

SOM with False-Neighbor Degree and its Behaviors

Haruna MATSUSHITA* and Yoshifumi NISHIO*

* Department of Electrical and Electronic Engineering, Tokushima University
2-1 Minami-Josanjima, Tokushima 770-8506, Japan
Telephone: +81-88-656-7470, Fax: +81-88-656-7471,
Email: {haruna, nishio}@ee.tokushima-u.ac.jp

Abstract—This study improves SOM with False-Neighbor degree between neurons (FN-SOM) which solves some problems of the original FN-SOM. The improved FN-SOM changes the neighborhood relationship more flexibly according to the situation and the shape of data. We investigate the behavior of FN-SOM in detail. Learning performance is evaluated both visually and quantitatively using the three measurements. We confirm effectively of the improved FN-SOM with comparing with the conventional SOM, Growing Grid and the original FN-SOM.

I. INTRODUCTION

The Self-Organizing Map (SOM) is an unsupervised neural network [1] and has attracted attention for its clustering properties. In the learning algorithm of SOM, a winner, which is a neuron closest to the input data, and its neighboring neurons are updated, regardless of the distance between the input data and the neighboring neurons. For this reason, if we apply SOM to clustering of the input data which includes some clusters located at distant location, there are some inactive neurons between clusters where without the input data.

Then, what are the “neighbors”? In the real world, it is not always true that the next-door house is close to my house. In other words, “neighbors” are not always “true neighbors”. In addition, the relationship between neighborhoods is not fixed, but keeps changing with time. It is important to change the neighborhood relationship flexibly according to the situation.

On the other side, the synaptic strength is not constant in the brain. So far, the Growing Grid network was proposed in 1985 [2]. Growing Grid increases the neighborhood distance between neurons by increasing the number of neurons. However, there is not much research changing the synaptic strength even though there are algorithms which increase the number of neurons or consider rival neurons [3].

In our past study, we proposed the algorithm which changes the neighborhood distance between neurons [4]. However, the algorithm used the rank order of the distances between the input data and weight vectors of neurons in addition to changing the neighborhood distance. Thus the algorithm did not work well if the positions of all the weight vectors of the neurons were not taken into consideration. Moreover, the algorithm needs a lot of calculation amount because we have to calculate the rank order at every updating of the weight vector. Next, we proposed the SOM with False-Neighbor degree between neurons (called FN-SOM) [5]. False-neighbor degrees are allocated between adjacent rows and adjacent columns of FN-SOM. The false-neighbor degrees act as a burden of the distance between map nodes when the weight

vectors of neurons are updated. However, the algorithm has some problems as follows: 1) Only one false-neighbor degree is increased in one considering false-neighbor step. This is disadvantage when FN-SOM is applied to data containing plural clusters. 2) One false-neighbor degree must be increased in one considering false-neighbor step, every time. The false-neighbor degree is a FALSE false-neighbor degree in case that the no inactive neurons exist or the input has only one clusters. 3) It is hard to determine a parameter which corresponds to an amount of increasing the false-neighbor degree. We must adjust the parameter according to the each input data even if the input data is high-dimensional. In addition to these things, in the study, the behavior of FN-SOM and the effect of the false-neighbor degree were not completely clear, and the clustering ability was not evaluated.

In order to solve these problems, this study improves FN-SOM. In the algorithm of the improved FN-SOM, some false-neighbors are found in the one considering false-neighbor steps. In addition, for the uniform data which has only one cluster, FN-SOM does not find the false-neighbors. For this reason, the improved FN-SOM needs fewer the considering steps than the original FN-SOM. It leads to reduce the amount of computation time. Furthermore, the improved FN-SOM does not need the parameter which correspond to an amount of increasing the false-neighbor degree, namely, the parameter is fixed. The improved FN-SOM changes the neighborhood relationship more flexibly according to the situation and the shape of data.

We investigate the behavior of the improved FN-SOM with applying a uniform input data. Learning performance is evaluated both visually and quantitatively using the three measurements. We confirm that the improved FN-SOM solves above problems and its effectively with comparing with the conventional SOM, Growing Grid and the original FN-SOM.

