushima-u.ac.jp

Abstract— The Self-Organizing Map (SOM) attracts attentions for clustering in these years. In this study, we propose a Reunifying Self-Organizing Map which is a new SOM algorithm. The initial state of all neurons of the proposed SOM are connected to no neuron. However, the neurons are connected gradually to other neurons as learning progressed. The behavior of the reunifying SOM is investigated with application to clustering. We can confirm that the result of using the reunifying SOM includes no inactive neuron.

1. Introduction

Since we can accumulate a huge amount of data in these years, it is important to investigate various clustering methods. Then, the Self-Organizing Map (SOM) attracts attentions in recent years. SOM is an unsupervised neural network introduced by Kohonen in 1982 [1] and is a model simplifying self-organization process of the brain. SOM obtains statistical feature of input data and is applied to a wide field of data classifications [2]-[5]. We can obtain the map reflecting the distribution state of input data using SOM. However, if we apply SOM to clustering of the input data which includes some clusters located at distant locations, there are some inactive neurons between clusters. Because inactive neurons are on a part without the input data, we are misled into thinking that there are some input data between clusters.

In this study, we propose a new SOM algorithm which is the Reunifying Self-Organizing Map. The initial state of all neurons of the proposed SOM are connected to no neuron. However, each neuron has its own physical location on the 2-dimensional grid, so, the neurons are connected gradually to other neurons as learning progressed: hence, the name "reunifying".

In Section II, we explain the learning algorithm of the conventional SOM. In Section III, the learning algorithm of the reunifying SOM in detail. The learning behaviors of the reunifying SOM are investigated in Section IV with applications to clustering of various 2-dimensional input data. Furthermore, clustering ability is evaluated by visualization of results. We can see that there are no inactive neurons using the reunifying SOM.

2. Self-Organizing Map (SOM)

We explain the learning algorithm of the conventional SOM. SOM consists of m neurons located at a regular low-dimensional grid, usually a 2-D grid. The basic SOM algorithm is iterative. Each neuron i has a d-dimensional weight vector $\boldsymbol{w}_i = (w_{i1}, w_{i2}, \cdots, w_{id})$ $(i = 1, 2, \cdots, m)$. The initial values of all the weight vectors are given over the input space at random. The range of the elements of d-dimensional input data $\boldsymbol{x}_j = (x_{j1}, x_{j2}, \cdots, x_{jd})$ $(j = 1, 2, \cdots, N)$ are assumed to be from 0 to 1.

(SOM1) An input vector $\boldsymbol{x}_j$ is inputted to all the neurons at the same time in parallel.

(SOM2) Distances between $\boldsymbol{x}_j$ and all the weight vectors are calculated. The winner, denoted by c, is the neuron with the weight vector closest to the input vector $\boldsymbol{x}_j$;

$$c = \arg\min_i \{\|\boldsymbol{w}_i - \boldsymbol{x}_j\|\}, \qquad (1)$$

where $\|\cdot\|$ is the Euclidean distance.

(SOM3) The weight vectors of the neurons are updated as;

$$\boldsymbol{w}_i(t+1) = \boldsymbol{w}_i(t) + h_{c,i}(t)(\boldsymbol{x}_j - \boldsymbol{w}_i(t)), \qquad (2)$$

where t is the learning step. $h_{c,i}(t)$ is called the neighborhood function and is described as a Gaussian function;

$$h_{c,i}(t) = \alpha(t)\exp\left(-\frac{\|\boldsymbol{r}_i - \boldsymbol{r}_c\|^2}{2\sigma^2(t)}\right), \qquad (3)$$

where $\|\boldsymbol{r}_i - \boldsymbol{r}_c\|$ is the distance between map nodes c and i on the map grid, $\alpha(t)$ is the learning rate, and $\sigma(t)$ corresponds to the width of the neighborhood function. Both $\alpha(t)$ and $\sigma(t)$ decrease monotonically with time as follows;

$$\alpha(t) = \alpha(0)(1 - t/T), \quad \sigma(t) = \sigma(0)(1 - t/T), \qquad (4)$$

where T is the maximum number of the learning.

(SOM4) The steps from (SOM1) to (SOM3) are repeated for all the input data.

