18-33

Behavior of Community Self-Organizing Map

Taku HARAGUCHI (Tokushima University) **Haruna MATSUSHITA** (Tokushima University) **Yoshifumi NISHIO** (Tokushima University)

1. Introduction

The Self-Organizing Map (SOM) has attracted attention for the study on clustering in recent years. Meanwhile, in human society, there is a fascinating story that human-beings belong to sub-society, which is called community, and are called social animals. Furthermore, the community is created around the leader of the community. In this study, we propose Community Self-Organizing Map (CSOM) which is a new SOM algorithm. In CSOM algorithm, the neurons create some communities according to their winning frequency. We apply CSOM to various input data for clustering.

2. Community Self-Organizing Map

In CSOM algorithm, the neurons create some communities according to their winning frequency. Furthermore, the community is created around the leader of the community. M neurons are arranged as a regular 2-dimensional grid. A winning frequency W_i $(i = 1, 2, \cdots, M)$ is associated with each neuron i and is initially set to zero: $W_i = 0$. The number of members in each community C_k and the number of communities n are zero. Before learning, the all neurons do not belong to any community, however, they gradually belong to a community with learning.

(CSOM1) An input is given to all the neurons at the same time in parallel.
(CSOM2) Find a winner c by calculating a distance between the input vector $\boldsymbol{x}_j$ $(j = 1, 2, \cdots, N)$ and the weight vector $\boldsymbol{w}_i$ of each neuron i. The winner neuron c is the neuron with the weight vector nearest to the input vector $\boldsymbol{x}_j$. Update the weight vectors of all the neurons. If half time of the learning is over, increase the winning frequency of the winner c by

$$W_c^{\mathrm{new}} = W_c^{\mathrm{old}} + 1, \tag{1}$$

and perform (CSOM3). If not, perform (CSOM8).
(CSOM3) Evaluate whether the winner c satisfies the condition of the winning frequency to update the community information. If $W_c > W_{\mathrm{th}}(t)$ is satisfied, perform (CSOM4). If not, perform (CSOM8) without updating the community information. $W_{\mathrm{th}}(t)$ is the threshold value and increases with learning as

$$W_{\mathrm{th}}(t) = \frac{t}{2M}. \tag{2}$$

(CSOM4) Find a community C_k including the winner c. If winner c does not belong to any community, create a new community, $n^{\mathrm{new}} = n^{\mathrm{old}} + 1$, and affiliate the winner c to new community C_k as $c \in C_k$ (where $k = n^{\mathrm{new}}$). If not, c remains in its community C_k.
(CSOM5) Find a leader l_k which has become the winner most frequently among the all neurons belonging to C_k, according to Eq. (3).

$$l_k = \arg\max_i \{W_i\}, \quad i \in C_k. \tag{3}$$

(CSOM6) Find the neurons, whose winning frequencies are higher than $W_{\mathrm{th}}(t)$, in 1-neighborhood of the winner c, then consider whether they belong to any community. If this neighborhood neuron belongs to any community, perform (CSOM7). If not, affiliate it to the community C_k including the winner c, update the leader l_k as (CSOM5) and perform (CSOM8).
(CSOM7) Compare the winning frequencies of two leaders between the community including the winner and the community including winner's neighborhood neuron. Without loss of generality, assume that the winner c belongs to C_1 and its neighborhood neuron belongs to C_2. The leaders of C_1 and C_2 are assumed as l_1 and l_2, respectively. If $W_{l_2} \geq W_{l_1}$, the neighborhood neuron keeps on belonging to C_2. If not, the neighborhood neuron belonging to C_2 is absorbed into C_1. Then, in a specific case, if the neighborhood neuron is the leader l_2 in the community C_2, all the neurons belonging to C_2 are absorbed into C_1 and decrease the number of communities as $n^{\mathrm{new}} = n^{\mathrm{old}} - 1$.
(CSOM8) Repeat the steps from (CSOM1) to (CSOM7) for all the input data.
(CSOM9) After all learning are finished, check whether $W_i > 0.8 \times T/M$ for each particle i. If this is not satisfied, remove the particle i from the community including it.

3. Application to Clustering

We consider 2-dimensional input data shown in Fig. 1(a). The simulation results of the conventional SOM and CSOM are shown in Figs. 1(b) and (c), respectively. In Fig. 1(c), we can see that the number of communities is the same as the number of clusters. Furthermore, only the neurons, which self-organize the area where the input data are concentrated, create the communities. Therefore, we can see the number of clusters by investigating the number of communities.

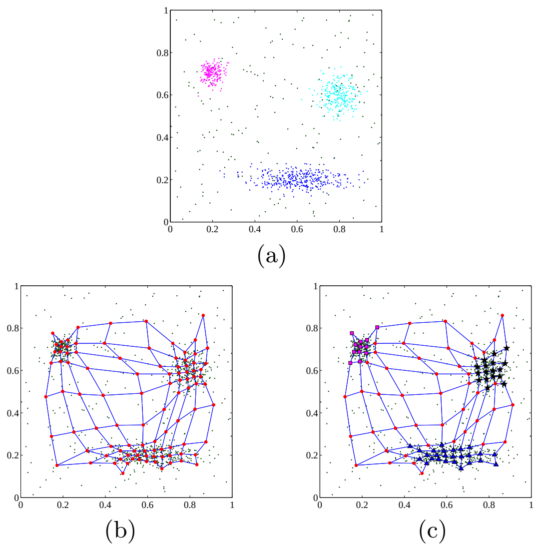

Figure 1: Simulation results for 2-dimensional input data. (a) Input Data. (b) Conventional SOM. (c) CSOM. ▲, ★ and ■ denote neurons belonging to the largest community C_1, the second community C_2 and the third largest community C_3, respectively.

4. Conclusions

In this study, we have proposed CSOM. We have investigated its behavior and have confirmed that the number of communities is the same as the number of clusters.