II. SOM WITH FALSE-NEIGHBOR DEGREE (FN-SOM)

We explain improved SOM with False-Neighbor Degree (FN-SOM). In this study, we improve considering false-neighbor steps of FN-SOM. FN-SOM consist of $n \times m$ neurons located at 2-dimensional rectangular grid. Each neuron i has a d -dimensional weight vector $w_i = (w_{i1}, w_{i2}, \dots, w_{id})$ ($i = 1, 2, \dots, nm$). The range of the elements of d -dimensional input data $x_j = (x_{j1}, x_{j2}, \dots, x_{jd})$ ($j = 1, 2, \dots, N$) are assumed to be from 0 to 1. False-neighbor degrees of rows R_r ($1 \leq r \leq n - 1$) are allocated between adjacent rows of FN-SOM with the size of $n \times m$ grid. Likewise, false-neighbor

degrees of columns C_k ($1 \leq k \leq m-1$) are allocated between adjacent columns of FN-SOM. The initial values of the false-neighbor degrees are set to zero.

Learning Step

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

(FN-SOM2) Distances between \mathbf{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 \mathbf{x}_j .

(FN-SOM3) Increment the winning frequency of winner c by $\gamma_c^{\text{new}} = \gamma_c^{\text{old}} + 1$.

(FN-SOM4) The neighboring distances between the winner c and the other neurons are calculated. For instance, for two neurons s_1 , which is located at r_1 -th row and k_1 -th column, and s_2 , which is located at r_2 -th row and k_2 -th column, the neighboring distance is defined as the following measure;

$$d_f(s_1, s_2) = (|r_1 - r_2| + \sum_{r=r_1}^{r_2-1} R_r)^2 + (|k_1 - k_2| + \sum_{k=k_1}^{k_2-1} C_k)^2, \quad (1)$$

where $r_1 < r_2$, $k_1 < k_2$, namely, $\sum_{r=r_1}^{r_2-1} R_r$ means the sum of the false-neighbor degrees between the rows r_1 and r_2 , and $\sum_{k=k_1}^{k_2-1} C_k$ means the sum of the false-neighbor degrees between the column k_1 and k_2 .

(FN-SOM5) The weight vectors of the neurons are updated:

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

where $h_{F_{c,i}}(t)$ is the neighborhood function of FN-SOM:

$$h_{F_{c,i}}(t) = \alpha(t) \exp\left(-\frac{d_f(c, i)}{2\sigma^2(t)}\right). \quad (3)$$

Both $\alpha(t)$ and $\sigma(t)$ decrease with time.

(FN-SOM6) If $\sum_{i=1}^{nm} \gamma_i \geq \lambda$ is satisfied, we find the false-neighbors and increase the false-neighboring degree, according to steps from (FN-SOM7) to (FN-SOM10). If not, we perform step (FN-SOM11). In other words, we consider the false-neighbors every time when the learning steps are performed for λ input data.

Considering False-Neighbors Step

(FN-SOM7) We find a set of neurons S which have never become the winner: $S = \{i \mid \gamma_i = 0\}$. Because some false-neighbors are found in one considering false-neighbor step, the improved FN-SOM need fewer the considering steps than the original FN-SOM. If the neuron, which have never become the winner, does not exist, namely $S = \emptyset$, we return to (FN-SOM1) without considering the false-neighbors. It leads to reduce the amount of computation time, and the improved FN-SOM can change the neighborhood relationship more flexibly.

(FN-SOM8) A false-neighbor f_q of each neuron q in S is chosen from the set of direct topological neighbors of q denoted as N_{q1} . f_q is the neuron whose weight vector is most distant from q :

$$f_q = \arg \max_i \{\|\mathbf{w}_i - \mathbf{w}_q\|\}, \quad q \in S, \quad i \in N_{q1}. \quad (4)$$

(FN-SOM9) A false-neighbor degree between each q and its

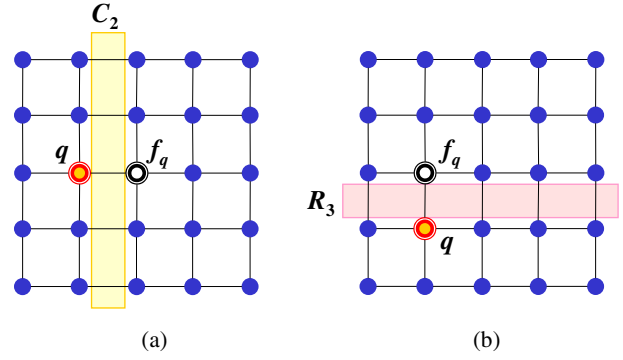


Fig. 1. Increment of the false-neighbor degree. (a) q and its false-neighbor f_q are in the 3rd row and in the 2nd and 3rd column, respectively. Then, the false-neighbor degree C_2 between columns 2 and 3 is increased by Eq. (5). (b) q and f_q are in the 2nd column and in the 4th and 3rd row, respectively. The false-neighbor degree R_3 between rows 3 and 4 is increased by Eq. (6).