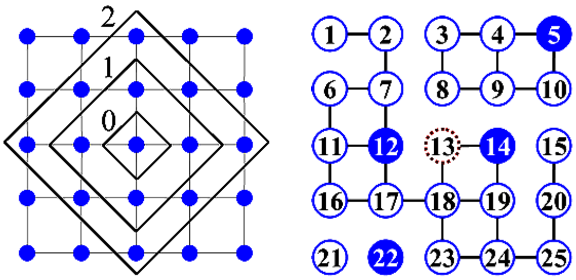
3. Reunifying Self-Organizing Map

In this study, we propose a new algorithm of SOM, Reunifying Self-Organizing Map. The initial state of all neurons of the reunifying SOM are connected to no neuron, but each neurons have its own physical location on the 2-D grid. The initial value of the weight vectors are given at orderly position.

(RSOM1) An input data $\boldsymbol{x}_j$ is inputted to all the neurons at the same time in parallel.

(RSOM2) The winner c is found according to Eq. (1).

(RSOM3) The neighborhood distance between c and each neuron i, denoted by $n_{c,i}$, is calculated. The



(a) (b)

Figure 1: Neighborhood distances of the rectangular grid. (a) Discrete neighborhoods (size 0, 1 and 2) of the centermost neuron when all the neurons are connected. (b) Neighborhood distances $n_{c,i}$ of $c = 13$. $n_{c,14} = 1$, $n_{c,12} = 3$, $n_{c,5} = n_{c,22} = m$ (namely, 25).

neighborhood distances are defined as shortest-path distances between connected map nodes as Fig. 1. If a neuron i is not connected directly or indirectly to the winner c, $n_{c,i}$ is equal to the number of neurons m.

(RSOM4) The weight vectors of the neurons are updated as;

$$\boldsymbol{w}_i(t+1) = \boldsymbol{w}_i(t) + h_{Rc,i}(t)(\boldsymbol{x}_j - \boldsymbol{w}_i(t)), \quad (5)$$

where $h_{Rc,i}(t)$ is the neighborhood function of the reunifying SOM and described as;

$$h_{Rc,i}(t) = \alpha(t) \exp\left(-\frac{(n_{c,i}/m + \|\boldsymbol{w}_i - \boldsymbol{x}_j\|)^2}{2\sigma^2(t)}\right). \quad (6)$$

(RSOM5) A set of 1-neighborhood neurons of winner c, on the assumption that all the neurons are connected (as Fig. 1(a)), is denoted as Nc_1. The number of Nc_1 is between two and four when the neurons are located on the 2-D rectangular grid.

A set of the neurons, whose neighborhood distance is the longest in Nc_1, are denoted as Sq (as Fig. 2).

$$Sq = \arg\max_i \{n_{c,i}\}, i \in Nc_1. \quad (7)$$

If $n_{c,Sq} = 1$ (as Fig. 2(c)), we perform (RSOM8). However, if not, we perform (RSOM6).

(RSOM6) The connecting neuron q is chosen from Sq, according to;

$$q = \arg\min_i \{\|\boldsymbol{w}_i - \boldsymbol{x}_j\|\}, i \in Sq. \quad (8)$$

In other words, the connecting neuron q is the neuron with the weight vector closest to $\boldsymbol{x}_j$ in Sq.

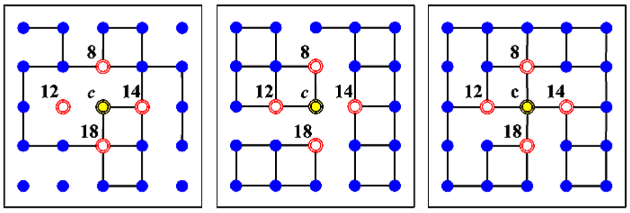
If $\|\boldsymbol{w}_q - \boldsymbol{w}_c\| \le D(t)$, we perform (RSOM7). However, if not, we except the neuron q from Nc_1, and perform the simulation again from (RSOM5).

$D(t)$ increases monotonically with time according to;

$$D(t) = \frac{D_{max}}{T} t + D_{min}. \quad (9)$$

(RSOM7) The winner c is directly connected to a connecting neuron q, namely, $n_{c,q}$ becomes 1.

(RSOM8) The steps is returned to (RSOM1) and is repeated for all the input data.



