17 - 26 Reunifying Self-Organizing Map for Effective Clustering

Haruna MATSUSHITA (Tokushima University)
Yoshifumi NISHIO (Tokushima University)

1. Introduction

The Self-Organizing Map[1] (SOM) attracts attentions for clustering in these years. 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. 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.

2. Reunifying Self-Organizing Map

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 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. 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.

(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.
(RSOM3) The weight vectors of the neurons are updated;

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

where $h_{Rc,i}(t)$ is described as;

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

where where $\|\cdot\|$ is the Euclidean distance, and $n_{c,i}$ is the neighborhood distance between c and each neuron i. The neighborhood distances are defined as shortest-path distances between connected map nodes. 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. $\alpha(t)$ 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;

$$\alpha(t) = \alpha(0)(1 - t/T), \quad \sigma(t) = \sigma(0)(1 - t/T), \qquad (3)$$

(RSOM4) A set of 1-neighborhood neurons Nc_1 of winner c, on the assumption that all the neurons are connected.

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

$$Sq = \arg\max_i \{n_{c,i}\}, i \in Nc_1. \qquad (4)$$

If $n_{c,Sq} = 1$, we perform (RSOM7). However, if not, we perform (RSOM5).
(RSOM5) The connecting neuron q is chosen from Sq. q is the neuron with the weight vector closest to $\boldsymbol{x}_j$ in Sq.

If $\|\boldsymbol{w}_q - \boldsymbol{w}_c\| \leq D(t)$, we perform (RSOM6). However, if not, we except the neuron q from Nc_1, and perform the simulation again from (RSOM4).

$D(t)$ increases monotonically with time;

$$D(t) = \frac{D_{max}}{T} t + D_{min}. \qquad (5)$$

(RSOM6) The winner c is directly connected to a connecting neuron q, namely, $n_{c,q}$ becomes 1.
(RSOM7) The steps is returned to (RSOM1) and is repeated for all the input data.

3. Application to Clustering

We consider 2-dimensional input data as shown in Fig. 1(a). The simulation result of the conventional SOM is shown in Fig. 1(b). We can see that there are some inactive neurons between input data. The other side, the result of the reunifying SOM are shown in Fig. 1(c). We confirm that there are no inactive neurons between input data. As we can see from these figures, the clustering ability of using the reunifying SOM method is effective.

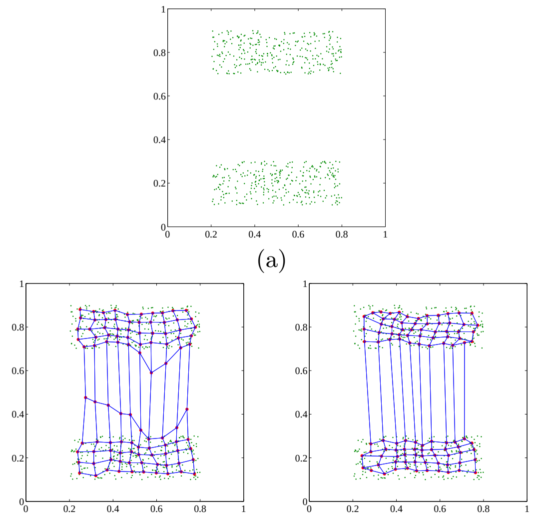

Figure 1: Clustering of 2-dimensional input data. (a) Input data. (b) Result of the conventional SOM. (c) Result of the reunifying SOM.

4. Conclusions

In this study, we have proposed the new SOM algorithm which is the Reunifying SOM. We have investigated its behaviors with applications to clustering, and have confirmed the efficiency.

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