false-neighbor f_q , R_r or C_k , is increased. If q and f_q are in the r -th row and in the k -th and $(k+1)$ -th column (as Fig. 1(a)), the false-neighbor degree C_k between columns k and $k+1$ is increased according to

$$C_k^{\text{new}} = C_k^{\text{old}} + \frac{n+m}{2nm}. \quad (5)$$

In the same way, if q and f_q are in the k -th column and in the $(r+1)$ -th and r -th row (as Fig. 1(b)), the false-neighbor degree R_r between rows r and $r+1$ is also increased according to

$$R_r^{\text{new}} = R_r^{\text{old}} + \frac{n+m}{2nm}. \quad (6)$$

These amounts of increasing the false-neighbor degree are derived by the number of neurons numerically and are fixed, namely, the improved FN-SOM does not need the parameter which the original FN-SOM needed.

(FN-SOM10) The winning frequency of all the neurons are reset to zero: $\gamma_i = 0$.

(FN-SOM11) The steps from (FN-SOM1) to (FN-SOM10) are repeated for all the input data.

III. EXPERIMENTAL RESULTS

We apply the improved FN-SOM to various input data and compare the improved FN-SOM with the conventional SOM, Growing Grid and the original FN-SOM. For all experiments, both SOM and FN-SOM have $nm = 100$ neurons (10×10). Growing Grid starts learning with a 2×2 neurons, and new rows and columns are inserted as long as the number of neurons is less than 100. $\lambda = 3000$ for the improved FN-SOM are chosen. We use the same decreasing equations of α and σ (in Eq. (3)) to SOM and the improved FN-SOM for the comparison and the confirmation of the false-neighbor degree effect. The input data are normalized and are sorted at random.

A. For uniform data

In order to confirm that the false-neighbor degree has a bad effect or not, we consider a 2-dimensional input data, which is uniformity and has no cluster. The total number of the input

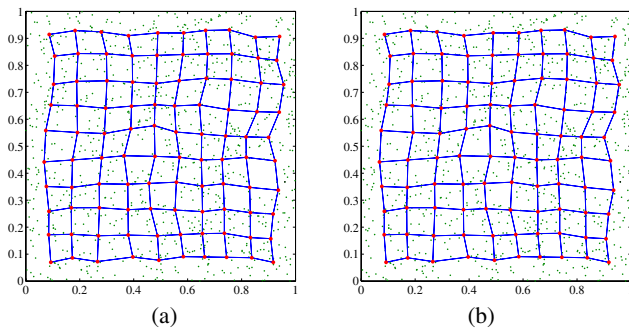


Fig. 2. Learning results of two algorithms for no cluster data. (a) Conventional SOM. (b) Improved FN-SOM.

data N is 1000. We repeat the learning 15 times for all input data.

Learning results of the two algorithms are shown in Fig. 2. We can see that the result of the improved FN-SOM is completely same as the conventional SOM.

Furthermore, in order to compare the learning performance of the improved FN-SOM with the conventional SOM numerically, we use the following well-used three measurements to evaluate the training performance.

Quantization Error Q_e measures the average distance between each input vector and its winner [1]. The small value Q_e is more desirable.

Topographic Error T_e describes how well the SOM preserves the topology of the studied data set [6]. The small value T_e is more desirable. Unlike the quantization error, it considers the structure of the map. For a strangely twisted map, the topographic error is big even if the quantization error is small.

Neuron Utilization U measures the percentage of neurons that are the winner of one or more input vector in the map [3]. Thus, U nearer 1.0 is more desirable.

The calculated three measurements are shown in Table. I. We can confirm that the all measurements of the improved FN-SOM are same as the conventional SOM. This is because all neurons become winner, namely false-neighbors are not chosen, for the uniform input data. We can say that the false-neighbor degree does not have a bad effect.

Furthermore, the improved FN-SOM reduces the amount of computation time 27.45% from using the original FN-SOM. This is because, if the neuron, which have never become the winner, does not exist, the improved FN-SOM does not find the false-neighbors.

B. For 2-dimensional problem

Next, we consider 2-dimensional input data as shown in Fig. 3. The input data is Target data set, which has a clustering problem of outliers [7]. The total number of the input data N is 770, and the input data has six clusters which include 4 outliers. We repeat the learning 20 times for all input data.