(a) (b) (c)

Figure 2: Example of connection of neurons. $c = 13$. $Nc_1 = [8, 12, 14, 18]$ (a) $Sq = 12$. $q = 12$. (b) $Sq = [14, 18]$. q is the neuron with the weight vector smaller distance between the input vector. (c) $Sq = [8, 12, 14, 18] = Nc_1$. q is not chosen.

4. Application to Clustering

4.1. Simulation 1

First, we consider 2-dimensional input data as shown in Fig. 3(a). The input data is generated artificially as follows. Total number of the input data N is 600, and the input data include two clusters. 300 data are distributed within a range from 0.2 to 0.8 horizontally and from 0.1 to 0.3 vertically. The remaining 300 data are distributed within a range from 0.2 to 0.8 horizontally and from 0.7 to 0.9 vertically. All the input data are sorted by random.

Both the conventional SOM and the reunifying SOM has $m = 100$ neurons (10×10). We repeat the learning 20 times for all input data, namely $T = 12000$. The parameters of the learning are chosen as follows;

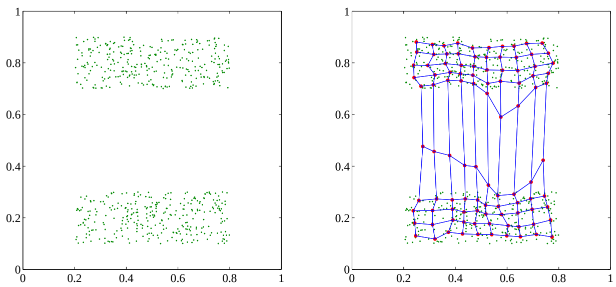
(For SOM)
$$\alpha(0) = 0.5, \sigma(0) = 3,$$

(For Reunifying SOM)

$$\alpha(0) = 0.5, \sigma(0) = 0.8, D_{min} = \sqrt{2}/10, D_{max} = \sqrt{2}/2.$$

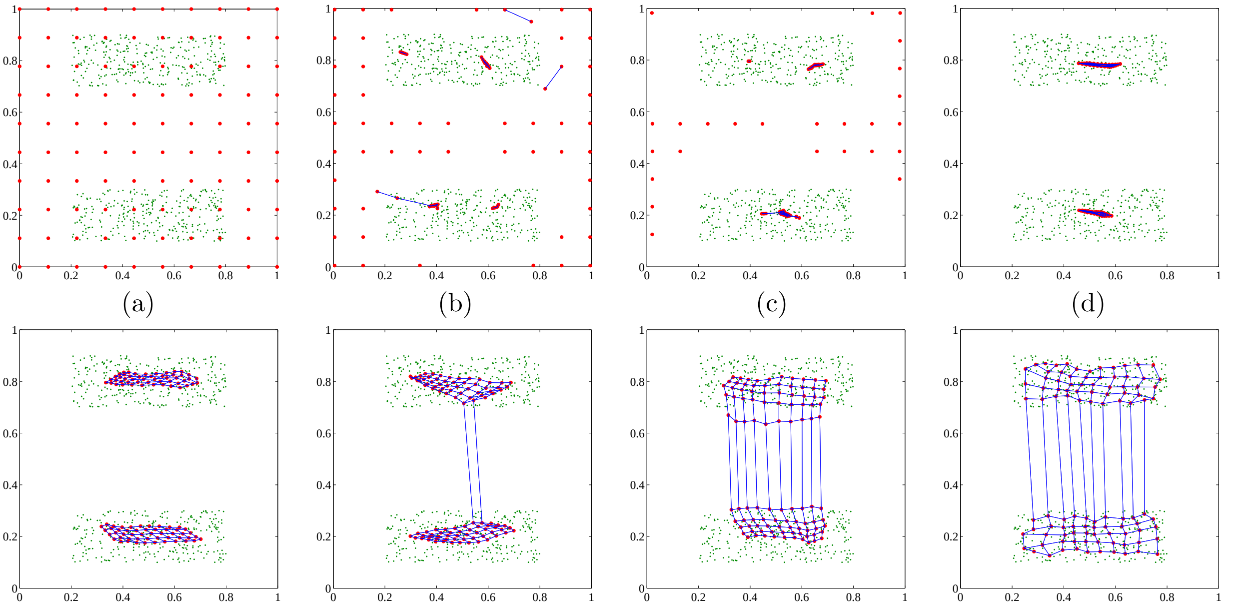
The simulation result of the conventional SOM is shown in Fig. 3(b). We can see that there are some inactive neurons between two clusters.