The learning results of the conventional SOM, Growing Grid and the original FN-SOM are shown in Figs. 3(a), (b)

TABLE I
THREE MEASUREMENTS FOR NO CLUSTER DATA.

	SOM	Improved FN-SOM
Q_e	0.0379	0.0379
T_e	0.01	0.01
U	1	1

TABLE II
THREE MEASUREMENTS FOR TARGET DATA.

	SOM	Growing Grid	FN-SOM	Improved FN-SOM
Q_e	0.0207	0.0237	0.0200	0.0191
T_e	0.0740	0.2455	0.0494	0.0442
U	0.8100	0.8137	0.8200	0.9100

and (c), respectively. We can see that there are some inactive neurons between clusters. The other side, the result of the improved FN-SOM is shown in Fig. 3(d). We can see from this figure that there are just a few inactive neurons between clusters, and the improved FN-SOM can obtain the more effective map reflecting the distribution state of input data than SOM, Growing Grid and FN-SOM.

The calculated three measurements are shown in Table. II. The quantization error Q_e of the improved FN-SOM is the smallest value among the four algorithms, and by using the improved FN-SOM, the quantization error Q_e has improved 7.7% from using the conventional SOM. This is because the result of the improved FN-SOM has few inactive neurons, therefore, more neurons can self-organize the input data. This is confirmed by the neuron utilization U . The neuron utilization U of the improved FN-SOM is the largest value in the four algorithms. It means that 90% of the neurons of the improved FN-SOM are the winner of one or more input data, namely, there are few inactive neurons. On the other hand, the topographic error T_e of the improved FN-SOM is the smallest value although Q_e and U are the best values. It means that the improved FN-SOM self-organizes most effectively with maintenance of top quality topology.

Furthermore, the improved FN-SOM can reduce the amount of computation 21.01% from using the original FN-SOM. This is because, the improved FN-SOM need fewer the considering false-neighbor steps than the original FN-SOM since some false-neighbors are found in the one steps.

C. For Iris data

Furthermore, we apply the improved FN-SOM to the real world clustering problem. We use the Iris plant data [8] as real data. This data is one of the best known databases to be found in the pattern recognition literature [9]. The data set contains three clusters of 50 instances respectively, where each class refers to a type of iris plant. The number of attributes is four as the sepal length, the sepal width, the petal length and the petal width, namely, the input data are 4-dimension. The three classes correspond to *Iris setosa*, *Iris versicolor* and *Iris virginica*, respectively. We repeat the learning 100 times for all input data.

Figure 4 shows distances between neighboring neurons and

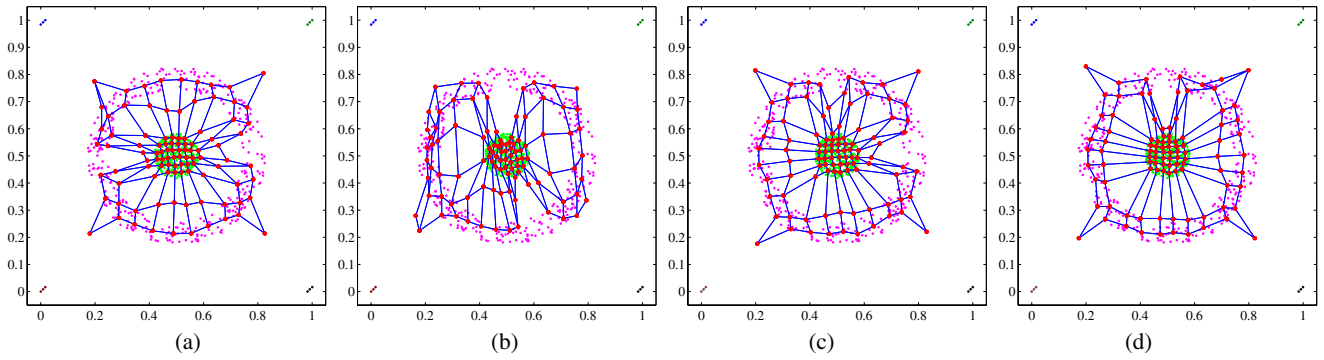


Fig. 3. Learning results of four algorithms for Target data. (a) Conventional SOM. (b) Growing Grid. (c) FN-SOM. (d) Improved FN-SOM.