The other side, the result of the reunifying SOM and its learning process are shown in Fig. 4. We can see from Fig. 4(h) that there are no inactive neurons between two clusters.



(a) (b)

Figure 3: Clustering of 2-dimensional input data. (a) Input data. (b) Simulation result of the conventional SOM.

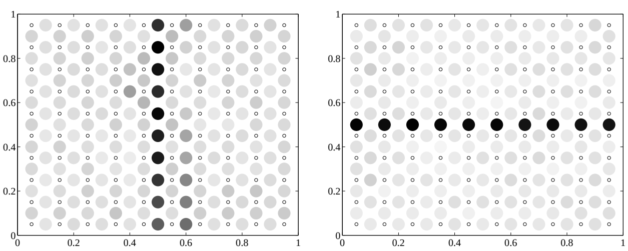


(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4: Clustering using the reunifying SOM. (a) Initial state ($t = 0$). (b) $t = 100$. (c) $t = 500$. (d) $t = 1700$. (e) $t = 6000$. (f) $t = 6200$. (g) $t = 6700$. (f) Simulation result ($t = 12000$).

Let us consider the learning process. The initial state of neurons are not connected to other neurons and are located at orderly position as Fig. 4(a). In the early-stage of learning as Figs. 4(b) and (c), we can see that some neurons are connected gradually to other neurons. However, there are some groups of neurons, because the connected neuron is preferentially-selected from the neuron which is connected to no neuron. Furthermore, as Fig. 4(d), there are no neurons which are connected to no neuron, and each two group of neurons self-organize respective two cluster without connecting each other. In the middle stage as Fig. 4(f), two groups of neurons are connected each other according to increase $D(t)$. However, neurons of each group self-organize the respective two clusters without being mingled as Fig. 4(g). in other words, the neurons of each group are not influenced by each other. This is because the learning rate $\alpha(t)$ and the neighborhood radius $\sigma(t)$ decrease according to Eq. (4), namely $\alpha(t)$ and $\sigma(t)$ are small value.



(a) (b)

Figure 5: Visualization of result. (a) Conventional SOM. (b) Reunifying SOM.

Figure 5 shows distances between neighboring neurons and thus visualizes the cluster structure of the map. Black circles on this figure mean large distance between neighboring map nodes. Clusters are typically uniform areas of white circles. We can see that the boundary line of the reunifying SOM is clear than the conventional SOM.

4.2. Simulation 2

Next, we carry out simulation for another 2-dimensional input data example shown in Fig. 6. The input data include two clusters. Total number of input data N is 600. We repeat the learning 20 times for all input data. The parameters of the learning are the same value used in the simulation 1 except for $D_{max} = \sqrt{2}$.

The results of SOM and the reunifying SOM are shown in Figs. 7(a) and (b), respectively. We can see that there are no inactive neurons in the reunifying SOM although there are some inactive neurons in conventional SOM.

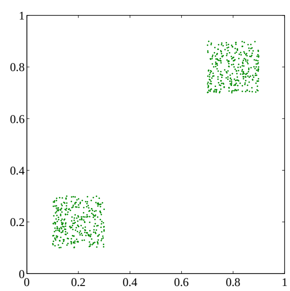


Figure 6: 2-dimensional input data.

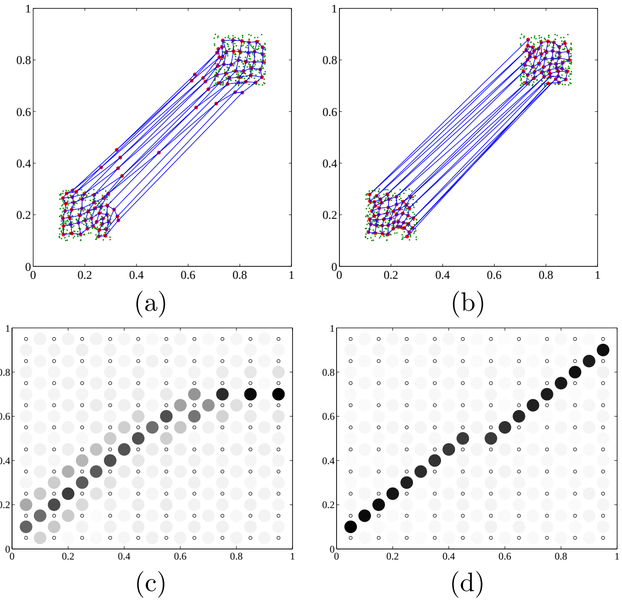


Figure 7: Results of clustering. (a) Conventional SOM. (b) Reunifying SOM. (c) Visualization of the conventional SOM. (d) Visualization of the reunifying SOM.

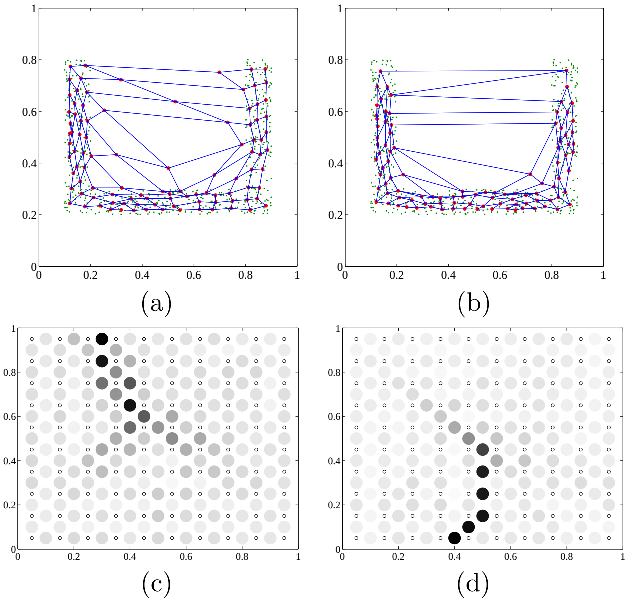


Figure 9: Clustering results for concave input data. (a) Conventional SOM. (b) Reunifying SOM. (c) Visualization of the conventional SOM. (d) Visualization of the reunifying SOM.

4.3. Simulation 3

Furthermore, we carry out simulation for the concave input data shown in Fig. 8. Total number of input data N is 600. We repeat the learning 20 times for all input data The parameters of the learning are the same values used in the simulation 1.

Figure 9 shows the result of SOM and the reunifying SOM. As we can see from these figures, the clustering ability of using the reunifying SOM method is effective.

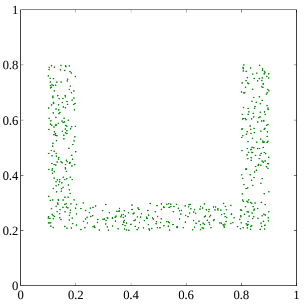


Figure 8: Concave input data.

5. Conclusions

In this study, we have proposed the new SOM algorithm which is the Reunifying SOM. The initial state of all neurons of the reunifying SOM are connected to no neuron. However, the neurons are connected gradually to other neurons as learning progressed. We have investigated its behaviors with applications to clustering, and have confirmed the efficiency.

References

[1] T. Kohonen, *Self-organizing Maps*, Berlin, Springer, vol. 30, 1995.

[2] Y. Cheng, "Clustering with Competing Self-Organizing Maps," *Proc. of IJCNN*, vol. IV, 1992, pp. 785-790.

[3] W. Wan and D. Fraser, "M2dSOMAP:Clustering and Classification of Remotely Sensed Imagery by Combining Multiple Kohonen Self-Organizing Maps and Associative Memory," *Proc. of IJCNN*, vol. III, 1993, pp. 2464-2467.

[4] J. Vesanto and E. Alhoniemi, "Clustering of the Self-Organizing Map," *IEEE Transactions on Neural Networks*, vol. 11, no. 3, 2002, pp. 586-600.

[5] I. Lapidot, H. Guterman and A. Cohen, "Unsupervised Speaker Recognition Based on Competition Between Self-Organizing Maps," *IEEE Transactions on Neural Networks*, 2002, vol. 13, no. 4, pp. 877-887.