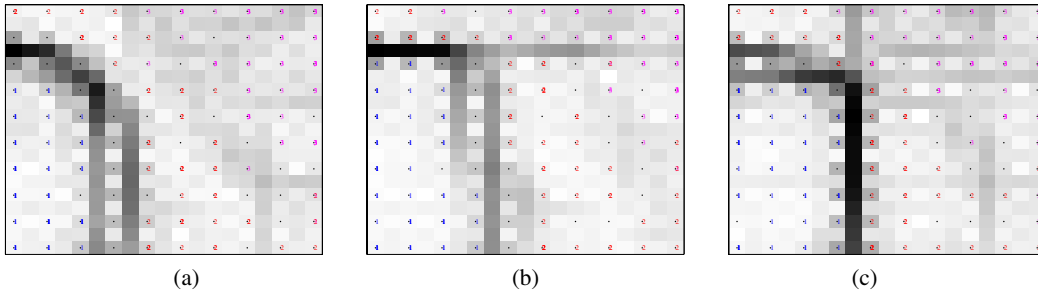


Fig. 4. U-Matrix of simulation results for Iris data. (a) Conventional SOM. (b) FN-SOM. (c) Improved FN-SOM.

TABLE III
THREE MEASUREMENTS FOR IRIS DATA.

	SOM	Growing Grid	FN-SOM	Improved FN-SOM
Qe	0.0374	0.0452	0.0368	0.0323
Te	0.2400	0.2267	0.1800	0.1667
U	0.7200	0.7048	0.7800	0.8200

thus visualizes the cluster structure of the map [7]. We can see that the boundary line of the improved FN-SOM is clearest.

The calculated measurements are shown in Table. III. We can confirm that the quantization error Qe and the topographic error Te of the improved FN-SOM are the smallest value among the four algorithms. Qe of the improved FN-SOM has improved 13.6% from using the conventional SOM and Te of the improved FN-SOM has also improved 30.5% from using the conventional SOM. This is because the result of the improved FN-SOM hardly has inactive neurons between *Iris setosa* and the other two, therefore, more neurons can self-organize the data of *Iris versicolor* and *Iris virginica*. Furthermore, U of the improved FN-SOM is also the best value. From these results, we can confirm the efficiency of the improved FN-SOM.

IV. CONCLUSIONS

In this study, we have improved SOM with False-Neighbor degree between neurons (called FN-SOM). In the algorithm of the improved FN-SOM, some false-neighbors are found in one considering false-neighbor step. However, for a uniform data which has only one cluster, the improved FN-SOM does

not find the false-neighbors. It leads to reduce the amount of computation time.

We have applied FN-SOM to a uniform data, 2-dimensional data and Iris data, and we have investigated the learning behaviors of FN-SOM. Furthermore, the results were compared with those obtained by the conventional SOM, Growing Grid and the original FN-SOM. We have confirmed that the quantization error and the topographic error of the improved FN-SOM were the smallest value among the four algorithms. Moreover, the neuron utilization of FN-SOM was the largest value in the four algorithms. The amount of computation time is significantly reduced from using the original FN-SOM. From these results, we have confirmed the efficiency of the improved FN-SOM.

REFERENCES

- [1] T. Kohonen, *Self-organizing Maps*, Berlin, Springer, 1995.
- [2] B. Fritzke, "Growing Grid – a self-organizing network with constant neighborhood range and adaptation strength," *Neural Processing Letters*, vol. 2, no. 5, pp. 9–13, 1995.
- [3] Y. Cheung and L. Law, "Rival-Model Penalized Self-Organizing Map," *IEEE Trans. Neural Networks*, vol. 18, no. 1, pp. 289–295, 2007.
- [4] H. Matsushita and Y. Nishio, "Self-Organizing Map Considering False Neighboring Neuron," *Proc. of ISCAS'07*, pp. 1533–1536, 2007.
- [5] H. Matsushita and Y. Nishio, "Self-Organizing Map with False Neighbor Degree between Neurons for Effective Self-Organization," *Proc. of WSOM*, pp. WeP-P-13, 2007.
- [6] K. Kiviluoto, "Topology Preservation in Self-Organizing Maps," *Proc. of ICNN*, pp. 294–299, 1996.
- [7] A. Ultsch, "Clustering with SOM: U*C," *Proc. of WSOM'05*, pp. 75–82, 2005.
- [8] D. J. Newman, S. Hettich, C. L. Blake and C. J. Merz, UCI Repository of Machine Learning Database, 1998.
- [9] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annual Eugenics*, no.7, part II, pp. 179–188, 1936.